MOUNTAIN HEMLOCK LANDSCAPE GENETICS ON THE KENAI PENINSULA, ALASKA

A Dissertation

by

JEREMY SCOTT JOHNSON

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	David M. Cairns
Committee Members,	Charles W. Lafon
	Oliver W. Frauenfeld
	Konstantin V. Krutovsky
Head of Department,	David M. Cairns

August 2016

Major Subject: Geography

ABSTRACT

Global ecological change is a serious threat to biodiversity. Changes in temperature, precipitation, and landscape fragmentation may reduce the ability of plants to persist within these changing environments. Knowing how plants have responded to past climate fluctuations, and accounting for alternative future responses will help conservationists, land managers, and policy makers plan for the future. This research takes a multiscale genomic approach to understand how a long lived conifer tree, mountain hemlock (*Tsuga mertensiana* Bong (Carr)), has responded to past climate variability, and assesses its potential to respond to future changes.

I used biogeographic and landscape genetic approaches, based on double digest Restriction Associated DNA sequencing (ddRADseq), to address three research questions: 1) Are isolated stands of mountain hemlock on the Kenai Peninsula of Alaska glacial relicts or are they the product of long distance dispersal following glacial retreat? 2) What is the dominant mode of reproduction at mountain hemlock treeline? 3) Are seeds arriving at the alpine treeline ecotone the product of local dispersal or are they arriving from more distant sources?

I found that genomic diversity and genetic structure differed between isolated stands of mountain hemlock and those found across the rest of the peninsula. A graph approach based on electrical circuit analysis identified high landscape connectivity and conductance across the peninsula. Genetic variation was primarily explained by landscape resistance and not geographic distance. These findings suggest that mountain

hemlock colonized the peninsula via long distance dispersal and repeated founding events accompanied by high levels of gene-flow. At the local scale, seed dispersal ranged from 1.44 to 326.85 m with a mean dispersal distance of 73 m. Most seeds arrived as cryptic gene flow from beyond treeline. Long distance dispersal was quantified at the 99th percentile of the dispersal curve and accounted for dispersal at distances greater than 450 m. This analysis showed that under similar landscape configurations, mountain hemlock has the capability to track its shifting climate niche in response to future climate change.

This research represents a novel integration of genomics and geography to answer a pertinent set of questions allowing us to have a deeper understanding of how plants may migrate under shifting climate conditions.

DEDICATION

For my children, Matilda and Miles, who have no idea what I'm doing, other than to think I could probably start practicing medicine, but all of this effort is for them.

Above all, this is for Lesley.

ACKNOWLEDGEMENTS

I don't know what the future holds, but when I started my PhD I figured on a fairly straight forward path. The true path was a little more crooked, but it was an amazing experience nonetheless. Now, five years later, I want to thank all those who helped me on this journey.

My research and academic training has been supported unquestionably by my advisor Dr. David Cairns. I am not sure what I was thinking when Dave told me to go see what landscape genetics was all about in my first semester as a PhD student, but his faith and support in me never wavered. Dave pushed me to apply for funding, to publish, and to take advantage of opportunities to travel and network around the world.

My committee guided me and challenged me throughout this process. Dr. Charles Lafon helped me to become a better writer by "Schimeling" my papers. He provided excellent perspective and interesting discussions on many matters related to biogeography. Dr. Oliver Frauenfeld and Dr. Konstantin Krutovsky both provided interesting discussion and perspective, and helped me shape my ideas and my research.

I am indebted to those who lent me a hand in this process. I learned many tricks in the laboratory from Dr. Clint Magill, Dr. Ghada Radwan, Dr. Keith Gaddis, and the folks in the molecular laboratory at Landcare Research in New Zealand (Kat and Caroline). I would like to thank Dr. Magill specifically, for without him and his generosity I would not have had laboratory space in which to conduct this research. Also, a special thanks goes out to Kranti Konganti and the Texas A&M Institute for

Genome Sciences and Society for providing so much help in the form of bioinformatics analysis.

I would like to thank the many people who helped me in the field. Dr. David Cairns and Dr. Charles Lafon explored the Kenai Peninsula with me on my first trip to Alaska in 2012. Hey Bear! In 2013 Ellen Gass and Trey Murphy accompanied me to collect the bulk of my samples (RNMK). Camping in the Chugach and sleeping on a boat in the Gulf of Alaska wouldn't have been the same without those two. Additionally the National Park Service, the US Forest Service, US Fish and Wildlife Service, and the Army all facilitated this research. In particular Dr. Amy Miller of the NPS allowed me to tag along on the NPS M/V Serac to collect samples in Kenai Fjords NP. We had many beneficial conversations on that boat.

The broader geography community at Texas A&M has been supportive for so many years now. I wouldn't have made it this far without them. I would like to thank the Biogeography research cluster specifically. This group had to listen to my jargon filled talks for too many years. I would like to thank Dr. Adam Naito and Parveen Chhetri for listening to my ideas and providing advice and insight. Many people have come and gone over my time at A&M, there are too many to thank individually, but I would like to say thank you to the entirety of the department in all of its configurations over my time here.

I was financially supported by the National Science Foundation [grant number BCS-1333527] an NSF East Asia and Pacific Summer Institute (EAPSI) fellowship and a Texas A&M Institute for Genome Sciences and Society fellowship in addition to many

other small research and travel grants. I am grateful to Kenai Fjords NP, Chugach NF, Kenai NWR, and the US Army Fort Richardson for providing in kind support of lodging and transportation and allowing me access to conduct fieldwork in spectacular places.

Thanks to my family who supported me through this process. My brothers and parents always checked in to see how I was doing. To my wife's family too, they provided so much help to my family while I was away at conferences, in Alaska, and in New Zealand.

Lastly, I thank my wife Lesley and kids, Matilda and Miles, for their support, patience, and accompaniment throughout graduate school. Without your support I would not be here. My family kept life in perspective when my research seemed to be going off of the rails. I can't wait for all the adventures ahead with you.

NOMENCLATURE

BP	Base Pair
ddRADseq	double digest Restriction Associated DNA sequencing
DNA	Deoxyribonucleic Acid
GBS	Genotyping-by-Sequencing
IBD	Isolation by Distance
IBR	Isolation by Resistance
LDD	Long Distance Dispersal
NGS	Next Generation Sequencing
SNP	Single Nucleotide Polymorphism

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

ECOLOGICAL BIOGEOGRAPHY

Ecological change is a rapidly growing problem with global implications. Changes in the abundance and distribution of organisms along with changes in their genetic variation have important consequences for biodiversity and the ecosystem functions and services that they provide to humanity. The response of organisms to these changes is still unknown. Plants in particular have different strategies allowing them to respond in alternative ways to changes in biotic and abiotic ecosystem components. For instance, they can move by dispersing to more suitable locations, they can stay put and evolve genetically to withstand the changes, or they can rely on currently available genetic variation to bend phenotypically. If these three strategies, or combinations thereof, are exhausted, then populations, or the species as a whole, will go extinct. This will lead to a critical loss of biodiversity at multiple scales.

Dispersal is one of the fastest ways for a plant to respond to rapid changes in its environment. For long lived plants, like forest trees, time lags associated with dispersal, establishment, and growth to reproductive maturity may be too long for the species to track rapid changes in their climate envelopes. Because of this it is important that we measure the capacity of tree species to move so that appropriate management and conservation policies can be implemented.

The difficulty in quantifying seed and pollen dispersal processes in plants results from the challenges of tracking highly mobile seeds and pollen across the landscape. A promising approach is to use genetics and the concept of effective dispersal to track the movement of seeds and pollen. Effective dispersal is the end result of seed and pollen movement and is the most biologically meaningful component of the dispersal process, and the only part that leads to the passing on of genes to the next generation and an increase in fitness. Because an organisms' genes are passed on during effective dispersal, comparisons of their genotypes and population genetic structure can reveal demographic process. The population genetic structure and genetic variation can then be correlated with climate and landscape configuration to provide inferences about movement potential and landscape connectivity. The demographic patterns associated with biotic and abiotic interactions provide a mechanism to address how plants will respond to global ecological change.

This dissertation addresses the larger biogeographic question of how plants will respond to global ecological change via gene flow and dispersal. I will broadly introduce the three primary ways that plants counter ecological change. Plant movement, adaptation, and phenotypic plasticity, will be discussed along with methods of addressing these responses using emerging genetics and genomic approaches. Secondly, the research presented here will explicitly address how a long lived tree species, mountain hemlock (*Tsuga mertensiana* (Bong.) Carr.), has responded to historic climate change at landscape scales, and its potential to respond to contemporary change at local

scales. This dissertation contributes to the broader understanding of plant movement within an ecological biogeographic context.

Landscape Genetics

Over the past ten years an analytical approach has emerged to address biological movement and adaptation and its relationship to landscape heterogeneity; a field described as landscape genetics (Manel et al., 2003). Landscape genetics is defined as a theoretical and methodological approach to investigate how landscape configuration, composition and matrix quality impact spatial genetic structure via gene flow, dispersal and adaptation. This analytical framework has grown rapidly with technological advances in geospatial analysis and high throughput genome sequencing (Holderegger et al., 2006; Holderegger and Wagner, 2008; Holderegger et al., 2010; Segelbacher et al., 2010; Sork and Waits, 2010; Storfer et al., 2010; Manel and Holderegger, 2013; Bolliger et al., 2014). Landscape genetics integrates the spatial and temporal components of geography and landscape ecology with the molecular and evolutionary approaches of population genetics (Manel et al., 2003; Storfer et al., 2007; Holderegger and Wagner, 2008; Storfer et al., 2010). Landscape genetics thus serves to generate hypotheses through data exploration and also provides a mechanism for testing these hypotheses with empirically collected data allowing spatial genetic structure to be compared to environmental and landscape variables. Generally, landscape genetics relies on presumably neutral genetic variation. Neutral genetic variation is the component of DNA

that is not involved directly in coding for a protein and thus is thought not to be under selection. Throughout this dissertation, landscape genetics will be used to address the research questions.

Landscape Genomics

Landscape genomics, as compared to landscape genetics, has emerged as a way to move beyond the assessment of neutral genetic variation and gene flow to the study of selection by looking at coding regions of DNA (Rellstab et al., 2015; Ćalić et al., 2015). In general, landscape genomics will screen a species' genome for outlier loci, usually defined as loci with very high F_{st} values, and then try to correlate the outliers with environmental and landscape features or gradients that might suggest a selective pressure. This framework is geared towards understanding the potential for species to adapt to global ecological change. The research presented in this dissertation is fundamentally a genomic approach to landscape genetics, but it is not a landscape genomic analysis itself.

Species Limits

Species limits are geographic locations or boundaries where species reach their maximum tolerance for growth and reproduction, usually because of abiotic stress or interspecific competition (Gaston, 2003). Species limits can occur across many scales.

For example, one form of a species limit is either its northern or southern range. These species limits may be characterized by the presence of other species with different life history traits and strategies that allow them to better use the resources available in that specific environment and thus possess a fitness advantage. Alternatively, species limits may be controlled by large scale climatic variation, such as along a latitudinal gradient. Small scale species limits are often found along ecotones or ecoclines (van der Maarel, 1990) where ecological gradients associated with geology and topography control the spatial occurrence of a species. Because range limits generally result in a transition of one community to another, they are useful for studying how species are responding to changes in their environment. This is because their movement at the ecotone or ecocline can be observed and measured more easily than within the core of the range. The term bioindicator is used to describe species at their limit where their movement can indicate if the species is responding to some change in its environment (Smith, 1994). Because of this, plant bioindicators can serve as the "canary in the coal mine" for climate change. This dissertation investigated landscape, climate, genotype interactions at two different spatial scales and at two different species limits: landscape level northern species range limit and site level altitudinal species limit at the alpine treeline ecoton.

MOUNTAIN HEMLOCK

Mountain hemlock (Fig 1-1) is a gymnosperm with many phenotypic similarities to other conifers in its life history traits. It is a slow growing, late successional species that

can live upwards of 800 - 1000 years in undisturbed locations (Means, 1990). It is monoecious, producing both seed and pollen on the same tree. Both seed and pollen are wind and gravity dispersed. Mountain hemlock is highly outcrossed with very little evidence of inbreeding (Ally et al., 2000). The species tends to inhabit sites that experience little disturbance and can be found in locations with poorly drained soils such as sloping peatlands or locations were snow pack tends to accumulate (Fig 1-2) (Peterson and Peterson, 2001; Taylor, 1995).

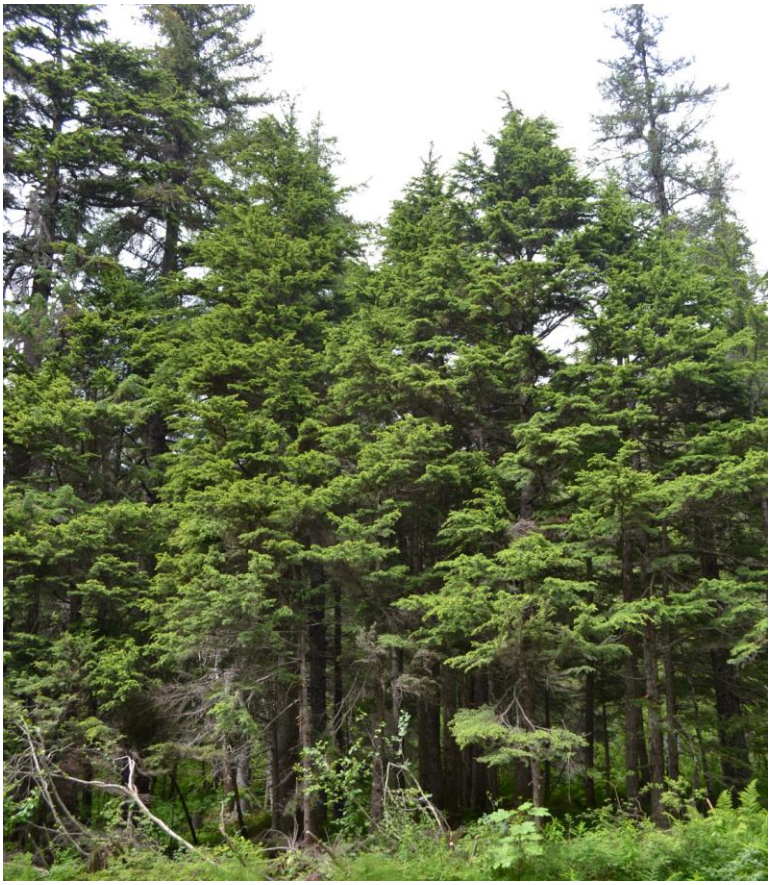

Figure 1-1: Typical mountain hemlock stand on the Kenai Peninsula, Alaska

Figure 1-2: Mountain hemlock growing (a) at alpine treeline in the Kenai Mountains (Palmer Creek), (b) Isolated stand (Discovery Well) in the Kenai Lowlands, and (c) slopping peatland in Kenai Fjords NP.

Mountain hemlock stand expansion and migration is related to length of growing season, a function of winter temperature and snow pack, and summer temperatures and moisture availability (Peterson and Peterson, 2001; Taylor, 1995). A past study of cone crop production indicated that warm July temperatures result in increased seed production (Woodward et al., 1994). The warming climate on the Kenai Peninsula is becoming more favorable for the growth of mountain hemlock (Fig 1-3) (Means, 1990) and new establishment of mountain hemlock within the treeline ecotone has been observed (Johnson pers. obs). Mountain hemlock is a dominant treeline species in wetter environments ranging from south central Alaska to the Olympic Mountains, and a component of treeline ecotones as far south as northern California. Therefore, this research has implications beyond the Kenai Peninsula. I chose mountain hemlock as the study species because it is the dominant treeline species on the Kenai Peninsula, Alaska and it is also at its northern range limit on the peninsula.

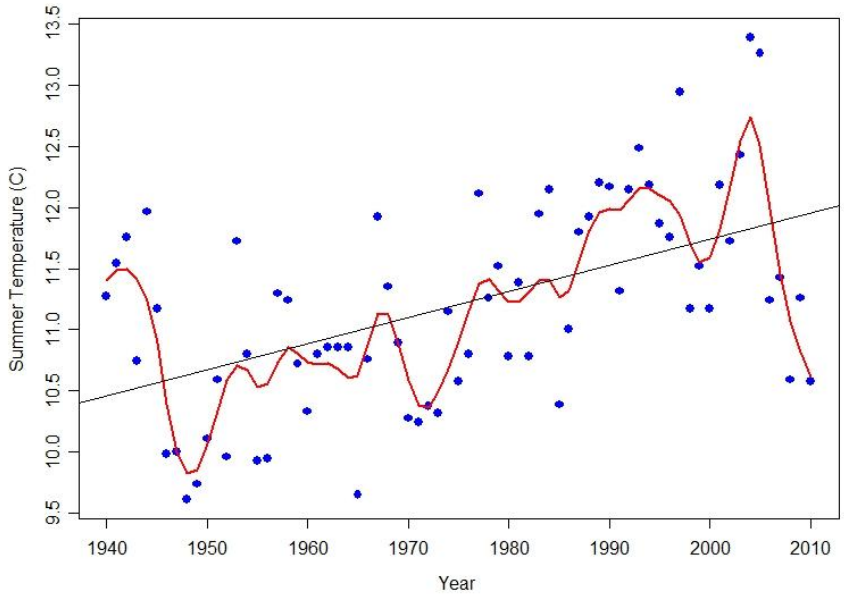

Figure 1-3 Trends in average summer temperature since 1940 at Homer, Alaska airport on the Kenai Peninsula. Red line is the mean summer temperature.

Population Genetics of Mountain Hemlock

Past studies have investigated the population genetics of mountain hemlock farther south in its range. For instance, Benowicz and El-Kassaby (1999) investigated the genetic variation in mountain hemlock to assess the quantitative and adaptive traits of the species that allow it to live in a wide diversity of environmental conditions. Using common garden trials, they found that variation in genetic structure was correlated with geographic position of originating population but that most of the variation was within populations and not between them.

Following up on the work of Benowicz and El-Kassaby (1999) Ally et al. (2000) used allozyme analysis to investigate the genetic diversity of mountain hemlock, finding higher genetic variation in isolated populations of the species than in continuously distributed forests, suggesting isolation by distance. Ally et al. (2000) also investigated historic migration patterns of mountain hemlock at 19 sampling locations along the west coast of British Columbia and found average expected heterozygosity increased with increasing elevation but decreased with increasing latitude. The higher heterozygosity at high elevations may be an adaptive trait that allows them to compensate for the extreme conditions or it may be a result of range shifts caused by a warming climate.

Ally and Ritland (2007) used microsatellite markers to look at the effects of forest fragmentation on genetic structure and dispersal of mountain hemlock. An old growth mountain hemlock stand with an adjacent regenerating clear-cut was used as a case study. The results showed that the fragmentation did not influence genetic diversity between the two sites significantly. As expected, the relatedness of individuals was a result of distance with the implication that local relatedness is a result of local seed dispersal limitation. Ally and Ritland's (2007) estimates of dispersal distance found that within an old growth Mountain Hemlock stand seeds dispersed 3.5-162m but when dispersing into a clear cut the distances were slightly higher, approximately 16.7-207m.

All of the previous research on mountain hemlock spatial genetic structure and dispersal has been based on only a handful of highly polymorphic neutral molecular markers or protein assays. This dissertation focuses on developing thousands of single

nucleotide polymorphisms (SNPs) to assess gene flow, dispersal, and spatial genetic structure across the entire genome of mountain hemlock.

KENAI PENINSULA, ALASKA

The Kenai Peninsula is located in south central Alaska between the Gulf of Alaska and the Cook Inlet (Figure 1-4). The peninsula consists of three regions: Cook Inlet Basin in the west (Kenai Lowlands), Chugach-St. Elias Mountains (Kenai Mountains) through the central portion of the peninsula, and the Gulf of Alaska Coast (Kenai Coast) along the southeast coast (Nowacki et al., 2001). The peninsula is a heterogeneous landscape with primary land cover types consisting of forests and persistent ice/snow. The Kenai Mountains are glaciated in many locations and the topography of the peninsula is typical of a glacially active area. Summer temperatures on the Kenai Peninsula have increased by 0.2 °C per decade over the past 70 years (Figure 1-3).

Figure 1-4: The Kenai Peninsula Alaska. Black lines delaminate the three ecoregions found on the peninsula. The Kenai Coast, the Kenai Mountains, and the Kenai Lowlands. Red stars indicate towns on the peninsula.

RESEARCH QUESTION AND OBJECTIVES

This dissertation addresses gene flow and dispersal process of mountain hemlock at multiple scales (northern range limit and alpine treeline). This research specifically address historic migration processes following Pleistocene glaciation and addresses the extent to which recruits at treeline are derived from local treeline populations or are arriving from more distant seed sources. A genomic analysis based on DNA SNPs was conducted to identify the demographic and genetic variation for mountain hemlock on the Kenai Peninsula in Alaska. The specific research questions are: 1) to what degree have microrefugial patches and long distance dispersal been responsible for the recolonization of the Kenai Peninsula following glacial retreat (Chapter III), 2) What is the primary mode of reproduction in the treeline forming conifer mountain hemlock (Chapter IV) and 3) are mountain hemlock recruits derived from local treeline populations or are they arriving from more distant seed sources (Chapter IV)?

DISSERTATION CHAPTER OUTLINE

This dissertation is organized around three papers. Chapters II-IV each represent a standalone body of research. Each chapter builds off of the previous chapter and represents a complete assessment of mountain hemlock dispersal and gene flow on the Kenai Peninsula. Briefly, the outline for the dissertation is as follows.

Chapter II

Chapter II takes a broad brushed look at plant responses to global ecological change. The chapter appraises how new high-throughput sequencing and associated technologies can be used by "next generation biogeographers" to assess plant responses to global ecological change. First I briefly review the recent advances in genome sequencing and available approaches of genomic analysis. And secondly, I discuss the three main plant responses to global ecological change: migration, adaptation, and phenotypic plasticity. The chapter is designed to place much of the research from chapters III and IV into context.

Chapter III

Chapter III untangles alternative historic dispersal pathways in mountain hemlock, using a landscape genetics approach, to better understand how northern tree species may respond to climatic changes. This chapter address the question to what degree have refugial patches and long distance gene flow been responsible for the colonization of mountain hemlock on the Alaskan Kenai Peninsula.

I used double digest Restriction Associated DNA sequencing (ddRADseq) to identify genetic variants across eight mountain hemlock sample sites on the Kenai Peninsula. I assessed genetic diversity and linkage (gametic phase) disequilibrium.

Alternative historic dispersal pathways were assessed using discriminant analysis of principle components and electrical circuit theory.

Chapter IV

Chapter IV assess seed dispersal and cryptic gene flow at a single alpine treeline ecotone site. This chapter addresses the following two questions: (1.) What is the primary mode of reproduction in the treeline forming conifer mountain hemlock, and (2.) are mountain hemlock recruits derived from local treeline populations or are they arriving from more distant seed sources? I investigate this question using a genomic dataset based on DNA SNPs. First I assess mode of reproduction by determining the proportion of sampled individuals that are the product of clonal reproduction. And second, I use a categorical allocation based parentage analysis to identify parent offspring pairs so that the proportion of treeline reproduction events can be quantified spatially and dispersal distance measured. The fine-scale parentage analysis of a single mountain slope allows for the characterization of dispersal over hundreds of meters, further investigating the degree to which plants can respond to rapid climate change.

Chapter V

Chapter V concludes and summarizes the findings of the dissertation. The analyses presented in this dissertation addressed historic and contemporary dispersal and gene

flow at both regional and local scales. The regional scale analysis relied on the classification of the mountain hemlocks in the south central Alaska region into distinct genetically differentiated populations. Additionally, the fine-scale parentage analysis of a single mountain slope allowed for the characterization of dispersal over hundreds of meters. This research represents a novel combination of genomics and biogeography to answer an important and as yet unresolved question pertinent to understanding how vegetation will respond to climate change.

CHAPTER II

PLANT RESPONSES TO GLOBAL CHANGE: NEXT GENERATION BIOGEOGRAPHY[*]

INTRODUCTION

Ecological biogeography is on the verge of a giant leap forward – if we can take advantage of the opportunities that are becoming available through advances in genetics and bioinformatics. Biogeographers are fundamentally concerned with the spatial and temporal distribution of living organisms, and the processes that structure these observable patterns. Phenotypic variation within and between organisms and across space is controlled through the interaction of genotype and environment; genetic variability that has in the past been difficult to expose. The processes that contribute to genotypic variation are expressed across space and time. Historic and contemporary approaches form the core of modern biogeography. Ecological biogeographers (Cox and Moore, 2010; Blumler et al., 2011) or geographical ecologists (Veblen, 1989) study contemporary physical processes acting on distribution patterns, while historical biogeographers study long-term changes in the distribution of organisms and past processes (Cox and Moore, 2010; Veblen, 1989; Blumler et al., 2011). In the past, genetics has been used to varying degrees to supplement biogeographic research (Parker

[*] Parts of this chapter are reprinted with permission from Johnson *et al*. (2016) "Plant responses to global change: Next generation biogeography" *Physical Geography* 37(2): 93-119. Copyright 2016 Taylor and Francis. http://www.tandfonline.com/doi/full/10.1080/02723646.2016.1162597

et al., 2001; Parker and Jorgenson, 2003; Parker and Markwith, 2007; Cwynar and MacDonald, 1987; Enright et al., 2003; Markwith and Scanlon, 2007a). In fact, a special edition of *Physical Geography* was devoted to highlighting genetic application in biogeography (Rigg, 2003). However, recent advances in genetics, reducing the cost and difficulty associated with extracting genotypic information, suggest that we should revisit this topic.

Global ecological change, the global-scale change in climate, sea-level, landscape heterogeneity and other planetary process, has become a major ecological concern that has the potential to dramatically alter plant distributions and assemblages worldwide. This threat has been driven by human activity and exacerbated by ongoing habitat fragmentation, pollution, and extensive changes in landscape heterogeneity. Within the last 100 years many organisms have spread to higher elevations and latitudes in response to these changes (Chen et al., 2011; Colwell et al., 2008; Walther, 2003; Lenoir et al., 2008; Root et al., 2003), while others have shifted phenological patterns to correspond to changes in the length of the growing season (Parmesan and Yohe, 2003; Menzel and Fabian, 1999). The specific response of plants to global ecological change will depend on their ability to move, evolve, or adjust through plasticity (Reusch and Wood, 2007; Aitken et al., 2008; Hoffmann and Sgro, 2011; Gienapp et al., 2008).

