Landscape genomic insights into the historic migration of mountain hemlock in response to Holocene climate change

Jeremy S. Johnson, Keith D. Gaddis, David M. Cairns, Kranti Konganti, and Konstantin V. Krutovsky

PREMISE OF THE STUDY: Untangling alternative historic dispersal pathways in long-lived tree species is critical to better understand how temperate tree species may respond to climatic change. However, disentangling these alternative pathways is often difficult. Emerging genomic technologies and landscape genetics techniques improve our ability to assess these pathways in natural systems. We address the question to what degree have microrefugial patches and long-distance dispersal been responsible for the colonization of mountain hemlock (Tsuga mertensiana) on the Alaskan Kenai Peninsula.

METHODS: We used double-digest restriction-associated DNA sequencing (ddRADseq) to identify genetic variants across eight mountain hemlock sample sites on the Kenai Peninsula, Alaska. We assessed genetic diversity and linkage disequilibrium using landscape and population genetics approaches. Alternative historic dispersal pathways were assessed using discriminant analysis of principle components and electrical circuit theory.

KEY RESULTS: A combination of decreasing diversity, high gene flow, and landscape connectivity indicates that mountain hemlock colonization on the Kenai Peninsula is the result of long-distance dispersal. We found that contemporary climate best explained gene flow patterns and that isolation by resistance was a better model explaining genetic variation than isolation by distance.

CONCLUSIONS: Our findings support the conclusion that mountain hemlock colonization is the result of several long-distance dispersal events following Pleistocene glaciation. The high dispersal capability suggests that mountain hemlock may be able to respond to future climate change and expand its range as new habitat opens along its northern distribution.

KEY WORDS: ddRAD; genotyping-by-sequencing; isolation by resistance; landscape genetics; long distance dispersal; microrefugia; Pinaceae; Pleistocene; Tsuga mertensiana

Quaternary glaciations are partially 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 their 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 (Bernabo and Webb, 1977; Davis, 1981). Although colonization from glacial macrorefugia in southern ranges or out of northern microrefugia can both lead to identical contemporary species distributions, the result of each migratory pathway foreshadows vastly different potential short and long-term futures for species in northern habitats. For instance, rare long-distance dispersal can lead to a genetic bottleneck, decreasing genetic variation in expanding populations (Hewitt, 2000). In the face of ongoing climatic change, this reduction in genetic diversity could limit the ability of northern species to adapt to shifting local environments. Nevertheless, disentangling alternative historic dispersal pathways of northern tree species is often difficult.
High-throughput genomic sequencing and molecular approaches have improved our ability to address migratory pathways in natural systems (Johnson et al., 2016). In the past, even when species were distributed in geographically isolated patches, suggesting preglacial refugia, the occurrence of rare long-distance dispersal leading to establishment in remote populations could not be discounted. The importance of long-distance dispersal as a mechanism explaining rapid recolonization of forest species after glacial retreat (sensu Clark et al., 1998) has been an active area of research in temperate forests (Cwynar and MacDonald, 1987; Hewitt, 1996; Clark et al., 1998), but difficulties in tracking highly dispersed propagules (e.g., seed and pollen) on the landscape have made a detailed understanding of long-distance dispersal difficult (Cain et al., 2000). High-throughput next-generation sequencing (NGS) technologies are able to improve 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 single nucleotide polymorphisms (SNPs) (Johnson et al., 2017). We can combine genomic information with improved species distribution modeling (SDM) to more effectively assess dispersal and gene flow pathways in natural systems.

Patches established by rare long-distance dispersal have several genetic characteristics that would distinguish them from those that are microrefugial remnants of past populations (Table 1) (Nichols and Hewitt, 1994; McLachlan et al., 2005). First, the establishment of a population via a single long-distance dispersal event would reduce genetic diversity overall (Ibrahim et al., 1996; Hewitt, 1996, 2000) 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 with those established by long-distance dispersal or distributed along the edge of patches that weathered past glaciations and later contributed to subsequent distributions of this species exist in the Kenai lowlands of the peninsula, 1500 yr ago (Cwynar, 1990; Jones et al., 2009). Nevertheless, populations of this species were distributed in geographically isolated patches, 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 giving it the potential for long-distance colonization and gene flow. Differentiating between microrefugial and long-distance colonization pathways presents an opportunity for us to gain insight into how northern tree species respond to climatic change (Roberts and Hamann, 2015).

