

This is to certify that the

thesis entitled

SPATIAL PATTERNS OF GENETIC VARIATION: OLD GROWTH
WHITE PINE

presented by

PAULA E. MARQUARDT

has been accepted towards fulfillment
of the requirements for

M.S. degree in FORESTRY


Major professor

Date April 17, 2002

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

SPATIAL PATTERNS OF GENETIC VARIATION: OLD GROWTH
WHITE PINE

By

Paula E. Marquardt

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Forestry

2002

ABSTRACT

SPATIAL PATTERNS OF GENETIC VARIATION: OLD GROWTH WHITE PINE

By

Paula E. Marquardt

Natural populations of tree species are dynamic systems: population demographics such as age, density and dispersal distance interact with evolutionary processes to determine spatial genetic structure. Old growth and second growth populations were evaluated, located within Hartwick Pines State Park, Grayling, MI. From each population, 120-122 contiguous trees were sampled for genetic analysis at seven SSR (simple sequence repeat) DNA loci. Genetic diversity was high and inbreeding levels were low for both populations. There was little divergence in allele frequency between populations. Spatial autocorrelation analysis suggested that individual genotypes were randomly distributed in the second growth population whereas weak positive spatial structure was observed at short distances for individuals in the old growth population. Logging may have decreased the degree of spatial structuring at the second growth site, suggesting that silvicultural practices may alter "natural" spatial patterns. Spatial structure is the main factor increasing levels of biparental inbreeding.

DEDICATION

To Emily and Jennie

ACKNOWLEDGMENTS

I appreciate my major professor, Dr. Bryan Epperson for his advice, guidance, and support in conducting this research and preparing this thesis. I also thank the rest of my thesis committee, Dr. Daniel Keathley and Dr. Kim Scribner, for their helpful comments in the final version of the manuscript. I am grateful to Dr. Jud Isebrands, Dr. Mark Kubiske, and Dr. Charles Michler, all of the USDA Forest Service, for their support throughout all aspects of my Masters program. The USDA Forest Service RWU-NC-4152 and 4155, and the McIntire - Stennis Cooperative Forestry Research Program (project #1774), funded this project.

I thank Ann Stephens, of the Hartwick State Park, for her help and aid in collecting sample materials.

I thank my daughter Jennie May, our students Jenny Dionne and Rebecca McFadden, and my coworker Ray Lange for helping to maintain my literature database and reprint files, and for various other thesis related support.

I thank my friends for their support of my research and studies, in particular Joeseeph Zeleznik and Kenichro Shimatani, for their help in sample collection, and Robert Bloye, Eric Myers and Erin Smith, for their discussions and moral support. I gratefully appreciate my friends and

colleagues, Dr. John Erpelding and Dr. Paul Anderson, for their encouragement and mentorship.

I am extremely grateful to my friend, Jerry Rayala, for his support and encouragement during the final few months of my thesis and defense preparations.

Finally, I would especially like to thank my daughters, Emily and Jennie, for their love and confidence, and for standing by me through the difficult times.

TABLE OF CONTENTS

LIST OF TABLES.....	viii
LIST OF FIGURES.....	xi
INTRODUCTION.....	1
LITERATURE REVIEW.....	7
Genetic Diversity.....	7
Methods Of Measurement.....	8
Spatial Processes.....	12
Population And Spatial Genetic Structure.....	15
Maintenance Of Genetic Diversity.....	19
METHODS.....	24
Site Description.....	24
Sampling.....	29
DNA Isolation.....	34
Marker Analysis.....	34
Quality Control.....	39
Statistical Analyses.....	41
<i>General Diversity</i>	41
<i>Population Structure</i>	43
<i>Spatial Structure</i>	45
RESULTS.....	50
Characterization Of Loci.....	50
<i>Polymorphism</i>	50
<i>Allele Frequency</i>	56
<i>Perfect Repeat Diversity</i>	59
Characterization Of Populations.....	61
<i>Allele Frequency And Class</i>	61
<i>Genetic Diversity</i>	64
<i>Population Structure</i>	64
<i>Spatial Structure</i>	67
DISCUSSION.....	72
Characterization Of Loci.....	72
<i>Polymorphism</i>	72
<i>Allele Frequency</i>	75
<i>Perfect Repeat Diversity</i>	77
<i>Null Alleles</i>	82
<i>Locus Rps6/34b</i>	89

Characterization Of Populations.....	91
<i>Allele Frequency And Class</i>	91
<i>Genetic Diversity</i>	94
<i>Population Structure</i>	96
<i>Spatial Structure</i>	99
CONCLUSION.....	103
APPENDIX.....	105
Spatial Autocorrelation Coefficients.....	106
LITERATURE CITED.....	119

LIST OF TABLES

Table		Page
1	Primer sequences for 8 nuclear (CA) _n simple sequence repeat markers. Forward then reverse sequences are listed for each primer pair. Rps6 and Rps34b amplify the same locus.....	36
2	Seven nuclear (CA) _n simple sequence repeat (nSSR) loci characterized for mature white pine. A total of 242 trees were assayed. Reported are numbers of chromosomes sampled, observed number of alleles per locus (k), observed allele sizes in base pairs (bp), change in number of base pairs between the smallest and largest alleles, amplified size of most common allele in basepairs (bp), and frequency of most common allele (x _i).....	51
3	a = OG, b = SS. Allele frequencies measured in two mature populations of <i>Pinus strobus</i> . The number of individuals sampled was 122 and 120 individuals in OG and SS respectively. Alleles are ordered according to size. The smallest allele is listed first (allele 1).....	57
4	Repeat sequence, category, and number of repeat units (m), in seven eastern white pine nSSR markers.....	60
5	Diversity measures for four nSSR perfect CA repeats: number of repeat units (μ), Nei's genetic diversity (H _e), observed heterozygosity H _o , observed number of alleles per locus (k), and effective number of alleles (A _e).....	60

6	Distribution of allele frequencies in mature white pine populations: total alleles, common and uncommon alleles, low frequency alleles, and rare alleles. The haploid sample size was 244 and 240 for OG and SS respectively. There were seven loci sampled. The total number of alleles present in both populations was 54. The number of alleles present in each class is in parentheses.....	62
7	Private alleles and common alleles present in two mature populations of white pine. The haploid sample size was 244 and 240 for OG and SS respectively. There were seven loci sampled. The number of alleles is in parentheses.....	63
8	General diversity measures for Old Growth and Second Growth <i>Pinus strobus</i> : number of trees sampled, percentage polymorphic loci, total number of alleles, allele richness (A), effective number of alleles per locus (A_e), observed heterozygosity (H_o), and expected heterozygosity or Nei's genetic diversity (H_e). Standard deviations are in parentheses.....	65
9	Population indices F and F_{st} for each nSSR locus and as a mean across all loci as measured in two populations of white pine. Seven loci were sampled. Sample size was 122 and 120 individuals in OG and SS, respectively.....	66
10	a = old growth (OG), b = second growth (SS). Mean spatial autocorrelation coefficients (Moran's I) in two mature populations of <i>Pinus strobus</i> for 9 distance classes, $2n = 244^1$ for OG and $2n = 240$ for SS.....	69

11	a = OG and b = SS. Markers Rps6 and 34b amplify the same locus. Characterized are the large allele heterozygotes for Rps6 (alleles 177, 179, 180 and 186) with the corresponding genotypes for Rps34b where the large allele is a null allele. Rps34b homozygotes are 17 bp smaller than the small Rps6 alleles.....	88
12	Estimating the rate of miscalls for the locus defined by markers Rps6 and Rps34b. Characterized for the Second Growth (SS) population are the homozygous genotypes scored for marker Rps6 that present heterozygous genotypes when scored for marker Rps34b. Rps34b peak 1 alleles are 17 bp smaller than Rps6 peak 1 alleles. The alleles with discrepancies are in bold typeface.	90
13	Spatial autocorrelation coefficients (Moran's I) for 8 loci in the Old Growth population (OG) of <i>Pinus strobus</i> for 10 distance classes, n = 122 ¹	107
14	Spatial autocorrelation coefficients (Moran's I) for 8 loci in the Second Site population (SS) of <i>Pinus Strobus</i> for 10 distance classes, n = 120 ¹	113

LIST OF FIGURES

Figure		Page
1	Map of the Old Growth (OG) and Second Growth (SS) study sites in the Hartwick Pines State Park in Crawford County, Michigan.....	26
2	Diameter at breast height distributions for both adult populations of white pine at Hartwick Pines State Park. (mean $\pm$ 1 SD).....	28
3	a = Old Growth (OG) and b = Second Growth (SS). The spatial distributions of individual trees located within two mature populations of white pine. Study sites were approximately 1.2 hectares in size. The distance between successive tick marks is 20 meters.....	33
4	a = Rps1b, b = Rps2, c = Rps6, d = Rps39, e = Rps50, f = Rps84, and g = Rps127. Distribution of allele frequencies detected at seven nSSR loci in 2 mature populations of white pine.....	52
5	Mean Moran's (I) indices measured spatial structure in two populations of white pine. The I values for the first three distance class were plotted.....	71
6	Sequence data for two cloned white pine fragments containing (CA) ₁₄ repeats. Marker Rps6 was derived from the plasmid clone pPs6 and Rps34b from pPs34 with a 17 bp difference in sequence size. Forward and reverse primers are underlined. The repeat sequences are in bold typeface. Lower case letters indicate differences between the sequences.....	84
7	a = Rps6 and b = Rps34b. Displayed are the allele frequency distributions for markers Rps6 and Rps34b. These markers amplify the same (CA) _n locus, with Rps 34b having four null alleles corresponding to the Rps6 alleles 177, 179, 180 and 186.....	85

INTRODUCTION

Natural populations of tree species are dynamic and evolving systems influenced by numerous genetic and environmental factors. Population genetics examines changes in gene frequencies on the shortest of evolutionary time scales. Furthermore, species can be subdivided into local breeding units, which allows us to evaluate the amount and distribution of genetic diversity. Population demographics such as age, density and dispersal distance interact with evolutionary processes such as genetic drift, natural selection, mating system, gene flow and mutation to determine spatial structure, i.e. the spatial distribution of genotypes and their degree of relationship. Because the magnitude and direction of correlation between distance and genotype is influenced by stand demographics, silvicultural practices such as thinning, harvesting, or no treatment can alter stand level spatial patterns. Spatial structure in turn influences other aspects of population genetic structure such as inbreeding.

An understanding of the level of genetic diversity, how it is partitioned, and the spatial patterns of its distribution in both natural and managed stands, will aid foresters in making informed decisions about stand management and reforestation. Thus, population genetic

evaluation provides them with additional tools for maintaining genetic diversity and reducing inbreeding. For example, spatial structure can help anticipate bi-parental inbreeding levels and quantify population genetic structure since limited seed and pollen dispersal may result in mating between related individuals.

Eastern white pine (*Pinus strobus* L.) is a species of both historical and economic value in the northeastern United States and Canada. It graces the landscape of our public and private forests with aesthetic beauty and is a valuable tree species for the lumber and paper industry. Despite its importance, little is known about the population and spatial genetic structure of white pine, its mating system, and how its genetic diversity is affected by changes in the forest ecosystem.

Although white pine is primarily out-crossing, self-fertilization and bi-parental inbreeding can occur, resulting in a reduction in genetic diversity from values expected under random mating. The loss of genetic variation may reduce the fitness of the present population adapted to the current or past environmental conditions that have contributed to the development of the population. Moreover, loss of genetic diversity may lead to the loss of latent genetic potential, which could degrade adaptation to

future environmental change, and sustainability of the forest resource. Future management will need to consider potential losses to the diversity of regenerating progeny when designing and applying silvicultural prescriptions to the parent population. One major concern is the loss of alleles present in low frequency (i.e., rare alleles). Rare alleles could be a crucial class of alleles for natural selection to act upon for future adaptation. Human activities have fragmented the landscape, introduced exotic pests, and created environmental changes such as pollution, which alters environmental chemistry (acid rain), the earth's temperature and precipitation patterns (global warming), and plant physiology (rate of plant growth, leaf chemistry affecting nutrition and defense to insects, and tree health). All of these changes in the forest ecosystem will require adaptive changes on the part of natural populations in order for life to continue to evolve and exist.

The design of effective management strategies can be facilitated by the use of population and spatial genetic structure to predict the effects of silvicultural treatments on genetic variation. Silvicultural systems can minimize inbreeding and maximize genetic diversity by three methods: 1) ensuring the effective breeding population is

large enough to retain sufficient numbers of pollen and seed parent trees representing most gene variants including sampling of gametes containing rare alleles; 2) minimizing related neighborhoods by maintaining proper spacing and reducing patchiness by keeping migration routes open; and 3) keeping stand densities high enough to ensure cross-fertilization while reducing self-fertilization and mating between related trees.

Because of its economic value, white pine will be more actively managed and manipulated in the future to improve growth, form, and disease resistance. Therefore it is imperative to study evolutionary relationships before they are lost with domestication. Of equal importance to its value as a crop, is its value for habitat, aesthetics, and history when maintained as old growth forests. Although the old growth in the United States is not in danger of becoming extinct, it has been made extremely fragmented, and if not managed using sound population genetic principles, losses in the unique variation found in these untouched gene pools will occur. Therefore, this study was conducted to offer preliminary insight into identifying appropriate resource management practices for maintaining levels of genetic variation characteristic of old growth forest ecosystems.

The study involved the evaluation of two *P. strobus* populations located within Hartwick Pines State Park, near Grayling, Michigan. The old growth population is approximately 240 years old and has been unmanaged, except for fire exclusion. The second growth population regenerated naturally after being harvested in the late 1800's. Both populations were minimally disturbed. From each population, 120-122 contiguous trees were sampled for genetic evaluation at seven SSR (simple sequence repeat) DNA loci.

The objectives in this study were to: 1) assess fine scale genetic diversity, population structure, and underlying spatial genetic patterns in the populations; and 2) explore the possible effects of logging on the spatial genetic structure of a naturally regenerated second growth stand as compared to a virgin old growth stand.

Specific research hypotheses were:

1) There will be little genetic difference between the two populations of white pine, i.e. they will have similar allele frequency distributions.

2) Genetic diversity will be high for both populations.

3) The second growth population may have less genetic diversity than the old growth population.

4) Population genetic structure and spatial genetic structure will be weak for both populations.

5) Spatial patterns will be different for the two study populations.

LITERATURE REVIEW

Genetic Diversity

Plants are genetically very diverse. For allozymes, average heterozygosity within populations ($H_e = 0.11$) is greater than that for invertebrates ($H_e = 0.10$) or vertebrates ($H_e = 0.05$) (Hamrick & Godt 1989). Gymnosperm is one of the most genetically diverse group of plant taxa ($H_e = 0.15$) (Hamrick et al. 1992). Eastern white pine is quite variable among individual trees and across its geographic range. Its diversity is maintained by mechanisms favoring an outcrossing mating system and other components of the pine genetic system i.e. the reproductive, recombination and mutation systems (Ledig 1998). There is abundant variation in white pine for morphological characters (Beaulieu & Simon 1995) in provenance seed trials (Pauley et al. 1955) (Mergen 1963) (Fowler & Heimburger 1969) (Wright 1970) (Li et al. 1997) allozyme loci (Eckert et al. 1981) (Eckert & Ryu 1982) (Brym & Eckert 1986) (Beaulieu & Simon 1994a) (Chagala 1996), DNA loci (Echt & Nelson 1997), quantitative traits (Cornelius 1994), and secondary plant compounds (Smith et al. 1969) (Zavarin et al. 1969). High heterozygosity is found for allozyme loci in eastern white pine (Beaulieu & Simon 1994b) (Buchert et al. 1997) (Rajora et al.

1998) (Epperson & Chung 2001), and for DNA loci (Echt et al. 1996) (Rajora et al. 2000).

Methods Of Measurement

Genetic evaluation with all marker techniques suffers limitations. Morphological traits are common first descriptors of phenotype, but often show little variation and can be difficult to measure. Quantitative traits are used for the genetic improvement of forest trees. They are often polygenic (Hamrick et al. 1992) so do not provide simple allele frequency data for calculating genetic diversity parameters. Early biochemical methods used secondary plant compounds such as terpenes to measure genetic variation in forest trees, but their genetic basis was and remains poorly understood (Hamrick et al. 1979). In addition, these three genetic systems (morphology, quantitative traits, and secondary chemistry) are influenced by environmental heterogeneity, and this decreases their usefulness in population genetic studies. Because estimates of genetic diversity and genetic structure are based on gene frequencies, the most appropriate marker systems are those that determine allele frequencies directly. More recent molecular based biochemical techniques accomplish this objective by surveying for Mendelian variation in protein expression

(allozymes) or at the DNA level. In addition, DNA markers are regarded as selectively neutral, serving as valuable measures of gene flow.

High levels of genetic variation are correlated with plant life history traits (Hamrick et al. 1979) (Hamrick et al. 1992). White pine is highly diverse, and combines many of the traits associated with greater diversity: gymnosperm taxonomic status, long lived perennial, widespread geographic range, wind pollination and seed dispersal, high chromosome number, long generation times, sexual reproduction, high fecundity, and late successional status.

One of the more discriminating molecular based marker systems is required for genetic analyses of eastern white pine, where high levels of variation are anticipated. Allozyme studies (Hubby & Lewontin 1966) are simple, inexpensive, and robust, but are limited by few enzyme systems, tissue specific expression, low levels of variability for some species, and variation revealed only in protein coding genes. In addition, they often require destructive sampling and therefore are not useful for evaluating endangered species or other populations when study organisms cannot be sacrificed i.e. mark and recapture studies. Allozymes evaluate the products of gene

expression, but it is better to evaluate genetic diversity at the DNA level, which offers greater polymorphism.

Four popular DNA based marker systems that overcome the limitation of few variable loci, include RFLPs (Restriction Fragment Length Polymorphism), RAPDs (Random Amplified Polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphism), and SSR's (Simple Sequence Repeat). RFLP's (Botstein et al. 1980) examine size differences among restriction fragments of DNA, where differences in banding patterns reflect genetic differences. RFLP's are simple, reproducible and co-dominant (heterozygotes can be distinguished from homozygotes), but require large amounts of high quality DNA, and are time consuming and expensive. The RAPD technique (Williams et al. 1990) is fast, simple, and an inexpensive PCR (Polymerase Chain Reaction) based marker system (Mullis & Faloona 1987), using arbitrary primers and little DNA. However, it suffers from problems of reproducibility. AFLP's (Vos et al. 1995) combine the RFLP and PCR based marker systems, creating highly informative fingerprints. Although AFLP markers are sensitive and reproducible, they are expensive to develop and technically demanding to use. Another shortcoming of both RAPDs and AFLPs is that they are dominant marker systems. Since a dominant gene (A) is expressed whether a

tree is homozygous (AA) or heterozygous (Aa) for the gene, dominant markers are unable to distinguish between these two genotypes, making them decidedly less powerful for population studies.