Plants have responded to changing climate in the past, but the current rate of environmental change may outpace the response rate of many species. During the extreme climatic shifts of the Quaternary, plant ranges both expanded and contracted over broad areas (Davis and Shaw, 2001; Hewitt, 1999; Hewitt, 2000; MacDonald et al.,

1993a). Plant species have relied on existing phenotypic plasticity, capable of altering the expression of a given genetic region in response to environmental cues (Nicotra et al., 2010). Case in point, studies of thale cress (*Arabidopsis thaliana*) (Salomé et al., 2011), chicory (*Cichorium intybus*) (Locascio et al., 2009) and sugar beet (*Beta vulgaris*) (Reeves et al., 2007) have all shown well established plastic changes in flowering time in response to shifts in temperature. Some species survive through evolutionary change, relying on existing or new genetic variation that shifts the characteristics of the species, and allows for survival in the new environment. The evolution of the C_4 pathway in angiosperms is an example (Ehleringer et al., 1991). In spite of these cases, current velocities of climate change may prove too fast to allow the same types of responses among contemporary plant species (Loarie et al., 2009; Dobrowski et al., 2013; Corlett and Westcott, 2013; Kremer et al., 2012; Malanson and Cairns, 1997). Current estimates indicate that up to 48% of terrestrial organisms will experience a change in local climate by 2100 (Williams et al., 2007). Additional work has demonstrated an ongoing loss of global biodiversity in response to contemporary climate change (Cardinale et al., 2012; Butchart et al., 2010; Parmesan and Yohe, 2003; Parmesan, 2006; Pereira et al., 2010) Given the scale of these changes, some species may be able to respond through one or a combination of these mechanisms, but many likely will not, and it is important that we further assess how species will respond in the future.

Improvements in population genetic tools have increased the ease through which we can determine a species' potential to follow any of these three avenues. Population genetics investigates spatial and temporal distributions of genetic variation within and

between groups of interbreeding individuals who all share a common gene pool, trying to understand the factors that affect observed genetic variation. Until recently, the main experimental tools used in population genetics were a limited number of molecular genetic markers, specifically the microsatellite. Their use in biogeography has been limited primarily to measuring demographic properties associated with gene flow (Cain et al., 2000; Nathan et al., 2003) and climatic association studies looking at phenotypic traits (Callaway et al., 1994; Chapin et al., 1996). More advanced genomic tools have been widely used in the biological and biomedical fields, however, their use has been out of reach for many biogeographers due to cost, lab availability, and specialized skills needed to conduct these studies. The advent of Next Generation Sequencing (NGS) technology and streamlined bioinformatics analysis is quickly changing this situation.

Next Generation Sequencing technology has reduced the cost of sequencing an organism's genome by 100 fold in the last 15 years, along with a concomitant increase in DNA sequences archived each year (Wetterstrand, 2014; Porter et al., 2012; Benson et al., 2011). This increase in raw sequence data is increasing our potential to answer complex biogeographic questions (Holderegger et al., 2010; Sork et al., 2013; Storfer et al., 2010; Allendorf et al., 2010; Puritz et al., 2014b). NGS can provide access to study both selectively neutral genetic information important for understanding demographic processes as well as adaptive genetic variation important in determining phenotypes (González-Martínez et al., 2011). Moreover, an increase in third party sequencing centers offering library preparation and bioinformatic assistance allows these tools to be used even by novices to genetic analysis (Rossetto and Henry, 2014). In addition to

generating new sequences, the increasing availability of previously published genomes and sequences will allow researchers to incorporate genomic data into their research (Reddy et al., 2015). The combination of genomics and biogeography will create a clearer global picture of the future, and will allow for improved biodiversity conservation and planning measures.

The primary focus of this chapter is to appraise how new high-throughput sequencing and associated technologies can be used by next generation biogeographers to assess plant responses to both contemporary and historic global ecological change. First I briefly review the recent advances in genome sequencing and available approaches of genomic analysis. Secondly I discuss the three main plant responses to global ecological change: range expansion/migration, in situ adaptation/evolution, and phenotypic plasticity. This chapter is intended to spark the interest of biogeographers who do not traditionally use genetics but could benefit from including genetic and genomic tools into their research. The overarching goal is to illustrate that their use can contribute new perspective and add an additional level of resolution into future biogeographic research.

Genetic Tools: Increases in Genetic Resolution

During the last decade Sanger sequencing (Sanger et al., 1977), also known as 1st generation sequencing, has rapidly been replaced by fast high throughput, sequencing technology, commonly termed 2nd generation or Next Generation Sequencing (NGS) (for

reviews of NGS platforms see Mardis, 2008; Glenn, 2011; Schatz et al., 2012; Barba et al., 2014; van Dijk et al., 2014). The development of Sanger sequencing, and its 30 year dominance in genetics, was critical in the development of whole systems genomics allowing for the promise of whole genome sequencing to be realized (Metzker, 2010). Despite its enormous utility, Sanger sequencing is expensive and time consuming and does not easily allow large genomic datasets to be built. Newer iterations of sequencing technology now allow for the investigation of an organism's "-omics": genome, proteome, metabolome, and transcriptome (Losos et al., 2013; Peñuelas et al., 2013; Ellegren, 2014) by allowing parallel high throughput sequencing without the need for the costly and time consuming cloning steps. NGS methods usually generate noticeably shorter genomic fragments (approximately 50- 300 base pair (bp)) compared to Sanger sequencing (1000bp or more) so novel bioinformatic and computational tools are used to assemble the short reads and to analyze the data (Schatz et al., 2012). The benefits of this new technology are its low cost per base pair and high output that allows us to generate enormous sequence datasets across larger numbers of individuals and to untangle a multitude of biological mechanisms (Metzker, 2010).

By and large, the genomic markers developed through NGS are SNPs. SNPs allow genetic variation to be assessed at the nucleotide level and are by far the most common form of genomic variation (Brooks, 2003). Another significant source of genome variation are copy number variants (CNVs) (Beckmann et al., 2007). CNVs and single base differences between individuals can be identified and assessed for their varying

contribution to phenotypic or demographic differences (Seeb et al., 2011; Marroni et al., 2014).

No longer is it necessary to be a molecular biologist with access to a genetics laboratory in order to use these new tools (Rossetto and Henry, 2014). Scientists in cognate fields can now benefit from more affordable efficient genome sequences from commercial sequencing facilities. Linking phenotypes and environments with genotypes and potential genes that are associated with them is now achievable. Because of the ease of generating molecular datasets and the rapidly improving bioinformatic tools available for assembly and analysis, genomics has the capacity to create a bridge between biogeography, ecology, policy, conservation management, and many other diverse fields to facilitate robust interdisciplinary global ecological change research (Hoffmann et al., 2015; Stillman and Armstrong, 2015).

It should be stressed that although the methods discussed below are easy to operationalize, users should still take some time to become familiar with foundational population genetic theory, proper sampling strategies, and, importantly, the assumptions of the bioinformatics and statistical analysis being used to interpret the sequencing output (Andrews and Luikart, 2014).

With respect to plant responses to global ecological change, high-throughput sequencing technology allows ecologists to obtain whole genome DNA sequences from non-model (e.g. organisms without whole genome sequences or developed molecular markers), natural, wild organisms so all biogeographers can take advantage of its power to collect molecular data (Stapley et al., 2010). Full genome sequences of most plants are

not common, but their numbers are growing (Reddy et al., 2015). For example, see the GOLD database (https://gold.jgi.doe.gov/) and Genebank (http://www.ncbi.nlm.nih.gov/genbank/) which are excellent source to assess complete and ongoing genome sequencing projects in many different organisms. Despite the lack of whole genome references for most organisms novel approaches are being used to create partial references *de novo*. Reduced genome representation methods are available to reduce the size of the genome without sacrificing complexity (Davey et al., 2011; Davey et al., 2013). For example, transcriptome sequencing (Van Tassell et al., 2008; McCoy et al., 2014), restriction site-associated DNA discovery (RADseq) (Peterson et al., 2012; Davey and Blaxter, 2010; Miller et al., 2007; Baird et al., 2008), and genotyping by sequencing (GBS) (Elshire et al., 2011; Peterson et al., 2014) all reduce the size of plant genomes for more efficient analysis.

Transcriptome sequencing, more frequently called RNA-seq (Strickler et al., 2012), allows all of the expressed, and thus coding, genes in a tissue sample to be analyzed (Wang et al., 2009). The sequencing of transcripts provides an efficient genome reduction mechanism to study adaptive genomics in non-model organisms. This approach allows specific tissue or organs to be targeted.

RADseq (Miller et al., 2007; Baird et al., 2008), double digest RADseq (ddRADseq) (Peterson et al., 2012), and other emerging RAD methods (Wang et al., 2012; Toonen et al., 2013; Andrews et al., 2014; Puritz et al., 2014b) reduce genome size by targeting all sites associated with one (RAD) or two (ddRAD) restriction enzymes. These sites are randomly spread throughout the genome facilitating a genome wide sampling strategy

for non-model organisms. RAD approaches allow reproducible identification of a large number of SNPs which greatly facilitate population, landscape and evolutionary studies (Baird et al., 2008; González-Martínez et al., 2006; Steiner et al., 2013). Further, the specific choice of restriction enzymes allows one to tailor the density of markers discovered to a particular research question. For example, 200-600 SNPs for a population structure study or 20,000 SNPs for an association mapping study (Peterson et al., 2012).

Genotyping by sequencing is a collection of approaches that creates reduced representation libraries based either on target enrichment using specific hybridization oligonucleotide probes (Dasgupta et al., 2015) or by using one (Elshire et al., 2011) or two restriction enzymes (Poland et al., 2012; Parchman et al., 2012) in a fashion similar to RADseq. GBS protocols have been deployed in parallel to RAD based approaches as a way to minimize sequencing of highly repetitive regions (Elshire et al., 2011; Parchman et al., 2012) and as an additional way to address large complex genomes. These approaches are also cost effective as they remove the random shearing and size selection steps found in RAD (Davey et al., 2011). Both the RAD and GBS approaches can be considered one step genotyping process, where both marker discovery and genotyping occur at the same time (Poland and Rife, 2012; Narum et al., 2013) thus simplifying a traditionally multi-stepped process.

PLANT RESPONSE TO GLOBAL CHANGE

Movement

Understanding the abundance and distribution of species is one of the cornerstones of biogeography (Hall et al., 1992; Hutchinson, 1959; May, 1986; Andrewartha and Birch, 1954). The size and location of a species' range is primarily determined by internal population dynamics and external environmental limitations (Brown et al., 1996; Gaston, 2003). Internal dynamics include population growth rate (Antonovics and Levin, 1980), dispersal ability (del Mar Delgado et al., 2011), available genetic diversity (Gaston, 2003), phenotypic plasticity (Hamrick, 2004; Chevin et al., 2013), and biotic interactions (Mitchell et al., 2006). External limitations include heterogeneity of geographic barriers (i.e. rivers, mountain ranges and man-made structures) and variation in biotic and abiotic environmental factors (such as precipitation, temperature, soil structure, species competition and pollution). Therefore, we would expect species with extensive dispersal ability to have a better chance of long term survival in the face of rapid climatic change. Nevertheless, a species may be influenced by conflicting pressures favoring geographic stability, such as synchronization with photoperiod (Körner and Basler, 2010), mutualistic relationships with other organisms (Inouye, 2008; Miller-Rushing et al., 2010), or intertrophic interactions (Post and Pedersen, 2008).

One possible mechanism for a shift in range area under climate change is expansion at the leading edge (Aitken et al., 2008). Gene flow from range centers to the periphery

is often associated with a reduction in the adaptive potential of individuals at the edge, because the influx of individuals adapted to the center of the range counters the impact of selection for traits suitable to the surrounding environment (e.g. gene swamping) (Kirkpatrick and Barton, 1997; Gaston, 2009; Kubisch et al., 2014; Lenormand, 2002). Conversely, when the conditions at the edge are rapidly changing, as at present, some adjacent areas likely will become increasingly similar to the population center allowing for the establishment of new individuals in an area previously unsuitable (Davis and Shaw, 2001; Alleaume-Benharira et al., 2006). Contraction of range margins, on the other hand, occurs when the landscape within the existing range becomes unsuitable and populations and individuals do not contain sufficient adaptive potential or phenotypic variation to compete for scarce resources or withstand rapidly changing environments and thus decline (Hampe and Petit, 2005; Thomas, 2010).

Plants, being fixed in space, depend on their propagules for movement, either through seed and pollen dispersal, or vegetative spread. The ability for a plant to produce viable propagules and disperse them from their natal location to suitable habitat depends on a diverse set of adaptations (e.g. control over masting), dispersal apparatuses (e.g. development of barbs or winged seeds), and chance (Kokko and López-Sepulcre, 2006; Howe and Smallwood, 1982; Willson and Traveset, 2000). The density of propagules dispersed from a given individual is expected to decrease with increasing geographic distance. Because of this, individuals have a similarly decreasing probability of reproducing with increasing distance between them and thus populations that are in close geographic proximity will be more genetically similar than those far away. This pattern

of genetic relatedness is called isolation by distance (IBD) (Wright, 1943), and is commonly observed in nature (Meirmans, 2012). IBD fits into a 'neutralist' expectation for genetic pattern (Orsini et al., 2013) that assumes uniform dispersal ability and does not account for the limitations the landscape might have on impeding movement, or the selective advantage individuals might have in establishing in a given area if they come from an area already adapted to the environment of the destination. The growing field of landscape genetics and landscape genomics has attempted to address this issue by incorporating geographic toolsets alongside the growing availability of genetic data (Manel et al., 2003; Storfer et al., 2007; Sork and Waits, 2010; Manel and Holderegger, 2013).

Isolation by distance is not the only process that can lead to variation in allele frequencies. Other explanations for dispersal limitation include Isolation by Resistance (McRae, 2006), Isolation by Adaptation (Nosil et al., 2008), Isolation by Environment (Sexton et al., 2014) or Isolation by Colonization (Orsini et al., 2013). McRae (2006) suggested Isolation by Resistance (IBR) as an alternative to IBD. Specifically, barriers to gene flow and dispersal between geographically close populations may lead them to have a greater degree of genetic differentiation than between more distant populations that have no geographic barriers preventing gene flow between them. Also, densities tend to decrease from central populations to the edge. In the context of adaptation and range expansion small population sizes on the edge may suffer from increased genetic drift and gene swamping from asymmetrical gene flow from the range center (Lenormand, 2002; Alleaume-Benharira et al., 2006); potentially reducing adaptive

potential. As dispersal pertains to range size and location, the current, or potential, ability of a species to track its climatic niche through range expansion is ultimately tied to dispersal ability which may potentially evolve in response to the changing environment.

Genomic approaches can begin to address questions of range expansion. For instance, there are generally two scenarios for re-establishment of northern forests following Pleistocene glaciation (sensu Clark et al., 1998a): long-distance dispersal from southern glacial refugia, or in situ survival of small populations within the extent of glaciated regions known as microrefugia (Elias, 2013; Rull, 2009). Population genomic approaches can be used to assess historical gene flow, and thus effective dispersal, by analyzing variation in neutral portions of the genome (see chapter III). The incorporation of selectively neutral highly polymorphic molecular markers, generally microsatellites (SSRs) or amplified length polymorphisms (AFLPs), into parentage and gene flow studies has greatly enhanced our understanding of seed and pollen movement (Krutovsky et al., 2012). There is, however, considerable effort and cost associated with the development of these markers and usually only tens of markers are available to study any given species. High throughput approaches allow the development of 100s to 1000s of informative SNPs, which have been shown to provide equal, if not better, ability to detect kinship and population structure, primarily due to their increased genomic coverage. Several recent studies have shown an increased ability of SNP markers to identify parentage (Anderson and Garza, 2006; Hauser et al., 2011) and kinship patterns (Leslie et al., 2015; Müller et al., 2015) compared to microsatellite or AFLP markers.

Genomic approaches have been used to identify population structure in many forest species (Eckert et al., 2010; Holliday et al., 2010; Poncet et al., 2010; Chhatre et al., 2013; Müller et al., 2015). Population genetic structure is the characterization of a population's gene pool which is described by the geographic distribution of alleles and genotypes found in the population. Population structure can be used to help identify how geographic and landscape features impact gene flow and individual movement. NGS studies can increase the number of samples and loci used to identify population structure providing higher resolution and clarity to the research question. Population genomics approaches can assess the amount of genetic diversity within and between species and individuals, and assess how ecological and geographic structure shape genomic variation (González-Martínez et al., 2011; González-Martínez et al., 2006).

Historic rates of dispersal that have been used to track past range expansion and contraction may not reflect the current ability of plants to track the high velocity of contemporary global ecological change. Therefore, contemporary seed and pollen movement must also be evaluated. This too can be accomplished using genomic approaches. For example, a high resolution dataset of molecular markers can be employed in parentage analysis to quantify effective dispersal distance intergenerationally (see Chapter IV). Parentage studies have been useful for assessing mean dispersal distance in plants and parameterizing dispersal kernels that can be used to model range expansion (Robledo-Arnuncio and García, 2007), identify populations that may become genetically isolated because of landscape fragmentation (Young et al., 1996; Jump and Peñuelas, 2006; Jump and Peñuelas, 2005), and to assess long distance

dispersal (LDD) of plants in response to rapid global ecological change (Sork and Smouse, 2006; Ouborg et al., 1999; Cain et al., 2000). In addition to parentage analysis, other tools are also available to measure historic and contemporary dispersal and gene flow, and should be amenable to NGS datasets. For example using network and graph theory individuals or populations connectivity can be assessed (Dyer and Nason, 2004; Dyer, 2015). In short, individuals or populations are treated as nodes and the connection between nodes can be assessed as genetic distance. Using this framework the degree of graph connectivity can be used to identify how gene flow is occurring on the landscape and test hypotheses related to gene and individual movement (Dyer and Nason, 2004; Dyer, 2007; Dyer et al., 2012; Garroway et al., 2008; Dyer, 2015). Electrical circuit theory is another approach to assess landscape genetic connectivity, where, similarly to graph approaches, electrical nodes are connected by a series of conductors. More numerous or larger conductor connections enhance electrical current flow. Electric circuit theory can be translated into individuals and populations being represented by the nodes and landscape resistance being measured as the degree of connectivity between nodes (McRae, 2006; McRae and Beier, 2007; McRae et al., 2008; Shah and McRae, 2008; McRae et al., 2013). Isolation by resistance can then be compared to alternative hypothesis, such as isolation by distance, to assess how landscape configuration influences genomic variation (Orsini et al., 2013; Sexton et al., 2014; Ruiz-Gonzalez et al., 2015).

Evolutionary Response

Evolution will play an important role in determining how plants will respond to future climates and landscapes. Within a relatively short time-scale, natural selection can quickly shift the mean characters of a population through directional selection, short-term opportunity for population growth, and differential dispersal costs (Bonte et al., 2012; Cheptou et al., 2008; Reznick and Ghalambor, 2001; Carroll et al., 2007). For instance, *Fagus sylvatica* allele frequencies were found to vary rapidly (~40 years) with changes in temperature indicating that long lived forest species can, in some cases, rapidly adapt to climate change (Jump et al., 2006). Similarly, the annual *Brassica rapa* evolved earlier flowering time, in only a few generations, in response to drought conditions, and this change was shown to be heritable (Franks et al., 2007). These two examples illustrate an important discovery in plants with both short and long generation times. Specifically, rapid evolutionary change in plant phenotypes are possible as a function of climate change and shifts in disturbance patterns (Fakheran et al., 2010; Baythavong et al., 2009; Travis et al., 2013). Despite findings of rapid changes in allele frequency and phenotype, it is still largely unknown to what degree plants can respond by evolving. This response may be compounded in plants with long generation times and low gene flow (Aitken et al., 2008) and they may experience migrational lags, that is because of long generation times the ability of the plant to respond to change is not immediate, and an evolutionary response may take many generations (Petit et al., 2008).

Seed dispersal capability can change through evolution, either enhancing or restricting seed movement as a result of changes in the environmental and natural selection. For example, fruiting characteristics may evolve in response to the presence of dispersers allowing for mutualism (Encinas-Viso et al., 2014; Galetti et al., 2013), seed morphology including barbs which allow seeds to attach to animals or modified wings enhancing wind dispersal (Fischer et al., 1996) and evolution of seasonal phenology such as earlier flowering time allowing dispersal to occur earlier (or later) in the year (Franks et al., 2007) can dramatically impact how, if, or when a seed is dispersed. In order for dispersal to respond to changing environments, there must be heritable genetic variation associated with dispersal traits coupled with the environment favoring those traits (Ronce, 2007). Empirical studies in plants have found rapid evolution of adaptive traits by changing seasonal phenology and dispersal strategies. Already mentioned was Franks *et al.* (2007) study demonstrating the evolution of flowering time, but it has also been demonstrated that evolution of plant height can contribute to increased dispersal capability. Thomson *et al.* (2011) conducted a meta-analysis of dispersal distance and found that although seed size was an important factor influencing how far a seed would disperse, plant height was a more powerful predictor of long dispersal distances. In the presence of increasing habitat fragmentation some plants may evolve reduced dispersal ability (Cheptou et al., 2008; Colas et al., 1997; Zohary, 1937) to avoid wasting reproductive output on unsuitable habitats, potentially resulting in evolutionary suicide (Bonte et al., 2012) and evolutionary traps (Wilson, 1959; Mayer, 1942; Mayer, 1954). Evolutionary suicide occurs when fitness decreases occur as a result of reduced dispersal

capabilities. This pattern can arise when the costs of dispersal are greater than the costs of staying put. Colas *et al*. (1997) demonstrated such an evolutionary response in *Centaurea corymbosa*, a cliff dwelling species which has evolved reduced dispersal capabilities and increased inbreeding.

Local adaptation

Within the study of evolutionary responses to global ecological change, adaptive genomics in plants is clearly a ripe area of research. The ability to identify ecologically relevant loci responsible for traits such as growth (Thavamanikumar et al., 2014a; Palle et al., 2013; Beaulieu et al., 2011), drought tolerance (Müller et al., 2012; Homolka et al., 2013), cold hardiness (Chen et al., 2014; Holliday et al., 2010), and seasonal phenology (Holliday et al., 2010; Alberto et al., 2013), in natural populations, and then associate them with changing global conditions is rapidly advancing. Even more importantly, as the number of whole genome sequences in plants increases the prospect of genome re-sequencing, simply sequencing the whole genomes of many individuals to identify large scale genomic differences, becomes viable (Begun et al., 2007; Ellegren, 2014). Past approaches in understanding adaptation in plants relied on quantitative trait loci (QTL). QTL analysis is a method of looking at phenotypic variation in many individuals and identifying the numerous genes contributing to that phenotype (Freeland et al., 2011; Ritland et al., 2011). In general, QTL approaches have been able to identify loci with large phenotypic effect, but have not been as successful at identifying loci with

small phenotypic effect (Stapley et al., 2010; Orr, 2005; Wheeler et al., 2005). For example, Bradshaw and Stettler (1995) found that nearly 50% of genetic variance for stem volume in *Populus* was explained by just two QTLs. Similar findings have been made in Loblolly pine (Kaya et al., 1999). Most phenotypes important to global ecological change adaptation, such as growth, water use efficiency, or drought tolerance, are complex, and though a large portion of their genomic variance may be explained by large effect QTLs, much of the variation is likely additive and the result of many genes in combination that contribute to a very small portion of the phenotypic expression (Neale and Savolainen, 2004; Ritland et al., 2011). NGS and genome-wide association studies (GWAS) provide improved genomic resolution potentially allowing loci with small phenotypic effect to be identified and correlated with environmental variables allowing us to better predict how plants will respond to changing climates though an increased understanding of gene function and adaptive potential.

The implementation of reduced representation strategies such as GBS and RAD approaches facilitates the generation of large genomic datasets without the need for a reference genome (Davey et al., 2013). A growing number of studies are integrating genomics into studies of adaptation in forest species (Sork et al., 2013; González-Martínez et al., 2006; González-Martínez et al., 2011; Ćalić et al., 2015). Outlier analysis can identify specific SNPs that are likely under selection (González-Martínez et al., 2011). Outlier analysis has identified SNPs that are correlated to climate variables (Bashalkhanov et al., 2013; Prunier et al., 2011; Jones et al., 2013) wood traits (Porth et al., 2013; Krutovsky and Neale, 2005), cold hardiness (Eckert et al., 2009a; Namroud et

al., 2008; Holliday et al., 2010; Eckert et al., 2009b), aridity (Eckert et al., 2010; Steane et al., 2014), budset timing (Holliday et al., 2010) and breeding traits (Chhatre et al., 2013; Parchman et al., 2012). These studies demonstrate how high-throughput methods can extend our understanding of plant responses to global ecological change by identifying genomic regions that are likely under selection. The identified regions can then be screened for phenotypes correlating to changing environmental variables. High throughput phenotyping, in conjunction with the identification of adaptive genomic regions will allow us to better assess the degree to which plants can respond to environmental change in place or will need to respond through other mechanisms.

Past approaches used to associate phenotypic variation with genotypic variation in plants relied on using common garden experiments to identify variation in observed phenotypes and then screening the individual's QTL for genetic variation. This approach has worked quite well for clones and selfing species with short generation times. As an example, using *A. thaliana*, QTL have been linked to phenotypes associated with root length (Reymond et al., 2006), biomass (Lisec et al., 2008), whole plant physiology (Juenger et al., 2005), and variation in flowering time (Salomé et al., 2011). However, the resolution of QTL mapping usually does not allow for the identification of particular genes, but rather chromosome regions that are responsible for measured phenotypic variation (unless candidate genes are used in the mapping (e.g. Wheeler et al., 2005)). QTL mapping has been difficult to use in long lived species, such as forest trees, because of the need for long-term data collection strategies and additional expenses associated with maintaining long term experiments (Sork et al., 2013; Neale and Kremer, 2011).

Because of the need to better understand genetic control of phenotypic variation in a large number of plant populations, a move from QTL studies to association studies, or GWAS is occurring. GWAS is used to identify associations between phenotypes and genetic variation at the genome-wide level in natural and experimental populations (Ingvarsson and Street, 2011; Sork et al., 2013; Nordborg and Weigel, 2008). The identified SNPs can then be screened against ecological and climate variables and mapped to identify genomic regions contributing to phenotypic variation. NGS sequencing combined with GWAS has identified associations between genotype and phenotype such as changes in shade tolerance phenotypes (Filiault and Maloof, 2012) variation in root phenotype associated with nutrient availability (Gifford et al., 2013) and xylem development in *Pinus taeda* (Palle et al., 2013).

In plant breeding studies, high throughput phenotyping has been suggested as a way to quickly collect phenotypic data non-invasively for multiple traits and multiple individuals (Cabrera-Bosquet et al., 2012). This technology includes a variety of remote sensing platforms such as spectroscopy, LiDAR, and use of infrared cameras that can be used to collect high throughput phenotype data on traits such as canopy height, biomass, dry matter content, leaf area, water content, photosynthetic output, nitrogen content and much more (Montes et al., 2007; Finkel, 2009; Cabrera-Bosquet et al., 2012; Fahlgren et al., 2015). The phenotype information can then be screened for statistical associations with genetic variation. In this way, phenotype by genotype interactions may more easily be detected.

Recently a GWAS was combined with high throughput phenotyping in rice to demonstrate the utility of combining the two high throughput approaches (Yang et al., 2014). Yang et al. (2014) developed an automated phenotyping platform that used a color imaging device and linear X-ray to collect phenotype data: including plant height, green leaf area, plant compactness and fresh and dry shoot weight for every individual plant. GWAS was conducted on a large number of individuals and screened against the collected phenotypes to identify associations. They then compared the automated phenotyping to manual phenotyping methods and showed that the automated method provided better results and allowed for genomic architecture to be dissected more completely. Though this study was conducted in a controlled laboratory setting, there are future applications based on this technology that may allow large scale phenotyping and genotyping to be conducted in natural environments.