**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 postglacial retreat? We investigated this question...
by using two approaches to evaluate genomic data collected using a genotyping-by-sequencing (GBS) approach based on double-digest restriction-associated DNA sequencing (ddRADseq) (Peterson et al., 2012). First, we compared the genetic pattern of a geographically isolated patch of the study species in the Kenai lowlands to other patches spread throughout the Kenai Peninsula. If isolated lowland patches of mountain hemlock represent microrefugia, they are expected to be in linkage equilibrium, have higher genetic diversity, and equivalent or higher pairwise population genetic difference when compared with other populations in the sampling area. On the other hand, if they were more recently colonized through long-distance dispersal, we would expect to see lower genetic diversity, higher LD, and little difference in population genetic structure. Second, we evaluated the degree to which connectivity within the Kenai Peninsula has been structured by past glacial climates or more contemporary conditions. If populations are remnants of preglacial environments, we would expect past climates to better explain observed genetic differences among sampling locations, whereas a more recent colonization would be better explained by a contemporary climate.

**MATERIALS AND METHODS**

**Study species**—Mountain hemlock (Bong.) Carrière (Pinaceae) 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 along the Kenai coastal, alpine, and subalpine 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 (Means, 1990; Taylor, 1995). 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 are related to length of growing season, a function of winter temperature and snow pack, and summer temperatures and moisture availability (Taylor, 1995; Peterson and Peterson, 2001). The growth of high-elevation mountain hemlock correlates 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**—The Kenai Peninsula, Alaska 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), we 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 (Carlstrom, 1964). The Kenai Mountains are composed of white spruce and mountain hemlock, where the latter dominates the alpine tree line (800 m a.s.l.). 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 tree line. 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 not covered by the Harding Ice Field, 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 yr ago (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**—We sampled 8–10 mountain hemlock individuals at each of eight different sampling sites covering the three ecoregions of the Kenai Peninsula during the summers of 2012 and 2013 (Fig. 1A, Table 2). Within the coastal ecoregion, we collected from three sites spaced between 35–50 km apart (N = 30). Within the mounain ecoregion, we collected from three sites spaced 30–48 km apart (N = 28). Within the Kenai lowlands, we 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 with other sampling sites in subsequent analyses to examine for characteristics of microrefugia, or long-distance dispersal establishment. In addition to the sites collected from the three Kenai ecoregions, we sampled one site north of the peninsula, near Anchorage (N = 9), which was 73 km from Discovery Well and located 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 because it is separated from the Kenai Peninsula by the Turnagain arm of Cook Inlet. At each location, we collected foliage (approximately 15 cm long branch tips with needles) for DNA extraction. We desiccated fresh needle tissue in silica gel (Colpaert et al., 2005), then placed it in a −20°C freezer to await DNA extraction and sequencing. The location of each sample was recorded using a handheld global positioning system (±3 m).

**Genomic marker development**—We used a GBS approach, ddRADseq (Peterson et al., 2012), to genotype the mountain hemlock individuals. DNA was extracted from ground tissue of sampled mountain hemlock needles using a modified CTAB protocol (Doyle and Doyle, 1987). We tested different restriction enzyme (RE) pairs using frequent cutter enzymes targeting ~4-bp long restriction sites (MspI, MluCI, NlaIII) in combination with a less-frequent
cutter enzyme targeting longer (~6 bp) restriction sites that are less frequent in a genome (SphI and EcoRI). Three samples (one from each ecoregion) were digested separately with five RE pairs (SbfI-EcoRI, SphI-EcoRI, EcoRI-MspI, SphI-MluCI, and NlaIII-MluCI) and fragments were assessed on an Agilent Bioanalyzer (Agilent Technologies) to adjust the fraction of the genome sampled to approximately 1%. On the basis of the bioanalyzer reports, we chose SphI-MluCI with a fragment size of 350 bp for paired-end (PE) sequencing using the Illumina HiSEQ. 2000 sequencer (Illumina). We 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. GBS allowed us to simultaneously sequence and identify variants across all individuals using the selected RE pair and 350-bp fragments. All library preparation and sequencing were conducted at the University of Texas Genomic Analysis and Sequencing Facility.