For this reason, SSRs i.e. microsatellite markers (Litt & Luty 1989) (Weber & May 1989) are attractive for stand level population studies since they are co-dominant markers. Microsatellite sequences are preferentially located in non-coding regions of the genome, within introns or between genes (Weber & May 1989) (Hancock 1999) and are assumed to be neutral markers. Moreover, microsatellites are single locus markers with high mutation rates, offering high levels of reproducible polymorphism. Length variations in microsatellite sequences can easily be detected from small amounts of DNA by PCR amplification of the repeat region with unique flanking primers. Microsatellite markers do have some limitations. They are expensive and time consuming to develop, technically demanding, and because of their sensitivity, suffer from cross-contamination problems. PCR primers bind a specific target; therefore, cross-contamination between species would not be a concern unless they were very closely related. Generally, sources of cross-contamination would be intra-specific (mixing tissue samples from the same

species) or from PCR products (previous amplification reactions with complimentary priming sites). These shortcomings do not negate the strengths and usefulness of microsatellites. Thus, they remain a popular marker system for population genetic studies, which evaluate allele frequency changes on a micro-evolutionary scale. Since successful amplification is based on relatedness (Fields & Scribner 1997) (Echt et al. 1999) time scale needs to be considered when addressing macro-evolutionary questions. For example, microsatellites did not amplify across more distantly related species in pine (Echt et al. 1999) (Karhu et al. 2000) (Mariette et al. 2001). An exception is when they occur in highly conserved loci, which has resulted in successful cross amplification for more closely related species in various taxa, including waterfowl (Fields & Scribner 1997), salmon (Scribner et al. 1996), mammals (Moore et al. 1991), whales (Schlötterer et al. 1991), and primates (Blanquer-Maumont & Crouau-Roy 1995) (Garza et al. 1995), as well as in pines (Echt et al. 1999).

Spatial Processes

A major assumption in population genetics theory is that individuals close to one another in space are more genetically alike (Wright 1943). Therefore, individual genes in populations may not be randomly distributed, but

distributed spatially in a patchy structure. Population geneticists examine changes in allele frequencies among populations, the spatial patterns of these genes, and the micro-evolutionary forces and population demographics that determine genetic patterns of spatial variation. Papers that emphasize population genetic inferences include Sokal & Oden (1978a), Sokal & Oden (1978b), and Sokal et al. (1989).

Spatial autocorrelation analyses test for the presence, sign and strength of spatial structure in ecological and genetic data. These methods also provide description by predicting the underlying cause of structure based on the shape of the correlogram i.e. plot of correlation coefficients (y) against distance (x) (Sokal & Oden 1978a) (Sokal 1979) (Legendre & Fortin 1989). Two common correlation coefficients used by geneticists for estimating spatial autocorrelation are the Moran's I index and the join counts (Sokal & Oden 1978a) (Sokal 1979) (Cliff & Ord 1981). The most popular is the Moran's coefficient (Moran 1950) for interval data. Genotypes are converted to allele frequencies (0 if the genotype doesn't contain the allele, 0.5 if it is a heterozygote, and 1.0 if it contains two copies). The analysis of nominal data (individuals

representing different genotypes) is performed using the join counts.

Many interacting factors may underlie observed patterns of spatial variation in allele frequencies. Stand demographics such as low stand density, short dispersal distances, and population age; all increase spatial genetic structure. Low stand density serves to enhance spatial structure by increasing homozygosity through the effects of inbreeding and genetic drift. This is accomplished in three ways: 1) decreasing the effective population size; 2) increasing self-fertilization and mating between relatives; and 3) decreasing gene flow (fewer migration routes). Short distance seed dispersal increases spatial structure through the formation of family groups (Schnabel & Hamrick 1990) (Berg & Hamrick 1995). Spatial structure can either increase or decrease with population age. Spatial structure builds through successive generations of consanguineous mating (Schnabel & Hamrick 1990). This generational effect results in the newest (youngest) generation having more positive structure (larger groups of related individuals) than older generations. Over time, structure increases in the population as a whole. In comparison, if the genetic load is high, natural selection will remove spatial structure. Genetic load is the number

of deleterious recessive genes carried in a population. When inbreeding brings two copies of a deleterious recessive gene together in one individual, this homozygous individual will have lower fitness, and may be removed by natural selection. Because natural selection has more opportunity to eliminate inbred individuals from the population as it ages (Yazdani et al. 1985), structure will decrease over time, with the older trees being less structured than younger age classes.

As mentioned above, the evolutionary forces of genetic drift, mating system, gene flow, and natural selection all interact with population demographics, influencing the patterns of spatial variation. Mutation is the final evolutionary force to consider. High levels of mutation and recombination will decrease spatial structure through the introduction of novel genotypes. Finally, in other mating system examples, it has been shown, that small amounts of clonal reproduction increases spatial structure (Schnabel & Hamrick 1990) (Berg & Hamrick 1995) (Chung & Epperson 1999).

Population And Spatial Genetic Structure

The genetic system of conifers is both remarkable and complex. Three separate genomes with three different modes of inheritance provide a model system for investigating

population and evolutionary genetic questions. Nuclear DNA markers follow Mendelian inheritance, whereas haploid organelle inheritance occurs primarily through one parent. The chloroplast DNA (cpDNA) is passed through the male lineage (pollen), whereas the opposite, sole maternal inheritance occurs for mitochondrial DNA (mtDNA) through seed (Stine & Keathley 1990) (Wagner 1992) (Dong & Wagner 1994).

Because mutation rates, dispersal distances, migration rates and effective population sizes interact and vary among the genomes, measures of diversity and population and spatial structure also vary accordingly. For example, the haploid nature of the mitochondria and chloroplast organelles would tend to create more structure than the nuclear genome. The smaller effective population size decreases diversity through increases in genetic drift, and inbreeding, although high levels of gene flow can disrupt and reduce this structure. In three studies evaluating organelle and nuclear markers in pine, it was evident that the mitochondrial genome has more structure than either the chloroplast or nuclear genomes. Population differentiation for maternal mtDNA markers is greater than for the more similarly structured paternal cpDNA markers and nuclear allozymes. Differences in allele frequencies among

populations were 68% for mtDNA, 1% for cpDNA, and 2% for the nuclear allozymes (Latta & Mitton 1997). These results are consistent with two other evaluations that showed mtDNA differentiation varied from 87-93% (Wu et al. 1998) and cpDNA ranged from 2-4% (Dong & Wagner 1994). The authors conclude the difference between the mitochondrial, and the chloroplast and nuclear population structure measures, indicate evidence of high gene flow through wind-dispersed pollen. Additional allozyme studies have also found little relative population divergence, with just 1.5 - 7.6% of diversity partitioned among pine populations. F_{st} values were 0.024 for *P. rigida* (Guries & Ledig 1979, 1982), 0.04 for *P. ponderosa* (Linhart et al. 1981), from 0.015 - 0.061 for *P. strobus* (Beaulieu & Simon 1994a) (Rajora et al. 1998) (Epperson & Chung 2001), 0.044 for *P. thunbergii* (Kim et al. 1997), 0.076 for *P. pinaster* (Salvador et al. 2000), and 0.021 for *P. brutia* (Panetsos et al. 1998).

In addition to reducing population structure, high gene flow also reduces spatial genetic structure and minimizes biparental inbreeding. Pines primarily outcross through wind-dispersed pollen. Inbreeding levels in non-isolated natural populations of adult pine are typically close to zero (Guries & Ledig 1979, 1982) (Linhart et al. 1981) (Yazdani et al. 1985) (Beaulieu & Simon 1994a) (Kim et

al. 1997) (Panetsos et al. 1998) (Rajora et al. 1998) (Epperson & Chung 2001). However, structuring of genotypes and inbreeding can still occur over time due to limited seed dispersal distances and mating between relatives. Spatial structure would negatively affect the population through inbreeding induced decreases in genetic diversity, which in turn may reduce the fitness of the population.

Conifers in general have a high genetic load i.e. number of recessive lethal genes (Ledig 1998). Therefore eastern white pine, like other conifers, suffers from strong inbreeding depression (Johnson 1945) (Fowler 1965), resulting in slower growth, reduced vigor and greater chlorosis. Other effects of inbreeding in pine are aborted embryos, low germination, stunted seedlings and reduced survival (Williams & Savolainen 1996) (Wu et al. 1998). In the only mating system study of eastern white pine (Beaulieu & Simon 1995c) the mean single locus outcrossing rate was close to 1.0 in two old growth populations, suggesting absence of self-fertilization and consanguineous matings (inbreeding resulting from mating between relatives). These results are consistent with average single locus estimates reported in lodgepole pine (Epperson & Allard 1984). In comparison, outcrossing rates were

slightly lower for western white pine with a mean single locus estimate of 0.94 (El-Kassaby et al. 1993) indicating possible spatial family structure, a major factor in biparental inbreeding.

Few studies have examined the spatial genetic structure of pines within stands. Although spatial genetic structure is expected to be weak in continuous populations, it can be the most important factor controlling levels of inbreeding. Epperson & Allard (1989) reported lodgepole pine (*P. contorta* ssp. *latifolia*) genotypes to be nearly randomly distributed suggesting absence of patch structure in the presence of high gene flow. These findings are consistent with Knowles (1991) spatial evaluation of old growth black spruce (*Picea mariana* (Mill.) B.S.P.). In comparison, Epperson and Chung (2001) provided evidence for reduced structure in old growth white pine, with weak positive structuring of alleles at 15 m (Moran's $I = 0.05$).

Maintenance Of Genetic Diversity

The sustainability of genetic variability will be instrumental in the ability of forests to adaptively respond to the often rapid, human induced changes in the environment. For example, pollution causes changes in plant physiological processes such as rate of tree growth, (Isebrands et al. 2001) photosynthesis (Kubiske & Pregitzer

1996) (Kubiske et al. 1997) (Takeuchi et al. 2001) tree health (Rundel & Yoder 1998) the introduction of exotic pests and changes in tree defenses to insects (Lodge 1993) (Niemela & Mattson 1996) (Mattson 1998) and leaf chemistry affecting nutrition and levels of insect herbivory (Herms et al. 1996). However, range expansion through plant migration in addition to adaptation has been crucial to the ability of plants to respond to past climate change (Davis & Shaw 2001). It is likely that habitat fragmentation, again a result of human activity, will interfere with plant migration (Lodge 1993) (Davis & Shaw 2001). Moreover, future predicted rates of climate change are much more rapid than historical climate changes of the past (Etterson & Shaw 2001). Together, these two factors may not provide adequate opportunity for plant adaptation to occur through range shifts and refugia. Examining the changes in gene frequencies of natural populations in response to the current changes in climate and habitat fragmentation is an exciting and relatively untouched area of research. Previous management goals have been focused at the community or species levels. Now we have the tools necessary to manage populations at the genetic level, for the sake of maintaining species health in the face of a continually changing environment.

Wright (1943) summarized the levels of neighborhood differentiation that could be anticipated in a continuous population as a result of isolation by distance only, from neighborhood size and predicted levels of seed and pollen dispersal associated with various mating systems. Wright's (1946,1969) definitions of neighborhood, neighborhood area, and neighborhood size are useful in understanding potential applications of spatial autocorrelation measures to stand management. Wright defined a neighborhood as the surrounding area from which parents are likely to have been drawn. Wright's neighborhood may not be panmictic, i.e. all male-female pairs have equal probability of mating. Although a population may be distributed over a large area, parents are restricted to a neighborhood of limited distance, which often results in the chance mating between relatives. Mathematically, the neighborhood area is contained in a circle with radius 2σ , where σ is the parent-offspring (combined seed and pollen) dispersal distance. Finally, the neighborhood size is the number of trees within a neighborhood, and is calculated from the standard deviation σ of the parent-offspring dispersal variance σ^2 . Both dispersal distance and density influence the neighborhood size. For example, neighborhood size will be small due to short pollen or seed dispersion. Shorter

dispersal (slower migration) will allow greater opportunity for genetic differentiation to occur (and spatial structure to build). Low population density will also slow migration rates (fewer paths) with a decrease in neighborhood size. Approximating Wright's neighborhood size from a measure of spatial autocorrelation, the Moran's I statistic (Epperson et al. 1999) affords a method to estimate the combined levels of seed (σ_s) and pollen (σ_p) dispersal (Epperson & Chung 2001) using Crawford's (1984) dispersal variances. Upon empirical validation, this approach shows promise for selecting pollen and seed sources for maintaining the diversity of regenerating progeny.

This is the first published microsatellite study to quantify spatial structure within stands of eastern white pine, using Wright's neighborhood size and Moran's I statistic. The complementary allozyme study (Epperson & Chung 2001) will allow comparisons between allozyme and microsatellite markers. As the number of alleles for the microsatellites will be much greater than for the allozyme markers, their power to distinguish genetic structure would be expected to be greater. Because stands are the management units of foresters, by examining the spatial patterns of genetic variation within natural pine populations, we can better anticipate inbreeding levels in

eastern white pine. Greater knowledge of inbreeding will aid in the design of effective management strategies for reducing family structure and maintaining in-situ eastern white pine genetic diversity.

METHODS

Site Description

Two mature populations of eastern white pine (*Pinus strobus* L.) from Hartwick Pines State Park (Hartwick Pines) were selected for genetic analysis. The park study site was chosen for several reasons: (1) the forest is unmanaged; (2) the trees are over 100 years in age with some being over 250 years old; and (3) both old growth and second growth forests are within close proximity to each other. Hartwick Pines is located in Crawford county of Michigan's lower peninsula, 11.3 km northeast of the town of Grayling (44°39'N / 84°42'N). The park is situated on an outwash plain, formed 10,000 to 15,000 years ago by the Wisconsin Pleistocene glaciers (Sommers et al. 1984), (Whitney 1986). The elevation of the region is 347.5 m above sea level (NCDN, June 2000). Spodosol soils are common to Crawford County, which are sandy, well drained and infertile (Sommers et al. 1984), (Wheeler 1898). The climate of Grayling is described as humid continental (Strahler 1965). Mean annual precipitation is 833 mm and fairly evenly distributed throughout the year. The average annual temperature is 5.6°C with an average growing season of 98 days. (NCDC, June 2000).

Hartwick Pines includes 16 hectares of virgin old growth forest (NW ¼ of Sec. 15, T.27N, R.3W). The old growth forest will be referred to as OG. The spodosol soil underlying this population can be further described as Rubicon sands (Robertson & Tiedje 1984). The OG stand is located along the old growth forest trail of the state park (Figure 1) and the dominant trees are primarily white pine with some red pine (*Pinus resinosa* Ait.). Although unmanaged, it is not entirely untouched by human activity. There are unpaved trails in addition to the major paved trail through the forest. It is thought that this virgin stand was established from a few seed trees following a wind disturbance in the 1750's (Wackerman 1924). Based on this history, the OG population was approximately 240 years old at the time of sampling. In 1978, increment cores (Rose 1984) placed the mean age of the OG white pine at 177 years with a range from 101 to 229 years, which agrees well with Wackerman's (1924) estimate. The mean diameter at breast height (DBH) was 58 cm and ranged from 25 to 110 cm. Today the age distribution of trees would be 123-251 years.

White pine study areas

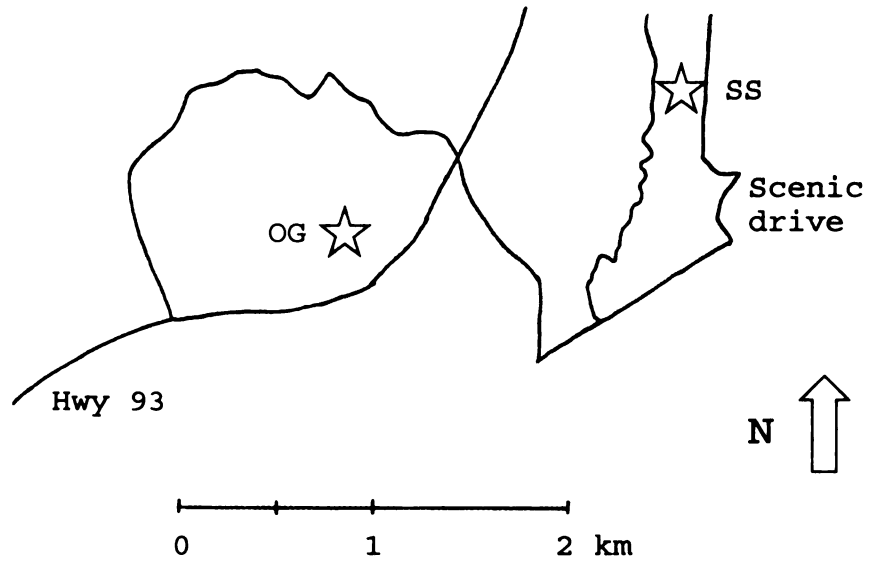


Figure 1. Map of the Old Growth (OG) and Second Growth (SS) study sites in the Hartwick Pines State Park in Crawford County, Michigan.

The mean DBH of the trees sampled in the study was 62 cm with a range from 33 to 99 cm (Figure 2). Sixty-three percent of the trees ranged from 50 to 69 cm in diameter. Other woody species found in a dense lower canopy of the forest are sugar maple (*Acer saccharum* Marsh), American beech (*Fagus grandifolia* Ehrh.) and eastern hemlock (*Tsuga canadensis* L.). There was no successful pine regeneration in the sparse undergrowth. Although white pine seedlings were numerous, they grew poorly and were less than 30 cm in height. Individuals greater than 3 years of age (Farrar 1995) and less than 33 cm in DBH, were completely absent. Red pine regeneration was also absent since no seedlings or small trees were observed in the study site.

The second growth stand, which will be referred to as SS, is located approximately 2 km east of the OG stand (SW $\frac{1}{4}$ of Sec. 11, T.27N, R.3W) and covers an area of approximately 6 hectares (Figure 1). This stand regenerated naturally following logging in the early 1890's and has been unmanaged since. White pine is the dominant species. Although all diameter size classes are represented in the population, we chose to sample trees larger than 20 cm in DBH. This approach allowed us to sample contiguous adult trees on the same scale (size of area sampled and density) as the OG, while avoiding the

Diameter at breast height (DBH)

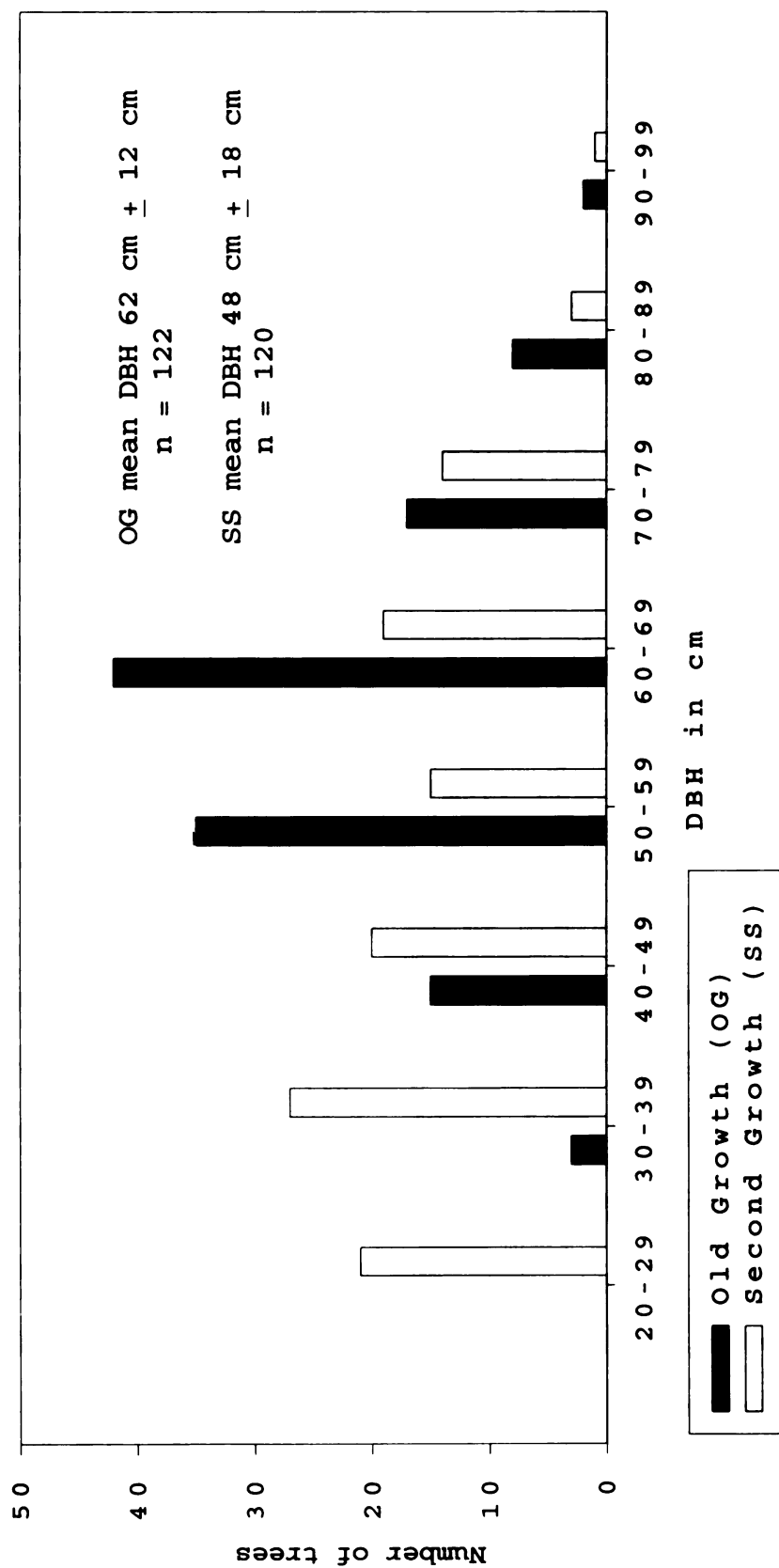


Figure 2. Diameter at breast height distributions for both adult populations of white pine at Hartwick Pines State Park. (mean $\pm$ 1 SD)

many, smaller trees that would not be contributing pollen and seed. The mean DBH of the sampled trees was 48 cm, ranging from 22 to 94 cm (Figure 2). The size distribution of the sampled trees was fairly uniform for the various diameter classes. Red pine was also present in the upper canopy of the SS stand. In addition to those woody species identified in the lower canopy at the OG study site, the lower canopy of the SS site included white pine, red pine, balsam fir (*Abies balsamea* (L.) Mill) and black spruce (*Picea mariana* (Mill.) BSP).