Transcriptomics

Transcriptomics studies gene expression by capturing the complete set of RNA transcripts (mRNA, tRNA, rRNA, siRNA, miRNA, piRNA, shRNA, and sRNA) produced by the genome, sequencing them using NGS (RNA-seq) and comparing their presence in cells of different tissues, after different treatments, or at a separate life stages (Wang et al., 2009). The transcriptomes are much smaller than entire genomes, therefore, their sequencing is more feasible and reduces the amount of DNA examined,

but has a greater likelihood of identifying genetic regions associated with phenotypic traits.

A range of plant studies have used RNA-seq to investigate phenotyipic variation and adaptive potential, such as wood traits (Beaulieu et al., 2011; Thavamanikumar et al., 2014b), temperature response (Chen et al., 2014; Holliday et al., 2008), growth (Raj Kumar Kullan et al., 2012; Geraldes et al., 2011), and phenology (Uddenberg et al., 2013). Most of these studies have been conducted in commercially important and model plant systems, but NGS will allow non-model and ecologically important systems to be studied quite easily (Neale and Kremer, 2011). The added benefit of transcriptomics in global ecological change studies includes the ability to target expressed portions of the genome in specific plant tissue allowing adaptive changes to be identified.

Studies of speciation between closely related plants have begun to incorporate genomic approaches (Payseur, 2010; Stölting et al., 2013; Wang et al., 2013). Next generation phylogenomics incorporates high resolution datasets to construct phylogenies that can inform our understanding of past evolutionary events and help assess the velocity of past migration and re-colonization of plants following glacial retreat and vicariance events (Burke et al., 2014). Within phylogenetics, because of the need to target many loci from many individuals of a focal, usually non-model, organism a reduced representation approach is likely to prove beneficial, especially as methods of selecting homologous regions and increased sequencer read length improves (McCormack et al., 2013; McCormack and Faircloth, 2013). Reduced representation approaches have been use to infer phylogenetic relationships. For example RAD

approaches (Emerson et al., 2010) and RNA-seq (Roeding et al., 2009; Weitemier et al., 2014; Hyun et al., 2014) have been used to construct phylogenetic trees with little prior genomic information. Moreover, the use of Ultra Conserved Genomic Elements (UCEs) and Conserved Ortholog Sets (COS) is allowing comparable genomic regions to be analyzed across species and across millions of years of evolutionary history (Faircloth et al., 2012; Krutovsky et al., 2006). Challenges, of course, still exist for phylogenomic studies. The ability to identify homologous regions is critical, and the need for methods to identify highly informative multilocus information, sometimes not available with SNPs, needs to be improved (McCormack et al., 2013).

Phenotypic Plasticity

Evolution, even when it occurs on ecological time scales, as some of the examples above illustrated, still requires changes in allele frequencies within a population over a generation or more. In contrast to evolution of novel traits, many species have a range of phenotypic traits, tolerances and response that can easily vary in responses to various climatic or ecological changes. It is called phenotypic plasticity, which is common in plants, allowing them to alter their response to global environmental change within their own lifespan, at least to a degree (Jump and Peñuelas, 2005; DeWitt et al., 1998; DeWitt and Scheiner, 2004; Kramer, 1995). In its most basic sense, phenotypic plasticity allows an individual genotype to tolerate broad variation in biotic and abiotic conditions (Ghalambor et al., 2007) by expressing a range of characters (Franks et al., 2014). For

example, temperature has been linked to variation in flowering time (Franks and Weis, 2008; Anderson et al., 2012), bud burst (Caffarra and Donnelly, 2011), and leaf phenology (Morin et al., 2009; Polgar and Primack, 2011), and water availability has been shown to influence plastic dispersal traits, such as seed size and plant height (Teller et al., 2014). Epigenetics plays the most important role in phenotypic plasticity. Epigenomics helps us to understand plant epigenetic landscapes and epigenetically regulated traits, such as flowering time. Additionally, epigenomics provides us with modern methods and techniques to study epigentic phenomenon, such as genomic imprinting, histone controlled gene silencing and activation, and DNA methylation (Spillane and McKeown, 2015). As the climate and environment change, the existence of these underlying tolerances may allow many plant species to maintain their current ranges, or occupy new territory acting as a buffer against global ecological change while providing an avenue for selection and future evolution.

CONCLUSION

Future studies in biogeography, facilitated by the ease of commercial sequencing services and low costs, will greatly enhance our understanding of how plants will respond to rapid global ecological change. This paper examined how plants are responding to rapid global ecological change through range expansion, adaptation and phenotypic plasticity, focusing on how new genomic approaches, specifically NGS technologies, can further our understanding of these primary plant responses. I examined

how genome wide high density molecular markers (SNPs) improve our ability to classify populations and identify kinship in order to better understand future range movement and migration. I suggest that NGS approaches can be used to enhance the resolution of marker loci and allow higher sample numbers to be included in phylogenetic studies that will help us understand past plant response. I demonstrated that association mapping and outlier analysis of genomic regions can help identify ecologically relevant loci that can provide insight into contemporary local adaptation potential. Additionally, reduced representation genomics can serve as a powerful method of genomic analysis of non-model species with large complex genomes that lack references genomes. I suggest that new high throughput phenotyping in conjunction with NGS sequencing may further improve our understanding of phenotypic plasticity. These methods in conjunction with improved data sharing and archiving will improve our ability to answer next generation biogeographic questions. Genomics should not replace field and experimental biogeographic investigation, but should instead be integrated into our toolbox of techniques that help biogeographers link pattern and process to improve our understanding of the world and the impacts we have on it.

CHAPTER III

HOLOCENE MIGRATION PROCESSES OF MOUNTAIN HEMLOCK (TSUGA MERTENSIANA) PROVIDE NEW INSIGHTS INTO THE MIGRATION POTENTIAL OF LONG LIVED TREE SPECIES IN RESPONES TO CLIMATE CHANGE: A GENOMICS PERSPECTIVE

INTRODUCTION

Quaternary glaciations are principally responsible for structuring the present day landscape at northern latitudes (Provan and Bennett, 2008). In many cases, advancing glaciers forced species to retreat into southern portions of their range where climates were more suitable (Webb, 1987), or reduced distribution locally into microrefugia, where climate remained within the species tolerance sustaining small populations while the remaining habitat was depopulated (Rull, 2009). As glaciers have receded, species have slowly recolonized their previous distributional range (Davis, 1981; Bernabo and Webb III, 1977). Although, recolonization from glacial macro-refugia in southern ranges or out of northern microrefugia can both lead to identical contemporary species distributions, the result of each migratory pathway predict vastly different potential short and long term futures for species in northern habitats. For instance, LDD can lead to a genetic bottleneck, decreasing the standing genetic variation in expanding populations (Hewitt, 2000). In the face of ongoing climatic change this could limit the ability of

northern species to adapt to shifting local environments. Nevertheless, disentangling alternative historic dispersal pathways of northern species is often difficult.

Emerging genomic technology and molecular approaches have improved the ability to assess past migratory pathways in natural systems (Johnson et al., 2016). In the past, even when species were distributed in geographically isolated patches, suggesting pre-glaciation refugia, the ability of rare LDD to establish remote populations could not be discounted. The importance of LDD as a mechanism explaining rapid recolonization of forest species post glacial retreat (Sensu Clark et al., 1998a) has been an active area of research (Cwynar and MacDonald, 1987; Hewitt, 1996; Clark et al., 1998a), but difficulties in tracking highly dispersed organisms (e.g. seed and pollen) on the landscape have made a detailed understanding of LDD difficult (Cain et al., 2000). Current high-throughput next generation sequencing (NGS) technologies have the capacity to increase our understanding of dispersal processes by identifying a much higher number of polymorphic genomic sites that can be used as informative genetic markers, most commonly SNPs. This, along with improved species distribution modeling (SDM), allows us to more effectively assess past dispersal and gene flow pathways in natural systems.

Patches established by rare LDD have several genetic characteristics that would distinguish them from those that are microrefugial remnants of past populations (Nichols and Hewitt, 1994; McLachlan et al., 2005). First, the establishment of a population via a single LDD event would reduce genetic diversity (Hewitt, 1996; Hewitt, 2000; Ibrahim et al., 1995) and likely increase linkage (gametic phase) disequilibrium (LD) (Excoffier

et al., 2009) due to founders effect, as has been seen in North American populations of lodgepole pine (*Pinus contorta*) (Cwynar and MacDonald, 1987) and sugar maple (Acer saccharum) (Vargas-Rodriguez et al., 2015). Alternatively, microrefugial patches that weathered past glaciations would contain equivalent or higher genetic diversity and lower LD compared to those established by LDD or distributed along the expanding edge of a population (Hewitt, 2000; Petit et al., 2003; Flint-Garcia et al., 2003). Second, microrefugial patches are more likely to be genetically distinct from the rest of their range due to genetic drift and divergent selection driving differentiation among localities (Anderson et al., 2006; Opgenoorth et al., 2010; Mee and Moore, 2014). For example, Anderson (2006) found high genetic diversity and high population genetic structure at microrefugial populations of white spruce (*Picea glauca*) along their colonizing front in Alaska. This theoretical framework provides a basis on which to differentiate isolated populations that were founded by more recent LDD from those that are remnants of past vegetation structure.

New toolsets incorporating electrical circuit theory have provided an additional methodological approach to test alternative dispersal models (McRae et al., 2008; Shah and McRae, 2008). Under this approach, all potential pathways between two locations are examined relative to an underlying raster "resistance surface" that represents the permeability of any location based on some hypothesized restrictive variable, such as a climate niche. This generates a pairwise resistance value for each site pair that measures not just the distance between each location, but incorporates the ease with which a species might traverse a landscape via multiple routes (McRae and Beier, 2007). This

approach allows hypothesis to be tested about the influence of alternative resistance surfaces on propagule movement, by comparing a simple null isolation-by-distance (IBD) model, where dispersal is limited by the Euclidean distance between locations, to each alternative isolation-by-resistance (IBR) model. Recent work in landscape genetics has demonstrated the utility of this approach. In studies of animals and plants, circuit theory provided a clearer picture of movement across a landscape than did Euclidian distance or least cost path analysis alone (McRae and Beier, 2007; Fant et al., 2014; Amos et al., 2014; Andrew et al., 2012).

Mountain hemlock (*Tsuga mertensiana* Bong. Carr.) is a species thought to have recently colonized the Alaskan Kenai Peninsula. It has been suggested that the species has migrated north from a single southern glacial refugia near Vancover Island BC (Roberts and Hamann, 2015; Ally et al., 2000) along the pacific coast. Pollen and macro-fossil analysis indicate that this species has expanded its range within the last 14,000 years (Jones et al., 2009) with an estimated arrival on the Kenai between 6700 and 1500 ybp (Cwynar, 1990; Jones et al., 2009). However, geographically isolated populations of this species exist in the Kenai lowlands of the peninsula suggesting that isolated microrefugial populations may have weathered past glaciation and later contributed to subsequent recolonization. Alternatively, these isolated patches may have established via more recent long distance seed dispersal as the Kenai Peninsula has been recolonized. Mountain hemlock has both wind dispersed seeds and pollen opening potential for long distance colonization and gene flow. Differentiating between these alternative pathways presents an opportunity to understand how northern tree species respond to sweeping

climatic changes and estimate the dispersal potential for this and similar species, providing essential information for both researchers and land managers in this area.

Objective

This research is aimed at addressing the following question: To what degree have microrefugial patches and long distance dispersal been responsible for the recolonization of the Kenai Peninsula post glacial retreat? I investigated this question by using two approaches to evaluate genomic data collected using a relatively new genotyping-by-sequencing (GBS) approach based on double-digest restriction associated DNA sequencing (ddRADseq) (Peterson et al., 2012). First I compare the genetic pattern of a geographically isolated patch of the study species in the Kenai lowlands to other patches throughout the Kenai Peninsula. If lowland patches of mountain hemlock represent microrefugia, they are expected be in linkage equilibrium, have higher genetic diversity, and equivalent or higher pairwise genetic difference than other populations in the sampling area. Second, I evaluate the degree to which connectivity within the Kenai Peninsula has been structured by past glacial climates or more contemporary environments. If populations are remnants of pre-glacial environments, I would expect past climates to better explain contemporary genetic differences among sampling locations. This study will help determine the past history of this region and help predict the potential for future response in this area.

MATERIALS AND METHODS

Study Species – Mountain Hemlock

Mountain hemlock is a highly outcrossed, monoecious, wind pollinated species with large winged seeds and pollen (Owens and Molder, 1975; Means, 1990; Ally et al., 2000). On the Kenai Peninsula the tree species is found in cool wet environments, including sloping peatlands with poorly drained soils, along the Kenai coastal, alpine, and subapline zones. The species range extends from its northern limit on the Kenai Peninsula in south-central Alaska to its southern range along the Pacific coast of the Olympic and Cascade Mountains (Peterson and Peterson, 2001), with southern populations remaining a component of forest structure as far south as the Sierras in northern California (Taylor, 1995; Means, 1990). On the Kenai Peninsula, Alaska mountain hemlock is part of the 'spruce hemlock zone' (Miller and Walton, 2014) consisting of Sitka Spruce (*Picea sitchensis*) on the coast and white spruce (*Picea glauca*) in the Kenai Mountains.

Mountain hemlock stand expansion and migration is related to length of growing season, a function of winter temperature and snow pack, and summer temperatures and moisture availability (Peterson and Peterson, 2001; Taylor, 1995). High elevation mountain hemlock growth correlate negatively to spring snowpack depth and positively to summer growing season temperature (Taylor, 1995; Peterson and Peterson, 2001).

Additionally, warm July temperatures result in increased seed production (Woodward et al., 1994).

Study Area - Kenai Peninsula, Alaska

The Kenai Peninsula, AK is a biologically diverse ecosystem covering 2.1 million ha in South-central Alaska. The peninsula contains three distinct ecoregions: Cook Inlet Basin in the west, Chugach-St. Elias Mountains through the central portion of the peninsula, and the Gulf of Alaska Coast along the southeast coast (Nowacki et al., 2001). Following Boucher and Mead (2006) I will refer to these ecoregions based on their landform characteristics: Kenai Lowlands, Kenai Mountains, and Kenai Coast respectively. The primary land cover types are forests and persistent ice/snow. The Kenai Lowlands is typical of interior boreal forests in vegetation composition, dominated by white spruce, black spruce (*Picea mariana*), paper birch (*Betula papyrifera*), and black cottonwood (*Populus trichocarpa*). The upper soil layer consists of glacial loess and discontinuous ash (Karlstrom, 1964). The Kenai Mountains are composed of white spruce and mountain hemlock, where the latter dominates the alpine treeline (800 masl). The Kenai Coast region of the peninsula is a narrow ecoregion along the Gulf of Alaska, which contains mountain hemlock and sitka spruce forests from sea level to alpine treeline. Much of the Kenai Mountains and coast are glaciated and the topography of the peninsula is typical of a glacially structured landscape with U-shaped valleys and many poorly drained wetlands and small lakes (Boucher and Mead, 2006). The central portion

of the peninsula is covered by the Harding Ice Field. Both the Kenai Mountains and Kenai Coast, where un-glaciated, consist of glacial till and colluvium (Boucher and Mead, 2006).

The present climate is boreal maritime with both temperature and precipitation gradients from east to west. Nearly all of the Kenai was covered by the late Wisconsin Cordilleran ice sheet approximately 26,000-14,500 ybp (Rymer and Sims, 1982; Ager, 2007). A few microrefugia have been proposed to have harbored species during this glaciation in the northwest Kenai mountains and the eastern Kenai lowlands (Jones et al., 2009). Though there is no evidence of conifer species being present in these purported microrefugia during past glaciation, their survival there cannot currently be ruled out.

Sampling

Mountain hemlock individuals were sampled across 8 different sites covering the three ecoregions of the Kenai Peninsula during the summers of 2012 and 2013 (Fig. 3-1; Table 3-1). Within the coastal ecoregion, I collected from three sites spaced between 35-50 km apart (n = 30). Within the mountain ecoregion, I collected from three sites spaced 30-48 km apart (n = 28).

Within the Kenai lowlands, I collected from one isolated stand of mountain hemlock (*n* = 10) identified by Berg (2003) in the Kenai National Wildlife Refuge (NWR) spaced 46-150 km from the other sampling sites. This isolated site, called Discovery Well, was compared to other sampling sites in subsequent analyses to examine for characteristics of microrefugia, or LDD establishment. In addition to the sites collected from the three Kenai ecoregions, I sampled one site north of the peninsula, near Anchorage (*n* = 9), which was 73 km from Discovery Well and spaced 40-188 km from the other sampling sites. The Anchorage site falls into the Chugach-St. Elias Mountains ecoregion, but is treated as a separate ecoregion in this study. At each location, I collected foliage (approximately 15 cm long branch tips with needles) for DNA extraction. The fresh needle tissue was desiccated in silica gel (Colpaert et al., 2005) and placed in a -20 °C freezer to await DNA extraction and sequencing. The location of each sample was recorded using a handheld GPS.

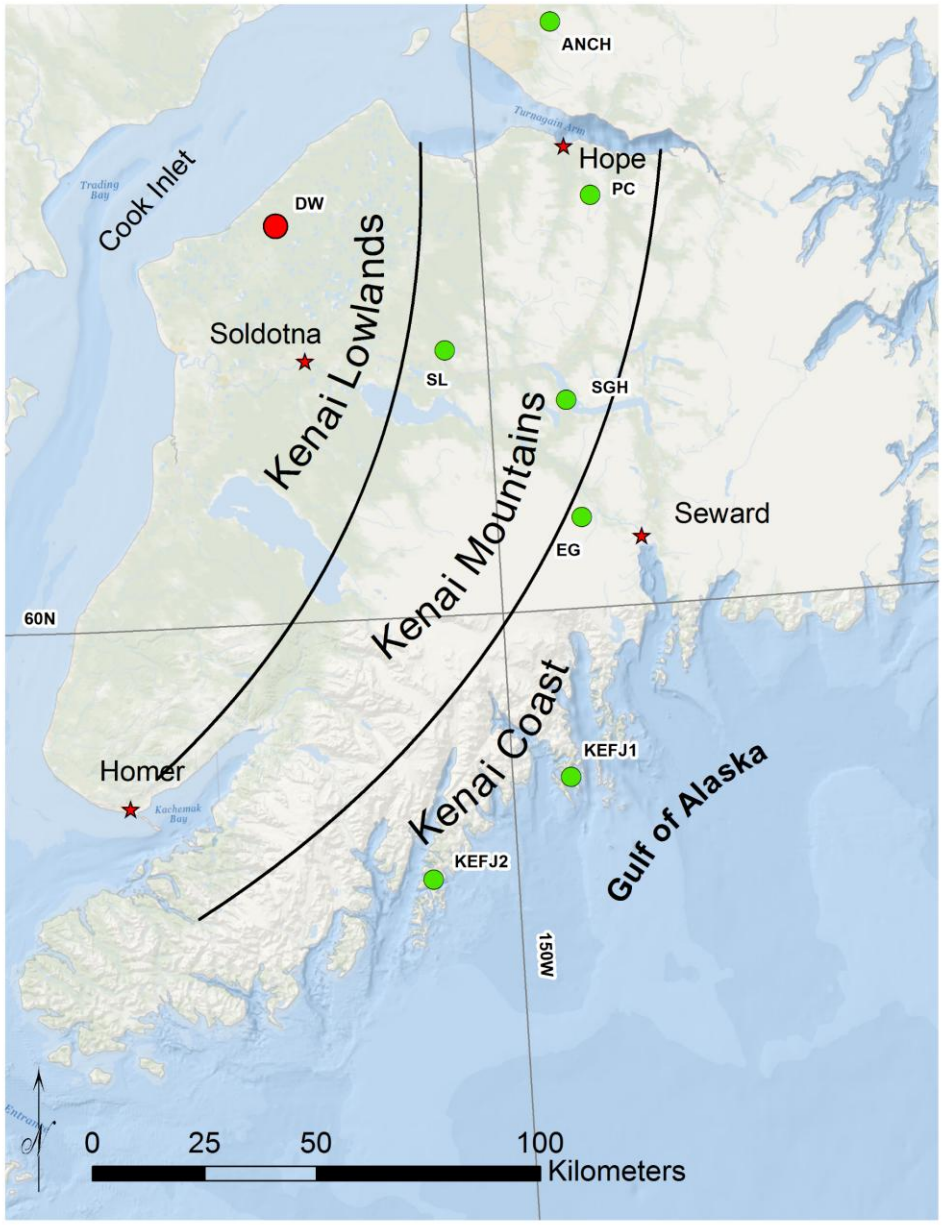

Figure 3-1: Mountain hemlock sample locations (green colored circles) on the Kenai Peninsula, AK. The Discovery Well (red colored circle) is a geographically isolated stand of hemlock. The Kenai Peninsula consists of three ecoregions: Kenai Coast, Kenai Mountains, and the Kenai Lowlands.

Table 3-1 Mountain hemlock sampled sites, abbreviation code, ecoregion, the number of individuals sequenced, Nei's unbiased average expected heterozygosity, H_e, and the standardized index of association, $\bar{r}_d$.

Sample Location	Code	Ecoregion	number sequenced	H_e	$\bar{r}_d$
Anchorage Bowl	ANCH	Anchorage Bowl	9	0.447	0.005
Discovery Well	DW	Kenai Lowlands	10	0.432	0.002
Exit Glacier	EG	Kenai Coast	10	0.460	0.011
Kenai Fjords NP	KEFJ1	Kenai Coast	10	0.465	0.006
Kenai Fjords NP	KEFJ2	Kenai Coast	10	0.450	0.021
Palmer Creek	PC	Kenai Mountains	10	0.440	0.011
Snug Harbor	SGH	Kenai Mountains	8	0.469	0.006
Sky Line	SL	Kenai Mountains	10	0.440	0.008

H_e was measured as a summary of gene diversity; $\bar{r}_d$ was measured as a summary of multilocus linkage disequilibrium (LD)

Genomic Marker Development

I used a GBS approach, ddRADseq (Peterson et al. 2012), to genotype the mountain hemlock individuals. Tissue grinding and DNA extraction of sampled mountain hemlock needles were performed using a modified CTAB protocol (Doyle and Doyle, 1987). I tested different restriction enzyme (RE) pairs using frequent cutter enzymes targeting~4bp long restriction sites (*Msp*I, *Mlu*CI, *Nla*III) in combination with a less frequent cutter enzyme targeting longer (~6bp) restriction sites that are less frequent in a genome (*Sph*I, and *Eco*RI). Three samples (one from each ecoregion) were digested with

five RE pairs (*Sbf*I-*Eco*RI, *Sph*I-*Eco*RI, *Eco*RI-*Msp*I, *Sph*I-*Mlu*CI, and *NIa*III-*Mlu*CI) and fragments were assessed on an Agilent Bioanalyzer (Agilent Technologies) to tune the fraction of the genome sampled to approximately 1%. Based on the bioanalyzer reports I have chosen *Sph*I-*Mlu*CI with a fragment size of 350bp for paired-end (PE) sequencing using the Illumina HiSeq 2000 sequencer (Illumina, Inc.). I generated individually barcoded PE sequencing libraries with ~ 350 bp long inserts from 77 mountain hemlock samples and sequenced them with 150 × 2 cycles in a single HiSeq 2000 lane. This allowed us to simultaneously sequence and identify variants across all individuals using the selected RE pair and 350bp fragments. All library preparation and sequencing were conducted at the University of Texas Genomic Analysis and Sequencing Facility.

For sequence analysis I used standard GBS bioinformatics analysis performed by the Texas A&M Institute for Genome Sciences and Society. This analysis assessed read quality, de-multiplexed samples, *de novo* contig assembly, aligned reads onto contigs, and called variant (SNPs) using the dDocent pipeline (Puritz et al., 2014a). This research used *de novo* assembly strategies for a non-model organism. Quality SNP filtering was accomplished using VCFtools (Danecek et al., 2011). I used a 10x coverage cutoff, and Phred quality score > 30. I filtered the SNPs to remove possible sequencing error and duplicate paralogous sequences. I removed SNPs that failed to genotype in greater than 80% of individuals and SNPs with a minor allele frequency less than 5%, which are difficult to distinguish from sequencing errors (Roesti et al., 2012). The identified SNPs were screened for outliers using Bayescan (Foll and Gaggiotti, 2008) to ensure that loci

remaining for further analysis were selectively neutral. Individuals missing greater than 15% of the identified SNPs were also removed.

Genetic Analysis

I examined the level of genetic variation and pattern of genetic diversity within each of our sampling locations. I estimated genetic diversity using Nei's unbiased average expected heterozygosity, H_e, as a measure of gene diversity (Nei, 1978), and the standardized index of association, $\bar{r}_d$, (Agapow and Burt, 2001) as a summary of multilocus LD. $\bar{r}_d$ is a modification of I_A, the original index of association developed by Brown *et al.* (1980), that is robust to variation in number of loci sampled. Hardy-Weinberg equilibrium was assessed using *pegas* (Paradis, 2010) and its significance was tested by permutation via 1000 simulations. All diversity values were calculated using the *R* package *poppr* (Kamvar et al., 2015; Kamvar et al., 2014).

I examined the amount of connectivity and partitioning of genetic variance among the sampling sites and the three ecoregions plus the Anchorage site. To determine the distribution of genetic variation among sampling sites I performed a hierarchical AMOVA. To determine the level of gene flow and genetic differences between sampling sites I used unbiased pairwise estimates of population differentiation (F_{ST}) (Weir and Cockerham, 1984). The results of the AMOVA were tested for significance using a randomization test (Excoffier et al., 1992). This analysis and pairwise genetic differences (F_{ST}) between the eight sample sites were all calculated in GenAlEx (Peakall

and Smouse, 2006). I used Discriminant Analysis of Principal Components (DAPC) (Jombart et al., 2010) to assess population subdivision and clustering, which is robust to violations of population subdivision rules and uncertain migration models that are required in other packages. This method summarizes the between population genetic variation while minimizing the within population variation by transforming multi-locus genotype information, using principal components analysis, into uncorrelated variables as an input into discriminant analysis. This approach is useful with extremely large genomic datasets made up of thousands of SNPs. I used a K-means clustering algorithm in the R package *adegenet* (Jombart et al., 2010) which allows the optimum number of population clusters to be assessed quantitatively using cross-validation.