For sequence analysis, we used a standard GBS bioinformatics analysis performed by the Texas A&M Institute for Genome Sciences and Society. This analysis assessed read quality, de-multiplexed samples, generated de novo contig assembly, aligned reads onto contigs, and called variants (SNPs) using the dDocent pipeline (Puritz et al., 2014). This research used de novo assembly strategies for a nonmodel organism. Quality SNP filtering was accomplished using VCFtools (Danecek et al., 2011). We used a 10x coverage cutoff, and Phred quality score > 30. We filtered the SNPs to remove possible sequencing error and duplicate paralogous sequences. We selected one SNP per RAD tag. We discarded SNPs that failed to genotype in greater than 80% of individuals and SNPs with a minor allele frequency less than 0.05.

<table>
<thead>
<tr>
<th>Sample location</th>
<th>Code</th>
<th>Ecoregion</th>
<th>Location (dd)</th>
<th>N</th>
<th>Nei’s unbiased average expected heterozygosity ( H_e )</th>
<th>Allelic diversity ( A )</th>
<th>Standardized index of association ( dr )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchorage Bowl</td>
<td>ANCH</td>
<td>Anchorage Bowl</td>
<td>−149.650W 61.174N</td>
<td>9</td>
<td>0.447</td>
<td>2.043</td>
<td>0.005</td>
</tr>
<tr>
<td>Discovery Well</td>
<td>DW</td>
<td>Kenai Lowlands</td>
<td>−150.837W 60.793N</td>
<td>10</td>
<td>0.432</td>
<td>2.045</td>
<td>0.002</td>
</tr>
<tr>
<td>Exit Glacier</td>
<td>EG</td>
<td>Kenai Coast</td>
<td>−149.658W 60.182N</td>
<td>10</td>
<td>0.460</td>
<td>2.046</td>
<td>0.011</td>
</tr>
<tr>
<td>Kenai Fjords NP</td>
<td>KEFJ1</td>
<td>Kenai Coast</td>
<td>−149.768W 59.664N</td>
<td>10</td>
<td>0.465</td>
<td>2.046</td>
<td>0.006</td>
</tr>
<tr>
<td>Kenai Fjords NP</td>
<td>KEFJ2</td>
<td>Kenai Coast</td>
<td>−150.171W 60.530N</td>
<td>10</td>
<td>0.450</td>
<td>2.045</td>
<td>0.021</td>
</tr>
<tr>
<td>Palmer Creek</td>
<td>PC</td>
<td>Kenai Mountains</td>
<td>−149.536W 60.822N</td>
<td>10</td>
<td>0.440</td>
<td>2.054</td>
<td>0.011</td>
</tr>
<tr>
<td>Snug Harbor</td>
<td>SGH</td>
<td>Kenai Mountains</td>
<td>−149.689W 60.416N</td>
<td>8</td>
<td>0.469</td>
<td>2.033</td>
<td>0.006</td>
</tr>
<tr>
<td>Sky Line</td>
<td>SL</td>
<td>Kenai Mountains</td>
<td>−150.171W 60.530N</td>
<td>10</td>
<td>0.440</td>
<td>2.058</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Figure 1. (A) Mountain hemlock sample locations (green circles) on the Kenai Peninsula, Alaska. The Discovery Well (DW) is a geographically isolated stand of hemlock. The Kenai Peninsula consists of three ecoregions: Kenai Coast, Kenai Mountains, and the Kenai Lowlands. (B) 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. (C) Electrical current map output from Mid-Holocene (~6000 yr ago) landscape resistance in Circuitscape. See Table 2 for site level abbreviations.
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**—We examined the level of genetic variation and pattern of genetic diversity within each of our sampling locations. We estimated genetic diversity using both Nei’s unbiased average expected heterozygosity, $H_e$, and allelic richness, $A$, as measures of gene diversity (Nei, 1978), and the standardized index of association, $r_{st}$ (Agapow and Burt, 2001), as a summary of multilocus LD. Linkage disequilibrium is simply the statistical correlation of alleles at different loci (Agapow and Burt, 2001; Flint-Garcia et al., 2003). The benefit of assessing LD is that it provides some insight into the biological and demographic processes shaping a population. For example, over time, LD is expected to decrease. However, if a population is under selection, experiencing drift or migration, then LD should be further from equilibrium (Flint-Garcia et al., 2003).