Sampling

The total number of sample genotypes is the main factor controlling experiment-wise statistical power for spatial analysis (Epperson & Li 1996). Sample size and proportion of the total number of trees sampled from the population (porosity) were chosen to optimize fine scale spatial statistical analyses for a wind-pollinated species with limited seed dispersal (Epperson et al. 1999). Our sampling scheme included a sample size of 120 individuals, a porosity of 1 (all adult individuals with DBH > 20 cm were sampled), and approximately 50 alleles surveyed at seven loci. The total number of alleles sampled per population was approximately 6000 (120 trees and 50 alleles). Statistical power is high for detecting spatial

structure at this level of sampling, based on theoretical simulations. Referring to Table 4 (Epperson et. al. 1999) the sampling situation (porosity = 1, sample size (n) = 10,000, and Wright's neighborhood size (N_e) = 125.7), the standard deviation is small (0.02) and statistical power high for rejecting the null hypothesis of a random distribution (99%). This situation is a good estimate for our study, where the sampling was identical to the above scheme except for a decrease in total sample size from 10,000 to approximately 6,000 (120 trees times 50 alleles). The predicted values of Moran's I statistic for individual genotypes are insensitive to sample size (Epperson et al. 1999). These authors found that for an identical sampling scheme, decreasing the size of the sample increases the standard deviation by the square root of the ratio of the original sample size to the decreased sample size. For example, a 100-fold decrease in sample size from 10,000 to 100, results in an approximate 10-fold increase in the standard deviation ($10000^{1/2} / 100^{1/2}$) (Epperson et al. 1999).

White pine is a shade intolerant species. Individuals less than 40 to 50 years old would not be expected to survive (Dallimore & Jackson 1984) or contribute to the gene pool of future generations. To determine the genetic

structure of the reproducing individuals in both stands, we sampled adult trees capable of producing both seed and pollen. The trees at the OG site were all very large with DBH's ≥ 33 cm. Adult trees from the SS were arbitrarily defined as individuals with DBH ≥ 20 cm. As mentioned above, this minimum DBH guideline allowed us to sample adult trees from an area approximately the same size and density as the OG, while avoiding the large number of non-reproducing trees. Sampling areas (Figure 3a and 3b) were approximately 1.2 hectares in size for OG (100m x 120m) and for SS (110m x 110m). The white pine tree density was approximately 100 adult trees per hectare for both populations. One plot was sampled per site, and within each plot, all contiguous adult individuals were tagged and mapped. Metal identification tags were attached near the base of the tree with a small aluminum nail. Trees were mapped to a neighboring study tree or a park location marker. The angle (θ) was measured to the nearest degree, distance (r) to the nearest 0.1 m, and DBH to the nearest cm. The x and y map coordinates were determined from the geometric relationships $x = r\cos\theta$ and $y = r\sin\theta$. The x and y coordinates were used when quantifying spatial structure. Triangulation was used to crosscheck the map.

Needle and bud tissues were collected from 122 OG and 120 SS trees from late August to November 1998, with total of 242 individuals being sampled. Harvesting tools included a pole pruner, flexible saw, and 222 rifle. Small branches were sampled from the tips of the lowest exposed branch, which sometimes was 30 meters or more up. Harvested tissue was kept on ice until transported to the laboratory and stored at 4°C for 1-3 weeks until bud and needle tissue could be separated. These tissues were then stored at -20°C prior to DNA extraction.

Tree spatial distributions

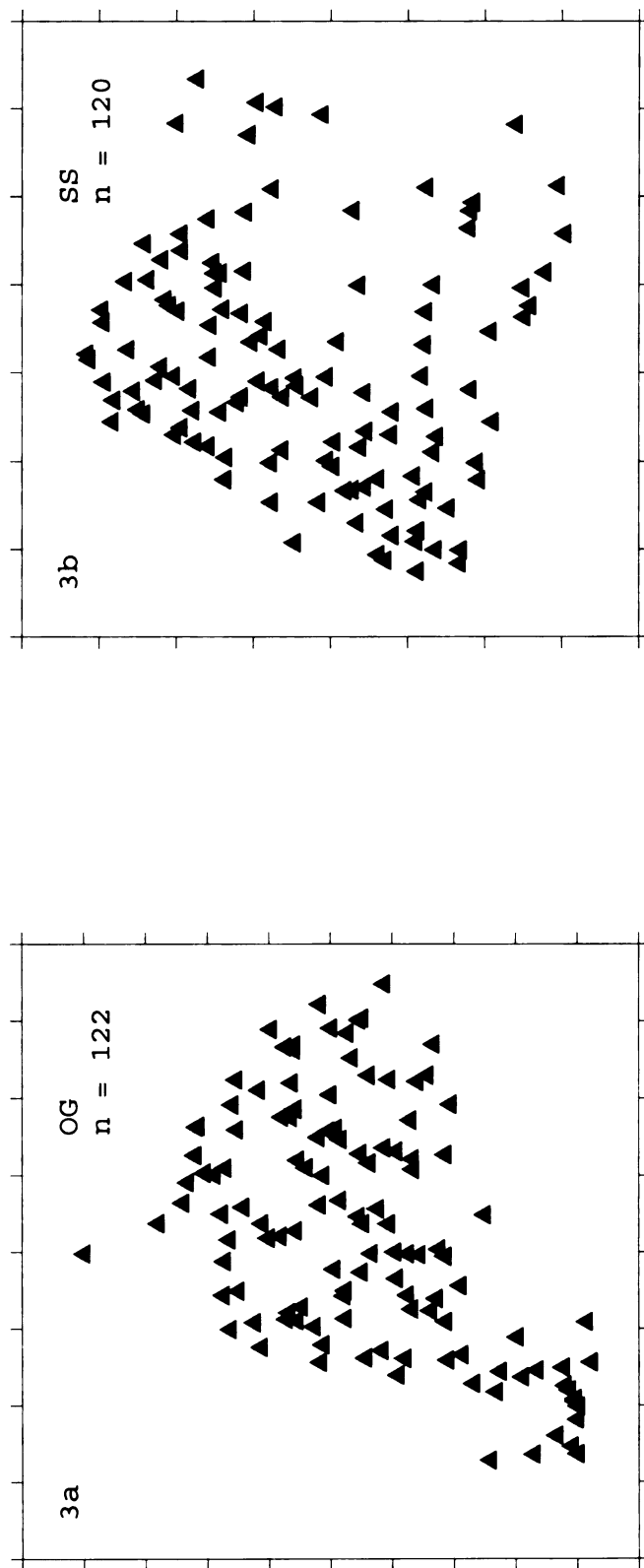


Figure 3a and 3b, a = Old Growth (OG) and b = Second Growth (SS). The spatial distributions of individual trees located within two mature populations of white pine. Study sites were approximately 1.2 hectares in size. The distance between successive tick marks is 20 meters.

DNA Isolation

Total genomic DNA was purified from 20-40 mg bud tissue with Qiagen Plant DNeasy isolation kits (Qiagen Inc., Valencia, CA). In situations where bud material was insufficient, 70 mg needle tissue was substituted for DNA extraction. Yields ranged from 20-160 µg DNA. DNA was quantified by DNA fluorometry and diluted to 2.5 ng/µl in T₁₀E₁ (10 mM Tris-Cl pH 8.0, 1 mM EDTA) for use in PCR amplifications. All DNA stocks and dilutions were kept at -20°C for long-term storage.

Marker Analysis

Eight microsatellite markers (Rps1b, 2, 6, 34b, 39, 50, 84 and 127) were used to survey genomic SSR variation at seven loci. After the initial analysis was completed, it was discovered that primer pairs for markers Rps6 and Rps34b were designed from the same sequence, and characterized the same locus; therefore data for Rps34b was removed from the final statistical analyses. In addition, it was determined that marker Rps34b contained null alleles. Echt et al. (1996) describes the primer sequences (Table 1) and marker characterization for the seven nuclear (CA)_n *P. strobus* loci. Unlabeled reverse primers were purchased from Research Genetics Inc., Huntsville, AL. The fluorescently labeled forward primers were purchased from

the Children's Hospital of Philadelphia, Philadelphia, PA or PE Biosystems, Foster City, CA.

The amplification reaction for each primer pair was conducted separately. The PCR reaction mixture contained 2 ng/ μ l DNA template in 10 μ l of reaction buffer. The reaction buffer consisted of 20 mM Tris-Cl pH 8.75, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 10 mM KCL, 2 mM MgSO_4 , 0.1% Triton X-100, 100 μ g/ μ l BSA, 6% sucrose, 0.1 mM cresol red, 200 μ M each dNTP, 200-800 nM each primer, and 0.025 U/ μ l AmpliTaq Gold DNA Polymerase (PE Biosystems, Foster City, CA). The final Mg concentration was adjusted to 4.5 mM with 25 mM MgCl_2 to improve yield of the amplified PCR products. The increased Mg^{2+} concentration also promotes non-template nucleotide addition to the 3' end of target sequences (Clark 1988) (PE Applied Biosystems 1996). The plus-nucleotide form

Table 1. Primer sequences for 8 nuclear (CA)_n simple sequence repeat markers. Forward then reverse sequences are listed for each primer pair. Rps6 and Rps34b amplify the same locus.

Rps locus	PCR primers
1b	GCCCACTATTCAAGATGTCA GATGTTAGCAGAAACATGAGG
2	CATGGTGTGGTCATTGTTCCA TGGAGGCTATCACGTATGCACC
6	TTTTCTAATCAGTGTGCGCTACA CACCGCTGCCCTATTTTACA
34b	CAGTGTTCTCTTATCACAGCG GCACTATAATGAAATAGCGCA
39	GCCAGCTCCAACCAGAATC GGCTCGCTGACCCAATAA
50	CCCAGAAATCTGTTTTAGAGC ACACATGAAATGTCAGAATGC
84	CCTTTGGTCATTGTATTTTGGAC CTTCCTTTTCCTTCTTGCTCCAC
127	ACTTCCTCCAAGTTACTATTGTCA CCTTGTCTTCTAAAAAACACTTTT

simplifies peak patterns and improves allele scoring by Genotyper software (PE Applied Biosystems 1994)

The cresol red and sucrose were included in the amplification reaction mix to eliminate the addition of loading dye prior to agarose gel electrophoresis (Routman & Cheverud J. 1994). The sucrose also improves amplification of weak reactions (Erpelding, personal communication).

A touchdown amplification protocol described by Echt et al. (1999) with a modified target annealing temperature of 55°C, was performed using PTC-100 or PTC-200 thermocyclers (MJ Research Inc., Watertown, MA). Other adjustments to the protocol included an initial denaturation of 92°C for 9 minutes to activate AmpliTaq Gold and a final extension of 70°C for 20 minutes to promote blunt ended, plus-nucleotide addition (Clark 1988) (PE Applied Biosystems 1996). The touchdown protocol included an initial denaturation step at 92°C for 9 minutes. The first two cycles of the protocol consisted of a denaturing step at 94°C for 1 minute, an annealing step at 65°C for 1 minute, and an extension step at 70°C for 35 seconds. The next 18 cycles comprised a denaturing step at 93°C for 45 seconds, an annealing step at 64°C for 45 seconds (which subsequently was decreased by 0.5°C every cycle until a final temperature of 55.5°C was reached), and an extension

step at 70°C for 45 seconds. Conditions for the last 20 cycles were 92°C for 30 seconds, 55°C for 30 seconds, and 70°C for 60 seconds, followed by a final extension at 70°C for 20 minutes. Touchdown PCR (Don et al. 1991) (Hecker & Roux 1996) and AmpliTaq Gold (Kebelmann-Betzing et al. 1998) (Moretti et al. 1998) were used to increase specificity and yield of amplification products.

Agarose gel electrophoresis was used to check the quality and quantity of the PCR amplified products. A 4 μ l sample of each amplification reaction was electrophoresed through 2% LE agarose gels using 1X TAE running buffer (Tris-acetate-EDTA) containing 0.2 μ g/ml ethidium bromide to determine the amplification efficiency for marker analysis. PCR products were diluted in ddH₂O in order to standardize and normalize fluorescent signal for accurate measurement of sample data peaks. Final PCR dilutions ranged from 1:1.2 to 1:15. Product dilutions for 2 to 3 loci were pooled for multiplexing on polyacrylamide gels, based on the fluorescent color of the marker label and size of amplified PCR product. The pooled, fluorescently labeled amplified fragments were resolved using denaturing polyacrylamide sequencing gels (6% acrylamide, 8.3M urea) with 1X TBE running buffer (Tris-borate-EDTA). Pre-prepared acrylamide gel mixes (Burst-Pac 6% sequencing gel)

were purchased from Owl Scientific, Portsmouth, NH. The resolved fragments were sized by GeneScan Analysis software. (PE Applied Biosystems 1996). Alterations to the manufacturer's GeneScan protocol included cleaning of glass plates with cerium oxide (Bunville et al. 1997) to remove fluorescing background haze and the use of CXR fluorescent ladder (Promega Corp., Madison, WI) as the internal size standard for assigning PCR fragment sizes. To aid in data interpretation, fragment lengths, scored as alleles and reported to 0.01 base pairs (bp) by GeneScan Analysis, were binned or grouped to 1 bp categories and assigned ordinal labels using Genotyper software.

Quality Control

To ensure consistency of PCR reactions, the mixing of PCR reaction buffer and aliquoting of DNA templates were conducted just once per population. For each primer pair, one large PCR master mix was constructed with enough reagents to amplify all DNA's for one population. These master mixes contained primer, enzyme and all reaction components necessary for PCR amplification except DNA and were stored at -20°C. DNA templates were aliquoted to the wells of the μ titer reaction plates, dried in a food dehydrator and stored at -20°C. Sample DNA's were aliquoted in a 2X-offset format using an eight-channel pipettor.

Enough DNA reaction plates were aliquoted at one time to complete the study.

To assess well-to-well gel variation in fragment mobility and fragment molecular weight size assignments, the CXR size standards from all lanes of all gels were combined in one Genotyper file. The fragment peaks were then binned to survey for accurate base-pair calling of the size standards by GeneScan analysis.

Several PCR amplification controls were run throughout the study. Running a negative PCR control, which was lacking template DNA and included for each primer pair master mix, monitored PCR product contamination. Internal and external PCR controls were developed to assess reproducibility of PCR generated genotypes. To measure PCR to PCR variability, an internal control procedure was designed in which 13% of the original PCR reactions were replicated. I was 95% confident that this level of re-sampling could detect a 1% departure from expected frequencies. Given the critical value (CV) for $\chi^2_{0.05, 1}$, and $\chi^2 = (Y - \mu)^2 / np(1 - p)$, then $n = (Y - \mu)^2 / p(1 - p)CV$, where $(Y - \mu)$ is the difference between the observed and expected values for the number of alleles, p the allele frequency, and n the re-sample size needed to detect a type I error, or rejection of H_0 when H_0 is true (Steel & Torrie

1980). A random number generator determined the DNA templates to be re-assayed and assigned the duplicate well positions in the μ titer reaction plate. The repeated PCR products were amplified, pooled, and evaluated under the same conditions as the original products. An external PCR control was evaluated to survey for lane variation between gels. The external control was constructed by amplifying loblolly pine (*Pinus taeda*, L.) DNA primed with the chloroplast marker Pt30204 (Vendramin et al. 1996). This amplification reaction generated a single PCR product for use in quality control, and was analyzed in the same lane on all GeneScan gels.

Statistical Analyses

Genetic diversity, population structure, and spatial structure were evaluated using conventional genetic measures. Each population was considered separately for all statistical analyses. POPGENE software, version 1.31 (Yeh et al. 1999), was used to calculate the diversity measures and F_{st} values. The spatial structure indices were computed by the SAAP - Spatial Autocorrelation Analysis Program written by Daniel Wartenberg (1989).

General Diversity:

Five indices measured genetic diversity: (1) observed number of alleles per locus (k); allele richness (A); (3)

allele frequency (x_i); (4) genetic diversity (H_e) or Hardy-Weinberg expected heterozygosity; and (5) effective number of alleles per locus (A_e). Observed number of alleles (k) was simply an arithmetic count of allelic classes for each locus. Allele richness (A) was the mean number of alleles over all loci. Allele frequency (x_i) was determined by $x_i = y_i/2n$ where y_i is the number of occurrences in the i th allelic class and n is the total sample size. To evaluate the distribution of allele frequencies within and between populations, allele morphs were arbitrarily assigned to three frequency classes. A two-tiered system first classified the allele frequencies as common ($x_i \geq 0.05$) and uncommon ($x_i < 0.05$). The uncommon class of alleles was further partitioned into alleles having low frequency ($0.005 < x_i < 0.05$) or rare alleles ($x_i \leq 0.005$) (Hartl & Clark 1997). The two populations were analyzed separately. A frequency of 0.005 meant there was only one copy of the allele in one population. Genetic diversity (H_e) was measured for each locus by $H_e = 1 - \sum x_i^2$ (Nei 1973), where x_i is the frequency of the i th allele, $\sum x_i^2$ the total homozygosity, and heterozygosity being $1 - \sum x_i^2$. Individual H_e values were averaged across all loci. The following arbitrary guidelines for H_e were used to assess the level of marker polymorphism:

$H_e \geq 0.75$	high polymorphism
$0.25 < H_e < 0.75$	moderate polymorphism
$H_e \leq 0.25$	low polymorphism

The effective number of alleles (A_e) for each locus was calculated by $A_e = 1/\sum x_i^2$ (Kimura & Crow 1964) or the reciprocal of homozygosity, where the x_i are defined as for H_e . Individual A_e values were averaged across all loci

Population Structure:

Deviations from Hardy-Weinberg equilibrium were measured for each locus by Wright's (1921, 1922) fixation index (F). $F = 1 - H_o/H_e$ (Nei 1977), where H_o is the observed heterozygosity and H_e the expected heterozygosity. $H_o = 1 - \sum X_{ii}/n$, where X_{ii} is the number of occurrences in the ii homozygous class, n is the total sample size, and X_{ii}/n is the proportion of homozygotes. Non-zero F values indicate either a deficit of heterozygosity (positive F) or an excess of heterozygosity (negative F) from the values expected when mating is random ($F = 0$). These indices were tested for significant differences from zero, given the $CV = \chi^2_{0.01, df}$, by $\chi^2 = F^2 n(k-1)$, $df = k(k-1)/2$, where n is the total sample size and k is the number of observed alleles (Li & Horvitz 1953). Summing the χ^2 -values and degrees of freedom provides a test of the mean F value over all loci.

