Genetic Divergence and Signatures of Natural Selection in Marginal Populations of a Keystone, Long-Lived Conifer, Eastern White Pine (*Pinus strobus*) from Northern Ontario

Vikram E. Chhatre¹*, Om P. Rajora¹,²

¹ Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada; ² Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, New Brunswick, Canada

### Abstract

Marginal populations are expected to provide the frontiers for adaptation, evolution and range shifts of plant species under the anticipated climate change conditions. Marginal populations are predicted to show genetic divergence from central populations due to their isolation, and divergent natural selection and genetic drift operating therein. Marginal populations are also expected to have lower genetic diversity and effective population size (*N_e*) and higher genetic differentiation than central populations. We tested these hypotheses using eastern white pine (*Pinus strobus*) as a model for keystone, long-lived widely-distributed plants. All 614 eastern white pine trees, in a complete census of two populations each of marginal old-growth, central old-growth, and central second-growth, were genotyped at 11 microsatellite loci. The central populations had significantly higher allelic and genotypic diversity, latent genetic potential (LGP) and *N_e* than the marginal populations. However, heterozygosity and fixation index were similar between them. The marginal populations were genetically diverged from the central populations. Model testing suggested predominant north to south gene flow in the study area with curtailed gene flow to northern marginal populations. Signatures of natural selection were detected at three loci in the marginal populations; two showing divergent selection with directional change in allele frequencies, and one balancing selection. Contrary to the general belief, no significant differences were observed in genetic diversity, differentiation, LGP, and *N_e* between old-growth and second-growth populations. Our study provides information on the dynamics of migration, genetic drift and selection in central versus marginal populations of a keystone long-lived plant species and has broad evolutionary, conservation and adaptation significance.

### Citation


### Editor

Nadia Singh, North Carolina State University, United States of America

### Received

September 13, 2013; Accepted April 17, 2014; Published May 23, 2014

### Copyright

© 2014 Chhatre, Rajora. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Funding

This study was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (RGPIN 170651) to Om P. Rajora. Vikram E. Chhatre was financially supported from the NSERC Discovery grant funds to Om P. Rajora and a graduate scholarship at Dalhousie University. Om P. Rajora held the Stora Enso Senior Chair in Forest Genetics and Biotechnology at Dalhousie University, which was supported by Stora Enso Port Hawkesbury Ltd., and the Senior Canada Research Chair in Forest and Conservation Genomics and Biotechnology at UNB, which was supported by the Canada Research Chair Program (CRC50-201869). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Competing Interests

The authors have declared that no competing interests exist.

* E-mail: Om.Rajora@unb.ca

† Current address: University of Maryland Center for Environmental Science, Appalachian Laboratory, Frostburg, Maryland, United States of America

### Introduction

Marginal populations are expected to provide the frontiers for adaptation, evolution and range shifts of plant species under the anticipated climate change conditions. Marginal populations are generally adapted to their sub-optimal habitats, and are predicted to be genetically diverged from central populations due to their isolation, and divergent natural selection and genetic drift operating therein [1,2]. The central-marginal hypothesis proposes that populations at the species’ range peripheries have lower genetic diversity and higher genetic differentiation than populations at the species’ center of abundance. This may in-part result from differential selection regimes, isolation, higher stochastic genetic drift, restricted gene flow, smaller census and effective population size (*N_e*) and sub-optimal habitats in marginal populations [1,2], and may consequently lead to genetic divergence from central populations. At range peripheries, organisms may experience a host of harsh environmental, climatic, edaphic, and nutrient conditions, and impediments to gene flow. In such environments, selection regimes different from those at the abundant center may operate. This may drive allele frequencies for selected genes, ultimately resulting in local adaptation. Thus, marginal populations are important for future evolution and adaptation of species and may serve as grounds for speciation [1,3]. On the other hand, intense competition for resources and abiotic stresses at the leading edge may cause marginal populations to have negative growth rates [4] and to become demographic sinks with reduced fecundity [3,5], as has been shown in lodgepole pine (*Pinus contorta*) [6]. Evolutionary success or failure of populations at range margins is also dependent upon the balance between gene flow from the abundant center and local adaptation [4,7]. For example, asymmetrical abundant gene flow from central to marginal populations can result in
maladaptation of populations at the range margins due to the infusion of maladaptive alleles from the center [7]. Naturally, the conservation significance and local adaptation of marginal populations have been debatable and depend upon their evolutionary potential for adaptation [1,3,9], particularly in the face of rapidly changing climate, emergence of new diseases and loss of habitat [9]. Therefore, it is crucial to understand patterns of genetic diversity, population structure and evolutionary processes such as natural selection, genetic drift and gene flow in central versus marginal populations, and genetic mechanisms underlying local adaptation in marginal populations, especially in long-lived, widely-distributed keystone plant species. This information is critical for forest trees as they are normally the keystone species in their ecosystems. Existence and survival of many flora and fauna in an ecosystem depend upon the existence of such keystone species.

While a number of studies have tested the central-marginal hypothesis in a diverse group of plants and animals, including forest trees [1,2], the issue remains unresolved for the lack of overwhelming evidence in favor of it. Forest trees, especially conifers, are particularly suitable for testing various central-marginal hypotheses, because of their longevity and wide geographical distribution accounting for environmental, ecological and selection heterogeneity over time and space. A number of studies in conifers have supported the central-marginal hypothesis, e.g., [10–12]; while many others did not, e.g., [13–15]. In recent years, landscape genetics of range margins has seen renewed interest, especially to understand local adaptation of marginal populations [16,17]. As a result, reports on detection of divergent selection in marginal populations of forest trees [18] and other plants [19] are beginning to emerge. However, the information is very scarce. We are not aware of any report on natural selection in in-situ natural marginal versus central populations of forest trees. Holliday et al. study [18] was conducted on samples from a common garden provenance test.

In long-lived plants, such as conifers, there is a general belief that old-growth populations harbor higher genetic diversity than second-growth populations. However, empirical data to support or contradict this view is scarce [20,21]. Nevertheless, it is important to take the population age into account when comparing central and marginal population genetic characteristics.