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., 2014, 2015).

We examined the partitioning of genetic variance among our sampling sites and the three ecoregions plus the Anchorage site to better understand demographic history. To determine the distribution of genetic variation among sampling sites, we performed a hierarchical analysis of molecular variance (AMOVA) (Excoffier et al., 1992). To determine the level of gene flow and genetic differences between sampling sites, we 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 $F_{st}$ between the eight sample sites were all calculated in GenALEX (Peakall and Smouse, 2006). We used discriminant analysis of principal components (DAPC) to assess population subdivision and clustering. This method is robust to violations of population subdivision rules and uncertain migration models that are required in other packages and provides an alternative method to assess population genetic structure (Jombart et al., 2010). It also summarizes the between-population genetic variation while minimizing the within-population variation by transforming multilocus genotype information, using principal component analysis, into uncorrelated variables as an input into discriminant analysis. This approach is useful with extremely large genomic data sets made up of thousands of SNPs. We 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**—We used maximum entropy modeling to create two SDMs representing the contemporary and mid-Holocene (6000 yr ago) 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.orl.gov) databases. To insure that botanical observations were of high quality, we filtered the herbarium data. We removed records that had uncertainly associated with georeferencing, duplicates, and nonwild individuals (Appendix S1, see Supplemental Data with the online version of this article). Contemporary (Hijmans et al., 2005) and mid-Holocene (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 Worldclim 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. We 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 us to model the climatic niche of mountain hemlock and thus identify the probability of their contemporary and mid-Holocene geographic distribution. This approach is robust to irregularly sampled data such as the herbarium records we used (Phillips et al., 2009). We ran the models using linear, quadratic, and product features. We 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 mid-Holocene bioclimatic variables. We parameterized the model with 15 replicates and 5000 iterations with 75% of data used to calibrate the model and 25% 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**—We tested our predictions of genetic characteristics of alternative population histories by comparing our genetic measures among sampling locations. We compared the mean values of $H_e$, $A$, and $r_{st}$ 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 whether the Discovery Well has been historically more separated from the rest of the region, we compared the pairwise $F_{st}$ values 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. To test whether long distance dispersal has a directional signal, we regressed site level $H_e$, $A$, and $r_{st}$ against distance from the proposed colonizing source of mountain hemlock in Prince William Sound. With increasing distance from the colonizing source, a significant decrease in $H_e$ and $A$ and an increase in $r_{st}$ will support our expectation for long-distance dispersal, whereas an increase in $H_e$ and a decrease in $r_{st}$ will indicate recolonization via microrefugia. As an additional level of analysis, we
used the computer program Migrate-n (Etterson et al., 2016) to estimate \( N_e \) and test directionality of gene flow between sample sites.

We tested whether genetic variation could be explained by three alternate models of landscape limitations: IBD, historic IBR, and contemporary IBR. We tested the three hypotheses using simple and partial Mantel’s test in the program PASSaGE 2.0.11.6 (Rosenberg and Anderson, 2011). This method tests the null hypothesis of no difference between two symmetrical matrices, in this case, distance matrices \((F_{ST})\), historic resistance, contemporary resistance, and geographic distance) (Mantel, 1967; McCune and Grace, 2002). Next, we used partial Mantel tests to compare the alternative resistance models to pairwise \( F_{ST} \) after controlling for geographic distance. Each test was permuted 1000 times. To avoid method dependent results, we chose to use a redundancy analysis as a secondary assessment of the influence of landscape resistance and geographic distance on genetic variation. We used RDA implemented in the R package vegan (Oksanen et al., 2016). RDA is a useful analysis where associations between a response variable (genetic variation) can be studied with respect to a variety of predictor variables (resistance surface and geographic distance). In essence, RDA is a form of multiple regression with a single response variable and multiple predictor variables.