Relative population divergence was measured at each locus by F_{st} (Wright 1943, 1951, 1965) using $F_{st} = 1 - H_s/H_t$ (Nei 1977, 1987) (Hartl & Clark 1997), where H_s and H_t are the expected average heterozygosities of the subpopulations (OG and SS) and total population (OG plus SS), respectively. H_s is calculated by averaging Nei's genetic diversity (H_e) over all subpopulations with $H_s = 1 - \sum_k \sum_i x_{ki}^2 / n$, where x_{ki} is the frequency of the i th allele in the k th subpopulation, and n is the number of subpopulations. H_t is calculated from the mean allele frequency across all subpopulations with $H_t = 1 - \sum \bar{x}_i^2$, where $\bar{x}_i$ is the mean frequency of the i th allele in the total population. The null hypothesis $F_{st} = 0$ was tested for significant differences in allele frequencies between populations by the χ^2 -test of heterogeneity given the CV = $\chi^2_{0.01, df}$, with $\chi^2 = 2nF_{st}(k-1)$, $df = k(k-1)(a-1)$, where n is the total sample size, k is the number of observed alleles, and a is the number of populations (Workman & Niswander 1970). Summing the χ^2 -values and degrees of freedom tested the mean F_{st} value over all loci. F_{st} values range from 0 for no divergence (populations share the same allele frequencies) to 1.0 for complete fixation of alternate alleles in different populations. Qualitative guidelines

for interpreting genetic differentiation are given by
(Hartl & Clark 1997):

0.00 - 0.05	little differentiation
0.05 - 0.15	moderate differentiation
0.15 - 0.25	great differentiation
> 0.25	very great differentiation

Spatial Structure:

The degree of relation between gene frequencies in individuals separated by distance, or spatial structure, (autocorrelation) was measured by the Moran's I-statistic (Sokal & Oden 1978a) for each of 10 distance classes.

Moran's I (Cliff & Ord 1981) is defined as

$$I = \frac{n \sum_{(2)} (w_{ij} z_i z_j)}{(\sum_{(2)} w_{ij} \sum z_i^2)^{-1}} \text{ where:}$$

- o n is the number of individuals.
- o w_{ij} is a join in the binary connectivity matrix.
I is weighted by the distance between individuals, with w_{ij} set to 1.0 if the ith and jth individuals are in the distance class and zero if they are not both present.
- o $z_i z_j$ is given by $z_i = x_i - \underline{x}$ and $z_j = x_j - \underline{x}$ where x_i and x_j are the genotypes for the ith and jth individuals expressed as frequencies (described below), and $\underline{x}$ is the mean score for all individuals in the population.

- o $W_{ij}Z_iZ_j$ is the matrix cross product or co-variance and $\sum Z_i^2$ the total variance.

Individual genotypes were converted to allele frequencies (Dewey & Heywood 1988) where 1.0 was assigned to X_iX_i homozygotes of the i th allele, 0.5 to X_iX_j heterozygotes of the i th allele ($j \neq i$), and 0 to genotypes with zero copies of the i th allele. The Euclidean distance between sampled trees was used to assign all pairs or joins of individuals to one of ten distance classes. The upper bound of the first distance class was set at 15 m, with upper bounds increased by 10 m for each successive class. To ensure that most pairs of nearest neighbors were included in the first distance class, the relationship that spacing is equal to the inverse of density (Epperson 1989) was used to approximate the minimum distance between any pair of trees. The upper bound of the first distance class was determined by $2^{1/2} * (A/T)^{1/2}$ (Epperson & Chung 2001) where A/T is the inverse of the adult tree density. The bounds for mutually exclusive distance classes were chosen to include 7-10 % of the total joins. Including approximately the same number of joins in each distance class is desired for uniformity of statistical power between classes (Epperson 1989). The total number of pairs or joins (J_t) is $n(n-1)/2$, where n is the total sample size.

After removing individuals with missing genotypes, separate I indices were calculated for each allele of each locus for each of the ten distance classes. Moran's I values roughly range from +1 for strong positive autocorrelation for pairs of trees that have identical genotypes, to -1 for strong negative autocorrelation for genotypes that are dissimilar. For each distance class, a 95% confidence interval was constructed from the standard error of the mean and compared to $E(I)$ to determine if the study populations differed significantly from the null hypothesis of no spatial structure. $E(I) = -1/n-1$ (Cliff & Ord 1981), where n is the sample size, and $E(I)$ approaches zero for a random distribution with no autocorrelation. To control for the accumulated risks associated with multiple type I errors, the Bonferroni's approximation was applied to each allelic class to adjust comparison-wise error and ensure overall experiment-wise (correlogram) protection. The Bonferroni's correction (Kuehl 1994) is given as $\alpha_c = \alpha_e / k$, where α_c is the comparison-wise error rate, α_e is the experiment-wise error rate, and k is the number of independent tests.

Many of the microsatellite alleles are low frequency alleles with $p < 0.05$, or less than 12 allele copies per population. The Moran's I values for the individual

alleles were highly variable and may have been influenced by these small sample sizes. Although the Moran I values for individual alleles were highly variable, for the final analyses, the mean population I indices provided meaningful indicators of spatial structure. Excluded from final analysis were alleles present in less than five copies. These low frequency alleles were considered to contain insufficient information for spatial analyses. The second allele of a bi-allelic locus was also not considered for further analysis since these alleles are correlated ($p = 1 - q$) and contribute identical information. The remaining individual I values were averaged over all alleles for each locus separately. Also calculated was the un-weighted average over all alleles and loci to generate a single Moran's I statistic for each distance class for each population. I compared the mean values of the first distance class to look for significant differences in spatial structure between the two populations.

In two closely related populations, allele frequencies are historically correlated through their ancestors. To compare population means in dependent samples, a paired t - test (Gonick & Smith 1993) was used to control for variability between loci by comparing the structure within a single locus (i). I calculated d_i , or the difference

between the mean I indices in the i th loci pair, as a single measure of difference for each locus. For example, the d_i for Rps2 (d_2) = $(\underline{I}_{OG}) - (\underline{I}_{SS})$, where $(\underline{I}_{OG})$ = the mean Moran's I for the old growth population at locus Rps 2 and $(\underline{I}_{SS})$ = the mean Moran's I for the second growth population at locus Rps 2. The expected value of the difference is $(d_i) = 0$. The paired t-test statistic was constructed from the mean and standard deviation of the differences and is defined as $t_{\alpha, n-1} = n^{1/2} * \underline{d} / S_d$, where $\underline{d}$ = the sample mean of the differences, n = number of allele pairs, and S_d = the standard deviation of the differences.

Autocorrelation is not the same for all distance classes. For a graphical display of the spatial patterns in each population, the mean I statistics were plotted against distance to produce a correlogram. Inferences can be formed about a population's structural pattern through the shape of its correlogram.

RESULTS

Characterization Of loci

Polymorphism:

The seven nuclear (CA)_n microsatellite loci used to evaluate the two white pine populations in the study detected polymorphisms in both populations. Two to 14 alleles were detected for the loci assayed for the 242 individual trees sampled, with an average of 7.7 alleles per locus (Table 2). The locus with the largest number of alleles (14) was Rps50. In contrast, the Rps127 locus had only two alleles. There were a total of 54 alleles for all loci combined. Allele sizes ranged from 144 to 213 base pairs (bp). Locus Rps50 had the largest range of allele sizes (32 bp) among the seven SSR markers and Rps127 had the smallest (2 bp).

The loci for Rps markers 1b, 2, and 6 produced variants with multiples of either 1 bp or 2 bp differences (Figure 4a-c). In contrast, for Rps loci 39, 50, 84 and 127, allele size differed by multiples of two bases, i.e. one repeat unit, as expected (Figure 4d-g). In most instances, the observed number of alleles was less than the total number of alleles possible based on the size range of the marker and predictions based on step-wise changes in the repeat sequence. For example, locus Rps39 exhibited

Table 2. Seven nuclear (CA)_n simple sequence repeat (nSSR) loci characterized for mature white pine. A total of 242 trees were assayed. Reported are numbers of chromosomes sampled, observed number of alleles per locus (k), observed allele sizes in base pairs (bp), change in number of base pairs between the smallest and largest alleles, amplified size of most common allele in basepairs (bp), and frequency of most common allele (x_i).

Rps locus	Sample size	Allele no. (k)	Allele sizes (bp)	Change in bases	Most common allele (bp)	Frequency (x _i)
1b	482	8	193 - 213	20	207	0.89
2	484	10	151 - 173	22	165	0.63
6	484	9	159 - 186	27	161	0.67
39	480	5	167 - 179	12	169	0.64
50	484	14	153 - 185	32	169	0.28
84	484	6	144 - 162	18	146	0.73
127	476	2	192 - 194	2	192	0.73
Mean	480	7.7				0.65
SD ¹		3.9				0.19
Total		54				

¹ Standard deviation

Allele frequency distributions

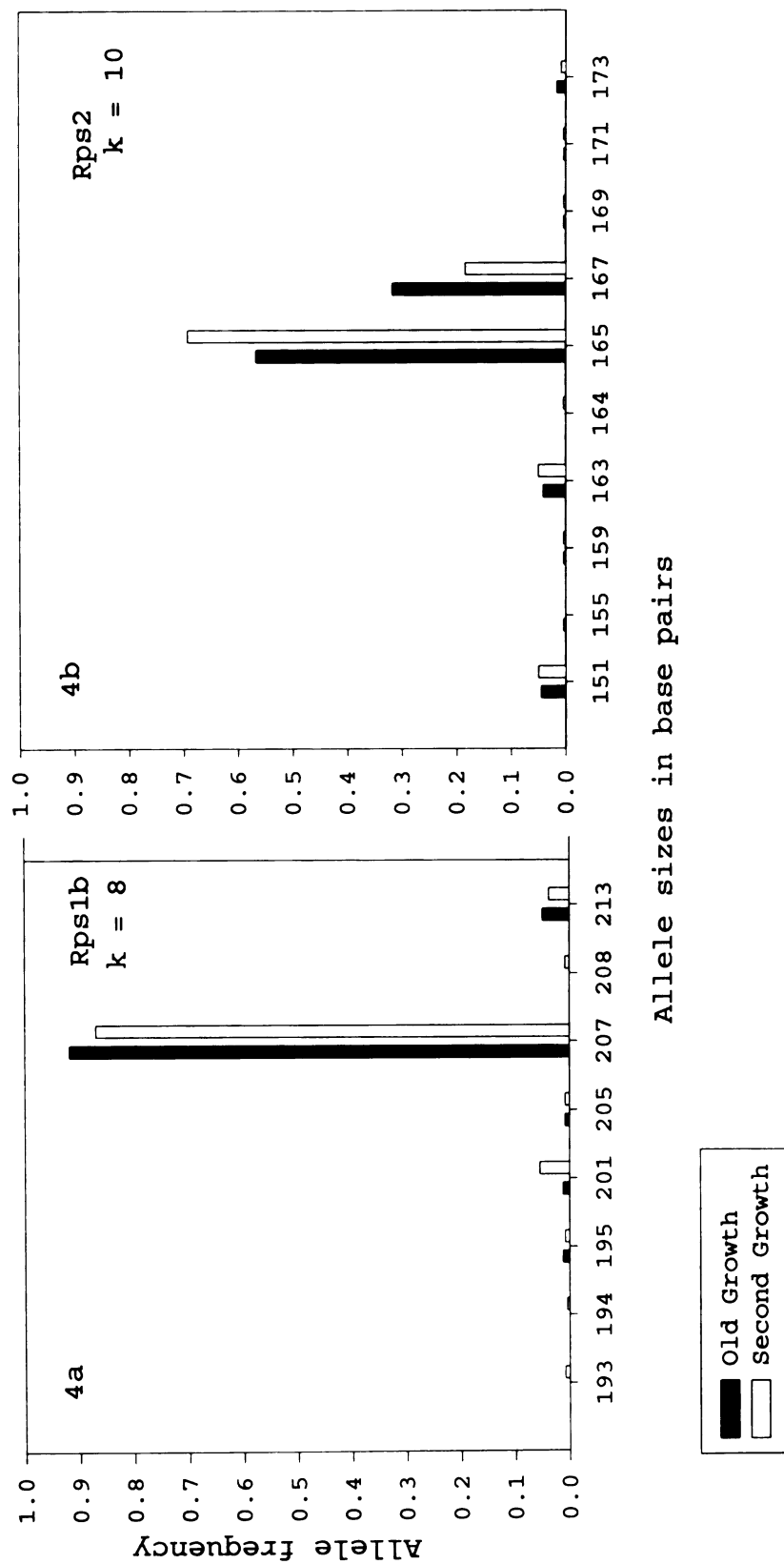


Figure 4a-g: a = Rps1b, b = Rps2, c = Rps6, d = Rps39, e = Rps50, f = Rps84, and g = Rps127. Distribution of allele frequencies detected at seven nSSR loci in 2 mature populations of white pine.

Allele frequency distributions

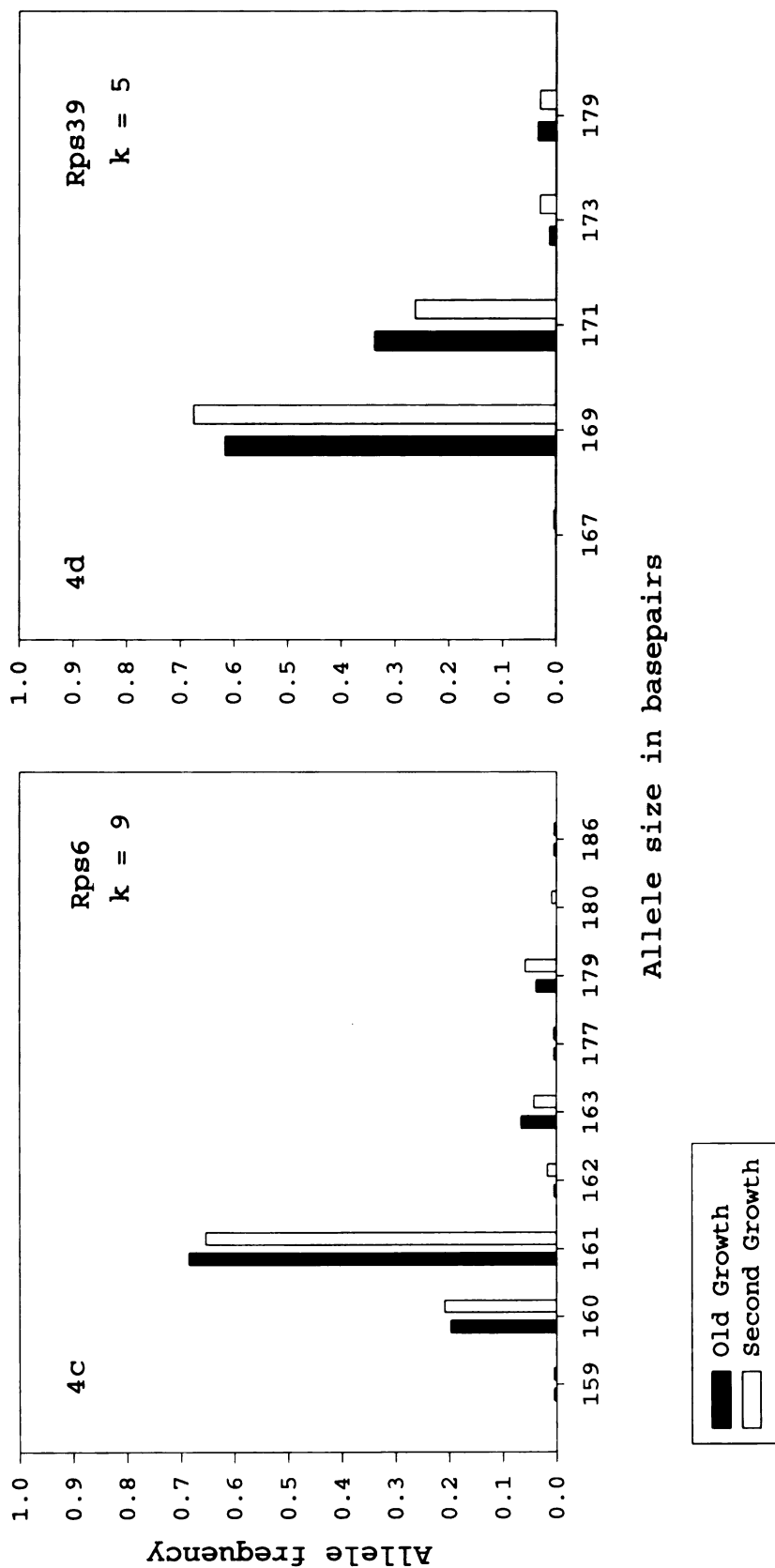


Figure 4 (continued)