Eastern white pine (Pinus strobus L.) provides an ideal organism to test the central- marginal and old-growth second-growth hypotheses. It is an ecologically important keystone species of white pine ecosystems, economically important for timber production, and long-lived (>400 years), with wide geographical distribution in North America [22,23]. Its natural range extends in southern Canada from Newfoundland to extreme southeastern Manitoba to southeast to northern Georgia and northwestern South Carolina [22]. In the pre-colonial era, eastern white pine covered much of the eastern North America [22]. However, it has undergone heavy exploitation during and after the colonial era [23], and post-glacial range expansion and retraction [24]. Much of the original old growth eastern white pine has vanished [23]. Ontario and Quebec still have some eastern white pine old-growth stands, where it exists mostly as second-growth forest. The range of eastern white pine is expected to shift northwards under anticipated climate change conditions, such as in northern Ontario.

Eastern white pine (EWP) is a predominantly outcrossing species [14,25] and has high inbreeding depression. It has moderate to high levels of genetic diversity [13,26–30]. Rajora et al. [13,14] reported similar levels of allozyme-based genetic diversity, population structure and outcrossing levels in EWP from central Ontario and about 2,000 km apart disjunct marginal Newfoundland populations, although inbreeding levels estimated from the empty seed data were higher in the Newfoundland populations. However, there is a lack of information on genetic diversity and differentiation of central and marginal populations from the same regional part and entire natural range, as well as on any selection regimes underlying local adaptation that may be operational in marginal populations of this species. Also, information on genetic diversity in old-growth versus second-growth populations is scarce [28].

Here, we address the hypothesis that geographically marginal populations of eastern white pine have lower genetic diversity and higher genetic differentiation than and are genetically distinct from geographically central populations. We also asked the question whether old-growth populations have higher genetic diversity than second-growth populations. We undertook comprehensive genetic diversity and population structure analyses in a complete census of four old-growth and two second-growth populations of eastern white pine from its central and marginal distribution in northern Ontario, Canada. We also tested various models for gene flow between populations, estimated \( N_m \) and tested loci for signatures of natural selection.

**Materials and Methods**

**Ethics Statement – Field Sampling**

The studied eastern white pine populations are located on public lands in Ontario, Canada, which are not designated as protected areas. The field sampling was done by the employees of the Ontario Ministry of Natural Resources, which manages public forests there. Therefore, no specific permission was required for field sampling from the studied locations. Also, our study did not involve any endangered species.

**Study Populations, Experimental Design and Sampling**

We studied six EWP natural populations from its natural distribution range in northern Ontario, Canada (Fig. 1; Table 1). While the French River (FR) and Rawhide Lake (RH) stands are part of the continuously distributed EWP central populations, Galloway Lake (GL) populations are located at the species’ northern boundary and are somewhat geographically isolated (Fig. 1). The average age of the populations at sampling was: GL – 250 years; RH – 235 years; and FR – 100 years.

We chose to conduct an in-depth genetic analysis for a fewer populations from the same regional part of the natural distribution range of EWP in northern Ontario; first because the northern populations in this part of the EWP range are expected to shift range northwards and second the selection pressures in different parts of the marginal range of this species are likely to be different. Two populations in one area separated by 2 to 10 km, were studied as replicates of a population type (marginal, central, old-growth, second-growth), in order to minimize confounding effects of major eco-site and climate conditions by sampling far-distant populations from a different part of the species range. In order to minimize the effect of population age on central-marginal comparisons, we sampled two old-growth (OG) marginal, two old-growth central and two second-growth (SG) central populations. We could not locate second-growth marginal populations in our study area. We sampled all trees in a complete census of the marginal populations from the GL area [27,31]. In the continuous RH and FR populations, plots of the same number of trees (~100) were established to keep the sampling uniform across the populations. All trees in the established plots were censused, so as to minimize missing rare alleles, which may be a subject or
product of natural selection and/or genetic drift and contribute to the latent genetic potential [32].

The two populations from the Rawhide Lake area are located 30 km north of Elliot Lake, Ontario, in the Rawhide Lake Conservation Reserve established in 2000 by Ontario Ministry of Natural Resources (OMNR). This area lies to the north of Mississagi Provincial Park. Classified as ecological site 5E, it is characterized by rocky landscape rich in glacial history and Moraine deposits. Old and ecologically mature pure eastern white pine stands are located in the western part of the reserve and are interspersed with sugar maple (*Acer saccharum*) and jack pine (*Pinus banksiana*) dominated stands (Information source: Rawhide Lake Conservation Reserve Fact Sheet, OMNR, 2000).

The two populations from the French River area are located near Grundy Lake Provincial Park in Ontario, Canada, occupying an area that was glaciated during Pleistocene and was subsequently dominated by EWP. Historical documents indicate three different periods of logging in this general area, the first of which was around 1870s by the Victoria Timber Company, which exported timber to the United States. The subsequent harvesting of mature eastern white pines occurred in two episodes starting in 1900 and 1930 respectively. The French River area population A is of post-fire origin, whereas French River B of post-harvest origin (Dianne Othmer, OMNR, pers. comm.).

The Galloway Lake old pine area, located roughly 100 km north of Sault Ste. Marie, Ontario, hosts the two old-growth EWP

![Figure 1. Eastern white pine distribution range and location of six studied populations. doi:10.1371/journal.pone.0097291.g001](image_url)

<table>
<thead>
<tr>
<th>Population</th>
<th>Pop ID</th>
<th>Region</th>
<th>Pop Type</th>
<th>Size</th>
<th>Latitude (N)</th>
<th>Longitude (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rawhide Lake A</td>
<td>RH-A</td>
<td>Central</td>
<td>Old-growth</td>
<td>99</td>
<td>46 40' 16&quot;</td>
<td>82 38' 18&quot;</td>
</tr>
<tr>
<td>Rawhide Lake B</td>
<td>RH-B</td>
<td>Central</td>
<td>Old-growth</td>
<td>97</td>
<td>46 40' 18&quot;</td>
<td>82 37' 16&quot;</td>
</tr>
<tr>
<td>French River A</td>
<td>FR-A</td>
<td>Central</td>
<td>Second-growth</td>
<td>100</td>
<td>45 52' 14&quot;</td>
<td>80 28' 42&quot;</td>
</tr>
<tr>
<td>French River B</td>
<td>FR-B</td>
<td>Central</td>
<td>Second-growth</td>
<td>94</td>
<td>45 52' 53&quot;</td>
<td>80 21' 31&quot;</td>
</tr>
<tr>
<td>Galloway Lake A</td>
<td>GL-A</td>
<td>Marginal</td>
<td>Old-growth</td>
<td>122</td>
<td>47 13' 40&quot;</td>
<td>83 55' 50&quot;</td>
</tr>
<tr>
<td>Galloway Lake B</td>
<td>GL-B</td>
<td>Marginal</td>
<td>Old-growth</td>
<td>102</td>
<td>47 13' 07&quot;</td>
<td>83 56' 30&quot;</td>
</tr>
</tbody>
</table>