**RESULTS**

**Genomic marker development**—For 77 individuals, GBS resulted in 41,057,267 unique sequencing reads and 50,952 de novo assembled contigs with an average per nucleotide read depth greater than 30x. Through the dDocent (Puritz et al., 2014) pipeline, 171,019 putative SNPs were identified throughout the mountain hemlock genome. Quality filtering reduced the number of SNPs to 6124 loci using VCFtools software (Danecek et al., 2011). All 6124 SNPs were found to be selectively neutral based on outlier \( F_{ST} \) analysis in the program 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 long-distance dispersal or through microrefugia. Discovery Well genetic diversity \((H_e)\) was statistically lower than the other seven sampling locations \((P = 0.005, 95\% CI 0.434–0.467)\) as would be expected under long-distance dispersal. \(H_e\) 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 2). 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 \(H_e\). Sites SGH \((H_e = 0.468, P = 0.005, 95\% CI 0.434–0.467)\) and KEFJ1 \((H_e = 0.464, P = 0.018, 95\% CI 0.434–0.467)\) had statistically higher \(H_e\) than in other sample sites (online Appendix S2). When allelic richness was assessed, the DW site was no different than the rest of the Kenai Peninsula \((A = 2.045, P = 0.6, 95\% CI 2.036–2.05)\).

All sites were in LD to some degree compared with neutral expectations. However, at the Discovery Well, LD was statistically lower than at the other sites \((P = 0.014, 95\% CI 0.001–0.016)\), which is indicative of the expectations we assumed for a microrefugia. The standardized index of association, \(r_{LDR} \), 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 2). Again, the magnitude of the signal was low across all sample sites, and only PC had a statistically higher LD than all other sample sites (Appendix S2).

The hierarchical AMOVA found no significant difference among regions \((F_{ST} = 0.003, P = 0.572)\) 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). 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 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 (Table 4). These results indicate 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. 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 we regressed \(H_e, \ A, \) and \(\overline{r}_{LDR}\) against distance from the predicted colonizing source near Prince William Sound, we found that \(H_e\) decreased significantly with increasing distance \((P = 0.022, R^2 = 0.6091, \text{Fig. 2})\). 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, neither \(A\) \((P = 0.771, R^2 = 0.0152)\) and \(\overline{r}_{LDR}\) \((P = 0.584, R^2 = 0.0528)\) showed any statistical trend. When we use the program Migrate-n (Etterson et al., 2016) to estimate \(N_e\) and test directionality of gene flow between sample sites, we failed to identify directionality in gene flow. The Migrate-n results were inconclusive and did not contribute to further interpretation.

**Population subdivision**—Discriminant analysis of principle components clustered individuals into three groups approximately corresponding to geographic position (Fig. 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 \((n \text{ samples sites} – 1)\). The proportion of conserved variance was 30%.

The DW clustered near ANCH and PC in the northwestern Kenai Mountains. The Kenai Mountain sites, SL and SGH, clustered together with the Kenai Coast site EG consisting of a southcentral Kenai Mountains cluster. The two Kenai Fjords coastal sites

<table>
<thead>
<tr>
<th>Site</th>
<th>ANCH</th>
<th>DW</th>
<th>EG</th>
<th>KEFJ1</th>
<th>KEFJ2</th>
<th>PC</th>
<th>SGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>DW</td>
<td>0.0123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EG</td>
<td>0.0119</td>
<td>0.0135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEFJ1</td>
<td>0.0120</td>
<td>0.0142</td>
<td>0.0093</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEFJ2</td>
<td>0.0119</td>
<td>0.0133</td>
<td>0.0106</td>
<td>0.0096</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>0.0127</td>
<td>0.0129</td>
<td>0.0124</td>
<td>0.0127</td>
<td>0.0113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGH</td>
<td>0.0148</td>
<td>0.0165</td>
<td>0.0105</td>
<td>0.0113</td>
<td>0.0122</td>
<td>0.0140</td>
<td></td>
</tr>
<tr>
<td>SL</td>
<td>0.0137</td>
<td>0.0130</td>
<td>0.0116</td>
<td>0.0127</td>
<td>0.0119</td>
<td>0.0121</td>
<td>0.0135</td>
</tr>
</tbody>
</table>
also clustered together as a Kenai Coast group. The northwestern 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 northwestern group. The overall pattern of clustering suggested a population structure consistent with the geography of three ecoregions on the Kenai.