Allele frequency distribution for locus Rps50

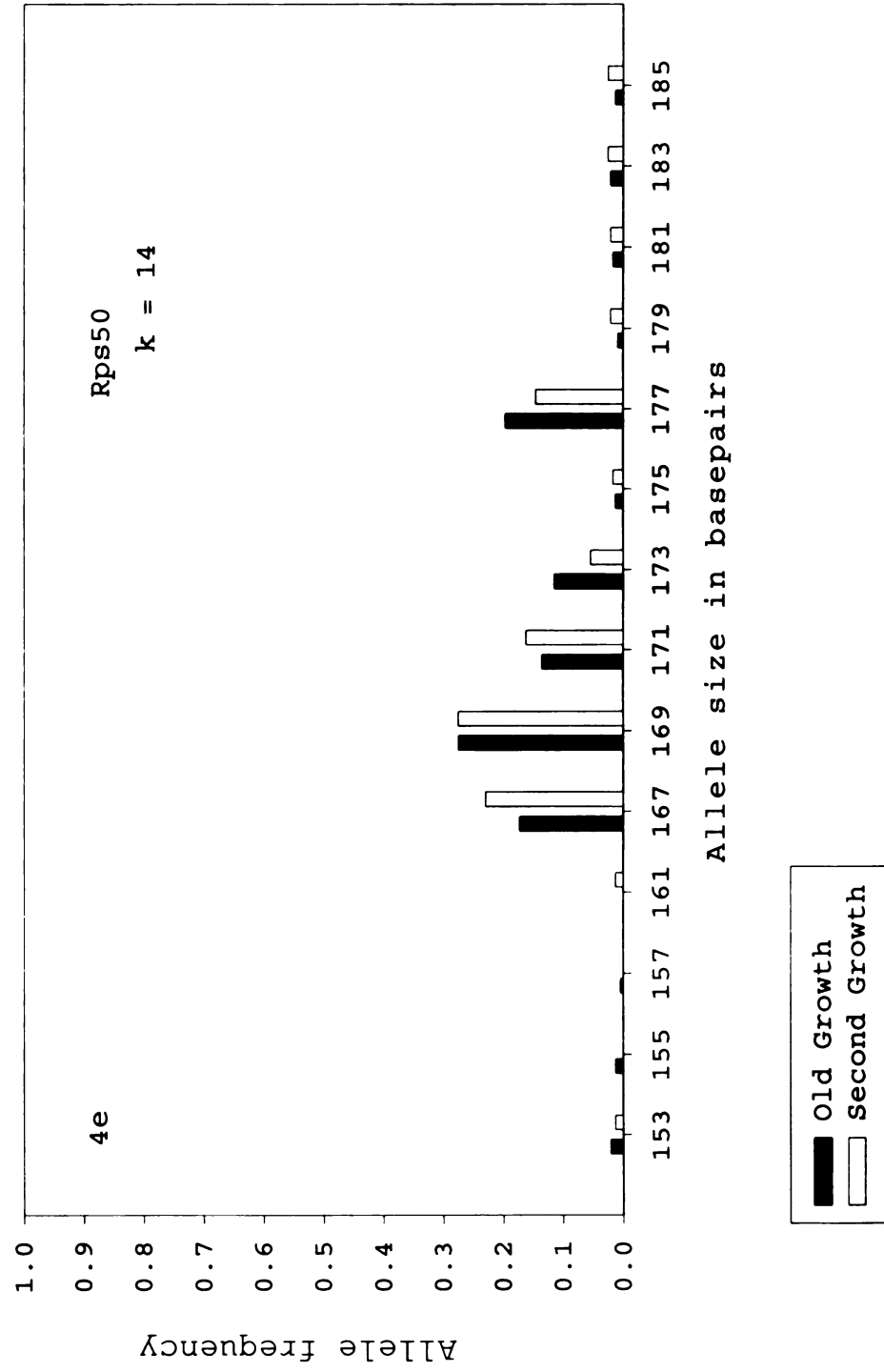


Figure 4 (continued)

Allele frequency distributions

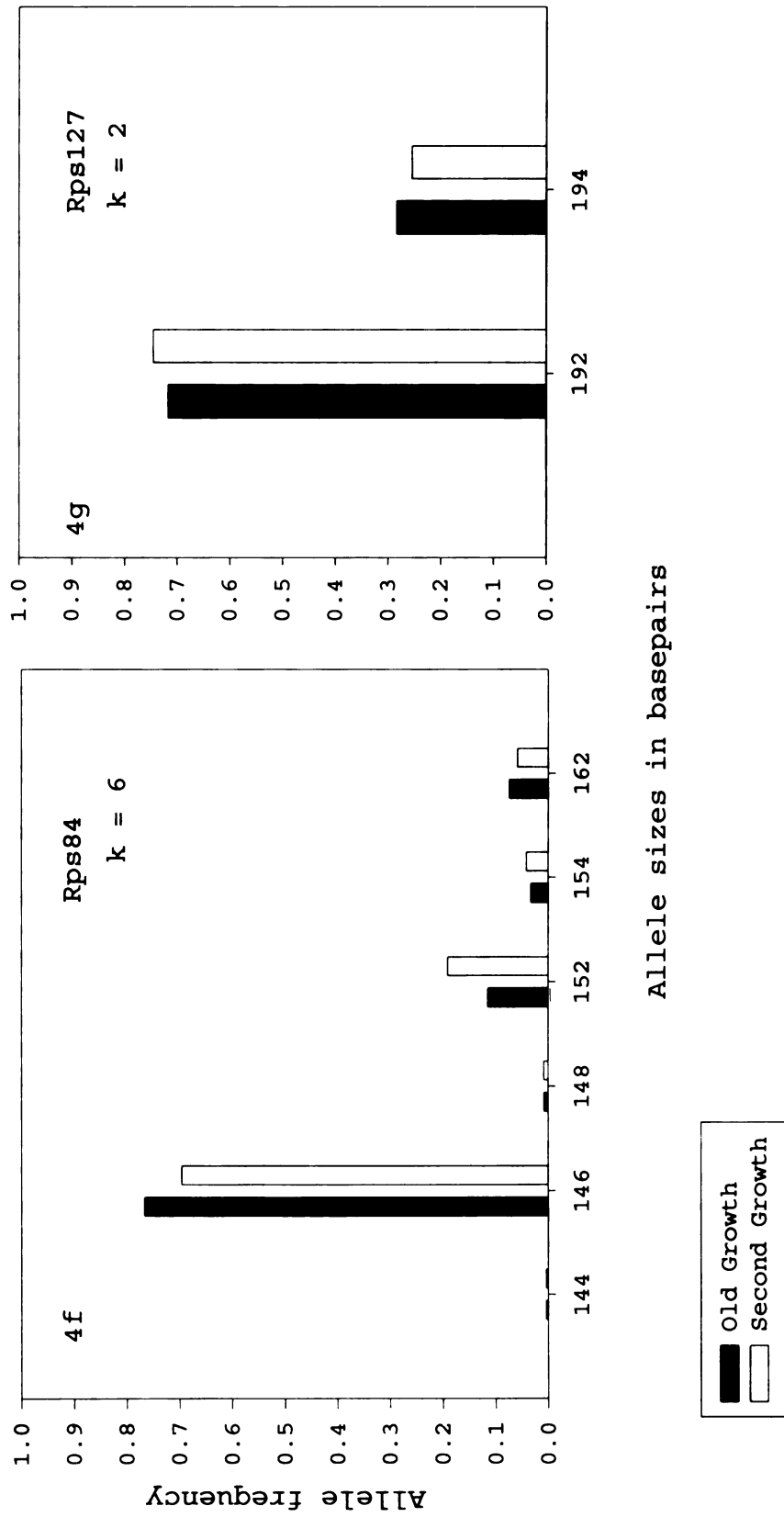


Figure 4 (continued)

variants that differed in size by multiples of two base pairs (Figure 4d). The change in size from the smallest allele of 167 bases to the largest allele of 179 bases was 12 bp (Table 2). The maximum number of alleles expected from a non-interrupted, two-step distribution is seven, but the observed number was only five. Only the locus Rps127 exhibited a non-interrupted allele frequency distribution, having two alleles separated by one repeat unit (Figure 4g). The other six distributions contained interruptions of two to seven repeat units. For example, locus Rps6 had a gap of 14 bp or seven repeat units between morphs 163 and 177, for both populations (Figure 4c). This was the largest distribution break found among the seven loci.

Allele Frequency:

Allele frequencies for each population were estimated from 232 to 244 sampled chromosomes (Table 3a-b). The frequency of the most common allele varied from 0.28 (OG and SS) for locus Rps50 to 0.87 (SS) and 0.92 (OG) for Rps1b, with an average frequency of 0.65 for all loci. For most loci the most common allele was in majority ($x_1 = 0.5656-0.9180$). The only exception was Rps50. This locus had both the largest number of alleles and the most even distribution of alleles frequencies, which ranged

Table 3a and 3b; a = OG, b = SS. Allele frequencies measured in two mature populations of *Pinus strobus*. The number of individuals sampled was 122 and 120 individuals in OG and SS respectively. Alleles are ordered according to size. The smallest allele is listed first (allele 1).

3a, OG

Allele \ Locus	Rps1b 244 ¹	Rps2 244	Rps6 244	Rps39 240	Rps50 244	Rps84 244	Rps127 244
Allele 1		0.0451	0.0041		0.0205	0.0041	0.7172
Allele 2		0.0041*	0.1967	0.6167	0.0123*	0.7664	0.2828
Allele 3	0.0123	0.0041	0.6844	0.3375	0.0041*	0.0082	
Allele 4	0.0123	0.0410	0.0041	0.0125		0.1148	
Allele 5	0.0082		0.0656	0.0333	0.1721	0.0328	
Allele 6	0.9180	0.5656	0.0041		0.2746	0.0738	
Allele 7		0.3156	0.0369		0.1352		
Allele 8	0.0492	0.0041			0.1148		
Allele 9		0.0041	0.0041		0.0123		
Allele 10		0.0164			0.1967		
Allele 11					0.0082		
Allele 12					0.0164		
Allele 13					0.0205		
Allele 14					0.0123		

¹ Number of chromosomes sampled

* Private allele

Table 3 (continued)

3b, SS

Allele \ Locus	Rps1b 238	Rps2 240	Rps6 240	Rps39 240	Rps50 240	Rps84 240	Rps127 232
Allele 1	0.0084 [*]	0.0500	0.0042	0.0042 [*]	0.0125	0.0042	0.7457
Allele 2	0.0042 [*]		0.2083	0.6750		0.6958	0.2543
Allele 3	0.0084	0.0042	0.6542	0.2625		0.0083	
Allele 4	0.0546	0.0500	0.0167	0.0292	0.0125 [*]	0.1917	
Allele 5	0.0084	0.0042 [*]	0.0417	0.0292	0.2292	0.0417	
Allele 6	0.8697	0.6917	0.0042		0.2750	0.0583	
Allele 7	0.0084 [*]	0.1833	0.0583		0.1625		
Allele 8	0.0378	0.0042	0.0083 [*]		0.0542		
Allele 9		0.0042	0.0042		0.0167		
Allele 10		0.0083			0.1458		
Allele 11					0.0208		
Allele 12					0.0208		
Allele 13					0.0250		
Allele 14					0.0250		

from 0.0041 to 0.2746. The frequency of Rps50's four most common alleles ranged from 0.1458 to 0.2750.

Perfect Repeat Diversity:

For the seven (CA)_n loci assayed, four were categorized as perfect repeats, one an imperfect repeat, and two were compound perfect repeat sequences (Table 4). Upon review of the text sequence for Rps39, I reclassified this locus as an imperfect repeat (AC)₁₃(TC)(AC)₃, which differs from (AC)₁₇, the original motif reported for this locus (Echt, May-Marquardt, et al. 1996). The chromatogram was not available for review. For all categories combined, the number of dinucleotide repeat units (μ) ranged from 11 to 21 with an average of 15.7 units. Nei's genetic diversity (H_e), in addition to other diversity measures, was used to evaluate if polymorphism increased as perfect repeat length increased among loci (Table 5). As the number of perfect repeat units (μ) increased from 11 to 17 ($\bar{\mu} = 14.2$), the level of polymorphism also increased. Genetic diversity or H_e increased from 0.20 (low polymorphism) to 0.82 (high polymorphism). The average H_e of 0.52 indicates moderate polymorphism for all loci. In addition, the observed heterozygosity (H_o) values increased from 0.18 to 0.80 ($\bar{H}_o = 0.49$), the total number of alleles (k) from 8 to 14

Table 4. Repeat sequence, category, and number of repeat units (μ), in seven eastern white pine nSSR markers.

Rps locus	Repeat sequence	Category	Repeat units (μ)
1b	(AC) ₁₁	perfect	11
2	(AC) ₁₅	perfect	15
6	(AC) ₁₄	perfect	14
39	(AC) ₁₃ (TC) (AC) ₃	imperfect	17
50	(AC) ₁₇	perfect	17
84	(CT) ₁₀ (AC) ₁₁	compound (perfect)	21
127	(AC) ₁₀ (AT) ₅	compound (perfect)	15
mean			15.7
SD			3.1

Table 5. Diversity measures for four nSSR perfect CA repeats: number of repeat units (μ), Nei's genetic diversity (H_e), observed heterozygosity H_o , observed number of alleles per locus (k), and effective number of alleles (A_e).

Rps locus	μ	H_e	H_o	k	A_e
1b	11	0.20	0.18	8	1.2
6	14	0.50	0.50	9	2.0
2	15	0.54	0.49	10	2.2
50	17	0.82	0.80	14	5.7
mean	14.2	0.52	0.49	10.2	2.8
SD	2.5	0.25	0.25	2.6	1.9

($\underline{k} = 10.2$), and the effective number of alleles (A_e) from 1.2 to 5.7 ($\underline{A_e} = 2.8$).

Characterization Of Populations

Allele frequency and class:

The distribution of allele frequencies at each of the seven loci was similar between the two study populations (Figure 4a-g). Allele frequencies ranged from 0.0041 to 0.9180 for OG, and from 0.0042 to 0.8697 for SS (Table 3a-b). The old growth population (OG) contained 87% of the total allelic diversity in both populations, and the second growth population (SS) contained 94% (Table 6). Thirty-three and 37% of alleles were in low frequency ($p < 0.05$) for OG and SS, respectively, with an additional 18% distributed as rare alleles ($p < 0.005$) for both populations.

Eighty-one percent of all alleles were shared between populations (Table 7). More private alleles were found in SS than OG, 13% of the 54 alleles were unique to SS, and only 5% unique to the OG population. The frequency of private alleles ranged from 0.0041 (one copy) to 0.0123 (three copies) in OG, and 0.0420 (one copy) to 0.0125 (three copies) in SS (Tables 3a-b), classifying them as either rare ($x_i \leq 0.005$) or low frequency alleles ($0.005 < x_i < 0.05$). Rps1b and Rps50 had the largest number of

Table 6. Distribution of allele frequencies in mature white pine populations: total alleles, common and uncommon alleles, low frequency alleles, and rare alleles. The haploid sample size was 244 and 240 for OG and SS respectively. There were seven loci sampled. The total number of alleles present in both populations was 54. The number of alleles present in each class is in parentheses.

Population Class	OG	SS
Total No. of alleles	87% (47)	94% (51)
Common ($x_i \geq 0.05$)	35% (19)	39% (21)
Uncommon ($x_i < 0.05$)	52% (28)	55% (30)
Low frequency ($0.005 < x_i < 0.05$)	33% (18)	37% (20)
Rare ($x_i \leq 0.005$)	18% (10)	18% (10)

Table 7. Private alleles and common alleles present in two mature populations of white pine. The haploid sample size was 244 and 240 for OG and SS respectively. There were seven loci sampled. The number of alleles is in parentheses.

Population	OG	SS
Private ¹	5% (3)	13% (7)
Shared ²	81% (44)	81% (44)
Total	87% (47)	94% (51)

^{1,2} Percentage of 54 alleles detected in both populations

private alleles, with each locus containing three. For Rps1b, all three private alleles were found in SS. At locus Rps50, two private alleles were found in OG with one in SS (Figures 4a and 4c). There were no private alleles for Rps84 and Rps127; the only two loci categorized as compound (perfect) repeat sequences (Table 4).

Genetic Diversity:

Each of two populations maintains a high level of polymorphism at the seven loci examined (Table 8). Diversity measures were similar for the two populations. All loci were polymorphic in both populations, with 47 and 51 alleles detected in OG and SS respectively. Respective values for the per-locus mean number of alleles or allele richness (A), and the effective number of alleles (A_e), were 6.7 and 2.4 for OG, and 7.3 and 2.3 for SS. On average, levels of heterozygosity were moderately high. Mean observed heterozygosities (H_o) were 2% to 6% lower than the probabilities expected under Hardy-Weinberg equilibrium (H_e). Respective values for H_o and H_e were 0.47 and 0.48 for OG and 0.46 and 0.49 for SS.

Population Structure:

The mean fixation indices F , were 0.01 (OG) and 0.05 (SS), indicating little inbreeding in either population (Table 9). For the OG genotype frequencies, 29% differed

Table 8. General diversity measures for Old Growth and Second Growth *Pinus strobus*: number of trees sampled, percentage polymorphic loci, total number of alleles, allele richness (A), effective number of alleles per locus (A_e), observed heterozygosity (H_o), and expected heterozygosity or Nei's genetic diversity (H_e). Standard deviations are in parentheses.

Population	Old Growth	Second Growth
Diversity statistic		
No. of trees	122	120
Percent loci polymorphic	100%	100%
Total No. of alleles over 7 loci	47	51
Allele richness (A)	6.7 (3.6)	7.3 (3.3)
Effective No. of alleles (A_e)	2.4 (1.5)	2.3 (1.4)
Observed heterozygosity (H_o)	0.47 (0.21)	0.46 (0.16)
Expected heterozygosity (H_e)	0.48 (0.20)	0.49 (0.18)

Table 9. Population indices F and F_{st} for each nSSR locus and as a mean across all loci as measured in two populations of white pine. Seven loci were sampled. Sample size was 122 and 120 individuals in OG and SS, respectively.

Population Statistic Rps locus	OG F	SS F	F_{st}
1b	-0.06	0.16	0.0056**
2	0.15	-0.02	0.0156**
6	0.01	0.01	0.0011
39	0.26**	0.17	0.0048**
50	-0.04	0.09	0.0033**
84	-0.04	0.05	0.0064**
127	-0.23**	-0.11	0.0010
Mean	0.01	0.05	0.0054**
SD	0.16	0.10	0.0050

**** Significant at $\alpha = 0.01$**

from Hardy-Weinberg expectations. Locus Rps127 had a significant excess of heterozygotes while locus Rps39 showed a significant heterozygote deficient. SS genotype frequencies did not vary significantly from values expected under Hardy-Weinberg equilibrium. There was little divergence between the two populations (mean F_{st} of 0.0054). Differences in allele frequencies between populations were statistically significant for 71% of loci.

Spatial Structure:

Spatial structure within both populations was weak. For each of two populations, Moran's I spatial autocorrelation coefficients (I) were calculated for each of 10 distance classes. The indices for the first nine distance classes for OG and SS are reported in Table 10a and 10b, respectively. Shown are mean indices for individual loci, and the overall mean values. Although no precise tests of significance are available for multiallelic - multilocus data, by comparing the overall mean I values for distance class one, the structure at OG ($\bar{I} = 0.015$) is 15 fold greater than that for SS ($\bar{I} = 0.001$). The theoretical expected value of I, when there is no spatial structure, is -0.008. The shape or pattern of spatial structure in the OG stand differs from that in the SS stand (Figure 5). OG had weak positive genetic correlation or similarity between

individuals separated by 15 meters or less, followed by negative correlations or dissimilarity at longer distances. This spatial pattern is in contrast to SS, which appears to have a completely random distribution.

Table 10a and 10b, a = old growth (OG), b = second growth (SS). Mean spatial autocorrelation coefficients (Moran's I) in two mature populations of *Pinus Strobus* for 9 distance classes, 2n = 244¹ for OG and 2n = 240 for SS.