Landscape Variables

I used maximum entropy modeling to create two SDM representing the contemporary and historic (Mid Holocene) climatic habitat suitability for mountain hemlock on the Kenai Peninsula. Two hundred and twenty five mountain hemlock observations were obtained from the Global Biodiversity Information Facility (GBIF) (www.gbif.org) and the Biodiversity Information Serving Our Nation (BISON) (www.bison.usgs.ornl.gov) databases. I used only high quality herbarium data. I removed records that had uncertainly associated with georeferencing, duplicates, and non-wild individuals. Contemporary (Hijmans et al., 2005) and historic (Community Climate System Model, Gent et al., 2011) bioclimatic variables were obtained from the Worldclim database

(www.worldclim.com) at approximately 1 km spatial resolution. The Woldclim database consists of 19 bioclimatic variables that have been interpolated globally. For example, the 19 bioclim layers include measures of annual mean temperature and maximum temperature of warmest month. I ran the computer program Maxent (Phillips et al., 2006), a freely available software, using all 19 bioclimatic variables to estimate the contemporary habitat suitability for mountain hemlock based on species occurrences. This approach allows me to model the climatic niche of mountain hemlock and thus identify the probability of their contemporary and historic geographic distribution. This approach is robust to irregularly sampled data such as the herbarium records I used (Phillips et al., 2009). I ran the models using linear, quadratic and product features. I used the contemporary model to estimate the habitat suitability of mountain hemlock in the mid-Holocene by projecting the predicted climate envelope based on the historic bioclimatic variables. I parameterized the model with 15 replicates and 5000 iterations with 75% of data used to calibrate the model and 25% percent withheld to test the models performance. The model was evaluated by assessing the area under the receiver operator curve (AUC) (Fielding and Bell, 1997). The output file from Maxent contained predicted occurrence values for every pixel in the study area that vary between 0 and 1 for low and high likelihoods, respectively.

Raster resistance surfaces were created from the habitat suitability maps generated in Maxent by inverting habitat suitability into resistance cost. Thus, habitats with high probability of suitability had low resistance to movement, while habitats with low probability of suitability had high resistance to movement based on the species climate

niche. Landscape connectivity was then assessed using a graph approach based on electrical circuit theory implemented in the program Circuitscape (McRae, 2006; Shah and McRae, 2008). Circuitscape uses electrical circuit theory to convert a raster of resistance values into a graph and calculates multiple pathways of least resistance by weighting landscape variables by the probability that they will facilitate or inhibit movement. Graphs were composed of nodes of sampled mountain hemlock patches with edges composed of all possible connections across the landscape.

Statistical Analysis

I tested the predictions of genetic characteristics of alternative population histories by comparing the genetic measures among sampling locations. I compared the mean values of H_e, and $\bar{r}_d$ within the Discovery Well sample location, to the rest of the sample locations across the Kenai Peninsula using a student's *t*-test. To determine if the Discovery Well has been historically more separated from the rest of the region, I compared the pairwise values of F_{ST} of the Discovery Well to all sites, with the pairwise F_{ST} values of all other sites among themselves again using a student's *t*-test. Additionally, I regressed H_e and $\bar{r}_d$ against distance from the proposed colonizing front of mountain hemlock, near Prince William Sound, to determine if there was a trend in genetic diversity.

I tested the alternate models of landscape limitations on dispersal, IBD and both contemporary and historic IBR, by comparing pairwise geographic distance to pairwise

F_{ST} (IBD) and pairwise resistance to pairwise F_{ST} (IBR) using a simple Mantel's test in PASSaGE (Rosenberg and Anderson, 2011). This method tests the null hypothesis of no difference between two symmetrical matrices, in this case distance matrices (McCune and Grace, 2002; Mantel, 1967). Finally, I used partial Mantel tests to look at IBR by comparing pairwise contemporary resistance to pairwise F_{ST} after controlling for geographic distance. Each test was permutated 1000 times.

RESULTS

Genomic Marker Development

From 77 individuals GBS resulted in 41,057,267 unique sequencing reads and 50,952 assembled contigs with an average per nucleotide read depth greater than 30x. 171,019 putative SNPs were identified throughout the mountain hemlock genome using the dDocent (Puritz et al., 2014a) pipeline. Quality filtering reduced the number of SNPs to 6124 loci using VCFtools (Danecek et al., 2011). All 6124 SNPs were found to be selectively neutral based on outlier F_{ST} analysis in Bayescan (Foll and Gaggiotti, 2008). No individuals had greater than 15% missing genotype information.

Genomic Diversity

The Discovery Well site was genetically distinct relative to other sampling sites, with a conflicting genetic pattern in terms of identifying an origin by LDD or through microrefugia (table 3-2). Discovery Well genetic diversity was statistically lower than the other seven sampling locations (P = 0.005, 95% CI 0.434–0.467) as would be expected under LDD. H_e within Discovery Well was equal to 0.432 and varied from 0.438 to 0.468 in the other sample sites, with a mean H_e across all samples of 0.449.

Table 3-2 Nei's unbiased average expected heterozygosity, H_e, and the standardized index of association, $\bar{r}_d$.

	He	r_d
DW Value	0.4324	0.0021
t statistic	3.9148	3.2458
P value	**0.0057**	**0.0141**
95% CI	0.4337-0.4671	0.0013-0.0161
mean	0.4504	0.0087

Student's t-tests were used to assess whether the isolated Discovery Well site is a glacial relict or the product of long distance dispersal. H_e was measured as a summary of gene diversity; $\bar{r}_d$ was measured as a summary of multilocus linkage disequilibrium (LD). In all cases, DW was distinct compared to the rest of the sampled sites (P <0.05)

Though the magnitude of difference in H_e is low among all sample sites, a comparison between all other sample sites showed no other site had statistically lower H_e, and only SH and KEFJ1 had statistically higher H_e.

All sites were in LD to some degree. However, at the Discovery Well, LD was statistically lower than the other sites (P = 0.014, 95% CI 0.001-0.016), which is indicative of the expectations I presented for a microrefugia. The standardized index of association, $\bar{r}_d$, within Discovery Well was equal to 0.002 and varied from 0.005 to 0.021 in the other sites, with a mean of 0.009 (table 3-2). Again, the magnitude of the signal was low across all sample sites.

The hierarchical AMOVA found no significant difference among regions (F_{RT} = 0.003, P = 0.372) or sample sites (F_{ST} = 0.005, P = 0.328). Pairwise F_{ST} between all sites varied from 0.007 to 0.018 (table 3-3). Student's t-tests found that the Discovery Well sample location had significantly higher F_{ST} values than those found among all other sites (P = 0.019) (table 3-4). Pairwise F_{ST} between the Discovery Well and all other sample locations varied between 0.012 and 0.017, whereas pairwise F_{ST} between all sampling locations except the Discovery Well varied between 0.009 and 0.015. This indicates that the Discovery Well has been generally less connected relative to other sample locations, however F_{ST} values were still within the observed range found in other samples sites. Although this result aligns with the predictions for a microrefugial site, the pattern observed here is weak, especially considering the distance between the Discovery Well and the rest of the sampling locations. F_{ST} values were lowest between the two Kenai Fjords sites (KEFJ1 and KEFJ2) (F_{ST} = 0.0096) and highest between the

Kenai Mountain site SGH and the DW in the Kenai Lowlands (F_{ST} = 0.0165) which was consistent with a weak signal of dispersal limitation.

When I regressed H_e and $\bar{r}_d$ against distance from the predicted colonizing front near Prince William Sound, I found that H_e decreased significantly with increasing distance (P = 0.022, R^2 = 0.6091 Fig. 3-2). In general the lowest H_e occurred in the Northwest Kenai Mountains and the Kenai Lowlands and the highest H_e diversity occurred on the Kenai Coast. On the other hand, $\bar{r}_d$ showed no statistical trend (P = 0.5842, R^2 = 0.0528).

Table 3-3 Pairwise F_{ST}

	ANCH	DW	EG	KEFJ1	KEFJ2	PC	SH
DW	0.0123						
EG	0.0119	0.0135					
KEFJ1	0.0120	0.0142	0.0093				
KEFJ2	0.0119	0.0133	0.0106	0.0096			
PC	0.0127	0.0129	0.0124	0.0127	0.0113		
SH	0.0148	0.0165	0.0105	0.0113	0.0122	0.0140	
SL	0.0137	0.0130	0.0116	0.0127	0.0119	0.0121	0.0135

Pairwise F_{ST} was used as a measure of among sampling site genetic divergence representative of historic gene flow. The hierarchical AMOVA found no significant difference among regions (F_{RT} = 0.003, P = 0.372) or sample sites (F_{ST} = 0.005, P = 0.328)

Table 3-4 Student's *t*-test of pairwise *F*st.

test	t	Means of Patch vs All	p	95% CI
ANCH vs All Other Sites	0.7587	0.0127 - 0.0123	0.4593	-0.0007 - 0.0015
DW vs All Other Sites	2.7486	0.0136 - 0.0120	**0.0199**	0.0003 - 0.0029
EG vs All Other Sites	-2.3033	0.0113 - 0.0127	**0.0427**	-0.0027 - 0.00006
KEFJ1 vs All Other Sites	-1.3917	0.0116 - 0.0126	0.1989	-0.0027 - 0.0006
KEFJ2 vs All Other Sites	-2.0934	0.0115 - 0.0127	0.0564	-0.0024 - 0.00003
PC vs All Other Sites	0.4267	0.0125 - 0.0123	0.6738	-0.0008 - 0.00123
SH vs All Other Sites	1.2864	0.0132 - 0.0121	0.2369	-0.0008 - 0.0030
SL vs All Other Sites	0.5641	0.0126 - 0.0123	0.5782	-0.0007 - 0.0012

F_{ST} was used as a measure of gene flow. Pairwise *F*st was compared between each sample site and all other sample sites against all sample sites without the focal patch included. A student's *t*-test was then used to compare the means between included and non-included focal patches. The DW patch and EG both had statistically significant means ($P < 0.05$). The DW had statistically higher *F*st values than the rest of the Kenai patches ($P = 0.019$) and EG had statistically lower *F*st compared the rest of the Kenai patches ($P = 0.042$). Bold values are significant at 95% confidence

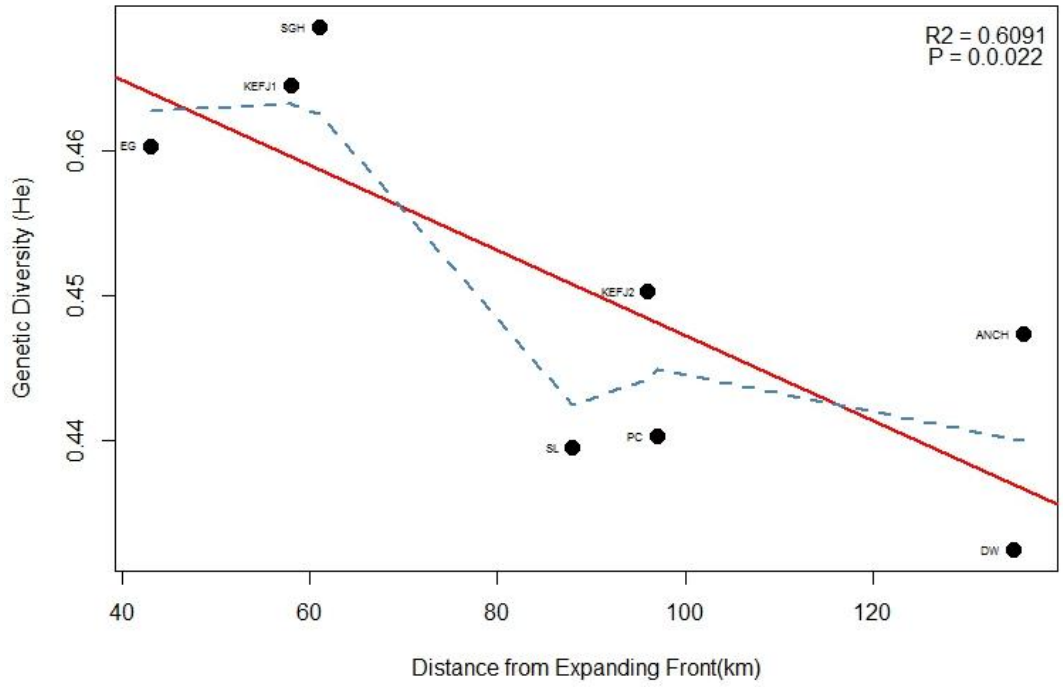

Figure 3- 2 Linear regression of H_e against distance from expanding front. H_e decreases significantly with increasing distance from Prince William sound (P=0.02, R^2 = 0.61). Solid line is the linear fit and the dotted line is the smothed lowess fit.

Population Discrimination

Discriminant analysis of principle components clustered individuals into three groups approximately corresponding to geographic position (Fig. 3-3). DAPC transformed the multivariate genetic information into 15 uncorrelated principle components. The number of retained PCs was chosen quantitatively based on *K*-means clustering and cross

validation. Seven discriminant functions were retained (*n* sampling sites – 1). The proportion of conserved variance was 30%.

The DW clustered near ANCH and PC in the northwest Kenai Mountains. The Kenai Mountain sites, SL and SGH, clustered together with the Kenai Coast site EG consisting of a south central Kenai Mountains cluster. The two Kenai Fjords coastal sites also clustered together as a Kenai Coast group. The northwest Kenai Mountain group was separated from the south central Kenai Mountain clusters by PC2. All Kenai Coast sites separated from the mountain and lowland sites by PC1. The south central Kenai Mountains clustered strongly together. The Discovery Well site showed some separation from the other sample locations despite its association with the northwest group. The overall pattern of clustering suggested a population structure consistent with the geography of three ecoregions on the Kenai.

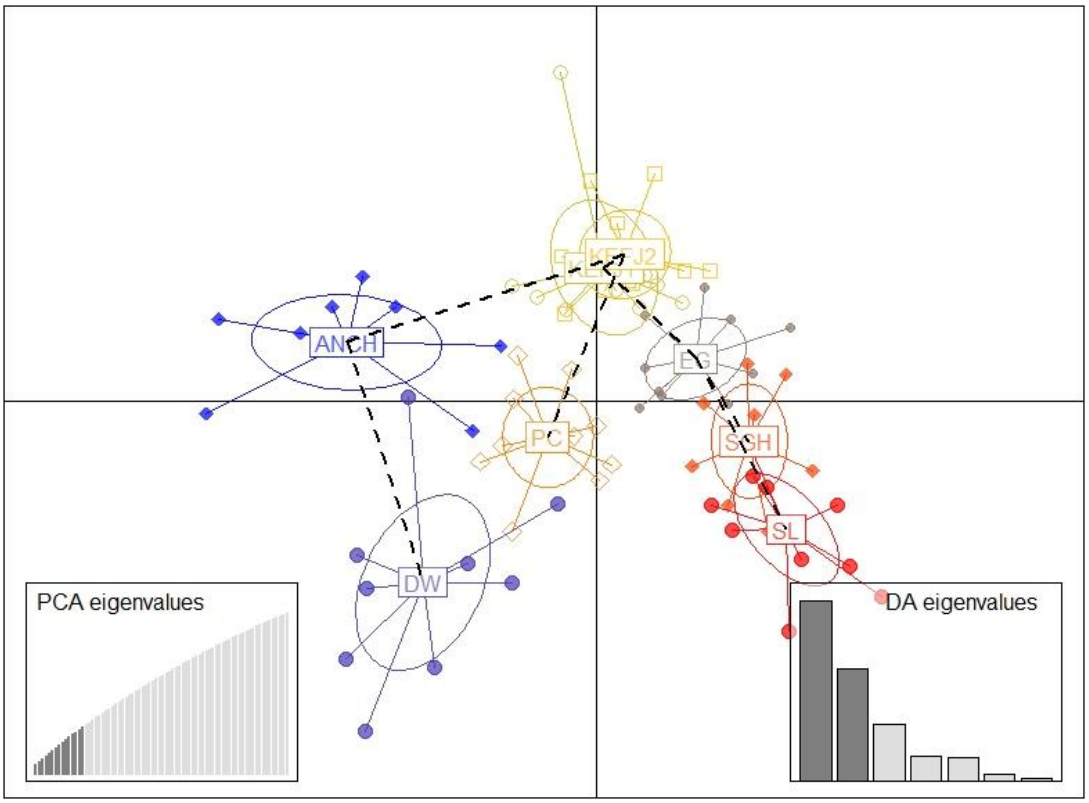

Figure 3-3: DAPC of 8 mountain hemlock sample regions clustered individuals into three groups approximately corresponding to geographic position. ANCH, PC, and DW clustered into a north west Kenai Mountains group. KEFJ1 and KEFJ2 clustered into a Kenai Coast group and SL, SGH, and EG clustered into a south central Kenai Mountains group. DW site showed some separation from the other sample locations despite its association with the northwest group. Site codes follow table 1. The dotted line represents the minimum spanning network between sites and serves as an approximate measure of similarity. Site codes follow abbreviations listed in table 3-1 above.

Landscape Variables

MaxEnt performance, as indicated by AUC, was good overall. The AUC score for mountain hemlock was 0.916 ∓ 0.012 (mean $\mp$ standard deviation) for the contemporary model and 0.912 ∓ 0.006 (mean $\mp$ standard deviation) for the mid Holocene model. All 19 bioclimatic variables were retained. Contribution of individual bioclimatic variables varied, with annual temperature seasonality and annual temperature range being the most important predictors of habitat suitability. All 19 bioclimatic variables were retained in our analysis as our interest was in the overall climate niche of mountain hemlock, and not an assessment of what climatic factors were important for determining that niche. The habitat suitability raster was inverted into resistance values and the resistance raster was used in Circuitscape to calculate pairwise resistance between all 8 sample locations (Fig. 3-4).

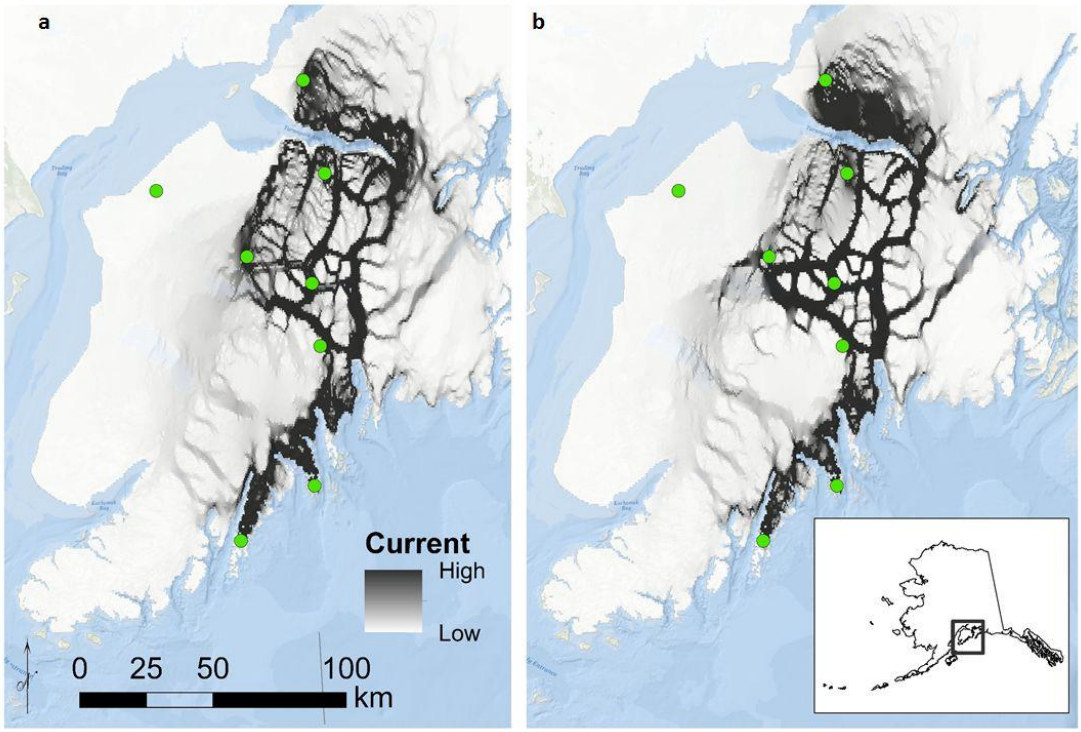

Figure 3-4: (a) Electrical current map output from contemporary landscape resistance in Circuitscape. Light color represents low current and low landscape permeability, dark colors represent high current and high landscape permeability. (b) Electrical current map output from Mid-Holocene landscape resistance in Circuitscape. Green circles are sampling sites. Inset map reflect the location of the Kenai Peninsula within Alaska.

Dispersal Limitation

I found no statistical effect of landscape on mountain hemlock dispersal limitation and historic connectivity on the Kenai Peninsula at 95% confidence. Tests of IBD found weak correlation between F_{ST} and geographic distance (r = 0.224, P = 0.181). Furthermore, tests of IBR found moderate correlation between F_{ST} and contemporary landscape resistance (r = 0.413, P = 0.083) and low correlation between F_{ST} and mid Holocene resistance (r = 0.178, P = 0.275). Despite the lack of statistical significance at 95% confidence, landscape resistance does look to be the most plausible explanation for the observed contemporary population genetic structure. A partial Mantel's test found contemporary IBR to explain moderate correlation when accounting for geographic distance (r = 0.362, P = 0.087) (Table 3-5). When low F_{ST} values are considered along with these results it appears that there is extensive movement of mountain hemlock described by contemporary climate. This analysis indicates that geneflow has been relatively unrestricted, and is best explained by contemporary climate when compared to a distance only or historic climate model. This pattern indicates that movement on the Kenai Peninsula has been more recent, supporting a model whereby remigration has occurred more recently by long distance dispersal from the southern range of mountain hemlock.

Table 3-5 Mantel's test.

Model	Hypothesis	Test	r	*P* value
Gene ~ Geog	IBD	Mantel	0.224	0.181
Gene ~ Contemporary Resistance	IBR	Mantel	0.413	0.083
Gene ~ Holocene Resistance	IBR	Mantel	0.178	0.275
Gene ~ Contemp | Geog	IBR	Partial Mantel	0.362	0.087
Gene ~ Holo | Geog	IBR	Partial Mantel	0.107	0.357

Simple Mantel's tests were run to test isolation by distance (IBD) and isolation by resistance (IBR). IBD compared pairwise geographic distance to pairwise $F_{ST.}$ IBR compared contemporary and mid Holocene pairwise resistance values, calculated using Circutscape, to pairwise $F_{ST.}$ A partial Mantel's test was used to test contemporary IBR while accounting for geographic distance. *P* values and Mantel's r are reported.

DISCUSSION

My results suggest that colonization of the Kenai Peninsula by mountain hemlock following the last glacial maximum was likely the results of a range expansion resulting from numerous LDD events and high levels of gene flow across a landscape with decreasing resistance as glaciers receded. I used a new GBS approach to assess if isolated stands of mountain hemlock in the Kenai Lowlands were glacial relicts and potential microrefugia, or if they were the result of random rare LDD. I answered my questions by looking at the differences in genetic diversity, population genetic structure, and LD across eight sample locations, which presented somewhat conflicting results.

Analysis assessing the trend in $H_{e,}$ and the level of historic gene flow connectivity on the Kenai Peninsula using DAPC and circuit theory gave solid evidence that colonization has occurred within recent geologic history following glacial retreat.

My conclusions are justified in part on past findings of the effects of range expansion and contraction on changes to the genetic structure of plants. Gene flow, via seed and pollen, mediates these genetic changes. Effective seed dispersal controls colonization and extinction processes (Baythavong et al., 2009), while pollen principally contributes to the amount of genetic diversity found within and between populations of plants (Ellstrand, 1992). Measures of historic gene flow (F_{ST}) can indicate that gene flow is more or less restricted across a particular landscape, but it cannot tell us how that landscape influences the movement of genes specifically. Electrical circuit theory, employed using Circuitscape, allowed us to better understand how climate of the Kenai has influenced dispersal. My results showed that contemporary, and not past, climate best explained the between site genetic patterns, despite being non-significant at 95% confidence. This combination of low F_{ST} values and circuit analysis has shown us that mountain hemlock is capable of rapid migration and extensive gene flow following climatic shifts via seed and pollen dispersal and likely can respond to future changes in its climatic niche. An additional line of evidence was based on the inability of past climate to explain contemporary differences among sampling locals and further indicates that the Discovery Well site is a product of rare LDD.

Origin of the Discovery Well

Boreal forest trees are commonly structured continuously in space and frequently do not conform to the island model of genetic connectivity (sensu Wright, 1931). More often, continuous forests have an IBD structure (Krutovsky et al., 2012), with continuous variation occurring in space. The life history traits of trees, such as long life, tall size, and outcrossing (Petit and Hampe, 2006), contribute to most genetic variation being distributed within individuals, resulting in populations only sharing a small fraction of genetic variation between them (low F_{ST} values) (Neale and Savolainen, 2004; Petit and Hampe, 2006).

Given the assumptions of genetic drift, small effective population size (N_e), and mutation in founding populations of forest trees, it is no small wonder that founding individuals survived LDD events at all. One premise in conservation genetics is that due to small population size and inbreeding, populations risk extinction (Allendorf and Lundquist, 2003; Simberloff, 2009). This is clearly not the case in scenarios of rapid range expansion relying on LDD. Ample evidence exists demonstrating the occurrence of rare LDD and gene flow events in plants allowing them to survive and persist (Alsos et al., 2007; Campbell et al., 1999; Kremer et al., 2012; Robledo-Arnuncio, 2011). The paradox, however, does give me pause. Simberloff (2009) addressed the problem of LDD and suggested that either 1.) inbreeding depression is not as much of a problem as has been argued, or 2.) even though the number of founders (seeds) in a LDD event may be small, continuous gene flow (mainly pollen) increases, or at least stabilizes, genetic

diversity in the short term. The stabilization of genetic diversity will stave off the effects of inbreeding depression and drift. Moreover, in the case where the number of founding individuals is large, this event may actually constitute a form of gene flow resulting in higher than expected genetic diversity in the founding population (Slatkin, 1977). In trees, where there is a lag of decades before reproduction, continuous arrival of seeds into a gap may build up the gene pool providing a buffer against inbreeding depression allowing a patch formed by LDD to survive more easily.

It is generally agreed that microrefugia should harbor higher genetic diversity, be in linkage equilibrium, and have greater genetic differences compared to the surrounding population (Hewitt, 1996; Hewitt, 2000; Nichols and Hewitt, 1994), though care should be taken when interpreting genetic patterns that may have nuanced explanations (Petit et al., 2003; Mee and Moore, 2014). My study has shown that the geographically isolated Discovery Well contains lower genetic diversity (lower H_e) compared to the surrounding mountain hemlock forests, and in fact there is a significant negative trend in genetic diversity with increasing latitude. In the case of the Kenai Peninsula, genetic diversity decreased from the pacific coast towards the lowlands. Despite these observations the magnitude of change in H_e was small, indicating diversity may be maintained by ongoing gene flow. Both the Kenai Mountains and Kenai Coast are cool and wet mountainous environments that are more suitable for mountain hemlocks. These locations could have served as source populations for seeds founding the initial Discovery Well population. The Kenai Lowlands, are boreal and have much higher incidence of disturbance, usually fire. The Discovery Well site likely originated after a

gap opening disturbance and a chance arrival of seed from the nearby Kenai Mountains that may have acted as a platform where winds from the Gulf of Alaska provide ample lift to transport the seed and pollen the 40 km distance to the Discovery Well. Subsequent and ongoing pollen gene flow many be maintaining genetic diversity in this geographically isolated patch.