**Landscape variables**—Maxent performance, as indicated by AUC, was good overall. The AUC score for mountain hemlock was 0.916 ± 0.012 (mean ± SD) for the contemporary model and 0.912 ± 0.006 (mean ± SD) for the mid Holocene model. 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 in assessing what climatic factors were important for determining that niche. The habitat suitability raster was inverted into resistance, and the resistance raster was used in Circuitscape to calculate pairwise resistance between all 8 sample locations (Fig. 1B, C). The two resistance surfaces were correlated ($R^2 = 0.74, P < 0.01$), suggesting a degree of temporal similarity.

**Dispersal limitation**—On the basis of the Mantel tests, we found no statistical effect of landscape on mountain hemlock dispersal limitation and mid-Holocene connectivity on the Kenai Peninsula at 95% confidence. Tests of IBD found weak correlation between $F_{ST}$ and geographic distance ($R^2 = 0.050, P = 0.181$). Furthermore, tests of IBR found moderate correlation between $F_{ST}$ and contemporary landscape resistance ($R^2 = 0.171, P = 0.083$) and low correlation between $F_{ST}$ and mid Holocene resistance ($R^2 = 0.032, P = 0.275$). On the basis of the redundancy analysis, we found a statistically significant effect of landscape on mountain hemlock genetic variability. Contemporary climate explained more variance ($R^2 = 0.289, P = 0.004$) than did mid Holocene climate ($R^2 = 0.256, P = 0.009$) (Table 5). In both tests, landscape resistance based on contemporary climate had the highest $r$ and lowest $P$ value, indicating, of the models considered, it looks to be the most plausible explanation for the observed population genetic structure. Partial Mantel ($R^2 = 0.131, P = 0.087$) and partial RDA ($R^2 = 0.218, P = 0.007$) tests found contemporary IBR to explain moderate correlation when accounting for geographic distance (Table 5). When low $F_{ST}$ values were considered along with these results, 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 mid-Holocene 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.

**DISCUSSION**

Our results suggest that colonization of the Kenai Peninsula by mountain hemlock following the last glacial maximum was likely the results of range expansion resulting from numerous long-distance dispersal events and high levels of gene flow across a landscape with decreasing resistance.

---

**FIGURE 2** Linear regression of $H_s$ against distance from expanding front. $H_s$ 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 smoothed lowess fit.

**TABLE 4.** Student’s t test of pairwise $F_{ST}$. Pairwise $F_{ST}$ was compared between each sample site and all other sample sites against all sample sites without the focal patch included. Bold values are significant at 95% confidence.

<table>
<thead>
<tr>
<th>Test</th>
<th>t</th>
<th>Means of patch vs. all</th>
<th>P</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANCH vs. all sites</td>
<td>0.7587</td>
<td>0.0127–0.0123</td>
<td>0.4593</td>
<td>−0.0007 to 0.0015</td>
</tr>
<tr>
<td>DW vs. all sites</td>
<td>2.7468</td>
<td>0.0136–0.0120</td>
<td>0.0199</td>
<td>0.0003 to 0.0029</td>
</tr>
<tr>
<td>EG vs all other sites</td>
<td>−2.3033</td>
<td>0.0113–0.0127</td>
<td>0.0427</td>
<td>−0.0027 to 0.0006</td>
</tr>
<tr>
<td>KEFJ1 vs. all other sites</td>
<td>−1.3917</td>
<td>0.0116–0.0126</td>
<td>0.1989</td>
<td>−0.0027 to 0.0006</td>
</tr>
<tr>
<td>KEFJ2 vs. all other sites</td>
<td>−2.0934</td>
<td>0.0115–0.0127</td>
<td>0.0564</td>
<td>−0.0024 to 0.0003</td>
</tr>
<tr>
<td>PC vs. all other sites</td>
<td>0.4267</td>
<td>0.0125–0.0123</td>
<td>0.6738</td>
<td>−0.0008 to 0.00123</td>
</tr>
<tr>
<td>SGH vs. all other sites</td>
<td>1.2864</td>
<td>0.0132–0.0121</td>
<td>0.2369</td>
<td>−0.0008 to 0.0030</td>
</tr>
<tr>
<td>SL vs. all other sites</td>
<td>0.5641</td>
<td>0.01260123</td>
<td>0.5782</td>
<td>−0.0007 to 0.0012</td>
</tr>
</tbody>
</table>