10a										
OG										
	15	25	35	45	55	65	75	85	95	
DistBound ²	1	2	3	4	5	6	7	8	9	
DistClass ³										
Rps1b ⁴	0.050	-0.010	-0.080	-0.035	0.050	0.010	0.045	-0.040	-0.065	
Rps2	0.033	-0.008	0.010	-0.045	-0.055	0.018	-0.005	0.023	0.060	
Rps6	0.003	-0.003	-0.005	-0.015	-0.018	0.008	0.025	-0.058	0.003	
Rps39	0.060	-0.010	0.000	-0.023	-0.030	-0.057	-0.010	0.027	0.070	
Rps50	-0.010	0.018	-0.042	-0.020	0.000	-0.016	0.006	-0.008	0.018	
Rps84	-0.015	-0.020	0.003	-0.033	-0.008	-0.008	-0.005	-0.025	0.018	
Rps127	0.030	-0.050	-0.010	0.040	-0.010	-0.020	0.020	-0.060	-0.050	
Mean ⁵	0.015	-0.006	-0.015	-0.025	-0.014	-0.008	0.007	-0.015	0.019	
SD	0.029	0.021	0.032	0.028	0.032	0.025	0.020	0.035	0.051	

¹ Haploid sample size

² Distance bounds (meters)

³ Distance class

⁴ Moran I values averaged over alleles

⁵ Arithmetic mean over alleles, weighted mean over loci

Table 10 (continued)

10b									
SS									
DistBound	15	25	35	45	55	65	75	85	95
DistClass	1	2	3	4	5	6	7	8	9
Rps1b	-0.007	-0.007	-0.023	-0.027	0.017	-0.013	0.003	-0.010	-0.023
Rps2	0.013	0.003	-0.035	-0.030	0.010	0.023	-0.008	-0.045	0.008
Rps6	0.010	0.020	0.025	-0.015	-0.020	-0.003	-0.043	-0.020	-0.043
Rps39	-0.005	-0.023	-0.005	-0.045	0.018	-0.008	0.018	0.000	-0.040
Rps50	-0.001	-0.012	-0.003	0.009	0.008	-0.037	-0.016	-0.004	-0.006
Rps84	-0.013	0.033	-0.008	-0.025	-0.030	-0.005	-0.030	0.000	-0.008
Rps127	0.030	0.050	0.050	-0.040	0.030	-0.030	-0.060	-0.050	-0.140
Mean	0.001	0.002	-0.005	-0.017	0.002	-0.013	-0.015	-0.013	-0.020
SD	0.015	0.026	0.029	0.018	0.022	0.019	0.027	0.021	0.049

Spatial Correlograms

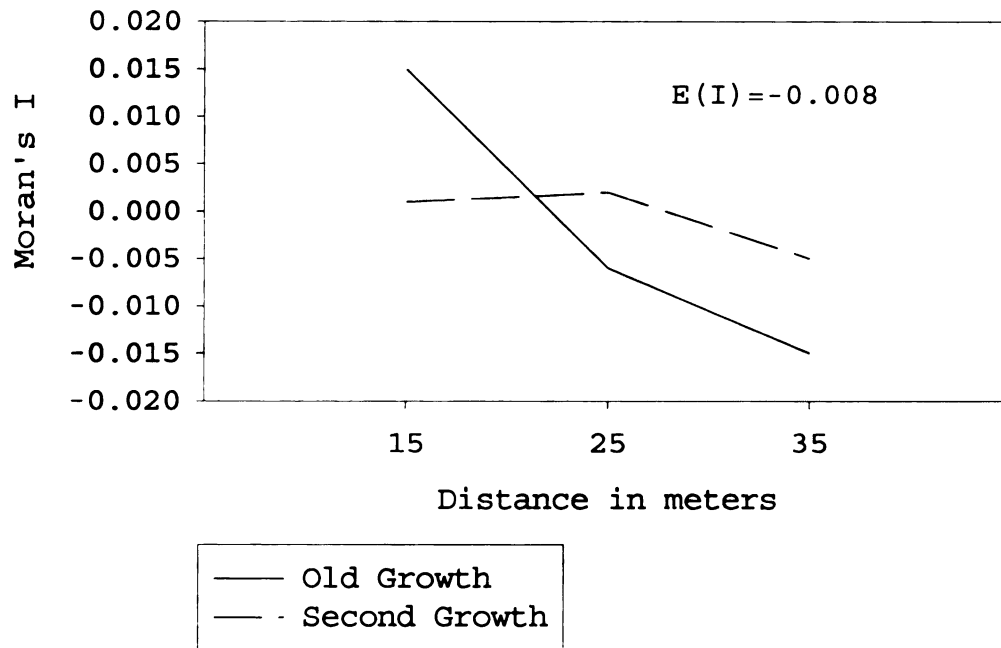


Figure 5. Mean Moran's (I) indices measured spatial structure in two populations of white pine. The I values for the first three distance classes were plotted.

DISCUSSION

Characterization Of Loci

Polymorphism:

The number of alleles was low to moderately high for the seven SSR loci studied. In total, fifty-four alleles were identified for both populations with an average of 7 alleles per locus and a range of 2 to 14 alleles. These measures were lower than values reported by Rajora et al. (2000), where two old growth white pine stands in Ontario, Canada, were evaluated at 13 loci, including 5 of the 7 loci assayed in this study. The latter study had an average of 10 alleles per locus and a range of 2 to 21 alleles. A similar number of trees were sampled for both studies: an average of 121 trees were sampled from the Hartwick Pines populations and 119 trees from the Ontario populations.

The mean number of alleles per locus is a measure of allele richness, independent of allele frequency. It is an important measure for surveying populations for purposes of conservation of unique genes (Marshall & Brown 1975) (Le Corre et al. 1997) or for quantifying effects of tree harvesting on gene pools (Rajora et al. 2000). For example, because allelic richness was moderate for the Ontario study, harvesting had a substantial impact on

allelic composition. A 75% reduction in the size of the breeding populations resulted in a loss of alleles of approximately 26%, and 44% of the lost alleles had frequencies less than 0.05 (Rajora et al. 2000). It is yet unknown how these genetic changes will affect future regeneration.

Allele frequency distributions at microsatellite loci are often predicted using the stepwise mutation model (SMM) (Valdes et al. 1993) (Goldstein et al. 1995) (Slatkin 1995) (Goodman 1997). SMM generally oversimplifies the mutational processes occurring at SSR loci. The strict SMM assumes that mutation only causes the repeat number to change by one unit, and thus alleles with similar sizes are much more related than alleles with very different sizes. However, mutations of SSRs can also result in large changes in allele size (Weber & Wong 1993). Transversions, transitions, indels and duplications occur both within and around the repeat sequence (Callen et al. 1993) (Estoup et al. 1995) (Paetkau & Strobeck 1995) (Grimaldi & Crouau-Roy 1997) (Karhu et al. 2000). The genotypes observed in the Hartwick Pines populations do not include all alleles predicted by SMM theory for neutral alleles. For instance, based on the model and the distribution of allele sizes, the total expected number of alleles would be approximately

73, which is considerably larger than the 54 alleles observed. Five premises can help explain the missing alleles or "distribution breaks" observed at all but one locus (Rps 127) of the study: (1) missing alleles occur in very low frequency and thus were not sampled; (2) missing alleles were lost by genetic drift; (3) more than one repeat was lost or gained during the mutation process; (4) other non-SMM mutations occurred; or (5) missing alleles are under selection or linked to genes under selection.

In addition, 1 bp allele size differences were observed and these cannot be produced under a dinucleotide SMM model. Moreover, it is unlikely that the 1 bp polymorphisms present in Rps1b, Rps2 and Rps6 initiated from the SSR repeat region of these loci. Thus, the number of alleles that are the result of mutation at the repeat region would be further reduced. For example, Rps1b would be characterized as having only six alleles rather than eight, if only size variants with multiples of 2 bp morphs were considered. Another mutation process is called plus-A. During plus-A activity, DNA polymerase catalyzes the addition of a single, non-template nucleotide to the 3' hydroxyl terminus of blunt ended template regions (Clark 1988). The extra nucleotide (usually adenosine) is added to both the stutter peaks and allele peaks of PCR products.

PCR amplification produces a characteristic pattern of 1 bp peaks, which is in contrast to the patterns of variation observed in Rps1b, Rps2 and Rps6. The primer pairs for these three loci produced 2 bp stutters with both 1 bp and 2 bp allele morphs, therefore, plus-A activity does not explain the observed 1 bp polymorphisms. A plausible explanation is the 1 bp variation originated from insertions or deletions of adenosine in the short poly A sequences that flank the repeat CA_n region (Primmer et al. 1998), but only perfect repeat sequences were affected. Since all of the sequences flanking the Rps markers contain poly A regions (data not shown), it is possible that the perfect SSR repeat, and / or the flanking sequence, increases the mutability of the poly A region.

Allele Frequency:

Allele frequency must be considered in addition to total number of alleles when evaluating the effectiveness of molecular markers for quantifying various aspects of population genetic structure. Diversity measures based on evenness of allele frequency are useful for measuring effective heterozygosity levels (H_e), effective population size (Hartl & Clark 1997), inbreeding (Nei 1973, 1977), and for identifying individuals and assigning parentage based

on DNA fingerprinting (Amos & Hoelzel 1992) (Adams 1992) (Blouin et al. 1996).

H_e is a popular diversity measure since it equates diversity with heterozygosity, assigning a direct genetic interpretation to diversity. Ideally for fine scale population genetic studies, a marker should possess a large number of alleles present in equal frequency. For example, Rps50 presented many alleles ($k = 14$), with the four most common alleles having similar frequencies. This locus had the largest A_e and H_e values of the study (5.7 and 0.82 respectively). Rps50 is the best suited of the 7 loci for fine scale population genetic analyses. But equally frequent alleles are usually not the case with microsatellite loci. Although SSR markers can be very polymorphic, a SSR locus is often characterized by one or a few high frequency alleles and many low frequency alleles (Streiff et al. 1998), which makes the allele richness index sensitive to sample size. In the Ontario study (Rajora et al. 2000) for example, allele richness was affected by harvesting, but H_e showed less than a 5% reduction. The mean H_e for both populations was approximately 0.5 both before and after harvest. For these reasons, expected heterozygosity is the diversity measure

of choice when study objectives are to maintain allele frequency evenness of common alleles.

Perfect Repeat Diversity:

Nei's genetic diversity measure, H_e , increased across loci as the perfect repeat length increased. These results agree with human studies in which a similar trend (Weber 1990) is observed for a related measure, the PIC (polymorphism information content) value (Botstein et al. 1980). The imperfect and compound perfect repeat sequences in white pine contain less information than the perfect repeat sequences. These loci have fewer alleles and lower heterozygosity levels than do perfect repeat sequences with the same number of repeat units. Weber (1990) suggested that the higher polymorphism of long, uninterrupted CA_n repeats was the result of higher mutation rates. Strand slippage is hypothesized as the mutational mechanism responsible for many of the new mutations formed in SSR repeat regions (Levinson & Gutman 1987) (Weber 1990), occurring more frequently when the number of repeats is larger.

Increased slippage explains the higher H_e values for alleles with larger repeat lengths. In 1992, Schlötterer and Tautz supported this hypothesis by in vitro experiments. They synthesized repetitive di- and

trinucleotide repeats sequences from short primers, dNTP's, and polymerase, estimating synthesis rates for the various motifs. The authors found that the rate of growth for DNA fragments was independent of the length of the template DNA fragment. Length independence suggests formation of transient single stranded regions within the simple repetitive sequence. Schlötterer and Tautz hypothesized that the unpaired regions could be substrate for repair mechanisms in vivo. The creation and elongation of simple sequence regions could contribute to the length variation observed in these repeat regions.

In non-perfect and compound sequences, the interruptions to the main repeat sequences may stabilize the repeat area by decreasing the number of tandem CA units (Weber 1990) (Jin et al. 1996). Interruptions should reduce slippage and improve proofreading (Taylor et al. 1999), lowering the mutation rate for interrupted sequences with fewer new alleles being formed.

Messler et al. (1996) proposes that the "birth" of a microsatellite is a two-step process. The first mutations are point mutations, creating enough repeat units for expansion to occur by replication slippage. The second mutations are strand slippage, producing repeat length variation. For substantial strand slippage to occur, a

minimum of 10 tandem CA_n repeats is necessary for the replicating DNA polymerase to falter and introduce more or less repeats into the newly synthesized strand than is found in the template strand (Weber 1990). All of the loci in this study have a sufficient number of tandem repeats, ranging from 10 to 17 CA_n units, to mutate further.

In addition to higher mutation rates, perfect repeat sequences may also have higher rates of homoplasy (Estoup et al. 1995). Homoplasy occurs when alleles are identical in state (IIS) but not identical by descent (IBD). Two alleles are IBD if they descend from the same ancestral allele without mutation. For microsatellite analyses, two alleles are homoplastic if they are the *same size* without being IBD. PCR products of the same size but different sequences (convergence) or even the same sequence (parallelism) can arise from independent mutational pathways. Two IIS alleles may also descend from the same ancestral allele if one of the alleles mutates and then back mutates to its original state (reversion), while the second allele remains unchanged. When allele size morphs contain a mix of IIS and IBD alleles, this has a homogenizing effect on genetic differences, decreasing the resolving power of SSR markers for detecting divergence of SSR loci (Nauta & Weissing 1996).

Although homoplasy commonly occurs at microsatellite loci (Estoup et al. 1995) (Grimaldi & Crouau-Roy 1997), it is unlikely to be a concern for the genetic analysis of closely related populations. Homoplasy could be problematic, for instance, if the study populations had been isolated for a long time and were not exchanging migrants.

In one study of honey-bee (*Apis mellifera*), homoplasy was not seen at the population level, or in populations belonging to the same subspecies, or in closely related subspecies, but length convergence was detected among distantly related subspecies (Estoup et al. 1995). Sequencing revealed that homoplastic alleles varied both in the number of TC repeats, and in the position and number of imperfections. In contrast, Grimaldi and Crouau-Ray (1997) demonstrated length convergence at the population level, in humans. Two allele morphs were identical in size but differed greatly in sequence. The unique size variation at this locus results not only from differences in the number of CA repeats, but also from differences in the flanking region. For example, alleles 7 and 8 differ in size by 2 bp. The expected variation is the addition of 1 repeat unit. The actual variation originated from a 2 repeat unit deletion and a 6 bp duplication occurring in the flanking

region, causing the observed 2 bp difference. Together, these studies suggest that in addition to operating in distantly related subspecies, homoplasy can also occur at the population level. However, it is important to note, that in both of these studies homoplasy only involves a fraction of the same size alleles common to the populations under study.

It is unknown if homoplasy operates at the population level for white pine. If it occurs at all, it is most likely rare, and should not affect data interpretation of this study. It may be that SSRs with moderate polymorphism, such as Rps2, Rps6, and Rps39 (and therefore often moderate rates of mutation and homoplasy), may be best suited for studying differences between populations. Markers with the highest polymorphism are the most useful for examining parentage or individual differences on the shortest of time scales. Estoup et al. (1995) suggests that interrupted SSRs, such as Rps84 and Rps127, may have lower homoplasy rates than perfect repeats making them better suited for investigating changes in allele frequency and evolutionary relationships between more distantly related populations.

S
t
t
c
m
s
t
R
o
b
W
O
O

Null Alleles:

At the population level, a greater area of concern for accurate data interpretation is the presence of null alleles (Callen et al. 1993) (Paetkau & Strobeck 1995) (Pemberton et al. 1995) (Jarne & Lagoda 1996) (Gullberg et al. 1997) (Becher & Griffiths 1998) (Van Treuren 1998). Null alleles result from mutations in the sequence at the priming site. Mispriming prevents PCR amplification of the target sequence; therefore genotypes are no longer visible through gel electrophoresis.

The primer pairs chosen for marker analysis can greatly affect the results, since the sequences flanking the SSR repeat region may be highly variable in addition to the repeat region. After the initial analysis was completed, it was discovered that the primer pairs for markers Rps6 and Rps34b were designed from the same sequence and flanked the same nSSR locus (Figure 6). Thus they amplify the same locus. I also suspected that marker Rps34b contained null alleles, which would explain the observed differences in the allele frequency distributions between the two markers. For these reasons, marker Rps34b was dropped from the final analyses. Nonetheless, the data on Rps6 and Rps34b was quite informative about the nature of alleles in *P. strobus*. Very little of the variation at

the Rps6/Rps34b locus appears to be in the dinucleotide repeat region. For example, the allele distribution pattern for Rps34b is straightforward and is characterized by five 1 bp morphs (142-146) (Figures 7a-b). In contrast, Rps6 has a greater allele number ($k = 9$) with a 14 bp break between the small alleles (159-163) and the large alleles (177-186). The allele frequency distribution for Rps6 is more complex than that of Rps 34b. Both 1 bp morphs (alleles ranging from 159 to 163) in addition to alleles differing by three repeat units (allele pair 180/186) were observed. It is also possible for one allele to contain both repeat (2 bp) and non-repeat (1 bp) polymorphisms. For example, allele 180 of Rps6 differed in size from the 179 bp allele by 1 bp, while allele 179 differed by 2 bp or 1 repeat unit from the 177 bp allele.

One premise for the observed genetic differences between these two primer pairs is the large alleles (177/179/180/186) of Rps6 are artifacts. A second explanation is mispriming at the Rps34b locus producing non-amplifying null alleles. If the large alleles of Rps6 were artifacts, we should occasionally see samples with three peaks, but this did not occur. If the suspected large alleles (160/162/163/169) of Rps34b are absent

pPs6¹ (marker sequence size 162 bp)

CCTATGACAACCTAACCCATGGGACGACTTACATAGTCGGATAATCCATGCCGGCCCTTG
cATGAATTTTTTAAAACACTGATTTTCTAATCAGTGTGCGCTACATAACCTAGCGCAC
CAGTGTCTCTTATCACAGCGCACCAAGCACATTTTGTATATAAAC**CACACACACACACAC**
ACACACACACACATATTATTTTATTTATTTTAAATATGTGCACCATATGTAAAATAGGG
CAGCGGTGCGCTATTTTATTATAGTGCATTACGCAaGT

pPs34 (marker sequence size 145 bp)

CCTATGACAACCTAACCCATGGGACGACTTACATAGTCGGATAATCCATGCCGGCCCTTG
nATGAATTTTTTAAAACACTGATTTTCTAATCAGTGTGCGCTACATAACCTAGCGCAC
CAGTGTCTCTTATCACAGCGCACCAAGCACATTTTGTATATAAAC**CACACACACACACAC**
ACACACACACACATATTATTTTATTTATTTTAAATATGTGCACCATATGTAAAATAGGG
CAGCGGTGCGCTATTTTATTATAGTGCATTACGCAtGT

Figure 6. Sequence data for two cloned white pine fragments containing (CA)₁₄ repeats. Marker Rps6 was derived from the plasmid clone pPs6 and Rps34b from pPs34 with a 17 bp difference in sequence size. Forward and reverse primers are underlined. The repeat sequences are in **bold** typeface. Lower case letters indicate differences between the sequences.