I found higher than expected LD in all sample locations as assessed against neutral expectations. The higher than expected LD is indicative of a recent bottleneck due to range expansion (Flint-Garcia et al., 2003) on the Kenai peninsula. LD is influenced by a broad range of factors. For instance, LD can increase under scenarios of inbreeding, small population size, founder events, genetic isolation and population subdivision (Gupta et al., 2005; Flint-Garcia et al., 2003). Conversely, factors such as outcrossing, and high recombination and mutation rates can decrease LD relatively quickly (Gupta et al., 2005; Flint-Garcia et al., 2003; Neale and Savolainen, 2004). In particular, the size of a plant's genome and its system of mating have a big influence over the level of LD and how quickly it will break down. For example Flint-Garcia (2003) pointed out that there was a 250-fold difference between maize, an outcrossed species, and *Arabidopsis*, a selfing species, with maize exhibiting much lower LD across its genome. Conifer species are predominantly outcrossed, and as expected tend to break down LD very rapidly (Neale and Savolainen, 2004; Petit and Hampe, 2006; Krutovsky and Neale, 2005; Chhatre et al., 2013). Estimates of LD in the conifer *Pinus taeda* showed r^2 values decreasing to 0.20 at less than 1500 bp (Brown et al., 2004). Similarly, *Pseudotsuga menziesii* data indicated that LD decayed >50% over relatively short segments from

r^2 = ~0.25 to ~0.10 within 2000 bp (Krutovsky and Neale, 2005). *Picea abies* exhibited low r^2 values decreasing over approximately 100 bp (Heuertz et al., 2006). Because LD, as measured by the metrics $\bar{r}_d$ and r^2 are correlation coefficients, values near 0 indicate no correlation between two SNPs. The average LD within the Kenai was similar to other studies of long lived conifers. Based on pollen analysis mountain hemlock arrived on the peninsula between 1,500 – 3,000 ybp (Jones et al., 2009; Reger et al., 2007) a sufficient amount of time for LD to begin to break down in the outcrossed species.

Interestingly, the Discovery Well site had statistically lower LD than the other sampled sites, which contradicts the idea of an origin due to LDD based on my initial predictions. This result was surprising given the degree to which lower genetic diversity at Discovery Well, the overall trend in genetic diversity on the peninsula, and generally low regional genetic structure indicated a range expansion into the Kenai as the origin of mountain hemlock in this area. The fact that all sites had some level of LD, even though $\bar{r}_d$ values were low, indicated that the entire region was likely remigrating following glacial retreat that occurred within the last 14,000 years. Despite the Discovery Well site having statistically lower LD than the other sampled locations I still contend that the level of LD was not consistent with the site being a microrefugium.

Historic Connectivity

The investigation into the historic connectivity of mountain hemlock on the Kenai Peninsula identified extensive gene flow and connectivity between sampling locations

based on DAPC, F_{ST} and circuit theory. Weak population structure was organized geographically among the Kenai Coast, Kenai Mountains, and Kenai Lowlands ecoregions. The use of DAPC proved to be useful for identifying the organization of weak population genetic structure where standard methods failed (e.g. Structure Pritchard et al., 2000). By distilling the large SNP dataset down into synthetic axes DAPC maximized between sampling site variation separating the lowland, mountain, and coastal hemlock sites. The differences between sampling sites based on the F_{ST} analysis supported the LDD explanation. The weak population structure indicated that there was some influence of geographic heterogeneity on genetic divergence. The geographic structure of genetic variation was consistent with a series of stepping stone founding events and further supports our conclusions about the origin of the Discovery Well. Values of F_{ST} among populations have long been used to identify the historic levels of gene flow (Krutovsky et al., 2012). Similar to other outcrossed, long-lived, tree species (Petit and Hampe, 2006), mountain hemlock demonstrated high dispersal and gene flow capabilities.

When looking at connectivity of mountain hemlock on the Kenai Peninsula based on historic and contemporary climates using circuit theory, I showed that contemporary climates best explained gene flow patterns on the Kenai. Without accounting for geographic distance, contemporary landscape resistance explained 17% of the genetic variation. Even though the finding was not statistically significant at 95% confidence, at $P = 0.08$ the effect is likely real and if ignored it would result in a type II error. When resistance based on historic climate was assessed it only explained 1.1% of the genetic

variation (P = 0.27). Similarly, IBD only explained 5% of the genetic variation at P = 0.18. When a partial Mantel test was used to assess the contemporary climate IBR model after accounting for geographic distance 13% of the variation was still explained at P = 0.08, suggesting that contemporary climate is a better explanation for population genetic structure than distance alone. Future research should target a larger geographic area along the pacific coast to improve our understanding of historic connectivity in mountain hemlock during range expansion.

CONCLUSION

This research addressed two questions pertaining to historic gene flow on the Kenai Peninsula, Alaska. First, I have shown that isolated stands of mountain hemlock found in the lowlands are the result of rare long distance founding events following Pleistocene glaciation. This conclusion is based on decreased levels of genetic diversity in the Kenai Lowlands and weak population genetic structure overall. It is plausible that the geographically isolated stands in the Kenai Lowlands are just the most recent colonizers in an expanding range of the species concomitant with a warming arctic. Secondly, an assessment of historic gene flow as a response to past climate fluctuations suggests that high dispersal capability, as determined by F_{ST} analysis and AMOVA, and moderate levels of diversity will allow mountain hemlock to respond rapidly to future climate change. It is likely that the species will begin to appear more frequently north of its current range and at higher elevations. A better understanding of landscape resistance as

characterized by the species climate niche may prove to be a powerful predictor of future species distribution.

The results suggest that mountain hemlock is a recent colonizer to the Kenai Peninsula and that the species is capable of long distance seed and pollen dispersal in a heterogeneous environment. Future research focusing on clearing up questions about the importance of landscape structure and composition near an expanding range will provide improved explanation about the demographic history of long lived conifer species at high latitudes.

CHAPTER IV

SEED DISPERSAL AT ALPINE TREELINE: INSIGHTS FROM THE KENAI PENINSULA, ALASKA

INTRODUCTION

Over relatively long periods, vegetation is assumed to be in equilibrium with climate (Cole, 2010). Late Pleistocene and Holocene climate changes resulted in the migration of major vegetation zones in approximate equilibrium with climate, but the temporal resolution of those changes is usually restricted to century-scale analysis (MacDonald et al., 1993b; Webb et al., 1993). Now, as future climates are expected to change more rapidly than they have in the past, it is important to understand the migration potential of species to predict what future ecosystems may look like. This is important for sensitive ecosystems that may become severely restricted in spatial extent or disappear altogether. The alpine tundra ecosystem is one that could become severely diminished in extent and diversity as subalpine forests migrate upslope (Greenwood and Jump, 2014; Jump et al., 2012).

Recent research has found that not all treelines respond in the same way to a warming climate (Harsch et al., 2009; Cieraad and McGlone, 2014; Harsch and Bader, 2011; Chhetri and Cairns, 2015; Mamet and Kershaw, 2012). In a tightly coupled vegetation-climate system, we would expect that the treeline should track climate relatively closely over periods as short as a decade (Stueve et al., 2009; Danby and Hik,

2007). However, there are many confounding conditions that could potentially decouple the treeline ecotone from climate (Stueve et al., 2009; Cairns et al., 2007; Cairns and Moen, 2004; Germino and Smith, 1999; Sullivan and Sveinbjornsson, 2010; Resler et al., 2005; Brown and Vellend, 2014). Unlike animals that can respond quickly to changing environmental conditions, plants are fixed in space and must respond inter-generationally. Therefore, the response of the treeline to changing climate must depend on the ability of seeds dispersing beyond the current treeline, yet the degree to which new treeline seedling recruitment depends on local or lower elevation seed sources remains unresolved (Körner, 2012). We understand some of the statistical characteristics of short distance dispersal (Greene and Calogeropoulos, 2002; Greene and Johnson, 1997; Greene and Johnson, 1996), but LDD, sometimes quantified as the 99^{th} percentile of a dispersal kernel (Nathan, 2006), has been an intractable area of inquiry. LDD is thought to be necessary for rapid responses of vegetation communities to climate change. The importance of LDD in temperate plant migration has been codified as Reid's paradox (Clark et al., 1998b) which states that the rapid migration rates observed for tree species in the paleoecological record are far too high to have been produced by the traditional thoughts on tree life history and restricted dispersal. To fully understand the importance of LDD, the hypothesized "fat-tailed" dispersal kernels suggested by Clark et al. (1998b), where the probability distribution function is leptokurtic with clustering of dispersal events and high kurtosis near the seed parent and rare long distance dispersers accumulating at the ends of the tail, will have to be validated against empirically derived data. We are still not capable of the unambiguous determination of

every seed source at treeline, but advances in genetic analysis have now made it possible to determine the extent to which local seed production is responsible for the establishment of new recruits (Piotti et al., 2009; Johnson et al., 2016).

Two broad research methods can be used to investigate dispersal: observational or genetic (Lowe et al., 2004). Observational methods require the investigator to physically capture, tag, and monitor the dispersal of seeds or pollen resulting in low sample size and little ability to observe LDD events. Genetic methods, using parentage analysis however, allow a researcher to identify the physical distance any established individual travelled from its maternal seed source by matching genotypes to likely parents across a given area. Genetic analysis of dispersal has been accomplished in a variety of environments including both animals and plants in terrestrial (Listle and Reisch, 2012; Piotti et al., 2009; Lian et al., 2008; Dow and Ashley, 1996; Weinman et al., 2015) and aquatic systems (Markwith and Scanlon, 2007b; Markwith and Scanlon, 2006; Hauser et al., 2011). In a recent review, Ashley (2010) identified fifty three papers in which parentage analysis was used to identify pollen and seed donors for forest trees. Most of these concentrated on pollen donors and relatively few were for conifers. The benefit of using parentage analysis stems from the ability to better understand the end result of dispersal and the resulting spatial pattern (Ashley, 2010).

At alpine treelines the mode of reproduction can either be clonal or sexual. Sexual reproduction (via pollen transfer and seed production) is the dominant mode of regeneration in conifers that occurs under optimum conditions (Viktora et al., 2011). However, suboptimal conditions, such as those found at alpine treelines, lend themselves

to asexual regeneration through layering (Arno and Hammerly, 1984), whereby lower branches can advantageously sprout roots when they come into contact with the soil; eventually separating into an autonomous individual. When seedling viability is reduced due to short growing season or establishment limitation due to permafrost, as we see at treeline, vegetative layering functions as an adaptation that improves recruitment (Holtmeier, 2003). The ability for tree species to migrate at a sufficient rate to track predicted climate change may depend on what mode of regeneration is dominant along their current habitat borders. If LDD of seeds is occurring, allowing individuals to establish in new distant habitat, then it is more likely treelines will be able to track contemporary climate change patterns. On the other hand, if reproduction is primarily vegetative along species borders, migration will occur much slower. Very few studies have addressed this issue empirically. A parentage analysis of Norway spruce (*Picea* abies) found that high levels of long distance gene flow (> 300 m) maintained treeline gene pools (Piotti et al., 2009). Similarly, Viktora et al. (2011) used molecular genetics techniques to identify mode of regeneration under varying conditions and found that in subalpine black spruce (*Picea mariana*) the mode of reproduction was mixed sexual-clonal at the treeline tundra ecotone.

In most cases seeds disperse only up to tens of meters (Howe and Smallwood, 1982), giving rare LDD events a disproportionately higher impact on population structure (Cain et al., 2000). It is, however, difficult to observe and quantify LDD of seeds alone, and assessing the successful dispersal of seeds plus their establishment, known as effective

dispersal (Cain et al., 2000; Nathan and Muller-Landau, 2000), is much easier and perhaps of greater biological importance (Nathan et al., 2003).

Understanding and quantifying LDD and reproductive mode at treeline is of great interest, particularly in the context of understanding global processes and species responses to climate change. If we can quantify the genetic variation at treeline, identify distinct populations, quantify LDD, and assess the influence of geographic structure on dispersal, we can assess the likelihood that treelines are strong bioindicators of climate change as well as assess the ability of treeline to migrate in a rapidly changing climate.

Objective

This research is aimed at addressing the following two questions: (1) What is the primary mode of reproduction in the treeline forming conifer mountain hemlock (*Tsuga mertensiana*), and (2) are mountain hemlock recruits derived from local treeline populations or are they arriving from more distant seed sources? I investigate this question using a genomic dataset based on DNA SNPs. First I assess mode of reproduction by determining the proportion of sampled individuals with identical multilocus genotypes that are the product of clonal reproduction. Second, I use a categorical allocation based parentage analysis to identify parent offspring pairs so that the proportion of treeline reproduction events can be quantified spatially and dispersal distance measured. I exhaustively sampled individuals along a single mountain slope in the Alaskan Kenai Peninsula to characterize dispersal over hundreds of meters further

elucidating the degree to which treeline forming species can respond to rapid climate change.

MATERIALS AND METHODS

Study Species – Mountain Hemlock

Mountain hemlock is a monoecious, wind pollinated species that has long distance seed dispersal capacity (Chapter III) and an outcrossing mating system (Owens and Molder, 1975; Means, 1990; Ally et al., 2000). The species is found principally in wet cool environments on the western cost of North America, extending from the Alaskan Kenai Peninsula to the Sierras of northern California (Taylor, 1995; Means, 1990). On the Kenai, mountain hemlock is the dominant conifer within the subapline zone of the Kenai Mountains, and is part of the 'spruce hemlock zone' (Miller and Walton, 2014) consisting of Sitka Spruce (*Picea sitchensis*) and white spruce (*Picea glauca*). Subalpine mountain hemlock growth is negatively correlated to spring snowpack depth and positively correlated to summer growing season temperature (Taylor, 1995; Peterson and Peterson, 2001), with warm July temperatures resulting in increased seed production (Woodward et al., 1994). The Kenai Peninsula has experienced recent warming over the past century, suggesting likely mountain hemlock treeline expansion.

Study Area – Palmer Creek Drainage: Kenai Peninsula, Alaska

I based the study along a west facing elevational gradient, from alpine treeline down to the valley floor in the Palmer Creek drainage of the Chugach National Forest on the Kenai Peninsula, Alaska (Fig. 4-1 Fig. 4-2 and Fig. 4-3). The study area is located in the north central portion of the Peninsula in the Chugach- St Elias Mountains ecoregion (Nowacki et al., 2001) at about 60° 47' 37" N and 149° 32' 11.35 W. The vegetation structure is typical of the Kenai Mountains and is composed of shrub communities at the lower portion of the ecotone, primarily willow (*Salix sp.*) and alder (*Alnus sp.*), embedded in a matrix of bluestem (*Calamagrostis sp.*) with white spruce and mountain hemlock occurring as the dominant conifer species. Mountain hemlock dominates the alpine treeline ecotone. The alpine treeline ecotone transitions into alpine lichen tundra at approximately 800 masl. Summer temperatures on the Kenai Peninsula have increased by 0.2 °C per decade over the past 70 years (temperature data from the Homer Airport climate station). These changes in temperature have resulted in both an increase in Kenai Peninsula treeline elevation on cool, moist north facing slopes and an increase in density of the treeline (but not an increase in elevation) on other aspects (Dial et al., 2007). Similar patterns of treeline response to climate change have been observed at other locations in Alaska (Lloyd et al., 2002).

Figure 4-1: Study transect in the Palmer Creek drainage of the Chugach NF on the Kenai Peninsula, Alaska. The red box deliniates the boundary of the transect. Individuals were sampled from the valley bottom to the highest mountain hemlock individual. Mountain hemlock appears as small dark green patches primarily distributed towards the top of the transect. The dense green near the lower portions is willow and alder.

Figure 4-2: Study site at Palmer Creek drainage. Photo of the transect illustrates the patchy nature of the mountain hemlock distribution. Moreover, the photo illustrates the lack of additional mountain hemlock near the transect. Dark green patches near the top right of the photo are mountain hemlock. The lighter green patches near the center and lower part of the photo are willow and alder. The suitability of the transect is a product of reduceing the likelihood of not sampling a potential parent candidate.

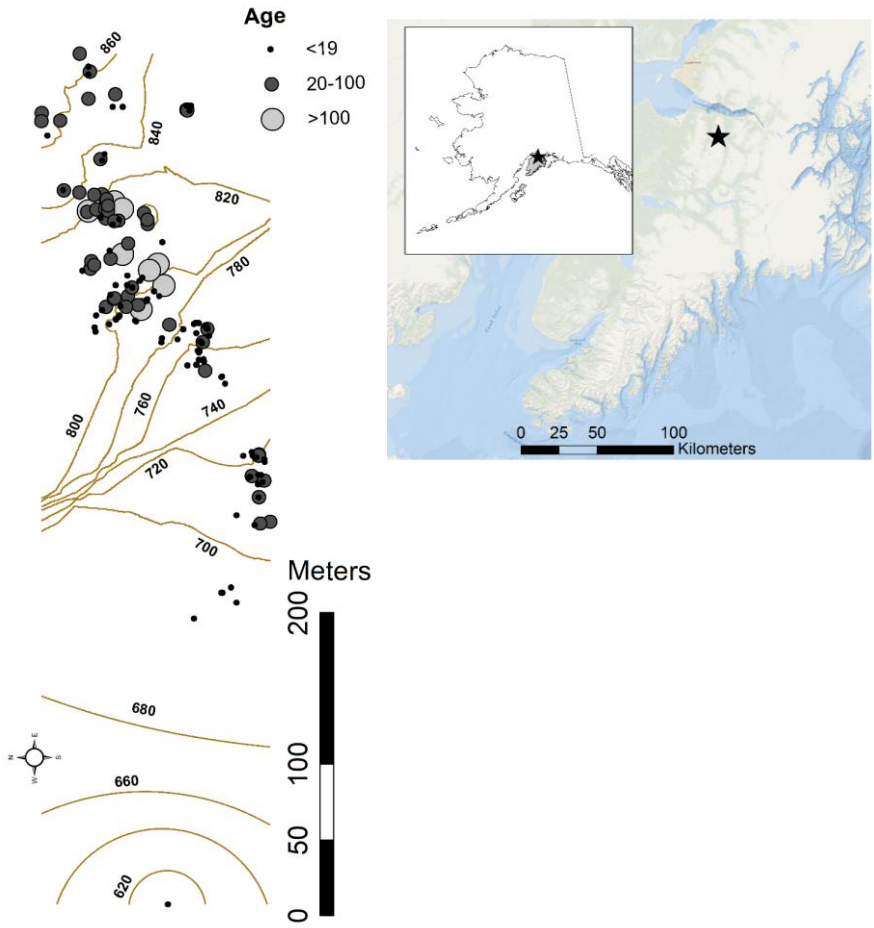

Figure 4-3: (a) Individual mountain hemlock locations along the study transect in the Palmer Creek drainage of the Chugach NF on the Kenai Peninsula, Alaska. Size of the circles represents the age of the sampled trees. Individuals were sampled from the valley bottom to the highest mountain hemlock individual. (b) The location of the transect is in the north central portion of the Kenai Peninsula near Hope, Alaska. The Kenai Peninsula is in south-central Alaska.

Sampling

I exhaustively sampled mountain hemlock along a 300m wide transect that extended from the highest mountain hemlock individual on the slope down to the valley floor (about 860 m in length between 610 to 880 masl. I observed no evidence of disturbance along the transect. The next closest patch of mountain hemlock to the study area was approximately 400 m away. The specific choice of study site and the study design reduced the likelihood that I would fail to sample a potential seed parent. Along the transect I collected foliage from individual mountain hemlock trees (approximately 15 cm long branch tips with young needles attached) for DNA extraction. I placed the twigs in a plastic zip-lock bag with silica gel for preservation (Colpaert et al., 2005), and stored bags at approximately 10° C in the field until being shipped back to the laboratory. I recorded the geographic location of each sample using a handheld GPS. Upon return from the field the samples were placed in a -20 C° freezer to await DNA extraction and sequencing. In addition, I collected a basal tree-core using an increment borer from all sampled trees with a basal diameter > 5 cm, and a stem cross-section for trees smaller than 5 cm basal diameter.

Genomic Marker Development

I used a genotyping-by-sequencing (GBS) approach, called double digest Restriction Associated DNA sequencing (ddRADseq) (Peterson et al., 2012), to genotype all

individuals on the transect. I processed tissue by first grinding approximately 30 mg of needle tissue, and extracted genomic DNA following a modified CTAB protocol (Doyle and Doyle, 1987). Genomic marker development is detailed in (chapter III). In brief, I targeted approximately 1% of the mountain hemlock genome using the restriction enzyme (RE) pair *SphI-MluCI*. Using the selected RE pair I generated PE sequencing libraries with ~ 350 bp long inserts from 163 mountain hemlock samples and sequenced them with 150 x 2 cycles in a single HiSeq 2000 lane (Illumina, Inc.). This approach allowed me to simultaneously sequence and identify variants across all sampled individuals. I conducted all library preparation and sequencing at the University of Texas Genomic Analysis and Sequencing Facility.

I used standard GBS bioinformatics analysis performed by the Texas A&M Institute for Genome Sciences and Society. This analysis assessed read quality, de-multiplexed samples, and aligned reads into contigs using the dDocent pipeline (Puritz et al., 2014a). I further filtered the SNPs identified in chapter III to select highly informative markers for parentage analysis. I performed quality filtering using VCFtools (Danecek et al., 2011). I used a 10x minimum coverage depth cutoff, and Phred quality score > 30. I removed SNP loci not in Hardy Weinberg equilibrium (HWE) at $P < 0.05$, that failed to genotype in greater than 95% of individuals, with a minor allele frequency less than 15%, and/or with expected heterozygosity (H_e) less than 0.2. Individuals missing greater than 15% of the identified SNPs were also removed.

Age Structure Analysis

To assess the age structure along the transect, I measured tree age from our collected increment cores and cross-sections following Lafon (2004). I dried cross-sections and cores in an oven for a minimum of 24 hours at 100 °C and then sanded them with progressively less abrasive sand paper (80 to 400 grit) to draw out the cellular structure of the annual rings (Fritts, 1976). I developed a master chronology (Stokes and Smiley, 1968) using 23 of the longest cores collected along the transect. This allowed me to identify significant marker rings which facilitated visual crossdating. I dated the remaining increment cores and cross-sections under varying magnification with a stereomicroscope. For any core not showing the pith, I estimated establishment date from the curvature and width of the innermost ring (Applequist, 1958). In most cases, I never added more than 5 years when estimating establishment date allowing me to bin trees into 5 year age classes. I created histograms to investigate the age structure of the transect, and disambiguate parent offspring relationships in our parentage analysis.

Clone Identification

To determine the major reproductive mode within the study site, I examined genotypic variation among the sampled trees. Using the computer program CERVUS (Kalinowski et al., 2007; Marshall et al., 1998), I conducted an identity analysis, where multilocus genotypes are compared to find near identical matches, which I labeled as clones. I

allowed for a 10% mismatch of loci in the multilocus genotypes based on calculated error rates of sequencing. I classified individuals with unique multilocus genotypes as the product of sexual reproduction rather than clonal.

Parentage Analysis

To determine the dispersal pattern at mountain hemlock treeline, I used parentage analysis to identify dispersal and recruitment by excluding all but the most likely parents based on the multilocus genotypes of the candidate offspring (Jones et al., 2010; Manel et al., 2005). The general principle in diploid organisms is that an offspring and a parent will share at least half of their alleles with each other at a single locus (Jones et al., 2010). A potential parent can be excluded if they do not have any alleles in common with the offspring. This analysis is limited by a need to sample all potential parents to reduce the likelihood of falsely assigning a parent, or missing the actual parent within a study area while sampling, restricting application to low density species within manageable study areas (Holderegger and Wagner, 2008). Although mountain hemlock can be locally gregarious, I chose a sample site with limited local density where I could exhaustively sample individuals along the slope. I conducted the analysis using the computer program CERVUS (Kalinowski et al., 2007; Marshall et al., 1998), which employs a categorical allocation approach based on a maximum-likelihood assignment tests to identify the most likely parent or parent pair from multilocus genotypes, useful when neither candidate parent is known *a priori* (Jones et al., 2010). In order to

determine the confidence of parentage assignments in the dataset, I calculated critical values of likelihood ratios using parentage simulation based on the allele frequencies in the SNP dataset. I simulated 10,000 offspring and 172 parents. I assumed that 10% of candidates were sampled and a 10% error rate in likelihood calculations based on my estimate of sequencing error. Parent offspring assignments were assessed using the logarithm of the odds (LOD) score. LOD scores were assigned to every single parent offspring combination along the transect. I estimated the critical LOD scores, used to assess statistical significance, using the simulated parentage analysis. I assigned offspring to parent or parent pairs by matching those individuals with the highest LOD score above the critical threshold. Confidence was assessed at strict, 95%, and relaxed, 80%, confidence levels as is commonly reported in studies using CERVUS. 95% confidence allows the most likely parent to be identified, while minimizing a wrong parent assignment, while 80% confidence increases the likelihood of type I errors but allows for a larger number of dispersal events to be identified and included in the analysis (Pluess et al., 2009).

I assigned seed parent status following the assumptions outlined by Dow and Ashley (1996) that has been applied in several studies (including spruce species (Piotti et al., 2009) (Table 4-1). When only one parent was identified, I assigned it as the seed parent. When I identified both likely parents, I assigned the parental candidate that was geographically closer to the offspring as the seed parent and the more distant parent as the pollen parent. When no parent was identified within the study site, I assumed that the seed was dispersed from beyond the surveyed area, which I defined as a LDD event. In

cases where the seed parent is identified I calculate dispersal distance based on the Euclidean distance between the parent and the offspring using ArcGIS (ESRI, 2011).

Table 4-1 Parentage assignments.

nDNA matches	Parent Identified	Long Distance Dispersal	Dispersal Distance
2	Maternal & Paternal	No	Yes
1	Maternal Only	No	Yes
0	Neither	Yes	No

Parentage assignment assumptions are based on Dow and Ashley (1996). If only one parent is identified than it will be classified as the seed parent. If two individuals are identified, then the individual that is geographically closer will be classified as the seed parent. If no parents are identified then cryptic gene flow and LDD will be assumed.

When I found a match between adults, I determined which of the pair was parent and which offspring based on the age of the two sampled trees with the older stem (based on tree-ring derived ages) being considered the parent. Conservatively, mountain hemlock reaches sexual maturity at about 20 years of age (Means, 1990). Thus, for a parent offspring assignment to be validated, the minimum difference in age between the two candidates must be greater than 20 years.

To determine the shape of the dispersal curve and average distance of dispersal, I fit normal, lognormal, Weibull, and gamma curves to the dispersal data. The distributions were fit based on the maximum likelihood approach using the *fitdistrplus* package in R (Delignette-Muller et al., 2014). I assessed the fit of each curve using Akaike's Information Criterion (AIC), whereby the model with the lowest AIC was chosen as the best model. This approach allowed me to assess the tail of the dispersal kernel (99^{th} percentile) to quantify LDD events following the suggestions of Piotti *et al.* (2009) and Nathan (2006).