---
geographic distance (Geog) to matrices of geographic distance (IBD) and isolation by resistance (IBR). IBD compared matrices of geographic distance (Geog) to matrices of \( F_{ST} \), IBR compared contemporary (Contemp) and mid-Holocene (Holo) (~6000 yr ago) resistance values, calculated using Circuitscape, to matrices of \( F_{ST} \). Partial RDA and Mantel’s test were used to test contemporary and mid-Holocene IBR while accounting for geographic distance. \( p \) values and \( R^2 \) are reported.

<table>
<thead>
<tr>
<th>Model</th>
<th>Hypothesis</th>
<th>Test</th>
<th>( R^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene ~ Geog</td>
<td>IBD</td>
<td>Mantel</td>
<td>0.050</td>
<td>0.181</td>
</tr>
<tr>
<td>Gene ~ Contemporary Resistance</td>
<td>IBD</td>
<td>Mantel</td>
<td>0.171</td>
<td>0.083</td>
</tr>
<tr>
<td>Gene ~ Holocene resistance</td>
<td>IBD</td>
<td>Partial Mantel</td>
<td>0.131</td>
<td>0.087</td>
</tr>
<tr>
<td>Gene ~ Contemp</td>
<td>Geog</td>
<td>IBD</td>
<td>Partial Mantel</td>
<td>0.011</td>
</tr>
<tr>
<td>Gene ~ Holo</td>
<td>Geog</td>
<td>IBD</td>
<td>Partial Mantel</td>
<td>0.289</td>
</tr>
<tr>
<td>Gene ~ Holoc + Geog</td>
<td>IBD + IBD</td>
<td>RDA</td>
<td>0.256</td>
<td>0.009</td>
</tr>
<tr>
<td>Gene ~ Contemp + Geog</td>
<td>IBD + IBD</td>
<td>RDA</td>
<td>0.218</td>
<td>0.007</td>
</tr>
<tr>
<td>Gene ~ Holoc</td>
<td>Geog</td>
<td>IBD</td>
<td>Partial RDA</td>
<td>0.185</td>
</tr>
<tr>
<td>Gene ~ Geog</td>
<td>IBD</td>
<td>RDA</td>
<td>0.067</td>
<td>0.11</td>
</tr>
</tbody>
</table>

FIGURE 3 Discriminant analysis of principle components (DAPC) of eight mountain hemlock sample regions individually into three groups approximately corresponding to geographic position. ANCH (blue), PC (light orange), and DW (purple) clustered into a northwestern Kenai Mountains group. KFJ1 (yellow) and KFJ2 (gold) clustered into a south-central Kenai Mountains group. The DW site showed some separation from the other sample locations despite its association with SGH (dark orange), and EG (light purple) clustered into a south-central Kenai Mountains group. The dotted line represents the minimum spanning network between sites and serves as an approximate measure of similarity.

that recolonization has occurred within recent geologic history following glacial retreat. Our 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 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. Our results showed that contemporary, and not mid-Holocene, climate best explained the between-site genetic patterns.

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 stature, and outcrossing (Petit and Hampe, 2006), contribute to most genetic variation being distributed within individuals and in populations that share a large fraction of total genetic variation between them (low \( F_{ST} \) values) (Neale and Savolainen, 2004; Petit and Hampe, 2006).