<p>| Table 1. Eastern white pine populations studied, their sizes and geographical coordinates. |
|---------------------------------------------------------------|---------------------------------------------------------------|</p>
<table>
<thead>
<tr>
<th>Population</th>
<th>Pop ID</th>
<th>Region</th>
<th>Pop Type</th>
<th>Size</th>
<th>Latitude (N)</th>
<th>Longitude (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rawhide Lake A</td>
<td>RH-A</td>
<td>Central</td>
<td>Old-growth</td>
<td>99</td>
<td>46 40' 16&quot;</td>
<td>82 38' 18&quot;</td>
</tr>
<tr>
<td>Rawhide Lake B</td>
<td>RH-B</td>
<td>Central</td>
<td>Old-growth</td>
<td>97</td>
<td>46 40' 18&quot;</td>
<td>82 37' 16&quot;</td>
</tr>
<tr>
<td>French River A</td>
<td>FR-A</td>
<td>Central</td>
<td>Second-growth</td>
<td>100</td>
<td>45 52' 14&quot;</td>
<td>80 28' 42&quot;</td>
</tr>
<tr>
<td>French River B</td>
<td>FR-B</td>
<td>Central</td>
<td>Second-growth</td>
<td>94</td>
<td>45 52' 53&quot;</td>
<td>80 21' 31&quot;</td>
</tr>
<tr>
<td>Galloway Lake A</td>
<td>GL-A</td>
<td>Marginal</td>
<td>Old-growth</td>
<td>122</td>
<td>47 13' 40&quot;</td>
<td>83 55' 50&quot;</td>
</tr>
<tr>
<td>Galloway Lake B</td>
<td>GL-B</td>
<td>Marginal</td>
<td>Old-growth</td>
<td>102</td>
<td>47 13' 07&quot;</td>
<td>83 56' 30&quot;</td>
</tr>
</tbody>
</table>

Pop = Population.
doi:10.1371/journal.pone.0097291.t001
populations used in this study. This area, classified as ecosite 4E [33], has complex topography consisting of rivers and lakes interspersed with rocky areas, with soil texture varying between fine sands and silt loams, and variable depth [23]. The mature EWP stands in this area are believed to have post-fire origin, mixed with hardwood species. The study populations A and B were small and had eastern white pine individuals of about 250 years, mixed with a variety of hardwood and conifer understory species [31].

Microsatellite Genotyping

We genotyped 390 trees from the four central populations at 11 nuclear microsatellite loci (Table S1) developed by Echt et al. [34]. The genotype data for 224 trees from the marginal GL populations was obtained from Rajora et al. [27]. To attain complete allele-size correspondence between RH, FR and GL datasets, we genotyped 10 individuals from each population in the same single run for each of the 11 microsatellites and confirmed that individual allele sizes were consistently identified correctly in all six populations. All microsatellite loci were in linkage equilibrium. The details of microsatellite genotyping methods, including PCR conditions, are provided in the Microsatellite Genotyping Text S1.

Data Analysis

Data analyses were performed in three sets: all six populations; central versus marginal populations; old-growth versus second-growth populations.

Genetic diversity. Standard measures of genetic diversity were calculated using GENEPOP [35]. In addition, number of private alleles per population, latent genetic potential (LGP) [36], and observed and expected genotypic additivity were calculated using GENEPOP [35]. In addition, one-sided and two-sided significance tests with 1000 permutations using FSTAT [43]. Significance of differences in genetic diversity estimates between groups of populations was tested using ANOVA and Duncan’s Multiple Range Test (DMRT) in the R environment [38].

Population Structure

The genetic structure of the studied populations was examined using several complementary methods: single-locus and multilocus \(F_{ST}\) using Weir and Cockerham’s method [37] in GENEPOP [35]; hierarchical AMOVA [39] in GENALEX [40]; and Bayesian modeling using STRUCTURE [41]. Nei’s [42] genetic distances between individual and groups of populations were calculated. To determine if \(F_{ST}\) between marginal populations was significantly higher than that between central populations, we applied one-sided and two-sided significance tests with 1000 permutations using FSTAT [43]. Significance of hierarchical AMOVA was tested with 999 permutations in GENALEX [40].

Allele frequency heterogeneity among pairs of populations was tested using Fisher’s exact test in GENEPOP [35]. Genetic relationships among populations were resolved using the Neighbor-Joining (NJ) method [44] in POPTREE [45]. We employed 5\( \times \)10\(^5\) bootstrap iterations for testing statistical support for clustering. Principal coordinate analysis (PCA) based on Nei’s [42] genetic distances was performed and the ordination of the populations on the first two principal components was plotted using GENALEX [40].

Bayesian analysis of the population structure was performed using the admixture model in STRUCTURE [41]. We assumed that all markers were unlinked (as they actually were), all trees were admixed and that allele frequencies were correlated between loci. The number of genetic clusters (\(K\)) tested varied from 1 through 6 when all six populations were analyzed, and from 1 through 4 when only central OG and SG populations were analyzed. We discarded the first \(10^5\) steps of each run as burn-in, to allow the parameters to converge, prior to collecting the data for the next \(10^5\) steps. Different combinations of burn-ins and MCMC chains (between \(0.25 \times 10^5\) and \(1 \times 10^5\)) were performed initially to find the point of convergence for our data set and \(1 \times 10^5\) steps of burn-in and MCMC chains each, per run were found to be sufficient. Twenty-five such runs for each putative value of \(K\) were performed to test for variance between the runs. To estimate the most optimal value of \(K\) for each level of population comparisons, we further processed the program output through STRUCTURE HARVESTER [46], which employs the \(\Delta K\) ad-hoc statistic of [47]. Once an optimal number of clusters was inferred, we performed \(10^5\) permutations for each of the 25 replicates of the chosen \(K\) using the Greedy algorithm in CLUMPP [48], in order to match the replicates as closely to each other as possible. Finally, barplots of cluster membership assignments were prepared with DISTRACT [49].

Isolation by distance was tested by regressing the logarithm of pair-wise geographic distance with pair-wise \(F_{ST}/1 - F_{ST}\) using Mantel test over 1,000 permutations. Additionally, to account for the bias introduced by hierarchical population structure, stratified and partial Mantel tests were performed using genetic clusters and geography as covariates following [50] using VEGAN R package [51].