RESULTS

Genomic Marker Development

From 163 individuals sequenced, GBS resulted in 41,057,267 unique sequencing reads and 50,952 assembled contigs with an average per nucleotide read depth greater than 30x. 171,019 putative SNPs were identified throughout the mountain hemlock genome using the dDocent (Puritz et al., 2014a) pipeline. Quality filtering reduced the number of SNPs to 5140 (chapter III) using VCFtools (Danecek et al., 2011). SNPs with H_e less than 0.2 were removed and loci not in HWE ($P < 0.05$) were removed. Additionally, any SNPs failing to sequence in more than 5% of individuals were removed. This process excluded an additional 4787 loci and resulted in a dataset of 353 highly informative genomic markers for clone identification and parentage analysis. Two individuals had

greater than 15% missing genotype information and were removed from the dataset. One hundred and sixty one individuals were retained for further analysis.

Age Structure Analysis

Age class distributions of mountain hemlock within the sample area show that the population is composed of a mixed age cohort establishing between 1900 and 2005 (Fig. 4-4). A pulse of young seedlings has established over the past decade. Additionally, the age structure suggests that there is a pulse of establishment in the early 1940s. The trees on the transect range in age from less than 5-year old seedlings to the oldest stem cored at 264 years of age (Table 4-2). The spatial pattern of the cohort indicates that the treeline ecotone has experienced infilling over the past two centuries with most uccessful establishment occurring between 700 and 800 masl (Fig. 4-3).

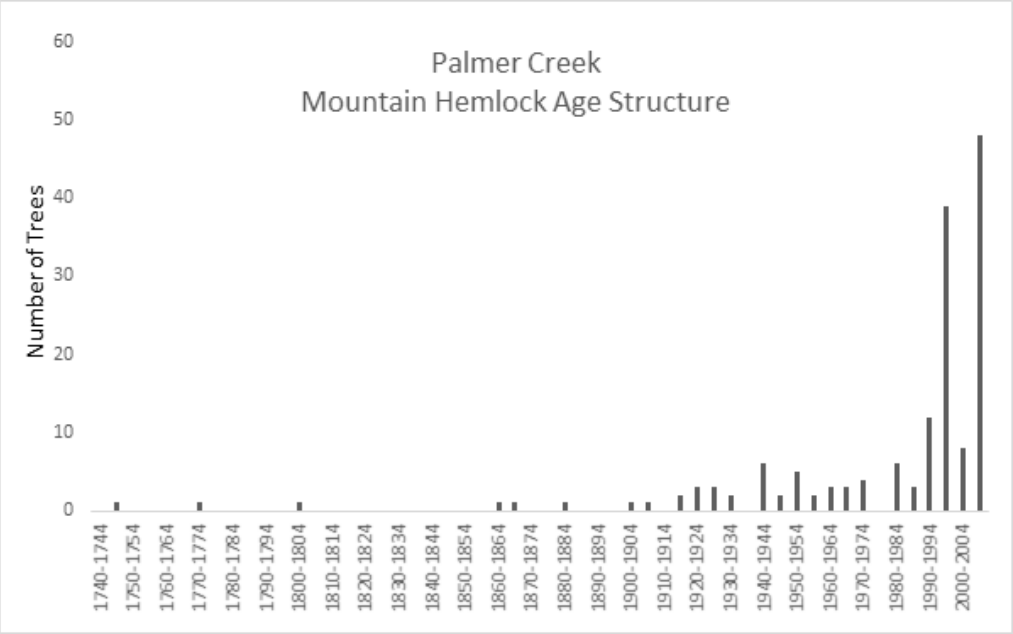

Figure 4-4: Histogram of tree ages based on establishment date binned into 5 year age classes. A recent pulse of establishmet can be seen in the early 2000s as well as in the 1940s.

Table 4-2 Age structure analysis.

1700s	Number of Trees	1800s	Number of Trees	1900s	Number of Trees	2000s	Number of Trees
1740-1744	0	1800-1804	1	1900-1904	1	2000-2004	8
1745-1749	1	1805-1809	0	1905-1909	1	2005-2009	48
1750-1754	0	1810-1814	0	1910-1914	0		
1755-1759	0	1815-1819	0	1915-1919	2		
1760-1764	0	1820-1824	0	1920-1924	3		
1765-1769	0	1825-1829	0	1925-1929	3		
1770-1774	1	1830-1834	0	1930-1934	2		
1775-1779	0	1835-1839	0	1935-1939	0		
1780-1784	0	1840-1844	0	1940-1944	6		
1785-1789	0	1845-1849	0	1945-1949	2		
1790-1794	0	1850-1854	0	1950-1954	5		
1795-1799	0	1855-1859	0	1955-1959	2		
		1860-1864	1	1960-1964	3		
		1865-1869	1	1965-1969	3		
		1870-1874	0	1970-1974	4		
		1875-1879	0	1975-1979	0		
		1880-1884	1	1980-1984	6		
		1885-1889	0	1985-1989	3		
		1890-1894	0	1990-1994	12		
		1895-1899	0	1995-1999	39		

Increment cores and cross sections of all sampled trees were aged and binned into 5 year age classes based on establishment date.

Clone Identification

I found no indication of clonal reproduction within the sample set. Clone analysis assessed all possible combinations of individual multilocus genotypes within the study site. Using the 353 identified SNPs, and allowing for 35 (10%) mismatched loci, 12,880 pairwise comparisons were made between the 161 analyzed individuals. All 161 individuals had unique multilocus genotypes after accounting for error. I conclude that sexual reproduction dominates the mountain hemlock alpine treeline ecotone within the Palmer Creek drainage of the Chugach National Forest.

Parentage Analysis

Simulation analysis identified critical LOD values of -1.50 for strict 95% confidence, and -4.50 for relaxed 80% confidence. 60 individuals were of reproductive age and classified as parental candidates, and 160 individuals were retained as offspring candidates. The offspring candidates included all individuals except the oldest tree on the transect. The parentage analysis assigned parent offspring pairs to 11.2% of sampled individuals. 18 of the offspring candidates had at least a single compatible parent match within the study area. In one case, two parents were identified within the transect signifying both seed and pollen dispersal events. 142 candidate offspring did not match to a compatible parent within the treeline ecotone. Paired LOD scores ranged from 6.319 to -1.440 at 95% confidence and -1.601 to -4.475 at 80% confidence.

Nineteen gametes were identified as being produced within the sample area (5.9%), while 301 gametes were potentially produced outside of it (142 none parent match x n2 - diploid gametes + 17 one parent match x n1 - haploid gametes). The immigration rate of seed into the transect was 88.8% and immigration of pollen into the transect was 99.3%. The lack of dual parental matches for all but one sampled individuals, gives a strong indication both seed and pollen gene flow are high within the study area, further suggesting that a large proportion of treeline establishment events are the result of extensive, cryptic, external gene flow and dispersal.

It is important to understand the direction and distance of dispersal to assess how this ecotone can respond to rapid global ecological change. From the 18 identified parent offspring matches, dispersal events occurred over distances of 1.44 to 326.85 m (mean 73 m) (Fig 4-5, Fig 4-6 and Table 4-3). All four dispersal events less than 10 m from the seed parent indicated upslope dispersal (the offspring was at a higher elevation than the parent). In contrast, ten (71.42%) of the 14 dispersal events over 10 m were downslope. There is some evidence that seed dispersal can occur to higher elevations over relatively long distances. Three identified dispersal events to higher elevations occurred between 50 and 100 m (max 97.9 m).

The shape of the dispersal curve was best approximated by the Weibull probability distribution function (AIC score 203.611) (Table 4-4). The Weibull, Gamma, and lognormal curves all represented the empirical data well, with only the normal distribution being an unlikely fit (Fig 4-7). LDD was quantified at the tail (99th percentile Nathan, 2006) of the Weibull curve as dispersal that occur at distances greater than 450 m. The transect did cover a distance greater than 450 m, however, I did not detect any dispersal events classified as LDD within the transect. The next closest stands of mountain hemlock were across the slope at distances greater than 400 m. This finding does suggest that LDD, as identified by cryptic dispersal and gene flow is extensive within the study area.

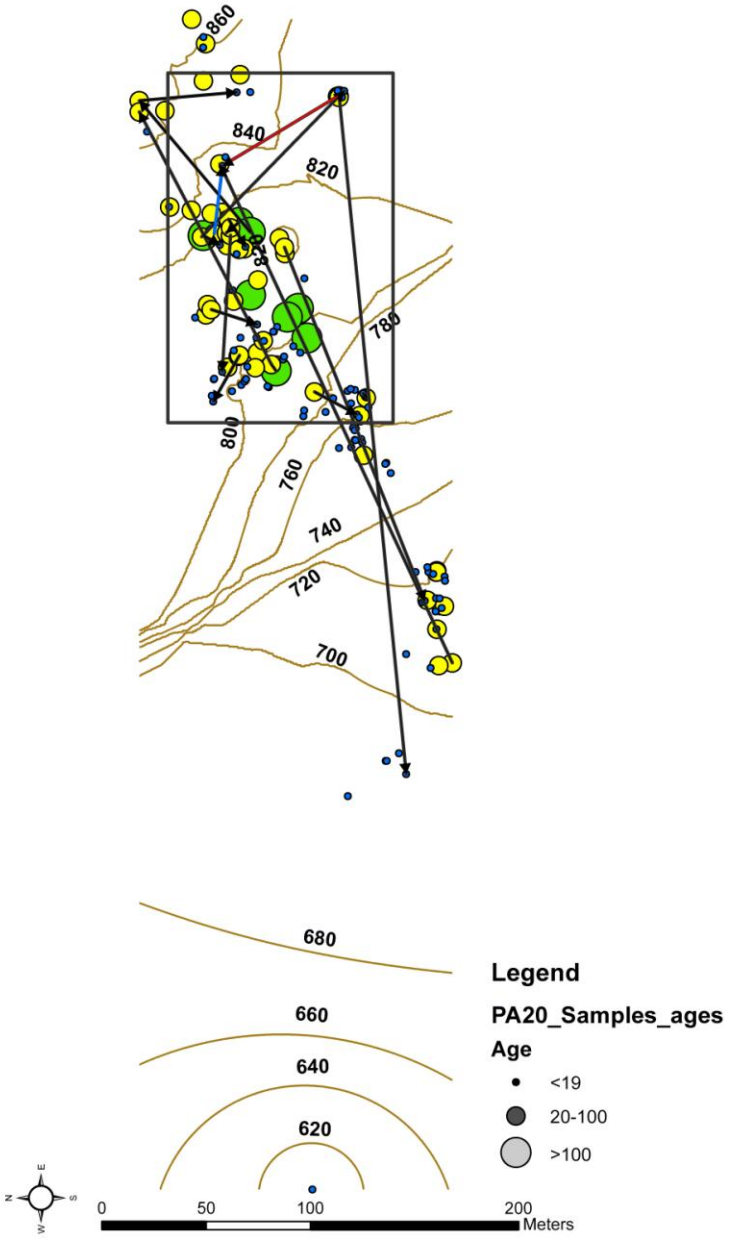

Figure 4-5 The study transect with individule parent offspring connections (thick black lines). The age of the trees is represented by circles. Large circles represent the oldest trees and smaller circles are progressivly younger. Black box indicates inset extent shown in figure 4-6

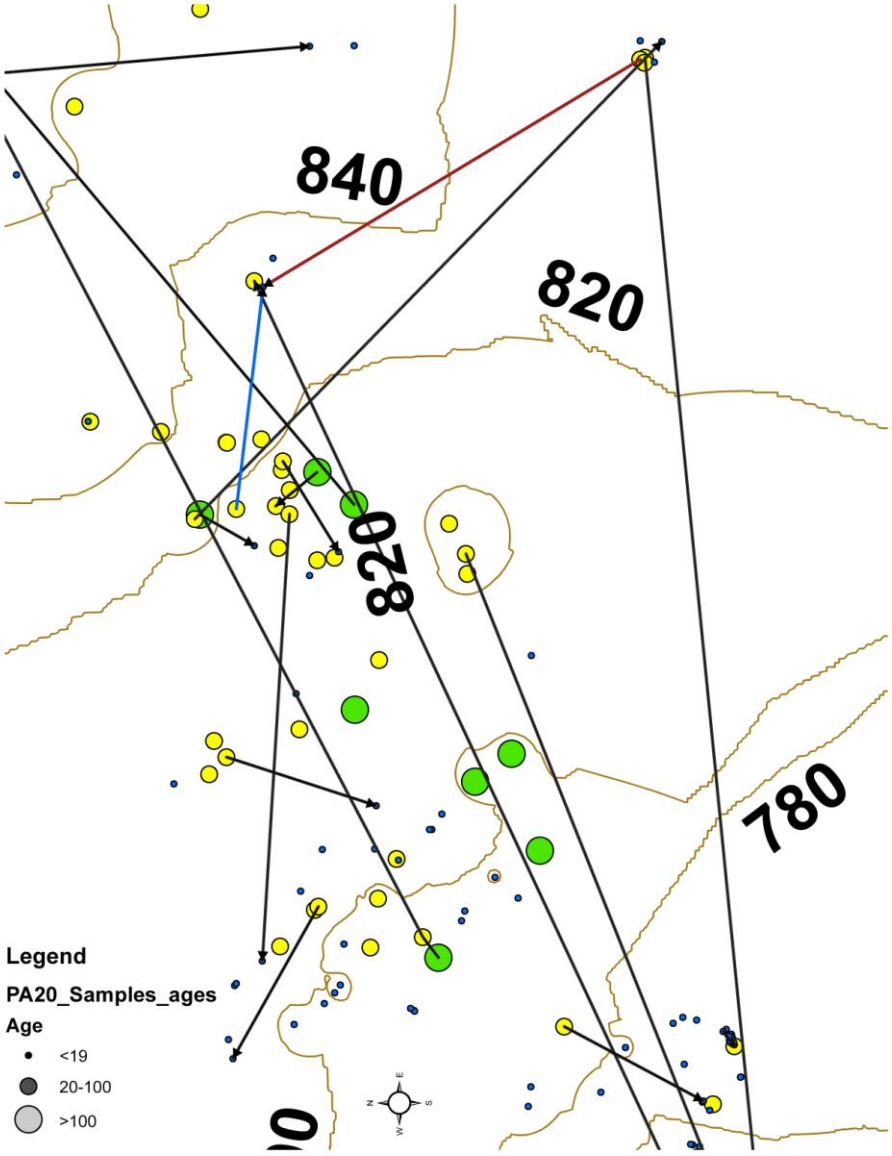

Figure 4-6 Inset of study transect with individule parent offspring connections (thick black lines). The age of the trees is represented by circles. Large circles represent the oldest trees and smaller circles are progressivly younger. Arrows indicate direction of dispersal. Red line is pollen and blue line is seed where both parents were identified.

Table 4-3 Parentage Analysis.

Offspring ID	Offspring Age	Parent ID	Parent Age	Paired LOD Score	Disp. Dist	Disp. Dir.
PA2080	88	PA2081	145	6.319	3.8	Higher
PA205	29	PA2034	106	2.575	82.55	Higher
PA20118	5	PA2091	91	1.918	23.13	Lower
PA2076	18	PA2065	72	1.149	25.54	Lower
PA2029	19	PA2027	129	0.953	9.18	Higher
PA20100	14	PA20107	72	0.685	1.44	Higher
PA2068	5	PA2040	30	0.526	65.61	Lower
PA20164	14	PA2021	64	0.069	326.85	Lower
PA2016	24	PA20157	95	-0.282	264.01	Lower
PA2017	11	PA2021	64	-0.873	64.53	Higher
PA2011	5	PA205	29	-0.904	46.84	Lower
PA2060	14	PA2044	49	-1.188	23.04	Lower
PA20149	12	PA2052	39	-1.601	182.48	Lower
PA2038	14	PA2032	39	-1.632	15.55	Lower
PA2017	11	PA2030	71	-1.647	32.78	Higher
PA2020	5	PA2028	73	-1.659	97.9	Higher
PA2041	43	PA2033	213	-1.697	7.86	Higher
PA206	29	PA2080	88	-1.75	136.77	Lower
PA2099	5	PA20107	72	-1.908	2.69	Higher

Mountain hemlock parent offspring matches assigned using Cervus. Offspring ID and Parent ID correspond to the individual tree on the transect. Offspring and Parent ages correspond to their age at the date of data collection. Paired LOD Score is the logarithm of the odds score used to assess confidence in assignment. LOD values greater than -1.5 have are assigned at 95% confidence and LOD values greater than -4.5 are assigned at 80% confidence. Disp. Dist is the Euclidian dispersal distance from parent to offspring and Disp. Dir identifies if the event was to a higher or lower elevation.

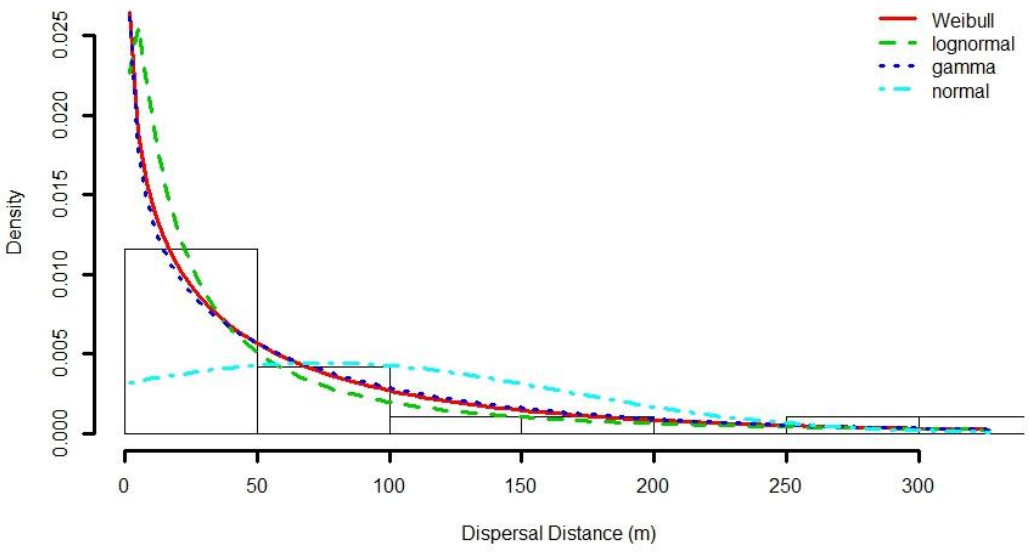

Figure 4-7 Normal, lognormal, Weibull, and Gamma probability distribution functions fit to the empirical dispersal data. The Weibull curve was chosen as the best curve based on Akaike's Information Criterion (AIC = 203.6).

Table 4-4 Dispersal curve fitting.

	effective distribution curves			
	Weibull	gamma	lognormal	normal
AIC	**203.611**	203.86	204.2006	228.9

Probability distribution functions (PDF) fit to the empirical dispersal data. Curves were assessed for goodness of fit using Akaike's Information Criterion (AIC). The PDF with the lowest AIC value is chosen as the best fit. The Weibull PDF (AIC = 203.611) was chosen as the best model to fit the dispersal data.

DISCUSSION

The results show that reproduction at the mountain hemlock alpine treeline ecotone in the Palmer Creek drainage is driven by sexual reproduction facilitated by extensive dispersal and gene flow into the ecotone. Both seed and pollen travels distances in excess of 450 m, constituting LDD. These dispersal events primarily contribute to infilling within the alpine treeline ecotone with only little evidence of elevation increases in the contemporary era (approximately the last 100 years). These findings are corroborated by earlier research on the Kenai Peninsula (Dial et al., 2007) and correspond with other dendroecological data showing similar patterns of infilling globally (Suarez et al., 1999; Chhetri and Cairns, 2015; Szeicz and Macdonald, 1995; Liang et al., 2011). There was a pronounced lack of observed seedlings near and above the species limit at the study location despite the availability of suitable sites.

The study site provided a natural transect from the bottom of the hill side all the way to the highest mountain hemlock individual at the tree species limit. The patchy distribution and floristic simplicity of the stand allowed me to exhaustively sample the species. Moreover, the next closest patch of mountain hemlock to the transect was in excess of 400 m, effectively reducing the likelihood that I would fail to sample all parent candidates within the treeline ecotone. Taken together, this configuration allowed me to assess both the mode of reproduction dominating the alpine treeline ecotone and the degree to which seeds are arriving from within the ecotone or from more distant seed sources.

My conclusions are supported by findings from previous conifer studies reporting seed dispersal and cryptic gene flow at large geographic distances (Piotti et al., 2009; Truong et al., 2007; Ally and Ritland, 2007; Lian et al., 2008; Robledo-Arnuncio and Gil, 2004). For instance, long distance wind dispersal of seeds have been reported in Hemlock (207.43 m) (Ally and Ritland, 2007), Pine (26.5-817.2 m) (González-Martínez et al., 2002; Gonzalez-MartÍnez et al., 2006; López de Heredia et al., 2015), Spruce (344.66 m) (Piotti et al., 2009) and Fir (236.9 m) (Lian et al., 2008). Animal dispersed seeds in oak species have been found to travel > 120 m (Moran and Clark, 2011; Dow and Ashley, 1996).

Similar to my findings, and in the only other study of conifer seed dispersal at alpine treeline, Piottie et al. (2009) recovered approximately 11% of parentage assignments in their study of Norway spruce and determined that cryptic gene flow accounted for over 66% of gametes. Their parentage analysis found extensive gene flow and dispersal at

distances in excess of 300 m and LDD was quantified at greater than 1.8 km. An early study of gene flow at alpine treeline assessed mountain birch (*Betula pubescens*) and also found extensive gene flow at the alpine treeline ecotone suggesting adequate capacity to track rapid warming (Truong et al., 2007).

Few studies have addressed the dominant form of reproduction at the alpine treeline ecotone in conifers. My finding of a dominant sexual mode of reproduction did, however, differ from the only other analysis conducted on a treeline forming conifer. Contrasting my finding is that of Viktora *et al.* (2011), who found that in relatively stable habitats black spruce reproduction was characterized as sexual, but at treeline reproduction was mixed clonal-sexual with the most dominant clone occurring geographically nearest alpine treeline. It is possible that the mountain hemlock treelines I studied do not experience the same degree of stress as those studied by Viktora *et al*. (2011) and as such have not needed to rely on clonal propagation to persist. Viktora's treeline study area was at higher latitude and elevation, as well as within a boreal continental climate. It is a curious difference, one that needs to be investigated further in a variety of conifer treeline species and locations to make general conclusions about its importance.

CONCLUSION

This study is the first to employ genome wide sampling of genetic variation to assess reproduction and dispersal at the alpine treeline ecotone. All previous studies to my

knowledge relied on only a handful of polymorphic microsatellite markers. Microsatellites have been shown to be quite effective at assigning parents, but they do represent only a small fraction of genetic variation. Studies in mammals (Tokarska et al., 2009) and fish (Hauser et al., 2011; Abadía-Cardoso et al., 2013; Anderson and Garza, 2006) show that SNPs perform just as well as microsatellites in parentage assignment studies.

Körner's Question

One of my overall objectives in this study was to assess the degree to which seed production within the alpine treeline ecotone is needed in order to maintain or advance treeline. Körner's (2012) question, though not stated explicitly, implies that local seed production may not be a dominant factor structuring the alpine treeline. My results corroborate this suggestion. The high levels of gene flow via effective seed and pollen dispersal from beyond treeline suggest that, at least in wind dispersed conifer species, the ability of the species to advance to higher elevations is not dependent on the availability of local seed and pollen. One observation worthy of comment is that with such effective seed and pollen dispersal capabilities why do we not see a gradual increase in treeline position as opposed to the observed infilling? In fact, it has been hypothesized that establishment is occurring well above most current treelines, but that frequent disturbance and mortality balances this high establishment rate (Slatyer and Nobel, 1992). The age data show that a large cohort of seedlings has established over the

past decade, but the age and pattern of older individuals leads me to conclude that a high mortality rate, due to harsh climate conditions, will likely reduce the surviving individuals substantially. It has been suggested that treeline advance cannot be assessed at less than half century scales due to time lags associated with mortality and age of reproductive maturity (Körner, 2012). As climates continue to warm, especially at high latitudes, the mortality rates associated with harsh climates will likely ameliorate. As this amelioration occurs the number of surviving seedlings will increase and we are likely to see an increase in elevation at mountain hemlock treelines on the Kenai Peninsula, Alaska.

CHAPTER V

CONCLUSION

A genome wide analysis of spatial genetic structure in mountain hemlock has found extensive long distance gene flow and dispersal ability at both local and landscape scales. Landscape level gene flow is facilitated by a high level of landscape connectivity as characterized by contemporary climate and landscape configuration. Local scale seed movement is extensive within the alpine treeline ecotone, with LDD characterized by cryptic gene flow on the order of 450 m. This analysis suggests that under similar landscape configurations, mountain hemlock has the capability to track its shifting climate envelope in response to future climate change.

While many wind dispersed trees, such as mountain hemlock, exhibit LDD capabilities, the relative rate of warming and landscapes fragmentation may limit the abundance and availability of safe site for establishment. Just having LDD capabilities alone may not translate into guaranteed persistence in the face of global ecological change. A combination of dispersal, adaptation, and phenotypic plasticity will be required for long term survival. Additional research will need to address how long lived trees at their range limits are adapting evolutionarily and bending phenotypically to changes in their environment.

The aim of this research was to investigate gene flow, mode of reproduction and dispersal process of a long lived tree species, mountain hemlock, at multiple scales (northern range limit and alpine treeline), in an effort to improve our understanding of

their capacity to respond to global ecological change. The objectives were (1) to specifically address historic migration process proceeding Pleistocene glaciation (Chapter III), (2) address dominant mode of reproduction at treeline (Chapter IV), and (3) address the extent to which recruits at treeline are derived from local treeline populations or are arriving from more distant seed sources (Chapter IV).

An important component of this research was to integrate high throughput genomic approaches into ecological biogeography (Chapter II). This approach has proven to be very informative as a way to assess fundamental biogeographic questions related to dispersal. Moreover, I would argue that the rapid technological advancement in the field of genomics can be incorporated much more easily into biogeographic research than it has previously. This is, in part, due to the proliferation of 3rd party sequencing facilities and bioidnformatics pipelines as suggested in chapter II. This dissertation represents the first use of genomics and parentage analysis to identify seed movement in a treeline forming conifer.

The research presented in this dissertation represents a novel integration of genomics and geography to answer a specific set of biogeographic questions allowing us to have a deeper understanding of how plants may migrate under altered climate conditions. The broader implications of this project include an improved understanding of the importance of LDD and the relative infrequency of clonal reproduction of mountain hemlock at high latitudes and at treeline. This study is broadly transferrable to other conifer treeline environments.

In addition to the general benefits of combining genetics and geography research,

treeline research in particular benefits from the knowledge gained here. The dispersal distances determined in this project can be used in future modeling studies to predict the form and rate of advance of the treeline under changing climate. This will expand future avenues of research.

The results of this study contribute to a stronger understanding of the influence of seed dispersal on treeline dynamics. Additionally, the methods can be transferred to other plant species such as arctic shrubs to investigate their expansion in the arctic.

REFERENCES

Abadía-Cardoso A, Anderson EC, Pearse DE, et al. (2013) Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (Oncorhynchus mykiss). *Molecular Ecology* 22: 4733-4746.