¹ plasmid *Pinus strobus* clone 6

Allele frequency distributions

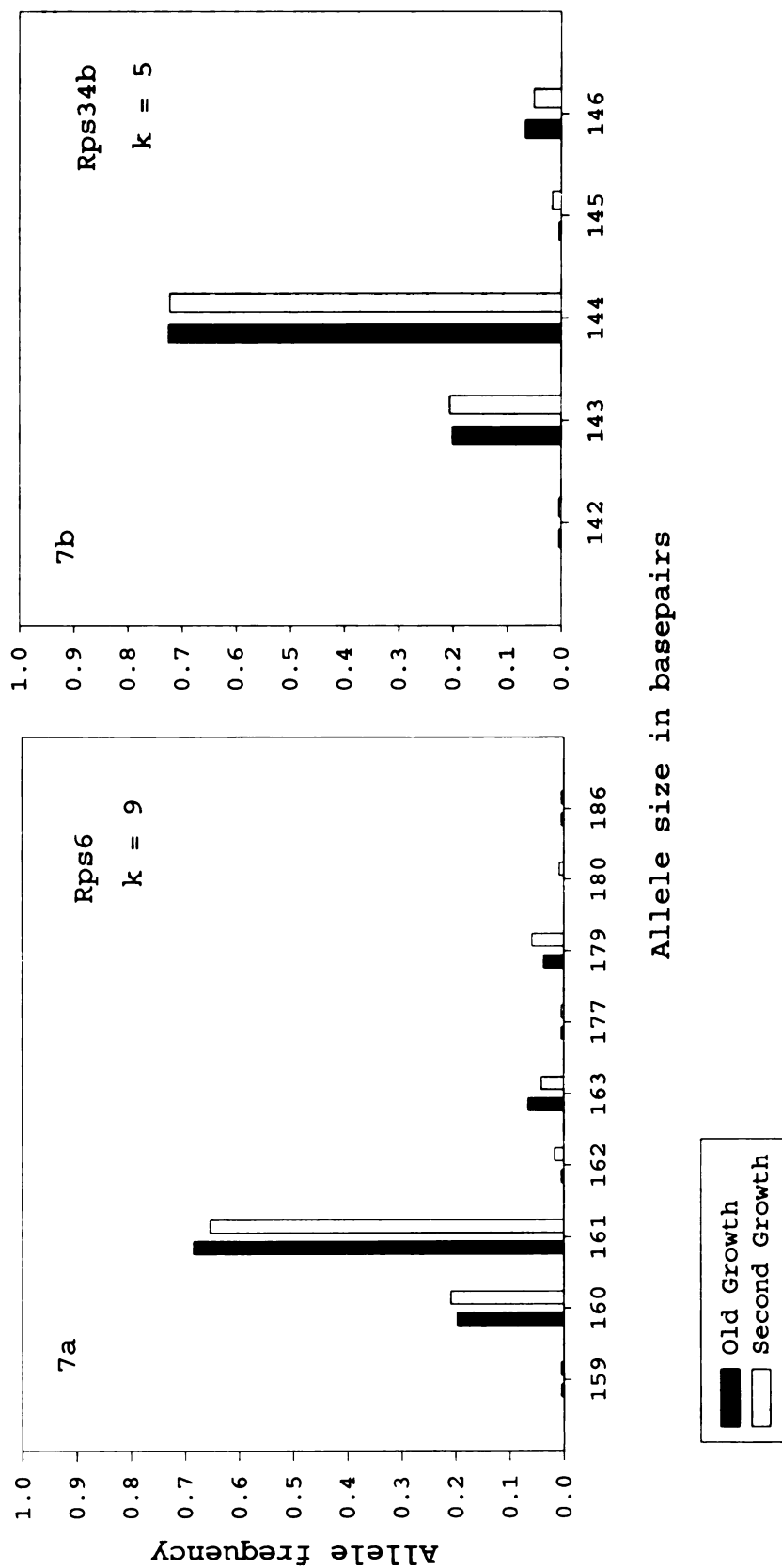


Figure 7a and 7b: a = Rps6 and b = Rps34b. Displayed are the allele frequency distributions for markers Rps6 and Rps34b. These markers amplify the same (CA)_n locus, with Rps 34b having four null alleles corresponding to the Rps6 alleles 177, 179, 180 and 186.

b

a

ai

ge

di

si

of

su

pr

th

(T

co

mi

Rps

lev

us

and

nul

mut

sit

hom

sho

resu

Call

because they are null alleles, then treating the large allele heterozygotes of Rps6 as small allele homozygotes and subtracting 17 bp from the Rps6 alleles should produce genotype scores identical to Rps34b. The expected 17 bp differences in allele sizes were based on the sequence sizes of the two markers (Figure 6), with a sequence size of 162 bp for Rps6, and 145 bp for Rps34b. The data supports the Rps34b null allele hypothesis with the second primer pair (Rps6) providing complete genotypic data for the locus. For the 28 trees that have large Rps6 alleles (Table 11a-b), all corrected genotypic assignments correspond to the genotypes for Rps34b. As with many microsatellite markers, within population variation for the Rps markers is not well characterized at the sequence level. The markers were isolated from genomic libraries using oligonucleotide probes (Echt et al. 1996). Cloning and sequencing the Rps34b null alleles would verify our null allele interpretation and determine the exact mutational changes that occurred at the Rps34b priming site. In addition, sequencing would enable us to look for homoplasy at the population level. Several studies have shown that variation at the priming site has occurred resulting in null alleles due to mispriming. For example, Callen et al. (1993) determined that an 8 bp deletion

within the priming sequence prevented the binding of one primer. Paetkau and Strobeck (1995) identified a G to C transversion at the 3' end, which prevented binding of the primer sequence and resulted in null alleles. For this study, a possible DNA insertion is suggested by the difference of 14 bp between the small and large alleles of Rps6. Since the Rps34b forward priming site lies interior to the Rps6 forward priming site (Figure 6), an insertion occurring downstream of the Rps6 site and interrupting the Rps34b forward priming site could produce the null allele pattern that we observed in our data set. Moreover, the mutation underlying the Rps34b null alleles seems to be ancient, since the large alleles of Rps6 vary in size according to predicted microsatellite patterns, with morphs changing by multiples of 1 or 2 bp repeat units.

The observed frequency of the Rps34b null alleles was 0.059 (29 of 484 alleles, see Table 11a-b) for both populations combined. In addition, one null homozygote in 242 trees was observed ($p = 0.004$), which agrees with the Hardy-Weinberg expected frequency of 0.003 (0.059^2) or 0.7 null homozygotes ($0.003 * 242$). It is estimated that 25-30% of all microsatellite markers contain null alleles

Table 11a and 11b: a = OG and b = SS. Markers Rps6 and 34b amplify the same locus. Characterized are the large allele heterozygotes for Rps6 (alleles 177, 179, 180 and 186) with the corresponding genotypes for Rps34b where the large allele is a null allele. Rps34b homozygotes are 17 bp smaller than the small Rps6 alleles.

11a

OG Sample	Rps6 peak 1	peak 2	Rps34b peak 1	peak 2
8	161	186	144	144
11	161	179	144	144
12	161	179	144	144
16	161	179	144	144
41	161	179	144	144
46	161	179	144	144
72	160	179	143	143
76	161	179	144	144
100	161	179	144	144
107	161	177	144	144
140	161	179	144	144

11b

SS Sample	Rps6 peak 1	peak 2	Rps34b peak 1	peak 2
136	161	177	144	144
163	161	179	144	144
176	161	179	144	144
177	161	179	144	144
200	160	179	143	143
201	161	179	144	144
202	161	179	144	144
208	179	180	0	0
212	161	179	144	144
214	161	179	144	144
220	161	179	144	144
223	160	179	143	143
228	160	179	143	143
242	161	179	144	144
244	161	180	144	144
264	161	186	144	144
303	161	179	144	144

(Callen et al. 1993) in frequencies up to 0.15 (Jarne & Lagoda 1996). If undetected and unaccounted, a high incidence of null alleles could bias conclusions drawn from a study. For example, allele frequencies will be inflated for the alleles opposite the null allele in a genotype and absent for the null allele itself. In addition, the presence of a null allele reduces the informativeness of a co-dominant marker because the null allele behaves like a recessive allele, making it impossible to distinguish a null allele heterozygote from a non-null allele homozygote. Increases in reported homozygosity levels may result in artificially significant inbreeding coefficients since the population will appear to have more homozygotes than expected at Hardy-Weinberg equilibrium.

Locus Rps6/34b:

The rate of miscalls for the Rps6/Rps34b locus was estimated by comparing the two complete sets of genotypes for these two markers. There were errors in four genotype assignments for the SS population (Table 12), resulting in a 2% (4 of 242 trees) discrepancy between the two markers. There were no similarly affected trees in the OG population. In all cases, the second larger peak of the Rps6 genotypes were either missing or weakly amplifying and not scored, resulting in the assignment of homozygous

gene

in R

stud

expe

mark

Tabl

defi

Seco

scor

when

17 b

disc

SS

Samp

155

185

254

258

genotypes for Rps6 as opposed to heterozygote assignments in Rps34b. Rps6 was the weakest amplifying loci of the study. Since this marker is more difficult to assay, it is expected to have a higher error rate than other Rps markers.

Table 12. Estimating the rate of miscalls for the locus defined by markers Rps6 and Rps34b. Characterized for the Second Growth (SS) population are the homozygous genotypes scored for marker Rps6 that present heterozygous genotypes when scored for marker Rps34b. Rps34b peak 1 alleles are 17 bp smaller than Rps6 peak 1 alleles. The alleles with discrepancies are in **bold** typeface.

SS Sample	Rps6 peak 1	peak 2	Rps34b peak 1	peak 2
155	160	160	143	144
185	160	160	143	146
254	161	161	144	146
258	160	160	143	144

Characterization of Populations

Allele Frequency and Class:

Populations that are close together exchange more migrants and are more similar in gene frequencies than populations that are far apart (Wright 1943). As expected from their close proximity, the two populations compared in this study had very similar allele frequency distributions.

Allele frequencies in a population typically range from near zero to near one. Rare alleles are more likely to be unique to a single population. For this reason, rare alleles ($p \leq 0.005$) (Hartl & Clark 1997) are useful for identifying individuals (Hartl & Clark 1997) and for distinguishing recent immigrants. The two loci with the largest number of rare alleles, and therefore most useful for identification in paternity, forensics and immigration analyses, were Rps2 and Rps6. Both loci are perfect repeat sequences. Five rare alleles were detected for Rps2 and four for Rps6. Rps6 morph 162 was rare in OG (1 copy) but had low frequency in SS (4 copies). Three rare alleles were detected for Rps1b, one each for Rps39, 50 and 84, and none for Rps127.

When rare and low frequency alleles are under selection or are very tightly linked to genes affecting fitness, they may represent the genetic potential necessary

for adaptation to future environmental change (Rajora et al. 2000). This is because more common alleles would have been selected for in the current environment or in the past. Thirty-five percent of alleles had frequencies less than 0.01 (2 copies), for both populations combined. These results were 19% higher than those of Rajora et al. (2000). An average of 16% of the alleles in the two Ontario white pine stands prior to harvest had frequencies less than 0.01. The difference in number of rare alleles between the two studies most likely reflects differences in the markers used to evaluate these populations. For example, there could be dissimilarities in allele frequency distributions among loci, some markers will have more rare alleles than others. The Ontario populations were characterized at 6 additional loci above the number used to characterize the Hartwick Pines study.

On average, 81% of alleles were shared between populations, similar to the 86% for the Ontario stands (Rajora et al. 2000). Private alleles are rare or low frequency alleles that are present in one population but absent in another. This allele class can be useful for estimating gene flow, or the rate of successful migrants exchanged among local populations (Slatkin 1985). Five of the seven loci contained private alleles. On average, 9%

o
c
S
a
g
o
o
P
t
r
m
b

p
c
f
m
d
a
s
r
t
f
p

of alleles were private in the Hartwick Pines study as compared to 7% for the Ontario stands (Rajora et al. 2000). Since both studies conducted a complete census with approximately the same number of trees sampled in each old growth population, the differences in the reported number of private alleles most likely reflects the larger number of loci sampled in the Ontario study. In both Hartwick Pines populations, the number of private alleles is too low to reliably estimate migration rates. Slatkin (1985) recommends 20 or more private alleles per sample for migration estimates to be close enough to the true value to be useful.

It is interesting to note that the only loci without private alleles (Rps84 and Rps127) are the only two classified as compound perfect repeats. One explanation for the absence of private alleles is the lack of new mutations. It is generally accepted that interruptions decrease SSR mutation rates (Weber & May 1989) (Taylor et al. 1999). Interruptions are also the hypothesized first step in the "death" of a microsatellite, stabilizing the repeat for subsequent deletion (Taylor et al. 1999). If the compound repeats in these two loci are recent formations, interruptions could help explain the absence of private alleles.

Ge

pr

lo

ch

ou

et

was

0.4

val

gen

rep

H_a ,

com

pre

div

dif

or

opp

The

for

turn

popu.

Genetic Diversity:

Conifers combine several life history traits that promote high genetic diversity: large geographic range, long lived perennial, late successional species, high chromosome number, sexual reproduction, high fecundity, outcrossing breeding system, and wind pollination (Hamrick et al. 1979) (Ledig 1998). Average expected heterozygosity was moderately high for the two populations combined ($H_e = 0.49$) and just 4% higher than the observed random mating values ($H_o = 0.47$), suggesting little inbreeding. Mean genetic diversity values were 20% lower than values reported by Rajora et al. (2000). For the Ontario study, H_e , averaged over both pre-harvest populations, was 0.61, as compared to 0.49 for the Hartwick Pines populations. Two premises could explain the decrease in the measured diversity as compared to the Ontario populations: (1) differences in marker polymorphism and allele frequencies; or (2) the Hartwick Pines populations are younger with less opportunity for selection against homozygous genotypes. The average stand age is 175 years as compared to 250 years for the Ontario populations.

Heavy logging occurred in the Hartwick region near the turn of the century. The SS was logged while the OG population was left undisturbed. Since logging reduces the

p

c

r

d

f

e

i

o

a

o

b

n

w

r

b

s

g

l

a

b

o

p

s

population size, it could be possible, under certain circumstances, for logging to decrease the diversity of the regenerating seedlings. A decrease in diversity would depend on many factors and would not occur every time a forest is logged. For example, a severe reduction in effective population size would leave just a few breeding individuals. If new seed parents were not introduced from outside of the stand, then just a few seed parents would be available to re-establish the stand, reducing the diversity of the regenerating progeny. This is because alleles would be lost since fewer parents would contribute genes to the new generation of white pine. In addition, genetic drift would cause further erosion of genetic diversity through random loss of alleles during gametic sampling.

In this study, there is little genetic difference between stands, with diversity measures (A , A_e , H_e and H_o) similar between populations. An explanation for similar genetic diversity between the OG and SS stands is the logging induced change in N_e was not severe enough to cause a decline in diversity levels. A second explanation would be the existence of a viable seedbed. Intense fires often occurred following logging. If the SS stand was spared a post logging burn, genetic diversity would be maintained since both seedbeds and younger uncut trees, along with the

s
m
h

F
i
c
l
o
s
s
p
4
p
f
a
c
g
-
wa
(F
si
po

seed trees, could have all contributed seed to the regeneration of the SS.

Population Structure:

In addition to high genetic diversity, we expect population structure to be weak for both populations. This is because white pine is primarily a wind pollinated, out-crossing species with little bi-parental inbreeding. The losses of heterozygosity from population structure are often measured by fixation indices or Sewall Wright's F-statistics. In support of this hypothesis of weak structure, mean F values indicated little inbreeding in the populations. The mean microsatellite fixation indices were 45% lower than the allozyme analysis of the same tree populations (Table 1) (Epperson & Chung 2001). Mean fixation indices, were 0.05 and 0.11 for the SSR and allozyme marker systems, respectively. These results are consistent with similar findings in the Ontario white pine genetic study. Using their reported H_o and H_e values ($F = 1 - (H_o / H_e)$), the mean microsatellite fixation index (0.141) was 20% lower than the corresponding allozyme value (0.179) (Rajora et al. 2000) (Buchert et al. 1997).

Twenty-nine percent of the OG loci differed significantly from Hardy-Weinberg expectations. The positive F value for Rps39 indicates an excess of

homozygotes. Inbreeding from mating relatives or as a result of population subdivision from geographic or temporal isolation (Wahlund effect), seems an unlikely explanation for the observed heterozygote deficiency. In theory, inbreeding affects all loci, but the observed significant deviation above H-W equilibrium for the OG population occurred at just 1 of 14 F values. One possibility for this decrease in heterozygosity would be the presence of a null allele, and there is some evidence to support this premise. Null homozygotes could explain the failed Rps39 PCR reactions for two trees in the OG population. The frequency of the Rps39 null allele was estimated from the observed deficiency of heterozygotes (Brookfield 1996). Using equation 4, individuals with no visible bands were treated as observations with the null allele frequency $(r) = (H_e - H_o)(1 + H_e)^{-1}$. The frequency of null alleles is estimated to be 0.07 in both populations combined. If assuming the two failed PCR reactions were the result of double null alleles, then the observed null homozygote frequency would be 0.008 (2 of 242 trees) with a null allele frequency of 0.09 ($0.008^{1/2}$), which is close to the frequency estimated from Brookfield's (1996) equation. As mentioned above, there is compelling evidence that null alleles also occur in marker Rps34b. Since marker Rps6

amplifies the same locus as Rps34b, and did not contain null alleles, Rps34b was removed from the final analysis.

Population divergence was negligible. Because statistical power was very high, the small differences in allele frequencies were statistically significant for 71% of loci. Spatial distance between populations or perhaps timber harvest at the SS, are reasonable explanations for the observed differences in allele frequencies between populations.

The effects of mutation and migration are similar in that they both introduce new genes into a population, but biologically the effects of these two evolutionary forces are very different. Mutation normally causes differentiation, in contrast to migration, which has a homogenizing affect. Typically, allozyme mutation rates are much lower than migration rates and would not affect F_{st} . But if mutation rates are high relative to migration, new mutations may increase variance among populations, by negating some of the homogenization of migration. This situation may apply to the relationship between F_{st} and the higher mutation rates found at some highly mutable microsatellite loci. The mutation rates of microsatellite markers are estimated from 10^{-4} - 10^{-2} , and are 100 to 10,000 times higher than for the 10^{-6} rate estimated for allozyme

markers (Voelker et al. 1980) (Jin et al. 1996). In addition, mutation rates vary among alleles at a microsatellite locus (Jin et al. 1996), increasing the variance within loci and may contribute to the heterogeneity in F_{st} values observed at individual loci.

In addition to differentiation, mutation can also have a homogenizing effect, analogous to migration. For example, if mutation rates are high enough, the chance of the same mutation occurring independently in two different populations increases (homoplasy). An increase in homoplasy would decrease the resolving power of microsatellites to detect differences among populations, decreasing the variance among populations.

Spatial Structure:

Dispersal is distance dependent with individuals proximal to one another being more closely related through local inbreeding (Wright 1943), if they are not removed by inbreeding depression. Wright (1943; 1946) describes these effects of isolation by distance in causing genetic differences between subpopulations of a continuous population. We expected spatial structure for the adults to be weak in both populations because white pine pollen and seed are highly dispersed by wind. Although seed dispersal can be much shorter than pollen dispersal,

extensive pollen flow or gene migration would disrupt clustering of related trees from low seed dispersal, preventing neighborhood structure from developing at short distances. In support of this premise of weak spatial structure, Moran's I analysis indicated a randomly distributed second growth population and weak positive spatial structure at short distances for the OG population. The results suggest that logging decreased spatial structure at the second growth stand.

The spatial genetic structure for OG at short distances is consistent with Moran's I values reported for other species with similar wind dispersed pollen and seed, i.e. 0.07 for *Maclura pomifera*, 0.04 for *Gleditsia trianthos* (Schnabel et al. 1991), and 0.05 for *Quercus laevis* (Berg & Hamrick 1995). The overall mean I value for distance class one was 0.02, compared to an isozyme value of 0.06 for this same population (Epperson & Chung 2001). The lower microsatellite Moran's I values could be a function of low frequency alleles, which may decrease measures of autocorrelation (Epperson, personal communication). It may also be that the high mutation rates at SSR loci directly decrease spatial structure (Epperson 1990).