Signatures of natural selection. We employed \(F_{ST}\) outlier approach in LOSITAN [52,53] to identify loci deviating from neutrality and potentially under selection. Mean neutral \(F_{ST}\) was estimated and applied for \(5 \times 10^5\) simulations to construct upper and lower 95% confidence limits. Markers with unusually high or low \(F_{ST}\) violating these thresholds were identified as candidates for selection. This analysis was done for three sets: all six populations, four central populations and three location-wise pooled populations. A built-in false positive correction was applied using a false discovery rate of 0.05.

Assuming the finite island model for detecting loci under selection can result in a large proportion of false positives due to confounding effects of shared history of populations and population structure or if the population samples are derived from hierarchically subdivided populations [54]. In such cases, a hierarchical island model [55] is better suited. In the hierarchical island model, Wright’s \(F\)-statistics is described by \(F_{ST}\) as variance among populations in-total and \(F_{CT}\) as variance among regions (central and marginal in this study) [54]. We used this model for detecting outlier loci potentially under selection using hierarchical \(F\)-statistics analysis following [54] by performing \(1 \times 10^5\) coalescent simulations in ARLEQUIN [56]. This model requires that the number of groups (regions) be equal to or larger than the actual number sampled and that the number of demes (populations within a region) simulated be more than the actual number sampled [54]. To satisfy these conditions, we chose 20 groups and 100 demes per group for the simulations. The resulting \(1 \times 10^5\) hierarchical fixation indices were used as null distributions to construct a 95% confidence interval and to identify loci outside the two-tailed significance threshold of \(P = 0.05\), as outliers.

For comparison, we also employed a Bayesian method to detect outlier loci based on differences in allele frequencies between four central and two marginal populations using the software BAYESCAN [57]. Analysis consisted of 30 pre-calculation pilot runs, each spanning 20,000 steps. The calculation stage included an initial burn-in period of \(2 \times 10^5\) steps followed by data collection...
for the next $10^3$ steps. Posterior probability $q$ values (FDR corrected $P$ values) were used to assign outlier status to markers. We performed a Genbank protein homology search (BlastX) to functionally annotate the sequences from which the markers found under natural selection were developed. The direction of selection on loci showing divergent selection was assessed by allele frequency changes in marginal populations.

**Gene flow and effective population size.** Bidirectional gene flow between pairs of populations and mutation-scaled $N_e$ were estimated using MIGRATE software [50]. This method assumes roughly equal distance between populations. Due to proximity of individual populations from the same study area, we selected one of the two study sites at each location as a representative sample for estimating gene flow. Bayesian model testing approach was applied to find the model that best explains the observed data at the three sampling locations. Three models were tested: north to south, south to north, both with unrestricted gene exchange between the two central populations, and a full symmetrical migration matrix encompassing all possible routes between all pairs of populations. Initial 20,000 Markov chain sweeps per locus were discarded as burn-in and data was then collected for another 20,000 sweeps at an increment of 100 sweeps. The analysis was repeated five times and estimates were combined over replicates. Relative geographical distance between pairs of populations was taken into consideration when estimating the rate of migration and $\theta$. The number of migrants per generation was estimated by multiplying parameter $M$ [50] by $\theta$. Mutation-scaled effective population sizes ($N_e$) for the six individual populations were estimated from a separate MIGRATE analysis under the following conditions. Two loci, RPS-20 and RPS-39 were removed from the analysis due to their $F_{ST}$ outlier status and 50 samples were randomly chosen per population. Ten short Markov chains and one long Markov chain were sampled per locus with the first 50,000 trees discarded as burn-in and the next 50,000 trees recorded over two replicates. Static heating was used and chains were allowed to swap. In addition, two replicate populations at each of the three sampling locations were combined into a pooled data set which was also analyzed with MIGRATE (50,000 trees discarded; 50,000 trees recorded). The $N_e$ was then estimated for individual and pooled populations from the $0$ parameter as follows: since $\theta = 4N_e \mu$, therefore, $N_e = 0.5 / \theta \mu$. The mutation rate ($\mu$) was assumed to be constant at $10^{-3}$ for all microsatellite loci [11,12,59]. To resolve the effect of environmental conditions on the dynamics of gene flow, information on historical wind patterns in the study area was obtained from the Atlas of Canada [60].

**Results**

**Genetic Diversity, Fixation Index, Latent Genetic Potential, and $N_e$.**

The results on genetic diversity measures, $F_{ST}$, LGP and $N_e$ for individual populations and means for central and marginal, OG and SG populations are reported in Table 2. The central SG population FR-A had the highest and the marginal OG population GL-B the lowest allelic and genotypic diversity and LGP (Table 2). In contrast, the marginal OG GL-B population had the highest and the central FR-B SG population the lowest observed ($H_e$) and expected ($H_e$) heterozygosity (Table 2). The central populations had 30 and marginal populations 6 private alleles ($A_p$). The central populations had significantly ($P<0.05$) higher allelic diversity ($A_T$, $A$ and $A_R$), $A_p$, expected genotypic additivity, and LGP than the marginal populations. However, the marginal populations had higher heterozygosity than the central populations but the differences were not significant ($P>0.03$).

The mean $F_{ST}$ ranged from 0.084 in GL-B to 0.183 in RH-B, with a mean of 0.139 over all populations. On average, $F_{ST}$ was lower in the marginal than central populations, but the differences were not significant (Table 2). Similar patterns of genetic diversity and LGP measures for central versus marginal populations were observed when we compared only two central OG populations with two marginal OG populations (Table 2). The central old-growth and central second-growth populations showed statistically similar allelic and genotypic diversity, LGP and $F_{ST}$ (Table 2).

$N_e$ ranged from 212 in the marginal OG GL-B population to 401 in the central SG FR-A population, with an overall mean of 307 (Table 2; Fig. S8). The central populations had significantly higher $N_e$ than the marginal populations ($P<0.01$), whereas central OG and central SG populations showed statistically similar $N_e$ (Table 2; Fig. S8). The confidence intervals for $N_e$ estimates did not overlap between the central and marginal populations (Fig. S8). Pooled population analysis also showed higher $N_e$ in the central populations (RH: 401; FR: 398) than in the marginal population (GL: 312). All $N_e$ estimates were within the bounds of 95% confidence intervals, which again did not overlap between the central and marginal populations (Fig. S8).