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

Ager TA. (2007) Vegetation response to climate change in Alaska: Examples from the fossil Record. U.S. Geological Survey, 44.

Aitken SN, Yeaman S, Holliday JA, et al. (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1: 95-111.

Alberto FJ, Derory J, Boury C, et al. (2013) Imprints of natural selection along environmental gradients in phenology-related genes of Quercus petraea. *Genetics* 195: 495-512.

Alleaume-Benharira M, Pen IR and Ronce O. (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal Of Evolutionary Biology* 19: 203-215.

Allendorf FW, Hohenlohe PA and Luikart G. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* 11: 697-709.

Allendorf FW and Lundquist LL. (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* 17: 24-30.

Ally D and Ritland K. (2007) A case study: looking at the effects of fragmentation on genetic structure in different life history stages of old-growth Mountain Hemlock (Tsuga mertensiana). *Journal of Heredity* 98: 73-78.

Ally D, Yousry E-K and Ritland K. (2000) Genetic diversity, differentiation and mating system in Mountain Hemlock (Tsuga mertensiana) across british columbia. *Forest Genetics* 7: 97-108.

Alsos IG, Eidesen PB, Ehrich D, et al. (2007) Frequent long-distance plant colonization in the changing arctic. *Science* 316: 1606-1609.

Amos JN, Harrisson KA, Radford JQ, et al. (2014) Species- and sex-specific connectivity effects of habitat fragmentation in a suite of woodland birds. *Ecology* 95: 1556-1568.

Anderson EC and Garza JC. (2006) The Power of Single-Nucleotide Polymorphisms for Large-Scale Parentage Inference. *Genetics* 172: 2567-2582.

Anderson JT, Inouye DW, McKinney AM, et al. (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences* 279: 3843-3852.

Anderson LL, Hu FS, Nelson DM, et al. (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences* 103: 12447-12450.

Andrew RL, Ostevik KL, Ebert DP, et al. (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology* 21: 2078-2091.

Andrewartha HG and Birch LC. (1954) *The distribution and abundance of animals,* Chicago: University of Chicago Press.

Andrews KR, Hohenlohe PA, Miller MR, et al. (2014) Trade-offs and utility of alternative RADseq methods. *Molecular Ecology* 23: 5943-5946.

Andrews KR and Luikart G. (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology* 23: 1661-1667.

Antonovics J and Levin DA. (1980) The ecological and genetic consequences of density-dependent regulation in plants. *Annual Review of Ecology and Systematics* 11: 411-452.

Applequist M. (1958) A simple pith locator for use with off-center increment cores *Journal of Forestry* 56: 141.

Arno SF and Hammerly RP. (1984) *Timberline: Mountain and Arctic Forest Frontiers,* Seattle: The Mountaineers.

Ashley MV. (2010) Plant Parentage, Pollination, and Dispersal: How DNA Microsatellites Have Altered the Landscape. *Critical Reviews in Plant Sciences* 29: 148-161.

Baird NA, Etter PD, Atwood TS, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376.

Barba M, Czosnek H and Hadidi A. (2014) Historical perspective, development and applications of Next-Generation Sequencing in plant virology. *Viruses (1999-4915)* 6: 106-136.

Bashalkhanov S, Eckert AJ and Rajora OP. (2013) Genetic signatures of natural selection in response to air pollution in red spruce (Picea rubens, Pinaceae). *Molecular Ecology* 22: 5877-5889.

Baythavong BS, Stanton ML and Rice KJ. (2009) Understanding the consequences of seed dispersal in a heterogeneous environment. *Ecology* 90: 2118-2128.

Beaulieu J, Doerksen T, Boyle B, et al. (2011) Association genetics of wood physical traits in the conifer White Spruce and relationships with gene expression. *Genetics* 188: 197-214.

Beckmann JS, Estivill X and Antonarakis SE. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Reviews Genetics* 8: 639-646.

Begun DJ, Holloway AK, Stevens K, et al. (2007) Population genomics: Whole-genome analysis of polymorphism and divergence in Drosophila simulans. *Plos Biology* 5: e310.

Benowicz A and A. El-Kassaby Y. (1999) Genetic variation in mountain hemlock (Tsuga mertensiana Bong.): quantitative and adaptive attributes. *Forest Ecology and Management* 123: 205-215.

Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. (2011) GenBank. *Nucleic Acids Research* 39: D32-D37.

Berg EE. (2003) Mountain hemlocks -- ghosts of the past or harbingers of the future? *Peninsula Clarion.* Kenai, Alaska.

Bernabo JC and Webb III T. (1977) Changing patterns in the Holocene pollen record of northeastern North America: A mapped summary. *Quaternary Research* 8: 64-96.

Blumler MA, Cole A, Flenley J, et al. (2011) History of biogeographical thought. In: Millington AC, Blumler MA and Schickhoff U (eds) *The Sage Handbook of Biogeography.* Los Angeles: SAGE.

Bolliger J, Lander T and Balkenhol N. (2014) Landscape genetics since 2003: status, challenges and future directions. *Landscape Ecology*: 1-6.

Bonte D, Van Dyck H, Bullock JM, et al. (2012) Costs of dispersal. *Biological Reviews Of The Cambridge Philosophical Society* 87: 290-312.

Boucher TV and Mead BR. (2006) Vegetation change and forest regeneration on the Kenai Peninsula, Alaska following a spruce beetle outbreak, 1987–2000. *Forest Ecology and Management* 227: 233-246.

Bradshaw HD and Stettler RF. (1995) Molecular genetics of growth and development in populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics* 139: 963-973.

Brooks LD. (2003) SNPs: Why do we care? In: Kwok P-Y (ed) *Single nucleotide polymorphisms methods and protocols.* Totowa, New Jersey: Humana Press Inc., 1-15.

Brown AHD, Feldman MW and Nevo E. (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* 96: 523-536.

Brown CD and Vellend M. (2014) Non-climatic constraints on upper elevational plant range expansion under climate change. *Proceedings of the Royal Society of London B: Biological Sciences* 281.

Brown GR, Gill GP, Kuntz RJ, et al. (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* 101: 15255-15260.

Brown JH, Stevens GC and Kaufman DM. (1996) The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* 27: 597-623.

Burke SV, Clark LG, Triplett JK, et al. (2014) Biogeography and phylogenomics of New World Bambusoideae (Poaceae), revisited. *American Journal of Botany* 101: 886-891.

Butchart SHM, Walpole M, Collen B, et al. (2010) Global biodiversity: Indicators of recent declines. *Science* 328: 1164-1168.

Cabrera-Bosquet L, Crossa J, von Zitzewitz J, et al. (2012) High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *Journal of Integrative Plant Biology* 54: 312-320.

Caffarra A and Donnelly A. (2011) The ecological significance of phenology in four different tree species: effects of light and temperature on bud burst. *International Journal of Biometeorology* 55: 711-721.

Cain ML, Milligan BG and Strand AE. (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany* 87: 1217-1227.

Cairns DM, Lafon CW, Moen J, et al. (2007) Influences of animal activity on treeline position and pattern: implications for treeline responses to climate change. *Physical Geography* 28: 419-433.

Cairns DM and Moen J. (2004) Herbivory influences tree lines. *Journal of Ecology* 92: 1019-1024.

Ćalić I, Bussotti F, Martínez-García PJ, et al. (2015) Recent landscape genomics studies in forest trees—what can we believe? *Tree Genetics & Genomes* 12: 1-7.

Callaway RM, DeLucia EH and Schlesinger WH. (1994) Biomass allocation of montane and desert Ponderosa Pine: An analog for response to climate change. *Ecology* 75: 1474-1481.

Campbell ID, McDonald K, Flannigan MD, et al. (1999) Long-distance transport of pollen into the Arctic. *Nature* 399: 29-30.

Cardinale BJ, Duffy JE, Gonzalez A, et al. (2012) Biodiversity loss and its impact on humanity. *Nature* 486: 59-67.

Carroll SP, Hendry AP, Reznick DN, et al. (2007) Editorial: Evolution on ecological time-scales. *Functional Ecology* 21: 387-393.

Chapin FS, Bret-Harte MS, Hobbie SE, et al. (1996) Plant functional types as predictors of transient responses of arctic vegetation to global change. *Journal of Vegetation Science* 7: 347-358.

Chen IC, Hill J, Ohlemuller R, et al. (2011) Rapid range shifts of species associated with high levels of climate warming. *Science* 333: 1024-1026.

Chen J, Tian Q, Pang T, et al. (2014) Deep-sequencing transcriptome analysis of low temperature perception in a desert tree, Populus euphratica. *BMC Genomics* 15: 326.

Cheptou P-O, Carrue O, Rouifed S, et al. (2008) Rapid evolution of seed dispersal in an urban environment in the weed Crepis sancta. *Proceedings of the National Academy of Sciences* 105: 3796-3799.

Chevin L-M, Collins S and Lefèvre F. (2013) Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field. *Functional Ecology* 27: 967-979.

Chhatre VE, Byram TD, Neale DB, et al. (2013) Genetic structure and association mapping of adaptive and selective traits in the east Texas loblolly pine (Pinus taeda L.) breeding populations. *Tree Genetics & Genomes* 9: 1161-1178.

Chhetri PK and Cairns DM. (2015) Contemporary and historic population structure of Abies spectabilis at treeline in Barun Valley, eastern Nepal Himalaya. *Journal of Mountain Science* 12: 558-570.

Cieraad E and McGlone MS. (2014) Thermal environment of New Zealand's gradual and abrupt treeline ecotones. *New Zealand Journal of Ecology* 38: 12-25.

Clark JS, Fastie C, Hurtt G, et al. (1998a) Reid's paradox of rapid plant migration. *BioScience* 48: 13-24.

Clark JS, Fastie C, Hurtt G, et al. (1998b) Reid's paradox of rapid plant migration: Dispersal theory and interpretation of paleoecological records. *BioScience* 48: 13-24.

Colas B, Olivieri I and Riba M. (1997) Centaurea corymbosa, a cliff-dwelling species tottering on the brink of extinction: A demographic and genetic study. *Proceedings of the National Academy of Sciences* 94: 3471-3476.

Cole KL. (2010) Vegetation Response to Early Holocene Warming as an Analog for Current and Future Changes. *Conservation Biology* 24: 29-37.

Colpaert N, Cavers S, Bandou E, et al. (2005) Sampling tissue for DNA analysis of trees: Trunk cambium as an alternative to canopy leaves. *Silvae Genetica* 54: 265-269.

Colwell RK, Brehm G, Cardelús CL, et al. (2008) Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* 322: 258-261.

Corlett RT and Westcott DA. (2013) Will plant movements keep up with climate change? *Trends in Ecology & Evolution* 28: 482-488.

Cox CB and Moore PD. (2010) *Biogeography an Ecological and Evolutionary Approach*, Hoboken NJ: John Wiley and Sons.

Cwynar LC. (1990) A late Quaternary vegetation history from Lily Lake, Chilkat Peninsula, southeast Alaska. *Canadian Journal of Botany* 68: 1106-1112.

Cwynar LC and MacDonald GM. (1987) Geographical variation of Lodgepole Pine in relation to population history. *The American Naturalist* 129: 463-469.

Danby RK and Hik DS. (2007) Responses of white spruce (Picea glauca) to experimental warmingat a subarctic alpine treeline. *Global Change Biology* 13: 437-451.

Danecek P, Auton A, Abecasis G, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156-2158.

Dasgupta MG, Dharanishanthi V, Agarwal I, et al. (2015) Development of genetic markers in Eucalyptus species by target enrichment and exome sequencing. *PLoS ONE,* 10. (accessed 2015).

Davey JW and Blaxter ML. (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics* 9: 416-423.

Davey JW, Cezard T, Fuentes-Utrilla P, et al. (2013) Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology* 22: 3151-3164.

Davey JW, Hohenlohe PA, Etter PD, et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics* 12: 499-510.

Davis M and Shaw RG. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* 292: 673-679.

Davis MB. (1981) Quaternary history and the stability of forest communities. In: West DC, Shugart HH and Botkin DB (eds) *Forest Succession.* Springer New York, 132-153.

del Mar Delgado M, Ratikainen II and Kokko H. (2011) Inertia: the discrepancy between individual and common good in dispersal and prospecting behaviour. *Biological Reviews* 86: 717-732.

Delignette-Muller M, Pouillot R, Denis J, et al. (2014) fitdistrplus: Help to fit a parametric distribution ot non-censored or censored data. *R package version 1.0-2*.

DeWitt TJ and Scheiner SM. (2004) *Phenotypic plasticity : functional and conceptual approaches*, New York: Oxford University Press.

DeWitt TJ, Sih A and Wilson DS. (1998) Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* 13: 77-81.

Dial RJ, Berg EE, Timm K, et al. (2007) Changes in the alpine forest-tundra ecotone commensurate with recent warming in southcentral Alaska: Evidence from orthophotos and field plots. *J. Geophys. Res.* 112: G04015.

Dobrowski SZ, Abatzoglou J, Swanson AK, et al. (2013) The climate velocity of the contiguous United States during the 20th century. *Global Change Biology* 19: 241-251.

Dow BD and Ashley MV. (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, Quercus macrocarpa. *Molecular Ecology* 5: 615-627.

Doyle J and Doyle J. (1987) A rapid procedure for DNA purification from small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11-15.

Dyer R, Chan D, Gardiakos V, et al. (2012) Pollination graphs: quantifying pollen pool covariance networks and the influence of intervening landscape on genetic

connectivity in the North American understory tree, Cornus florida L. *Landscape Ecology* 27: 239-251.

Dyer RJ. (2007) The evolution of genetic topologies. *Theoretical Population Biology* 71: 71-79.

Dyer RJ. (2015) Population Graphs and Landscape Genetics. *Annual Review of Ecology, Evolution, and Systematics* 46: null.

Dyer RJ and Nason JD. (2004) Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology* 13: 1713-1727.

Eckert AJ, Bower AD, Wegrzyn JL, et al. (2009a) Association genetics of coastal Douglas Fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. cold-hardiness related traits. *Genetics* 182: 1289-1302.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, et al. (2010) Patterns of population structure and environmental associations to aridity across the range of Loblolly Pine (Pinus taeda L., Pinaceae). *Genetics* 185: 969-982.

Eckert AJ, Wegrzyn JL, Pande B, et al. (2009b) Multilocus Patterns of Nucleotide Diversity and Divergence Reveal Positive Selection at Candidate Genes Related to Cold Hardiness in Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii). *Genetics* 183: 289-298.

Ehleringer JR, Sage RF, Flanagan LB, et al. (1991) Climate change and the evolution of C4 photosynthesis. *Trends in Ecology & Evolution* 6: 95-99.

Elias SA. (2013) The problem of conifer species migration lag in the Pacific Northwest region since the last glaciation. *Quaternary Science Reviews* 77: 55-69.

Ellegren H. (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* 29: 51-63.

Ellstrand NC. (1992) Gene flow by pollen: Implications for plant conservation genetics. *Oikos* 63: 77-86.

Elshire RJ, Glaubitz JC, Sun Q, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.

Emerson KJ, Clayton RM, Catchen JM, et al. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences* 107: 16196-16200.

Encinas-Viso F, Revilla TA, van Velzen E, et al. (2014) Frugivores and cheap fruits make fruiting fruitful. *Journal Of Evolutionary Biology* 27: 313-324.

Enright N, He T, Krauss S, et al. (2003) Genetic variation and biogeographic history in the restricted southwestern Australian shrub, Banksia Hookeriana. *Physical Geography* 24: 358-377.

ESRI. (2011) ArcGIS Desktop: Release 10.x. Redlands, CA: Environmental Systems Research Institute.

Excoffier L, Foll M and Petit R. (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* 40: 481-501.

Excoffier L, Smouse PE and Quattro JM. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Fahlgren N, Gehan MA and Baxter I. (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology* 24: 93-99.

Faircloth BC, McCormack JE, Crawford NG, et al. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*.

Fakheran S, Paul-Victor C, Heichinger C, et al. (2010) Adaptation and extinction in experimentally fragmented landscapes. *Proceedings of the National Academy of Sciences* 107: 19120-19125.

Fant JB, Havens K, Keller JM, et al. (2014) The influence of contemporary and historic landscape features on the genetic structure of the sand dune endemic, Cirsium pitcheri (Asteraceae). *Heredity* 112: 519-530.

Fielding AH and Bell JF. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.

Filiault DL and Maloof JN. (2012) A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. *PLoS Genetics* 8: 1-12.

Finkel E. (2009) With 'phenomics,' plant scientists hope to shift breeding into overdrive. *Science* 325: 380-381.

Fischer SF, Poschlod P and Beinlich B. (1996) Experimental studies on the dispersal of plants and animals on sheep in calcareous grasslands. *Journal of Applied Ecology* 33: 1206-1222.

Flint-Garcia SA, Thornsberry JM and Buckler ES. (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54: 357-374.

Foll M and Gaggiotti O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180: 977-993.

Franks SJ, Sim S and Weis AE. (2007) Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences* 104: 1278-1282.

Franks SJ, Weber JJ and Aitken SN. (2014) Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications* 7: 123-139.

Franks SJ and Weis AE. (2008) A change in climate causes rapid evolution of multiple life-history traits and their interactions in an annual plant. *Journal Of Evolutionary Biology* 21: 1321-1334.

Freeland JR, Kirk H and Petersen S. (2011) *Molecular Ecology*: John Wiley & Sons, Ltd.

Fritts HC. (1976) *Tree rings and climate,* New York: Academic Press.

Galetti M, Guevara R, Cortes MC, et al. (2013) Functional extinction of birds drives rapid evolutionary changes in seed size. *Science* 340: 1086-1090.

Garroway CJ, Bowman J, Carr D, et al. (2008) Applications of graph theory to landscape genetics. *Evolutionary Applications* 1: 620-630.

Gaston KJ. (2003) *The structure and dynamics of geographic ranges,* New York: Oxford University Press.

Gaston KJ. (2009) Geographic range limits: Achieving synthesis. *Proceedings: Biological Sciences* 276: 1395-1406.

Gent PR, Danabasoglu G, Donner LJ, et al. (2011) The Community Climate System Model Version 4. *Journal of Climate* 24: 4973-4991.

Geraldes A, Pang J, Thiessen N, et al. (2011) SNP discovery in black cottonwood (Populus trichocarpa) by population transcriptome resequencing. *Molecular Ecology Resources* 11: 81-92.

Germino MJ and Smith WK. (1999) Sky exposure, crown architecture, and low-temperature photoinhibition in conifer seedlings at alpine treeline. *Plant, Cell & Environment* 22: 407-415.

Ghalambor CK, McKay JK, Carroll SP, et al. (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* 21: 394-407.

Gienapp P, Teplitsky C, Alho JS, et al. (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology* 17: 167-178.

Gifford ML, Banta JA, Katari MS, et al. (2013) Plasticity regulators modulate specific root traits in discrete nitrogen environments. *PLoS Genetics* 9: 1-13.

Glenn TC. (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11: 759-769.

González-Martínez S, Dillon S, Garnier-Gere PH, et al. (2011) Patterns of nucleotide diversity and association mapping. In: CPlomion C, Bousquet J and Kole C (eds) *Genetics, Genomics and Breeding of Conifers.* Enfield, New Hampshire: CRC Press, Science Publishers, Inc., 239-275.

González-Martínez S, Gerber S, Cervera M, et al. (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (Pinus pinaster Ait.) using nuclear microsatellite markers. *Theoretical and Applied Genetics* 104: 1290-1297.

Gonzalez-MartÍnez SC, Burczyk J, Nathan RAN, et al. (2006) Effective gene dispersal and female reproductive success in Mediterranean maritime pine (Pinus pinaster Aiton). *Molecular Ecology* 15: 4577-4588.

González-Martínez SC, Krutovsky KV and Neale DB. (2006) Forest-tree population genomics and adaptive evolution. *New Phytologist* 170: 227-238.

Greene DF and Calogeropoulos C. (2002) Measuring and modelling seed dispersal of terrestrial plants. In: Bullock JM, Kenward RE and Hails RS (eds) *Dispersal Ecology.* Oxford: Blackwell Science Publ, 3-23.

Greene DF and Johnson EA. (1996) Wind Dispersal of Seeds from a Forest Into a Clearing. *Ecology* 77: 595-609.

Greene DF and Johnson EA. (1997) Secondary dispersal of tree seeds on snow. *Journal of Ecology* 85: 329-340.

Greenwood S and Jump AS. (2014) Consequences of Treeline Shifts for the Diversity and Function of High Altitude Ecosystems. *Arctic, Antarctic, and Alpine Research* 46: 829-840.

Gupta PK, Rustgi S and Kulwal PL. (2005) Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology* 57: 461-485.

Hall CAS, Stanford JA and Hauer FR. (1992) The distribution and abundance of organisms as a consequence of energy balances along multiple environmental gradients. *Oikos* 65: 377-390.

Hampe A and Petit RJ. (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* 8: 461-467.

Hamrick JL. (2004) Response of forest trees to global environmental changes. *Forest Ecology and Management* 197: 323-335.

Harsch MA and Bader MY. (2011) Treeline form – a potential key to understanding treeline dynamics. *Global Ecology and Biogeography* 20: 582-596.

Harsch MA, Hulme PE, McGlone MS, et al. (2009) Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecology Letters* 12: 1040-1049.

Hauser L, Baird M, Hilborn RAY, et al. (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. *Molecular Ecology Resources* 11: 150-161.

Heuertz M, De Paoli E, Källman T, et al. (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway Spruce [*Picea abies* (L.) Karst]. *Genetics* 174: 2095-2105.

Hewitt GM. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247-276.

Hewitt GM. (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68: 87-112.

Hewitt GM. (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405: 907-913.

Hijmans RJ, Cameron SE, Parra JL, et al. (2005) Very high resolution interpolated climate surfaces for glbal land areas. *International Journal of Climatology* 25: 1965-1978.

Hoffmann A, Griffin P, Dillon S, et al. (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* 2: 1-24.

Hoffmann AA and Sgro CM. (2011) Climate change and evolutionary adaptation. *Nature* 470: 479-485.

Holderegger R, Buehler D, Gugerli F, et al. (2010) Landscape genetics of plants. *Trends in Plant Science* 15: 675-683.

Holderegger R, Kamm U and Gugerli F. (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* 21: 797-807.

Holderegger R and Wagner HH. (2008) Landscape genetics. *BioScience* 58: 199-207.

Holliday JA, Ralph SG, White R, et al. (2008) Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (Picea sitchensis). *New Phytologist* 178: 103-122.

Holliday JA, Ritland K and Aitken SN. (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (Picea sitchensis). *New Phytologist* 188: 501-514.

Holtmeier F-K. (2003) *Mountain Timberlines: Ecology, Patchiness, and Dynamics*, Dordrecht: Kluwer Academic Publishers.

Homolka A, Schueler S, Burg K, et al. (2013) Insights into drought adaptation of two European oak species revealed by nucleotide diversity of candidate genes. *Tree Genetics & Genomes* 9: 1179-1192.

Howe HF and Smallwood J. (1982) Ecology of seed dispersal. *Annual Review of Ecology and Systematics* 13: 201-228.

Hutchinson GE. (1959) Homage to Santa Rosalia or Why are there so many kinds of animals? *American Naturalist* 92: 145-159.

Hyun T, Lee S, Kumar D, et al. (2014) RNA-seq analysis of Rubus idaeus cv. Nova: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. *Plant Cell Reports* 33: 1617-1628.

Ibrahim KM, Nichols RA and Hewitt GM. (1995) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77: 282-291.

Ingvarsson PK and Street NR. (2011) Association genetics of complex traits in plants. *The New Phytologist* 189: 909-922.

Inouye DW. (2008) Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology* 89: 353-362.

Johnson JS, Gaddis KD, Cairns DM, et al. (2016) Plant responses to global change: Next generation biogeography. *Physical Geography* 37: 93-119.

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

Jones A, Small C, Paczolt K, et al. (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources* 10: 6-30.

Jones MC, Peteet DM, Kurdyla D, et al. (2009) Climate and vegetation history from a 14,000-year peatland record, Kenai Peninsula, Alaska. *Quaternary Research* 72: 207-217.

Jones MR, Forester BR, Teufel AI, et al. (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution* 67: 3455-3468.

Juenger TE, McKay JK, Hausmann N, et al. (2005) Identification and characterization of QTL underlying whole-plant physiology in Arabidopsis thaliana: δ13C, stomatal conductance and transpiration efficiency. *Plant, Cell & Environment* 28: 697-708.

Jump AS, Huang T-J and Chou C-H. (2012) Rapid altitudinal migration of mountain plants in Taiwan and its implications for high altitude biodiversity. *Ecography* 35: 204-210.

Jump AS, Hunt JM, Martínez-Izquierdo JA, et al. (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in Fagus sylvatica. *Molecular Ecology* 15: 3469-3480.

Jump AS and Peñuelas J. (2005) Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters* 8: 1010-1020.

Jump AS and Peñuelas J. (2006) Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. *Proceedings of the National Academy of Sciences* 103: 8096-8100.

Kalinowski ST, Taper ML and Marshall TC. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1106.

Kamvar ZN, Brooks JC and Grünwald NJ. (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics* 6: 208.

Kamvar ZN, Tabima JF and Grünwald NJ. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.

Karlstrom TNV. (1964) Quaternary geology of the Kenai Lowland and glacial history of the Cook Inlet region, Alaska. *U. S. Geological Survey.* Washington, DC.

Kaya Z, Sewell MM and Neale DB. (1999) Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (Pinus taeda L.). *Theoretical and Applied Genetics* 98: 586-592.

Kirkpatrick M and Barton NH. (1997) Evolution of a species' range. *The American Naturalist* 150: 1-23.

Kokko H and López-Sepulcre A. (2006) From individual dispersal to species ranges: Perspectives for a changing world. *Science* 313: 789-791.

Körner C. (2012) *Alpine Treelines: Functional Ecology of the Global High Elevation Tree Limits,* Basel: Springer.

Körner C and Basler D. (2010) Phenology under global warming. *Science* 327: 1461-1462.

Kramer K. (1995) Phenotypic plasticity of the phenology of seven European tree species in relation to climatic warming. *Plant, Cell & Environment* 18: 93-104.

Kremer A, Ronce O, Robledo-Arnuncio JJ, et al. (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* 15: 378-392.

Krutovsky K, Elsik C, Matvienko M, et al. (2006) Conserved ortholog sets in forest trees. *Tree Genetics & Genomes* 3: 61-70.

Krutovsky KV, Burczyk J, Chybicki IJ, et al. (2012) Gene flow, spatial structure, local adaptation, assisted migration in trees. In: Schnell RJ and Priyadarshan PM (eds) *Genomics of Tree Crops.* New York, Heidelberg, Dordrecht, London: Springer, 71-116.

Krutovsky KV and Neale DB. (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas Fir. *Genetics* 171: 2029-2041.

Kubisch A, Holt RD, Poethke H-J, et al. (2014) Where am I and why? Synthesizing range biology and the eco-evolutionary dynamics of dispersal. *Oikos* 123: 5-22.