Given the assumptions of higher genetic drift in founding populations with a small effective population size (\( N_e \)) that can result in fixation of deleterious mutations in founding populations of forest trees, it is no small wonder that founding individuals survived long-distance dispersal 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 long-distance dispersal. Amele evidence exists demonstrating the occurrence of rare long-distance dispersal and gene flow events in plants, allowing them to survive and persist (Campbell et al., 1999; Alsos et al., 2007; Robledo-Arnuncio, 2011; Kremer et al., 2012). The paradox, however, does give us pause. Simberloff (2009) addressed the problem of long-distance dispersal 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 long-distance dispersal 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, when 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, when 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 and allowing a patch formed by long-distance dispersal to survive more easily. In addition, the constant arrival of pollen gene flow may mask the effects of a bottleneck and make its detection more difficult.

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 (Nichols and Hewitt, 1994; Hewitt, 1996, 2000). However, great care should be taken when interpreting genetic patterns that may have nuanced explanations (Petit et al., 2003; Mee and Moore, 2014). Our study has shown that the geographically isolated Discovery Well contains lower genetic diversity (lower $H_e$) compared with the surrounding mountain hemlock forests. 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 toward the lowlands ($R^2 = 0.61, P = 0.02$). 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 suitable for mountain hemlock, which could have served as source populations for seeds founding the initial Discovery Well population. The Kenai Lowlands are boreal in composition and have a much higher incidence of disturbance. 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 site. Subsequent and ongoing pollen gene flow likely proved sufficient to maintain genetic diversity in this geographically isolated patch.

We 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 (Flint-Garcia et al., 2003; Gupta et al., 2005). Conversely, factors such as outcrossing and high recombination and mutation rates can decrease LD relatively quickly (Flint-Garcia et al., 2003; Gupta et al., 2005). In particular, the size of a plant’s genome and its mating system have a large influence over the level of LD and how quickly it will break down. For example, Flint-Garcia et al. (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; Neale and Savolainen, 2004; 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 mid-Holocene and contemporary climates using circuit theory, we found that contemporary climates best explained gene flow patterns on the Kenai. Without accounting for geographic distance, resistance based on mid-Holocene climate was assessed, 22% of the genetic variation explained 26% of the genetic variation. When resistance based on mid-Holocene climate was assessed, 22% of the genetic variation ($P < 0.01$) was explained. IBD explained only 6% of the genetic variation at $P = 0.11$. When a partial RDA test was used to assess the contemporary climate IBR model after accounting for geographic distance, 22% of the variation was still explained at $P < 0.01$, 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**

Our research addressed two questions pertaining to historic gene flow on the Kenai Peninsula, Alaska. First, we have shown that mountain hemlock arrived on the peninsula between 1500 and 3000 yr ago (Reger et al., 2007; Jones et al., 2009), which should provide ample time for LD to begin to break down in the outcrossed species. Interestingly, the Discovery Well site had statistically lower LD than in our other sampled sites, which contradicts the idea of an origin due to long-distance dispersal based on our initial predictions. This result was surprising given the degree to which the 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 $r^2$ values were low, indicated that the entire region was likely remigrating following glacial retreat that occurred within the last 14,000 yr. This finding is evident despite the fact that the Discovery Well site had a statistically lower LD than the other sampled locations. We still argue that the level of LD was not consistent with the site being a microrefugium.

**Historic connectivity**—Our 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 genetic 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 data set down into synthetic axes, the 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 long-distance dispersal explanation. The weak population structure indicated 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.
isolated stands of mountain hemlock found in the lowlands are likely 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 Fst analysis and AMOVA, and moderate levels of diversity allowed mountain hemlock to respond to past 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.

Our 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 improve explanations about the demographic history of long-lived conifer species at high latitudes.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (grant number BCS-1333527) and a Texas A&M Institute for Genome Sciences and Society fellowship to J.S.J. We thank Ellen Gass, Trey Murphy, Charles Lafon, and Clint Magill for assistance in the field and laboratory. Logistical support was provided by Kenai Fjords NP, Kenai NWR, Chugach NF, and the U.S. Army. We thank two anonymous reviewers for constructive feedback on a previous version of this article.

DATA ACCESSIBILITY

Individual ddRAD sequences can be found in GenBank Short Read Archive accession numbers SAMN06342827 through SAMN06342903.

REFERENCES