**Population Genetic Structure and Differentiation**

Well-defined population genetic structure was observed among the six populations, where marginal populations showed significant genetic divergence from the central populations. The multi-locus $F_{ST}$ ranged from 0.03 between RH-A and FR-A to 0.134 between FR-B and GL-B, with an overall mean of 0.083 among all six populations (Table S2; Table 3). The $F_{ST}$ estimates between the central and marginal populations were high and highly significant (Table 3), whereas between central OG and SG populations were low but still significant. Although $F_{ST}$ among central populations (0.008) was lower than that between two marginal populations (0.021), the differences were not statistically significant (one-tailed $P = 0.42$, two-tailed $P = 0.99$). The genetic structure and differentiation patterns from hierarchical AMOVA mirrored those observed from $F_{ST}$, confirming highly significant genetic differentiation between marginal and central populations (Table 3). Fisher’s exact tests revealed highly significant allele frequency heterogeneity between central and marginal populations (data not shown).

Genetic distances between central and marginal populations were orders of magnitude higher (0.22 to 0.26) than among central populations (0.006 to 0.016) (Table S2). The NJ tree and PCA from Nei’s [42] genetic distances clustered six populations into two distinct groups, separating the marginal from the central populations (Figs. S1, S2). No particular grouping of OG versus SG populations was observed, when the analyses were performed for only central populations (data not shown).

The Bayesian STRUCTURE analysis identified two distinct clusters ($K = 2$) (Fig. S3) among six populations; clearly separating marginal populations from central populations (Fig. 2A). Individuals had more than 95% membership in their assigned cluster (Fig. 2A). Results from an independent analysis using only OG central and OG marginal populations also yielded highly similar results, with $K = 2$, differentiating the marginal from central populations (Fig. 2B). When population structure was analyzed for central OG and SG populations using identical conditions, the $K$ suggested the optimal number of clusters to be four. The membership assignments of individuals across these four populations showed admixture among the four clusters (Fig. 2C). Thus,
Table 2. Genetic diversity parameters, fixation index and effective population size of eastern white pine populations.

<table>
<thead>
<tr>
<th>Region</th>
<th>Type</th>
<th>Pop</th>
<th>$A_t$</th>
<th>$A$</th>
<th>$A_e$</th>
<th>$A_p$</th>
<th>$A_e$</th>
<th>$H_o$</th>
<th>$H_e$</th>
<th>$F_{IS}$</th>
<th>LGP</th>
<th>$GA_o$</th>
<th>$GA_e$</th>
<th>$N_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Populations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>OG</td>
<td>RH-A</td>
<td>130</td>
<td>11.82</td>
<td>4.23</td>
<td>6</td>
<td>11.55</td>
<td>0.535</td>
<td>0.603</td>
<td>0.113</td>
<td>83.49</td>
<td>246</td>
<td>1226</td>
<td>312.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RH-B</td>
<td>129</td>
<td>11.73</td>
<td>4.22</td>
<td>10</td>
<td>11.57</td>
<td>0.496</td>
<td>0.608</td>
<td>0.183</td>
<td>82.56</td>
<td>264</td>
<td>1046</td>
<td>328.5</td>
</tr>
<tr>
<td></td>
<td>SG</td>
<td>FR-A</td>
<td>133</td>
<td>12.10</td>
<td>4.50</td>
<td>9</td>
<td>11.83</td>
<td>0.531</td>
<td>0.612</td>
<td>0.133</td>
<td>83.53</td>
<td>284</td>
<td>1242</td>
<td>400.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FR-B</td>
<td>113</td>
<td>10.27</td>
<td>3.59</td>
<td>5</td>
<td>10.17</td>
<td>0.457</td>
<td>0.557</td>
<td>0.181</td>
<td>73.54</td>
<td>210</td>
<td>956</td>
<td>353.1</td>
</tr>
<tr>
<td>Marginal</td>
<td>OG</td>
<td>GL-A</td>
<td>104</td>
<td>9.45</td>
<td>3.86</td>
<td>3</td>
<td>9.24</td>
<td>0.538</td>
<td>0.626</td>
<td>0.142</td>
<td>61.57</td>
<td>225</td>
<td>733</td>
<td>235.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GL-B</td>
<td>102</td>
<td>9.27</td>
<td>4.40</td>
<td>3</td>
<td>9.12</td>
<td>0.590</td>
<td>0.644</td>
<td>0.084</td>
<td>53.55</td>
<td>217</td>
<td>701</td>
<td>211.9</td>
</tr>
<tr>
<td>Overall mean</td>
<td></td>
<td></td>
<td>118.5</td>
<td>10.77</td>
<td>4.13</td>
<td>4.3</td>
<td>10.58</td>
<td>0.525</td>
<td>0.608</td>
<td>0.139</td>
<td>73.04</td>
<td>241</td>
<td>984</td>
<td>306.7</td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central vs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Populations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>OG+SG</td>
<td>126</td>
<td>11.48</td>
<td>4.13</td>
<td>7.5</td>
<td>11.28</td>
<td>0.503</td>
<td>0.594</td>
<td>0.153</td>
<td>80.78</td>
<td>251</td>
<td>1117</td>
<td>348.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>103</td>
<td>9.37</td>
<td>4.13</td>
<td>3</td>
<td>9.18</td>
<td>0.564</td>
<td>0.635</td>
<td>0.113</td>
<td>57.56</td>
<td>221</td>
<td>717</td>
<td>223.5</td>
<td></td>
</tr>
<tr>
<td>Central vs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central OG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central SG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central OG vs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central SG vs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal OG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal OG vs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal OG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pop = Population; OG = Old-growth; SG = Second-growth; $A_t$ = Total number of alleles; $A$ = Mean number of alleles per locus; $A_e$ = Mean effective number of alleles per locus; $A_p$ = Number of private alleles; $A_R$ = Allelic richness; $H_o$ = Mean observed heterozygosity; $H_e$ = Mean expected heterozygosity; $F_{IS}$ = Mean fixation index; LGP = Latent genetic potential; $GA_o$ = Observed genotypic additivity; $GA_e$ = Expected genotypic additivity; $N_e$ = Effective population size. Means followed by different letters are significantly different at $P<0.01$, based on ANOVA and Duncan’s multiple range test (DMRT).

doi:10.1371/journal.pone.0097291.t002
we can infer that the central OG and SG populations exhibit weak
or no population structure.