Lafon CW. (2004) Stand dynamics of a yellow-poplar (Liriodendron tulipifera L.) forest in the Appalachian Mountains, Virginia, USA. *Dendrochronologia* 22: 43-52.

Lenoir J, Gégout JC, Marquet PA, et al. (2008) A significant upward shift in plant species optimum elevation during the 20th century. *Science* 320: 1768-1771.

Lenormand T. (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* 17: 183-189.

Leslie S, Winney B, Hellenthal G, et al. (2015) The fine-scale genetic structure of the British population. *Nature* 519: 309-314.

Lian C, Goto S, Kubo T, et al. (2008) Nuclear and chloroplast microsatellite analysis of Abies sachalinensis regeneration on fallen logs in a subboreal forest in Hokkaido, Japan. *Molecular Ecology* 17: 2948-2962.

Liang E, Wang Y, Eckstein D, et al. (2011) Little change in the fir tree-line position on the southeastern Tibetan Plateau after 200 years of warming. *New Phytologist* 190: 760-769.

Lisec J, Meyer RC, Steinfath M, et al. (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. *The Plant Journal* 53: 960-972.

Listle D and Reisch C. (2012) Spatial genetic structure of the sedge Carex nigra reflects hydrological conditions in an alpine fen. *Arctic, Antarctic and Alpine Research* 44: 350-358.

Lloyd A, Rupp TS, Fastie C, et al. (2002) Patterns and dynamics of treeline advance on the Seward Peninsula, Alaska. *Journal of Geophysical Reseach* 108: 8161.

Loarie SR, Duffy PB, Hamilton H, et al. (2009) The velocity of climate change. *Nature* 462: 1052-1055.

Locascio A, Lucchin M and Varotto S. (2009) Characterization of a MADS FLOWERING LOCUS C-LIKE (MFL) sequence in Cichorium intybus: a comparative study of CiMFL and AtFLC reveals homologies and divergences in gene function. *New Phytologist* 182: 630-643.

López de Heredia U, Nanos N, García-del-Rey E, et al. (2015) High seed dispersal ability of Pinus canariensis in stands of contrasting density inferred from genotypic data. *2015* 24.

Losos JB, Arnold SJ, Bejerano G, et al. (2013) Evolutionary biology for the 21st century. *Plos Biology* 11: e1001466-e1001466.

Lowe AJ, Harris S and Ashton P. (2004) *Ecological Genetics: Design, Analysis, and Application*, Malden, MA: Blackwell Publishing Co.

MacDonald G, Edwards TWD, Moser KA, et al. (1993a) Rapid response of treeline vegetation and lakes to past climate warming. *Nature* 361: 243-246.

MacDonald GM, Edwards TWD, Moser KA, et al. (1993b) Rapid response of treeline vegetation and lakes to past climate warming. *Nature* 361: 243-246.

Malanson GP and Cairns DM. (1997) Effects of dispersal, population delays, and forest fragmentation on tree migration rates. *Plant Ecology* 131: 67-79.

Mamet SD and Kershaw GP. (2012) Subarctic and alpine tree line dynamics during the last 400 years in north-western and central Canada. *Journal of Biogeography* 39: 855-868.

Manel S, Gaggiotti OE and Waples RS. (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution* 20: 136-142.

Manel S and Holderegger R. (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution* 28: 614-621.

Manel S, Schwartz MK, Luikart G, et al. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18: 189-197.

Mantel N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer research* 27: 209-220.

Mardis ER. (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24: 133-141.

Markwith SH and Scanlon MJ. (2006) Characterization of six polymorphic microsatellite loci isolated from Hymenocallis coronaria (J. LeConte) Kunth (Amaryllidaceae). *Molecular Ecology Notes* 6: 72-74.

Markwith SH and Scanlon MJ. (2007a) Multiscale analysis of Hymenocallis coronaria (Amaryllidaceae) genetic diversity, genetic structure, and gene movement under

the influence of unidirectional stream flow. *American Journal of Botany* 94: 151-160.

Markwith SH and Scanlon MJ. (2007b) Multiscale analysis of Hymenocallis coronaria (Amaryllidaceae) genetic diversity, genetic structure, and gene movement under the influence of unidirectional stream flow. *American Journal of Botany* 94: 151-160.

Marroni F, Pinosio S and Morgante M. (2014) Structural variation and genome complexity: is dispensable really dispensable? *Current Opinion in Plant Biology* 18: 31-36.

Marshall TC, Slate J, Kruuk LE, et al. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655.

May RM. (1986) The search for patterns in the balance of nature: Advances and retreats. *Ecology* 67: 1116-1126.

Mayer E. (1942) *Systematics and the origin of species*

New York: Columbia University Press.

Mayer E. (1954) Changes in genetic environment and evolution. In: Huxley J, C. HA and Ford EB (eds) *Evolution as a process.* London: Allen and Unwin, 157-180.

McCormack JE and Faircloth BC. (2013) Next-generation phylogenetics takes root. *Molecular Ecology* 22: 19-21.

McCormack JE, Hird SM, Zellmer AJ, et al. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66: 526-538.

McCoy RC, Garud NR, Kelley JL, et al. (2014) Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Molecular Ecology* 23: 136-150.

McCune B and Grace J. (2002) *Analysis of Ecological Communities* Mjm Software Design.

McLachlan JS, Clark JS and Manos PS. (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088-2098.

McRae BH. (2006) Isolation by resistance. *Evolution* 60: 1551-1561.

McRae BH and Beier P. (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences* 104: 19885-19890.

McRae BH, Dickson BG, Keitt TH, et al. (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* 89: 2712-2724.

McRae BH, Shirk A and Platt JT. (2013) Gnarly Landscape Utilities: Resistance and habitat calculator. Seattle, WA: The Nature Conservancy.

Means JE. (1990) Tsuga mertensiana. In: Burns RM and Honkala BH (eds) *Silvics of North America.* Washington DC: Forest Service.

Mee JA and Moore J-S. (2014) The ecological and evolutionary implications of microrefugia. *Journal of Biogeography* 41: 837-841.

Meirmans PG. (2012) The trouble with isolation by distance. *Molecular Ecology* 21: 2839-2846.

Menzel A and Fabian P. (1999) Growing season extended in Europe. *Nature* 397: 658-659.

Metzker ML. (2010) Sequencing technologies - the next generation. *Nature Reviews. Genetics* 11: 31-46.

Miller-Rushing AJ, Høye TT, Inouye DW, et al. (2010) The effects of phenological mismatches on demography. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 3177-3186.

Miller AE and Walton JK. (2014) Vegetation monitoring - Southwest Alaska Network: 2013 Annual report - Baseline sampling in coastal forests of Kenai Fjords. *Natural Resource Data Series NPS/SWAN/NRDS - 2014/675.* Fort Collins, Colorado: National Park Service.

Miller M, Dunham J, Amores A, et al. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research* 17: 240-248.

Mitchell CE, Agrawal AA, Bever JD, et al. (2006) Biotic interactions and plant invasions. *Ecology Letters* 9: 726-740.

Montes JM, Melchinger AE and Reif JC. (2007) Novel throughput phenotyping platforms in plant genetic studies. *Trends in Plant Science* 12: 433-436.

Moran EV and Clark JS. (2011) Estimating seed and pollen movement in a monoecious plant: a hierarchical Bayesian approach integrating genetic and ecological data. *Molecular Ecology* 20: 1248-1262.

Morin X, Lechowicz MJ, Augspurger C, et al. (2009) Leaf phenology in 22 North American tree species during the 21st century. *Global Change Biology* 15: 961-975.

Müller T, Ensminger I and Schmid K. (2012) A catalogue of putative unique transcripts from Douglas-fir (Pseudotsuga menziesii) based on 454 transcriptome sequencing of genetically diverse, drought stressed seedlings. *BMC Genomics* 13: 673.

Müller T, Freund F, Wildhagen H, et al. (2015) Targeted re-sequencing of five Douglas-fir provenances reveals population structure and putative target genes of positive selection. *Tree Genetics & Genomes* 11: 1-17.

Namroud M-C, Beaulieu J, Juge N, et al. (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* 17: 3599-3613.

Narum SR, Buerkle CA, Davey JW, et al. (2013) Genotyping-by-sequencing in ecological and conservation genetics. *Molecular Ecology* 22: 2841-2847.

Nathan R. (2006) Long-distance dispersal of plants. *Science* 313: 786-788.

Nathan R and Muller-Landau HC. (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution* 15: 278-285.

Nathan R, Perry G, Cronin JT, et al. (2003) Methods for estimating long-distance dispersal. *Oikos* 103: 261-273.

Neale DB and Kremer A. (2011) Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12: 111-122.

Neale DB and Savolainen O. (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* 9: 325-330.

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

Nichols RA and Hewitt GM. (1994) The genetic consequences of long distance dispersal during colonization. *Heredity* 72: 312-317.

Nicotra AB, Atkin OK, Bonser SP, et al. (2010) Plant phenotypic plasticity in a changing climate. *Trends in Plant Science* 15: 684-692.

Nordborg M and Weigel D. (2008) Next-generation genetics in plants. *Nature* 456: 720-723.

Nosil P, Egan SP and Funk DJ. (2008) Heterogeneous genomic differentiation between walking-stick ecotypes:"Isolation by Adaptation" and multiple roles for divergent selection. *Evolution* 62: 316-336.

Nowacki GJ, Spencer P, Brock T, et al. (2001) Unified ecoregions of Alaska and neighboring territories. *U. S. Geological Survey.* Reston, Virginia.

Opgenoorth L, Vendramin GG, Mao K, et al. (2010) Tree endurance on the Tibetan Plateau marks the world's highest known tree line of the Last Glacial Maximum. *New Phytologist* 185: 332-342.

Orr HA. (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* 6: 119-127.

Orsini L, Vanoverbeke J, Swillen I, et al. (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* 22: 5983-5999.

Ouborg NJ, Piquot Y and Groenendael JMv. (1999) Population genetics, molecular markers and the study of dispersal in plants. *Journal of Ecology* 87: 551-568.

Owens JN and Molder M. (1975) Sexual reproduction of mountain hemlock (Tsuga mertensiana). *Canadian Journal of Botany* 53: 1811-1826.

Palle SR, Seeve CM, Eckert AJ, et al. (2013) Association of loblolly pine xylem development gene expression with single-nucleotide polymorphisms. *Tree Physiology* 33: 763-774.

Paradis E. (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 26: 419-420.

Parchman TL, Gompert Z, Mudge J, et al. (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* 21: 2991-3005.

Parker KC, Hamrick JL, Parker A, et al. (2001) Fine-scale genetic structure in Pinus clausa (Pinaceae) populatons; Effects of disturbance history. *Heredity* 87: 99-113.

Parker KC and Jorgenson SM. (2003) Examples of the use of molecular markers in biogeographic research. *Physical Geography* 24: 378-398.

Parker KC and Markwith S. (2007) Expanding biogeographic horizons with genetic approaches. *Geography Compass* 1: 246-274.

Parmesan C. (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics* 37: 637-669.

Parmesan C and Yohe G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421: 37-42.

Payseur BA. (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* 10: 806-820.

Peakall R and Smouse PE. (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Peñuelas J, Sardans J, Estiarte M, et al. (2013) Evidence of current impact of climate change on life: a walk from genes to the biosphere. *Global Change Biology* 19: 2303-2338.

Pereira H, Leadley P, Proenca V, et al. (2010) Scenarios for global biodiversity in the 21st century. *Science* 330: 1496-1501.

Peterson BK, Weber JN, Kay EH, et al. (2012) Double Digest RADseq: An inexpensive method for *De Novo* SNP discovery and genotypin in model and non-model species. *PLoS ONE* 7: 1-11.

Peterson DW and Peterson DL. (2001) Mountain hemlock growth responds to climatic variability at annual and decadal time scales. *Ecology* 82: 3330-3345.

Peterson GW, Hong Y, Horbach C, et al. (2014) Genotyping-By-Sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. *Diversity* 6: 665-680.

Petit RJ, Aguinagalde I, de Beaulieu J-L, et al. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563-1565.

Petit RJ and Hampe A. (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology and Systematics* 37: 187-214.

Petit RJ, Hu FS and Dick CW. (2008) Forests of the past: A window to future changes. *Science* 320: 1450-1452.

Phillips SJ, Anderson RP and Schapire RE. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

Phillips SJ, Dudík M, Elith J, et al. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19: 181-197.

Piotti A, Leonardi S, Piovani P, et al. (2009) Spruce colonization at treeline: where do those seeds come from. *Heredity* 103: 136-145.

Pluess AR, Sork VL, Dolan B, et al. (2009) Short distance pollen movement in a wind-pollinated tree, Quercus lobata (Fagaceae). *Forest Ecology and Management* 258: 735-744.

Poland JA, Brown PJ, Sorrells ME, et al. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7: 1-8.

Poland JA and Rife TW. (2012) Genotyping-by-Sequencing for plant breeding and genetics. *The Plant Genome* 5: 92-102.

Polgar CA and Primack RB. (2011) Leaf-out phenology of temperate woody plants: from trees to ecosystems. *New Phytologist* 191: 926-941.

Poncet BN, Herrmann D, Gugerli F, et al. (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina. *Molecular Ecology* 19: 2896-2907.

Porter JH, Hanson PC and Lin C. (2012) Staying afloat in the sensor data deluge. *Trends in Ecology & Evolution* 27: 121-129.

Porth I, Klapšte J, Skyba O, et al. (2013) Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. *New Phytologist* 200: 710-726.

Post E and Pedersen C. (2008) Opposing plant community responses to warming with and without herbivores. *Proceedings of the National Academy of Sciences* 105: 12353-12358.

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

Provan J and Bennett KD. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* 23: 564-571.

Prunier J, Laroche J, Beaulieu J, et al. (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology* 20: 1702-1716.

Puritz JB, Hollenbeck CM and Gold JR. (2014a) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* 2: e431.

Puritz JB, Matz MV, Toonen RJ, et al. (2014b) Demystifying the RAD fad. *Molecular Ecology* 23: 5937-5942.

Raj Kumar Kullan A, van Dyk MM, Hefer CA, et al. (2012) Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of Eucalyptus grandis and E. urophylla. *BMC Genetics* 13: 60-71.

Reddy TBK, Thomas AD, Stamatis D, et al. (2015) The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Research* 43: D1099-D1106.

Reeves PA, He Y, Schmitz RJ, et al. (2007) Evolutionary conservation of the FLOWERING LOCUS C-mediated vernalization response: Evidence from the sugar beet (Beta vulgaris). *Genetics* 176: 295-307.

Reger RD, Sturmann AG, Berg EE, et al. (2007) A guide to the late Quaternary istory of northern and western Kenai Peninsula, Alaska: Guidebook 8. *Alaska Department of Natural Resources.*

Rellstab C, Gugerli F, Eckert AJ, et al. (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* 24: 4348-4370.

Resler LM, Butler DR and Malanson GP. (2005) Topographic shelter and conifer establishment and mortality in an alpine environment, Glacier National Park, Montana. *Physical Geography* 26: 112-125.

Reusch TBH and Wood TE. (2007) Molecular ecology of global change. *Molecular Ecology* 16: 3973-3992.

Reymond M, Svistoonoff S, Loudet O, et al. (2006) Identification of QTL controlling root growth response to phosphate starvation in Arabidopsis thaliana. *Plant, Cell & Environment* 29: 115-125.

Reznick D and Ghalambor C. (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* 112-113: 183-198.

Rigg LS. (2003) Genetic applications in biogeography. *Physical Geography* 24: 355-357.

Ritland K, Krutovskii KV, Tsumura Y, et al. (2011) Genetic mapping in conifers In: Plomion C, Bousquet J and Kole C (eds) *Genetics, Genomics, and Breeding of Conifers.* Enfield, New Hampshire: CRC Press, Science Publishers, Inc.

Roberts DR and Hamann A. (2015) *Glacial refugia and modern genetic diversity of 22 western North American tree species*.

Robledo-Arnuncio JJ. (2011) Wind pollination over mesoscale distances: an investigation with Scots pine. *New Phytologist* 190: 222-233.

Robledo-Arnuncio JJ and García C. (2007) Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology* 16: 5098-5109.

Robledo-Arnuncio JJ and Gil L. (2004) Patterns of pollen dispersal in a small population of Pinus sylvestris L. revealed by total-exclusion paternity analysis. *Heredity* 94: 13-22.

Roeding F, Borner J, Kube M, et al. (2009) A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (Pandinus imperator). *Molecular Phylogenetics and Evolution* 53: 826-834.

Roesti M, Salzburger W and Berner D. (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology* 12: 94.

Ronce O. (2007) How does It feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics* 38: 231-253.

Root TL, Price JT, Hall KR, et al. (2003) Fingerprints of global warming on wild animals and plants. *Nature* 421: 57-60.

Rosenberg MS and Anderson CD. (2011) PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods in Ecology and Evolution* 2: 229-232.

Rossetto M and Henry RJ. (2014) Escape from the laboratory: new horizons for plant genetics. *Trends in Plant Science* 19: 554-555.

Ruiz-Gonzalez A, Cushman SA, Madeira MJ, et al. (2015) Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Molecular Ecology* 24: 5110-5129.

Rull V. (2009) Microrefugia. *Journal of Biogeography* 36: 481-484.

Rymer MJ and Sims JD. (1982) Lake-sediment evidence for the date of deglaciation of the Hidden Lake area, Kenai Peninsula, Alaska. *Geology* 10: 314-316.

Salomé PA, Bomblies K, Laitinen R, et al. (2011) Genetic Architecture of Flowering-Time Variation in Arabidopsis thaliana. *Genetics* 188: 421.

Sanger F, Air GM, Barrell BG, et al. (1977) Nucleotide sequence of bacteriophage [phi]X174 DNA. *Nature* 265: 687-695.

Schatz MC, Witkowski J and McCombie WR. (2012) Current challenges in de novo plant genome sequencing and assembly. *Genome Biology* 13: 243-250.

Seeb JE, Carvalho G, Hauser L, et al. (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources* 11: 1-8.

Segelbacher G, Cushman S, Epperson B, et al. (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics* 11: 375-385.

Sexton JP, Hangartner SB and Hoffmann AA. (2014) Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* 68: 1-15.

Shah VB and McRae BH. (2008) Circuitscape: A tool for landscape ecology. *Proceedings of the 7th Python in Science Conference*: 62-66.

Simberloff D. (2009) The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics* 40: 81-102.

Slatkin M. (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* 12: 253-262.

Slatyer RO and Nobel IR. (1992) Dynamics of montane treelines. In: Hansen AJ and di Castri F (eds) *Landscape Boundaries: Consequences for Biotic Diversity and Ecological Flows.* New York: Springer-Verlag.

Smith RIL. (1994) Vascular plants as bioindicators of regional warming in Antarctica. *Oecologia* 99: 322-328.

Sork V and Smouse P. (2006) Genetic analysis of landscape connectivity in tree populations. *Landscape Ecology* 21: 821-836.

Sork V and Waits LP. (2010) Contributions of landscape genetics approaches, insights, and future potential. *Molecular Ecology* 19: 3489-3495.

Sork VL, Aitken SN, Dyer RJ, et al. (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* 9: 901-911.

Spillane C and McKeown PC. (2015) *Plant epigenetics and epigenomics*: Humana Press.

Stapley J, Reger J, Feulner PGD, et al. (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution* 25: 705-712.

Steane DA, Potts BM, McLean E, et al. (2014) Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology* 23: 2500-2513.

Steiner CC, Putnam AS, Hoeck PEA, et al. (2013) Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences* 1: 261-281.

Stillman JH and Armstrong E. (2015) Genomics are transforming our understanding of responses to climate change. *BioScience*: Advance Access on-line publication.

Stokes M and Smiley T. (1968) *An introduction to tree-ring dating,* Chicago: University of Chicago Press.

Stölting KN, Nipper R, Lindtke D, et al. (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* 22: 842-855.

Storfer A, Murphy MA, Evans JS, et al. (2007) Putting the 'landscape' in landscape genetics. *Heredity* 98: 128-142.

Storfer A, Murphy MA, Spear SF, et al. (2010) Landscape genetics: where are we now? *Molecular Ecology* 19: 3496-3514.

Strickler SR, Bombarely A and Mueller LA. (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. *American Journal of Botany* 99: 257-266.

Stueve KM, Cerney DL, Rochefort RM, et al. (2009) Post-fire tree establishment patterns at the alpine treeline ecotone: Mount Rainier National Park, Washington, USA. *Journal of Vegetation Science* 20: 107-120.

Suarez F, Binkley D, Kaye MW, et al. (1999) Expansion of forest stands into tundra in the Noatak National Preserve, northwest Alaska. *Ecoscience* 6: 465-470.

Sullivan PF and Sveinbjornsson B. (2010) Microtopographic control of treeline advance in Noatak National Preserve, Northwest Alaska. *Ecosystems* 13: 275-285.

Szeicz JM and Macdonald GM. (1995) Recent White Spruce dynamics at the subarctic alpine treeline of north-western Canada. *Journal of Ecology* 83: 873-885.

Taylor AH. (1995) Forest expansion and climate change in the Mountain Hemlock (Tsuga mertensiana) zone, Lassen Volcanic National Park, California, U.S.A. *Arctic and Alpine Research* 27: 207-216.

Teller BJ, Campbell C and Shea K. (2014) Dispersal under duress: Can stress enhance the performance of a passively dispersed species? *Ecology* 95: 2694-2698.

Thavamanikumar S, McManus L, Ades P, et al. (2014a) Association mapping for wood quality and growth traits in Eucalyptus globulus ssp. globulus Labill identifies nine stable marker-trait associations for seven traits. *Tree Genetics & Genomes* 10: 1661-1678.

Thavamanikumar S, Southerton S and Thumma B. (2014b) RNA-Seq using two populations reveals genes and alleles controlling wood traits and growth in Eucalyptus nitens. *PLoS ONE* 9: e101104.

Thomas CD. (2010) Climate, climate change and range boundaries. *Diversity and Distributions* 16: 488-495.

Thomson FJ, Moles AT, Auld TD, et al. (2011) Seed dispersal distance is more strongly correlated with plant height than with seed mass. *Journal of Ecology* 99: 1299-1307.

Tokarska M, Marshall T, Kowalczyk R, et al. (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity* 103: 326-332.

Toonen RJ, Puritz JB, Forsman ZH, et al. (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* 1: e203.

Travis JMJ, Delgado M, Bocedi G, et al. (2013) Dispersal and species' responses to climate change. *Oikos* 122: 1532-1540.

Truong C, Palmé AE and Felber F. (2007) Recent invasion of the mountain birch Betula pubescens ssp. tortuosa above the treeline due to climate change: genetic and ecological study in northern Sweden. *Journal Of Evolutionary Biology* 20: 369-380.

Uddenberg D, Reimegard J, Clapham D, et al. (2013) Early cone setting in Picea abies acrocona is associated with increased transpritional activity of a MADS Box transcription factor. *Plant Physiology* 161: 813-823.

van der Maarel E. (1990) Ecotones and ecoclines are different. *Journal of Vegetation Science* 1: 135-138.

van Dijk EL, Auger H, Jaszczyszyn Y, et al. (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* 30: 418-426.

Van Tassell CP, Smith TPL, Matukumalli LK, et al. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* 5: 247-252.

Vargas-Rodriguez YL, Platt WJ, Urbatsch LE, et al. (2015) Large scale patterns of genetic variation and differentiation in sugar maple from tropical Central America to temperate North America. *BMC Evolutionary Biology* 15: 1-14.

Veblen TT. (1989) Biogeography. In: Gaile GL and Willmott CJ (eds) *Geography in America.* Columbus, OH: Merrill, 28-46.

Viktora M, Savidge RA and Rajora OP. (2011) Clonal and nonclonal genetic structure of subarctic black spruce (Picea mariana) populations in Yukon territory. *Botany* 89: 133-140.

Walther G-R. (2003) Plants in a warmer world. *Perspectives in Plant Ecology, Evolution and Systematics* 6: 169-185.

Wang N, Thomson M, Bodles WJA, et al. (2013) Genome sequence of dwarf birch (Betula nana) and cross-species RAD markers. *Molecular Ecology* 22: 3098-3111.

Wang S, Meyer E, McKay JK, et al. (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* 9: 808-810.

Wang Z, Gerstein M and Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57-63.

Webb T, Bartlein PJ, Harrison SP, et al. (1993) Vegetation, lake levels, and climate in Eastern North America for the past 18,000 years. In: H.E. Wright J, Kutzbach JE, III TW, et al. (eds) *Global Climates since the Last Glacial Maximum.* Minneapolis: University of Minnesota Press, 415 - 467.

Webb T, III. (1987) The appearance and disappearance of major vegetational assemblages: Long-term vegetational dynamics in eastern North America. *Vegetatio* 69: 177-187.

Weinman LR, Solomon JW and Rubenstein DR. (2015) A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage

and kinship in a cooperatively breeding bird. *Molecular Ecology Resources* 15: 502-511.

Weir BS and Cockerham CC. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Weitemier K, Straub SCK, Cronn RC, et al. (2014) Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: apps.1400042.

Wetterstrand K. (2014) *DNA sequencing costs: Data from the NHGRI Genome Sequencing Program (GSP).* Available at: www.genome.gov/sequencingcosts.

Wheeler N, Jermstad K, Krutovsky K, et al. (2005) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Molecular Breeding* 15: 145-156.

Williams JW, Jackson ST and Kutzbach JE. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences* 104: 5738-5742.

Willson MF and Traveset A. (2000) The ecology of seed dispersal. In: Fenner M (ed) *Seeds: the ecology of regeneration in plant communities.* 2nd ed. New York: CABI.

Wilson EO. (1959) Adaptive shift and dispersal in a tropical ant fauna. *Evolution* 13: 122-144.

Woodward A, Silsbee DG, Schreiner EG, et al. (1994) Influence of climate on radial growth and cone production in subalpine fir (Abies lasiocarpa) and mountain

hemlock (Tsuga mertensiana). *Canadian Jounal of Forest Research* 24: 1133-1143.

Wright S. (1931) Evolution in mendelian populations. *Genetics* 16: 97-159.

Wright S. (1943) Isolation by distance. *Genetics* 28: 114-138.

Yang W, Guo Z, Huang C, et al. (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications* 5.

Young A, Boyle T and Brown T. (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* 11: 413-418.

Zohary M. (1937) Die Verbreitungsokologischen Verhaltnisse der Pflanzen Palastinas, I. *Beih Bot Zentralbl A* 56: 1-155.

APPENDIX

Appendix I Parentage Sensitivity Analysis.

Prop_Cand_Sample	Prop_Loci_Typed	Error Rate	Obs_Assignment_Rate
0.01	0.9867	0.1	0.21
0.01	0.9867	0.01	0.53
0.1	0.9867	0.1	0.26
0.2	0.9867	0.1	0.25
0.2	0.9867	0.01	0.53
0.3	0.9867	0.1	0.26

Sensitivity analysis of simulation parameters in Cervus. Proportion of candidate parents sampled was varied between 1 – 30%. Sequencing error rate was varied between 1-10%. Observed assignment rate is calculated at 80% confidence. At 95% confidence the identified parent offspring relationships were the same.