Various Mantel tests for isolation by distance were not
significant (regular Mantel test: \( P = 0.06 \); stratified Mantel test:
\( P = 0.09 \); Mantel test for hierarchical population structure using
genetic clusters: \( P = 0.09 \); and partial Mantel test with geography
as covariate: \( P = 0.07 \)).

Signatures of Natural Selection

The \( F_{ST} \) outlier test for six populations identified 5 loci violating
the 95% confidence interval expectations under neutrality (Fig.
S4). Two of these loci (RPS-20 and RPS-39) were candidates for
divergent and three (RPS-12, RPS-25b and RPS-50) for balancing
selection (Fig. S4a). When only central OG and marginal OG
populations were tested, the results were mostly similar except that
RPS-25b and RPS-50 were not candidates for balancing selection
(Fig. S5). The hierarchical \( F_{ST} \) analysis with all six populations also
confirmed these results where RPS-20 and RPS-39 loci were
detected as candidates for divergent selection (Fig. 3); however,
only RPS-12 was detected as candidate for balancing selection.

One more locus RPS-127 was identified as an outlier potentially
under balancing selection when \( F_{CT} \) was used to detect outliers in
place of \( F_{ST} \) from the hierarchical analysis (Fig. 3). In contrast, no
loci showed signatures of selection when only central group of OG
and SG populations were compared (Fig. S4b). At RPS-20 and
RPS-39, several high frequency alleles were either exclusive to
central or marginal populations or showed significant directional
change in the marginal populations (Fig. 4). When two populations

Table 3. Genetic differentiation of eastern white pine populations from \( F_{ST} \) and hierarchical AMOVA analyses.

<table>
<thead>
<tr>
<th>Region</th>
<th>( F_{ST} )</th>
<th>( P )</th>
<th>( \Phi_{ST}/\Phi_{RT} )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>0.008</td>
<td>0.001</td>
<td>0.014</td>
<td>0.001</td>
</tr>
<tr>
<td>Central OG</td>
<td>0.006</td>
<td>0.002</td>
<td>0.010</td>
<td>0.001</td>
</tr>
<tr>
<td>Central SG</td>
<td>0.010</td>
<td>0.001</td>
<td>0.017</td>
<td>0.001</td>
</tr>
<tr>
<td>Marginal</td>
<td>0.021</td>
<td>0.001</td>
<td>0.038</td>
<td>0.001</td>
</tr>
<tr>
<td>Central – Marginal</td>
<td>0.104</td>
<td>0.001</td>
<td>0.169</td>
<td>0.001</td>
</tr>
<tr>
<td>Central OG – SG</td>
<td>0.005</td>
<td>0.001</td>
<td>0.008</td>
<td>0.001</td>
</tr>
<tr>
<td>All populations</td>
<td>0.083</td>
<td>0.001</td>
<td>0.137</td>
<td>0.001</td>
</tr>
</tbody>
</table>

OG = Old-growth; SG = Second-growth.

Figure 2. Population structure in central and marginal populations of eastern white pine. Bar plot of estimated membership coefficient
(Q) of eastern white pine individuals from (A) all six central and marginal populations showing two groups corresponding to central and marginal
populations at \( K = 2 \), (B) two marginal old-growth and two central old-growth populations showing two groups corresponding to central and
marginal populations at \( K = 2 \), and (C) two central old-growth and two central second-growth populations showing no separate groups at \( K = 4 \).

doi:10.1371/journal.pone.0097291.g002
were pooled for each of the three study locations, the number of outlier loci decreased to two (divergent candidate: RPS-39; balancing candidate: RPS-12) (Fig. S6). Bayesian analysis using BAYESCAN showed very similar results (Fig. S7). Four loci were identified as outliers, with RPS-39 reconfirmed as putatively being under diversifying selection and RPS-12 and RPS-50 under balancing selection. In addition, locus RPS-2 was also identified as balancing selection candidate. Overall, RPS-39 and RPS-20 were consistently identified as candidate microsatellite loci under divergent selection and RPS-12 as a candidate for balancing selection by almost all analyses. Thus, we consider only these three microsatellite loci as putative loci under selection in the marginal populations. RPS-20 showed significant homology to a hypothetical protein of unknown function, and RPS-12 to KAOT1–09906 and EAI-06382 proteins (Table S3).

Genetic Diversity and Ne

Our results demonstrate that EWP central populations have significantly higher allelic diversity, genotypic diversity (Gt), Ne, and LGP than the marginal populations but statistically similar heterozygosity and FIS. These patterns were consistent when two OG marginal populations were compared with two OG central or

Gene Flow

Of the three gene flow models tested, the north to south (marginal to central) model had the highest probability (P = 1.00) based on the comparison of marginal likelihood Bayes factors (Table S4). The number of migrants per generation received from northern populations is depicted in Fig. 5. Although this model allowed for free exchange of migrants between the two central locations, French River population still received substantially more migrants from Rawhide Lake population than vice-versa. Estimates of migration parameters for individual populations were well within the confidence intervals for the tested model. The observed pattern of gene flow is probably a result of the prevailing north to south wind flow in the study area.

Discussion

Genetic Diversity and Ne

Our results demonstrate that EWP central populations have significantly higher allelic diversity, genotypic diversity (Gt), Ne, and LGP than the marginal populations but statistically similar heterozygosity and FIS. These patterns were consistent when two OG marginal populations were compared with two OG central or
The patterns of genetic diversity (allelic diversity and heterozygosity) and \( N_e \) between central and marginal populations are consistent with those recently reported for long-lived conifer, *Thuja occidentalis* [11,12]. Our results support the central-marginal hypothesis. However, the results are in contrast with the genetic diversity results previously reported for central Ontario and marginal Newfoundland populations of EWP [13]. This inconsistency may be due to differences in the parts of the range sampled, markers used and sampling design. Rajora *et al.* [13] study was based on selective sampling, allozyme markers and populations separated by about 2,000 km, whereas our study was based on population census, microsatellite markers and sampling from the same part of EWP range in northern Ontario. Also, our results are in contrast to some studies in other conifers that reported similar genetic diversity between marginal and central populations, e.g., [15].

The studied old-growth and second-growth EWP populations have similar levels of genetic diversity, \( N_e \), inbreeding and LGP. These results are in contrast with the general belief and limited empirical evidence [20,21] that old-growth populations have higher genetic diversity than the second-growth populations, but are in agreement with [28], who reported similar genetic diversity for one old-growth and one second-growth EWP stand from USA. Old-growth stands are expected to harbor higher genetic diversity due to genetic homeostasis, selection against inbreds and survival of the fittest over time in long-lived plants. The age of the studied old-growth populations was \( \sim 250 \) years and that of second-growth populations \( \sim 100 \) years, similar to that sampled in [28]. We do not know the exact reasons for similar genetic diversity and \( N_e \) observed in OG and SG populations. It may be possible that EWP goes through the processes that maintain high genetic diversity by age 100 years. EWP has predominantly outcrossing mating system [14,25] and severe inbreeding depression where, like other conifers [65], selection against inbreds probably occurs at a very early stage. These features will help in genetic diversity maintenance in \( \sim 100 \) years old SG EWP.

We sampled central and marginal and old-growth and second-growth populations from a relatively small part of EWP range in northern Ontario in order to avoid major regional variation confounding our results. Study of populations from other parts of EWP range may yield different results. Also, a range-wide central-marginal population genetic study in EWP is needed.

### Population Structure and Genetic Divergence

Bayesian structure, \( F_{ST} \), AMOVA, genetic distance, \( N_{ST} \), and PCA consistently demonstrated that the marginal populations are genetically diverged and separated from the central populations. This may be a result of isolation, impeded gene flow, genetic drift and divergent selection and their synergistic effects. Although different Mantel tests were not significant, the isolation by distance (IBD) contributing to the genetic differentiation of the marginal populations from the central populations could not be completely ruled out due to low statistical power to detect IBD with a limited number of populations examined in the present study. Nevertheless, other genetic differentiation measures such as AMOVA attributed hierarchical population structure to the divergence of marginal populations where a change in allele frequencies towards fixation or purging was observed at loci RPS-1b,
RPS-20, RPS-34b and RPS-39. Also curtailed gene flow to marginal populations (Fig. 5) and divergent selection at RPS-20 and RPS-39 (Fig. 3) were observed. Thus, genetic drift coupled with limited gene flow from central populations and possibly divergent selection seems to be the most plausible explanation for the genetic distinctness of the marginal populations. We also observed that EWP central old-growth populations are genetically similar to central second-growth populations. This may be due to large population size, continuous distribution and long-distance gene flow.

Marginal populations are expected to have higher differentiation among themselves than central populations [1,2]. While there is empirical support for this view [1,2,11], we did not find strong support for this pattern. We attribute these results to the geographical proximity of the two marginal GL populations, which are separated by about 2 km. Although these populations have private alleles [27,31], there are no known geographical barriers to gene flow between them. Likewise, our study suggests very little genetic differentiation between 190 km apart central OG and SG populations. This is consistent with the expectation for central populations of highly outcrossing species with long-distance gene flow.

Curtailed Asymmetrical Gene Flow to Marginal Populations

The Bayesian model-based approach demonstrated that north to south (marginal to central) migration was the most likely model explaining the observed data. This pattern coincides with the prevalent north to south wind flow in the area. Geographical isolation of GL populations also may have resulted in restricted gene flow to them. Thus the observed pattern of migration is in agreement with the theoretical central-marginal expectation of curtailed gene flow to marginal populations despite abundant center. Coalescent methods, such as MIGRATE, estimate migration rates accurately when the population divergence time is deep. Because the time of population split between the studied central and marginal populations is likely to be very recent (<200 generations) based on the geological and fossil evidence, some assumptions have likely been violated in the gene flow analysis using MIGRATE. Thus caution must be exercised when interpreting the rates of migration between the study populations.

Signatures of Divergent and Balancing Natural Selection in Marginal Populations

Our study provides molecular evidence for divergent and balancing selection operating in the marginal populations of EWP in northern Ontario. Two alleles at RPS-20 and four alleles at RPS-39, candidate loci for divergent selection, showed sharp directional change in their frequencies in the marginal populations (Fig. 4). Such a dramatic difference in gene frequencies may also result from strong genetic drift in marginal populations where rare alleles from core populations may sur the wave of population expansion and greatly decrease in frequency. This phenomenon, dubbed 'allele surfing' [66] has been found to account for many of the large gene frequency differences in human populations from different continents, previously attributed to local adaptation through selection following migration out of Africa [67]. We do not find allele surfing contributing to allele frequency spikes at RPS-20 and RPS-39 on account of several reasons. First, our marginal populations had once been a part of the continuous distribution of the species, which went through multiple cycles of range expansion and contraction, a proposition supported by fossil pollen evidence [24]. Therefore, the Galloway Lake marginal populations do not represent the de-novo leading edge of eastern white pine range expansion, a characteristic target of allele surfing [66]. Second, the probability of a rare allele to surf increases with the reduction in the size and connectivity (through gene flow) between the local demes [67] (and references therein), for which we found no evidence between the two marginal populations. Finally, surfing is expected to cause multiple rare alleles to increase in frequency in an expanding population [67]. In our study, increase in the frequency of rare alleles in the marginal populations was not across the loci but was limited to only one allele at RPS-20 and two alleles at RPS-39 (data for other loci not shown).

The divergent selection in the marginal populations is likely due to their local adaptation to different climatic and site conditions. The studied marginal and central EWP populations occur in two different ecocregions of Ontario. The GL populations occur in the ecoregion 4E (Tamagami Ecoregion) and the central RH and FR populations in the ecoregion 5E (Georgian Bay Ecoregion) [33]. The ecoregion 4E is characterized as Humid Low Boreal Ecoclimatic Region, whereas the ecoregion 5E as Humid High Moderate Ecoclimatic Region [33]. These ecoregions differ in mean annual temperatures, average growing season, rainfall and climate [33], with ecoregion 4E as Humid High Moderate Ecoclimatic Region [33]. Such climatic and ecological differences could result in different selection regimes in the GL populations thereby driving changes in frequencies of alleles. The selection pressures and regimes are most likely to be different in different parts of the geographical distribution range of EWP which has wide geographical distribution in North America encountering a variety of temperature, moisture, soil and other ecological conditions in its central and marginal populations. Under this scenario, the loci showing signatures of selection in marginal populations from different parts of the EWP range are likely to be different. The studied populations are from the northern Ontario part of the distribution range of EWP. Thus, broad inferences about range-wide selection pressures and loci showing signatures of selection in marginal populations cannot be drawn here. Nevertheless, our results provide robust inference of genetic divergence and natural selection in EWP marginal populations in the part of the EWP range in northern Ontario that we studied, where marginal populations are expected to expand northward.

The GL marginal populations demonstrated the presence of divergent selection at two candidate loci under both simple and hierarchical island models (Fig. 3; Fig. S4). Therefore, it is unlikely that the effects of shared population history or presence of hierarchical population structure have confounded the effects of natural selection on these loci. Moreover, our results of divergent selection in marginal populations are consistent with similar results reported for SNP markers in *Picea stichensis* [18] and for AFLP markers in *Bacillus laevigate* [19].

Under balancing selection, heterozygotes of beneficial alleles may be maintained preferentially over homozygotes. Indeed, for RPS-12 consistently showing the signatures of balancing selection, *H*<sub>b</sub> was high in marginal populations. The balancing selection may be due to heterozygous advantage in marginal harsher climatic conditions [64] that the sampled marginal populations experience.

Signatures of selection in response to salt tolerance have been reported for microsatellite markers in *Helianthus paradoxus* [63]. Microsatellite genetic variation is generally assumed to be selectively neutral. However, it is possible that the selective candidate microsatellite markers are in linkage disequilibrium with functional causative variation in the genome, responsible for local adaptation in marginal populations. This proposition is supported
by the fact that none of the three candidate loci were detected to be under selection when only central OG and SG populations were analyzed. The status of these markers will remain putative until linkage of these microsatellite loci with adaptive variation (e.g. SNPs) is ascertained. Nevertheless, we believe that our study provides the first evidence for divergent and balancing natural selection operating in *in situ* natural marginal populations of long-lived and widely distributed plants.

Although SNPs in candidate genes and genomic elements/sequences make powerful markers to detect natural selection, microsatellites provide more precise information on genetic diversity, population structure and demography [69], which is essential for disentangling the confounding effects of demographic processes and shared population history from that of selection. Microsatellite markers have allowed us to address the primary objective of examining differences in genetic diversity and population structure between central and marginal EWP populations from northern Ontario. In future, a large number of SNPs in candidate genes and other genomic elements along with microsatellites should be used to identify genes under selection in marginal populations.

**Conclusions**

Our study results conform to most of the theoretical expectations of central-marginal hypothesis. The marginal EWP populations have lower allelic diversity, $N_e$ and LGP than the central populations. The marginal populations have slightly higher but statistically similar heterozygosity to the central populations. The central and marginal EWP populations have similar levels of genetic divergence. The marginal EWP populations are genetically distinct from the central populations. Gene flow is asymmetrical with north to south migration fitting the observed data better than either south to north or panmixia models, consistent with the prevailing north-south wind pattern in the area. The marginal populations showed signatures of diversifying and balancing selection, probably in response to local adaptation. Curtailed gene flow and natural selection may be potential mechanisms underlying local adaptation of the GL marginal populations. The successional stage (old-growth and second-growth) of the populations apparently has no effect on the central-marginal population genetic patterns in EWP. The studied EWP old-growth and second-growth populations have similar genetic diversity and genetic constitution. Our study provides the original report on the dynamics of migration, genetic drift and selection in central and marginal populations of EWP, and perhaps of any long-lived plant species with wide geographical distribution. The results contribute to resolving the classical central-marginal debate and to the understanding of evolutionary genetic forces underlying local adaptation in marginal populations.

Conservation value of marginal populations is debatable. We strongly caution against discounting the genetic conservation importance of the Galloway Lake area EWP populations, considering their genetic distinctness and potential for evolutionary change and local adaptation for future range expansion under anticipated climate change conditions. Our study calls for more extensive range-wide investigations of genome-wide and adaptive genetic variation and population structure in EWP, which could offer insights into local adaptation of the marginal populations and effectively address conservation genetic issues.

**Supporting Information**

**Figure S1** Neighbor-Joining tree showing genetic relationships among central and marginal populations of *eastern white pine*. (PDF)

**Figure S2** Principal coordinates plot showing genetic relationships of *eastern white pine populations*. Ordination of the eastern white pine populations on principal coordinates 1 (PC1) and 2 (PC2) based on their Nei (1972) genetic distances. (PDF)

**Figure S3** Log probability of data and $\Delta K$ estimates for all central and marginal eastern white pine populations. (PDF)

**Figure S4** $F_{ST}$ outliers showing signatures of natural selection under simple island model. $F_{ST}$ outlier graphs showing microsatellite loci putatively under selection in (a) central vs marginal populations and, (b) central old-growth vs central second-growth populations. (PDF)

**Figure S5** $F_{ST}$ outlier test for old-growth populations under simple island model. $F_{ST}$ outlier test for detection of candidate loci under selection in two old-growth marginal and two old-growth central populations. (PDF)

**Figure S6** $F_{ST}$ outlier test for three pooled populations under simple island model. $F_{ST}$ outlier test for detection of candidate loci under selection in three pooled central and marginal populations of eastern white pine. (PDF)

**Figure S7** Detection of natural selection using Bayesian $F_{ST}$ method. Bayesian analysis for detection of candidate loci under selection in central and marginal populations. $F_{ST}$ is plotted against log of posterior probability $q$ values that indicate outlier status of markers. Filled circles represent markers under selection. (PDF)

**Figure S8** Effective population sizes and their 95% confidence intervals for *eastern white pine populations*. Effective population size estimates for six individual populations (A), and for three pooled populations (B). The K numbers indicate MCMC sweeps discarded as burn-in and recorded respectively. (PDF)

**Table S1** (DOCX)

**Table S2** (DOCX)

**Table S3** (DOCX)

**Table S4** (DOCX)

**Microsatellite Genotyping Text S1** (DOCX)

**Acknowledgments**

We thank Andree Morneault, Dianne Othmer, Dr. Tom Noland and Dr. Bill Parker from Ontario Ministry of Natural Resources (OMNR) for study plot establishments and sampling; Brian Brown (OMNR), Isla Langis and Robert Berry for locating population coordinates; Dr. George Buchert,
for valuable comments and suggestion, and Dr. Patrick Meirmans for useful discussions about the isolation by distance test. Data accessibility: The microsatellite genotype data has been deposited to Dryad (doi:10.5061/dryad.6f8p6).

Author Contributions
Conceived and designed the experiments: OPR. Performed the experiments: VEC. Analyzed the data: VEC. Contributed reagents/materials/analysis tools: OPR. Wrote the paper: VEC OPR. Overall guidance and supervision: OPR. Principal Investigator: OPR.

References


69. Hale ML, Burg TM, Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLOS ONE 7(9): e45170.