Wright's (1946) neighborhood size, or the number of mating individuals (N) drawn at random from within a circle of area $4\pi\sigma^2$ and radius 2σ , can be used to estimate total dispersal distance. Based on Epperson et al (1999) simulations the Moran's I value of 0.02 corresponds to a Wright's neighborhood size of approximately 300 trees for the OG population (Line 1, Table 4). Standardized for density, the combined seed and pollen dispersal distance (σ) was estimated using the neighborhood formula for a monocious population, $N = 4\pi\sigma^2d$ (Wright 1952). With a stand density (d) of 100 trees ha^{-1} , the standard deviation σ would be approximately 47 m and the average variance of the parent-offspring distance (σ^2) approximately 2210 m^2 . Seed dispersal can be shorter than pollen dispersal. To allow different offspring dispersal distance estimates for the male and female parents, Crawford (1984) proposed the combined female parent-offspring (σ^2_s) and the male parent-offspring (σ^2_p) dispersal variance to be $\sigma_T^2 = 1/2\sigma^2_p + \sigma^2_s$. Typical empirical measures of dispersal distances for pine pollen range from 17 meters (Wright 1952) to 69 meters (Wang et al. 1960), and for seed from 15 to 30 meters (Epperson & Allard 1989). Given the total parent variance of 2210 m^2 and the average pollen standard deviation of 43 m, the seed dispersal distance would be 27 m, which agrees

well with the average empirical seed dispersal distance of 23 m.

Anticipated levels of neighborhood differentiation can also be estimated from neighborhood size. At the OG site there is little differentiation with N approximately 300 trees. For the SS site, the mean $M(I) = 0.001$, which corresponds to randomness or no structure. With a mean F_{st} of 0.0054 indicating little genetic difference between populations, these microsatellite estimates of N and F_{st} agree reasonably well with Wright's (1943) predictions. Wright concludes that greater differentiation occurs when $N=10$, moderate differentiation at $N=100$, and approaching panmixia at $N=1000$. The allozyme study of these same two populations approximates a smaller neighborhood size of 100 trees for OG adults (Epperson & Chung 2001), with little genetic difference between populations (F_{st} of 0.008).

CONCLUSION

Mutational mechanisms responsible for generating microsatellite polymorphisms are important factors to consider in conducting genetic evaluations and population studies. The complex mutational processes do not negate the usefulness of these markers, but it is important to understand SSR mutation and how it affects the statistics used to describe population genetic and spatial structure.

Autocorrelation fits predicted levels for white pine, based on pollen and seed dispersal and density. The results for the microsatellites are slightly lower than for the allozymes, but are not inconsistent. In addition to statistical error, SSRs may directly affect spatial statistics through mutation and allele frequency.

Gene flow was high and population divergence negligible with the levels of genetic diversity similar between stands. Spatial distance between populations or timber harvest at the SS, are reasonable explanations for the observed differences in allele frequencies between populations. The weak positive structure at short distances observed for the OG site suggests spatial structure fits the isolation by distance model for predicted genetic differences in a continuous population.

Whereas, the results suggest that logging of the second growth stand removed spatial genetic structure.

APPENDIX

APPENDIX

Spatial Autocorrelation Coefficients

Table 13: Spatial autocorrelation coefficients (Moran's I) for 8 loci in the Old Growth population (OG) of *Pinus strobus* for 10 distance classes, $n = 122^1$.

DistBound ²	15	25	35	45	55	65
DistClass ³	1	2	3	4	5	6
<u>Rps1b</u> 2n ⁴ = 244						
No. Pairs ⁵	421	649	843	886	883	876
Locus1-03	-0.03	-0.04	-0.02	0.00	-0.04	0.03
Locus1-04	-0.02	-0.03	-0.02	0.01	-0.04	0.03
Locus1-05	-0.01	-0.02	-0.01	-0.00	-0.01	-0.01
Locus1-06	0.06	-0.04	-0.06	-0.03	0.04	0.01
Locus1-08	0.04	0.02	-0.10**	-0.04	0.06**	0.01
Average	0.01	-0.02	-0.04	-0.01	0.00	0.01
<u>Rps2</u> 2n = 244						
No. Pairs	421	649	843	886	883	876
Locus2-01	0.04	-0.02	0.02	-0.05	-0.02	-0.01
Locus2-02	0.00	-0.01	-0.01	-0.01	-0.00	-0.01
Locus2-03	-0.00	-0.02	-0.00	-0.00	-0.01	-0.01
Locus2-04	-0.01	-0.00	0.02	-0.01	-0.01	0.00
Locus2-06	0.06	0.01	-0.02	-0.06*	-0.10**	0.02
Locus2-07	0.04	-0.02	0.02	-0.06*	-0.09**	0.06*
Locus2-08	-0.01	-0.00	-0.01	-0.01	-0.00	-0.01
Locus2-09	-0.01	-0.00	-0.01	-0.01	-0.01	-0.01
Locus2-10	-0.03	-0.04	0.01	-0.01	-0.01	0.00
Average	0.01	-0.01	0.00	-0.02	-0.03	0.00

¹ Sample size of study population

² Distance bound (meters)

³ Distance class

⁴ Haploid sample size for each locus

⁵ Number of joins

* $p < 0.05$

** $p < 0.01$

Table 13 continued

DistBound	75	85	95	105		
DistClass	7	8	9	10		
<u>Rps1b</u>						
No. Pairs	684	627	428	1084	p ⁶	q ⁷
Locus1-03	0.04*	-0.01	-0.03	-0.01	0.498	0.0123
Locus1-04	0.02	-0.02	-0.03	0.00	0.812	0.0123
Locus1-05	-0.01	-0.01	-0.01	-0.00	1.000	0.0082
Locus1-06	0.05*	-0.02	-0.07	-0.02	0.498	0.9180
Locus1-08	0.04	-0.06	-0.06	-0.01	0.023	0.0492
Average	0.03	-0.02	-0.04	-0.01		
<u>Rps2</u>						
No. Pairs	684	627	428	1084	p	q
Locus2-01	-0.02	-0.00	0.06	-0.02	0.638	0.0451
Locus2-02	-0.01	-0.00	-0.01	-0.01	0.584	0.0041
Locus2-03	-0.02	-0.01	-0.00	-0.00	0.650	0.0041
Locus2-04	-0.00	-0.06	0.03	-0.03	0.689	0.0410
Locus2-06	0.03	0.07*	0.11**	-0.06*	0.023	0.5656
Locus2-07	-0.03	0.08*	0.04	-0.04	0.047	0.3156
Locus2-08	-0.01	-0.02	-0.02	-0.01	1.000	0.0041
Locus2-09	-0.01	-0.02	-0.01	-0.01	1.000	0.0041
Locus2-10	-0.00	0.02	-0.04	-0.01	1.000	0.0164
Average	-0.01	0.01	0.02	-0.02		

⁶ Overall correlogram significance (Bonferroni approx.)

⁷ Allele frequency

Table 13 continued

DistBound	15	25	35	45	55	65
DistClass	1	2	3	4	5	6
<u>Rps6</u> 2n = 244						
No. Pairs	421	649	843	886	883	876
Locus6-01	-0.01	-0.00	-0.01	-0.01	-0.01	-0.01
Locus6-02	-0.05	0.00	0.01	-0.05	0.01	0.01
Locus6-03	0.02	-0.03	0.03	0.01	-0.01	0.01
Locus6-04	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Locus6-05	0.06	0.07*	-0.06	-0.03	-0.01	0.00
Locus6-06	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Locus6-07	-0.02	-0.05	0.00	0.01	-0.06*	0.01
Locus6-09	0.00*	-0.00	-0.00	-0.00	-0.01	-0.00
Average	-0.00	-0.00	-0.01	-0.01	-0.01	-0.00
<u>Rps34b</u> 2n = 244						
No. Pairs	421	649	843	886	883	876
Locus34-01	-0.01	-0.00	-0.01	-0.01	-0.01	-0.01
Locus34-02	-0.06	-0.01	0.01	-0.05	0.01	0.01
Locus34-03	-0.05	-0.02	0.07**	-0.04	-0.00	0.02
Locus34-04	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Locus34-05	0.06	0.07*	-0.06	-0.03	-0.01	0.00
Average	-0.01	0.01	0.00	-0.03	-0.00	0.00
<u>Rps39</u> 2n = 240						
No. Pairs	412	630	821	865	868	848
Locus39-02	0.08*	-0.00	0.01	-0.04	-0.05	-0.05
Locus39-03	0.10**	-0.02	-0.01	-0.02	-0.01	-0.05
Locus39-04	0.02	-0.01	-0.00	-0.01	-0.03	-0.01
Locus39-05	-0.00	-0.01	-0.00	-0.01	-0.03	-0.07*
Average	0.04	-0.01	-0.00	-0.02	-0.03	-0.05

Table 13 continued

DistBound	75	85	95	105		
DistClass	7	8	9	10		
<u>Rps6</u>						
No. Pairs	684	627	428	1084	p	q
Locus6-01	-0.01	-0.01	-0.00	-0.01	1.000	0.0041
Locus6-02	0.00	-0.05	0.04	-0.01	0.715	0.1967
Locus6-03	0.00	-0.08*	-0.00	-0.03	0.267	0.6844
Locus6-04	-0.01	-0.00	0.00	-0.01	1.000	0.0041
Locus6-05	0.02	-0.06	-0.04	-0.01	0.119	0.0656
Locus6-06	-0.01	-0.00	0.00	-0.01	1.000	0.0041
Locus6-07	0.08**	-0.04	0.01	-0.00	0.074	0.0369
Locus6-09	-0.02	-0.02*	-0.03**	-0.01	0.083	0.0041
Average	0.01	-0.03	-0.00	-0.01		
<u>Rps34b</u>						
No. Pairs	684	627	428	1084	p	q
Locus34-01	-0.01	-0.01	-0.00	-0.01	1.000	0.0041
Locus34-02	-0.01	-0.04	0.04	-0.00	0.689	0.2008
Locus34-03	-0.01	-0.06	0.02	-0.02	0.076	0.7254
Locus34-04	-0.01	-0.00	0.00	-0.01	1.000	0.0041
Locus34-05	0.02	-0.06	-0.04	-0.01	0.119	0.0656
Average	-0.01	-0.04	0.01	-0.01		
<u>Rps39</u>						
No. Pairs	660	598	409	1029	p	q
Locus39-02	0.01	0.04	0.05	-0.03	0.253	0.6167
Locus39-03	-0.04	0.05	0.03	-0.02	0.094	0.3375
Locus39-04	-0.03	-0.04	0.04	0.02	0.698	0.0125
Locus39-05	-0.00	-0.01	0.13**	-0.01	0.006	0.0333
Average	-0.02	0.01	0.06	-0.01		

Table 13 continued

DistBound	15	25	35	45	55	65
DistClass	1	2	3	4	5	6
<u>Rps50</u> 2n = 244						
No. Pairs	421	649	843	886	883	876
Locus50-01	0.01	-0.03	-0.01	-0.02	-0.03	0.01
Locus50-02	-0.02	-0.01	0.00	-0.00	-0.00	-0.01
Locus50-03	0.01**	0.01**	0.01*	0.00*	-0.00	0.00
Locus50-05	-0.03	-0.02	-0.00	0.00	-0.02	-0.05
Locus50-06	0.02	0.05	-0.03	-0.06*	-0.03	0.02
Locus50-07	-0.05	-0.00	0.01	0.01	-0.01	-0.02
Locus50-08	0.04	0.03	-0.09**	-0.03	0.04	-0.08*
Locus50-09	-0.03	-0.05	0.01	0.04*	-0.04	-0.03
Locus50-10	-0.03	0.03	-0.10**	-0.02	0.02	0.05*
Locus50-11	-0.02	-0.02	-0.02	-0.03	-0.02	0.06**
Locus50-12	-0.02	-0.03	0.02	-0.03	0.00	0.02
Locus50-13	-0.05	-0.01	-0.03	-0.03	0.02	-0.03
Locus50-14	-0.02	-0.02	-0.03	0.02	0.02	-0.04
Average	-0.02	-0.01	-0.02	-0.01	-0.00	-0.01
<u>Rps84</u> 2n = 244						
No. Pairs	421	649	843	886	883	876
Locus84-01	0.00	-0.01	-0.00	-0.00	-0.01	-0.00
Locus84-02	0.01	-0.03	0.03	-0.04	-0.02	-0.04
Locus84-03	-0.03	-0.02	-0.02	-0.02	0.05**	-0.02
Locus84-04	-0.03	-0.02	0.03	-0.07*	-0.01	0.01
Locus84-05	-0.03	-0.00	-0.02	0.02	-0.04	0.00
Locus84-06	-0.01	-0.03	-0.03	-0.04	0.04*	-0.00
Average	-0.01	-0.02	-0.00	-0.03	0.00	-0.01
<u>Rps127</u> 2n = 244						
No. Pairs	421	649	843	886	883	876
Locus127-01	0.03	-0.05	-0.01	0.04	-0.01	-0.02
Locus127-02	0.03	-0.05	-0.01	0.04	-0.01	-0.02
Average	0.03	-0.05	-0.01	0.04	-0.01	-0.02

Table 13 continued

DistBound	75	85	95	105		
DistClass	7	8	9	10		
<u>Rps50</u>						
No. Pairs	684	627	428	1084	p	q
Locus50-01	-0.04	-0.00	0.09*	-0.02	0.107	0.0205
Locus50-02	-0.01	0.05**	-0.03	-0.05*	0.040	0.0123
Locus50-03	-0.01	-0.01	-0.02*	-0.05**	0.028	0.0041
Locus50-05	0.01	-0.00	0.05	-0.01	0.936	0.1721
Locus50-06	0.00	-0.06	-0.02	0.02	0.360	0.2746
Locus50-07	-0.02	-0.00	0.02	-0.01	1.000	0.1352
Locus50-08	0.06*	0.05	0.06	-0.04	0.029	0.1148
Locus50-09	-0.02	-0.02	0.01	0.01	0.215	0.0123
Locus50-10	-0.02	-0.03	-0.02	0.01	0.028	0.1967
Locus50-11	-0.02	-0.01	-0.00	0.00	0.024	0.0082
Locus50-12	-0.04	0.01	-0.03	-0.01	1.000	0.0164
Locus50-13	0.06*	-0.04	-0.03	0.02	0.149	0.0205
Locus50-14	-0.04	-0.04	0.07*	0.00	0.185	0.0123
Average	-0.01	-0.01	0.01	-0.01		
<u>Rps84</u>						
No. Pairs	684	627	428	1084	p	q
Locus84-01	-0.01	-0.02	-0.01	-0.01	0.584	0.0041
Locus84-02	0.01	-0.04	0.08*	0.00	0.235	0.7664
Locus84-03	-0.00	-0.01	-0.01	-0.02	0.045	0.0082
Locus84-04	-0.00	0.00	0.05	-0.03	0.239	0.1148
Locus84-05	-0.02	-0.01	-0.05	0.02	0.800	0.0328
Locus84-06	-0.01	-0.05	-0.01	0.03	0.481	0.0738
Average	-0.01	-0.02	0.01	-0.00		
<u>Rps127</u>						
No. Pairs	684	627	428	1084	p	q
Locus127-01	0.02	-0.06	-0.05	0.00	0.639	0.7172
Locus127-02	0.02	-0.06	-0.05	0.00	0.639	0.2828
Average	0.02	-0.06	-0.05	0.00		

Table 14: Spatial autocorrelation coefficients (Moran's I) for 8 loci in the Second Site population (SS) of *Pinus strobus* for 10 distance classes, n = 120¹.

DistBound ²	15	25	35	45	55	65
DistClass ³	1	2	3	4	5	6
<u>Rps1b</u> 2n ⁴ = 238						
No. Pairs ⁵	519	743	924	910	796	810
Locus1-01	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01
Locus1-02	-0.01	-0.02	-0.02	-0.01	-0.01	-0.00
Locus1-03	-0.01	-0.02	-0.02	-0.02	-0.01	-0.01
Locus1-04	0.01	-0.02	-0.02	-0.03	0.01	-0.01
Locus1-05	-0.04	-0.02	-0.03	0.05**	-0.02	-0.02
Locus1-06	-0.04	-0.01	-0.03	-0.02	-0.03	-0.03
Locus1-07	-0.03	-0.03	0.04*	-0.02	-0.01	-0.02
Locus1-08	0.01	0.01	-0.02	-0.03	0.01	-0.00
Average	-0.02	-0.01	-0.01	-0.01	-0.00	-0.01
<u>Rps2</u> 2n = 240						
No. Pairs	524	753	933	921	804	822
Locus2-01	0.13**	-0.04	-0.07*	-0.09**	0.01	
	0.08**					
Locus2-03	0.00	0.00	0.01*	0.00	0.00*	-0.00
Locus2-04	-0.06	0.01	-0.02	-0.03	0.03	-0.03
Locus2-05	-0.01	-0.02	0.00	-0.00	-0.01	-0.01
Locus2-06	0.01	-0.02	-0.00	-0.02	0.01	0.01
Locus2-07	-0.03	0.06*	-0.05	0.02	-0.01	0.03
Locus2-08	-0.02	-0.01	-0.02	-0.02	-0.01	-0.01
Locus2-09	0.00	-0.00	-0.00	-0.01	-0.02*	-0.02*
Locus2-10	-0.00	-0.00	-0.02	-0.03	0.03*	-0.02
Average	0.00	-0.00	-0.02	-0.02	0.00	0.00

- ¹ Sample size of study population
² Distance bound (meters)
³ Distance class
⁴ Haploid sample size for each locus
⁵ Number of joins
* p < 0.05
** p < 0.01

Table 14 continued

DistBound	75	85	95	105		
DistClass	7	8	9	10		
<u>Rps1b</u>						
No. Pairs	768	663	458	430	p ⁶	q ⁷
Locus1-01	-0.02	-0.03	0.09**	-0.02	0.013	0.0084
Locus1-02	-0.01	0.00	0.00	-0.00	0.743	0.0042
Locus1-03	0.06**	-0.02	-0.02	-0.03	0.016	0.0084
Locus1-04	-0.01	-0.01	-0.02	0.05	0.709	0.0546
Locus1-05	-0.02	-0.01	0.01	0.01	0.067	0.0084
Locus1-06	-0.01	0.02	-0.02	0.02	1.000	0.8697
Locus1-07	-0.01	-0.02	0.00	0.01	0.114	0.0084
Locus1-08	0.03	-0.04	0.03	-0.04	1.000	0.0378
Average	0.00	-0.01	0.00	-0.00		
<u>Rps2</u>						
No. Pairs	780	678	472	453	p	q
Locus2-01	-0.01	-0.03	0.01	0.02	0.003	0.0500
Locus2-03	-0.00	-0.03**	-0.03*	-0.07**	0.009	0.0042
Locus2-04	0.01	-0.00	0.02	-0.01	0.890	0.0500
Locus2-05	-0.01	-0.01	-0.01	0.00	0.523	0.0042
Locus2-06	0.00	-0.04	-0.02	-0.02	1.000	0.6917
Locus2-07	-0.03	-0.11**	0.02	0.00	0.013	0.1833
Locus2-08	-0.00	0.00	0.01	0.01	0.641	0.0042
Locus2-09	-0.01	-0.01	-0.00	0.00	0.106	0.0042
Locus2-10	-0.01	-0.00	-0.02	0.00	0.362	0.0083
Average	-0.01	-0.03	-0.00	-0.01		

⁶ Overall correlogram significance (Bonferroni approx.)

⁷ Allele frequency

Table 14 continued

DistBound ²	15	25	35	45	55	65
DistClass ³	1	2	3	4	5	6
<u>Rps6</u> 2n = 240						
No. Pairs	524	753	933	921	804	822
Locus6-01	-0.01	-0.00	-0.02	-0.01	-0.01	-0.00
Locus6-02	-0.10*	0.03	-0.01	-0.01	-0.02	0.01
Locus6-03	-0.05	-0.01	0.00	0.02	-0.02	-0.02
Locus6-04	0.03	0.00	0.00	-0.03	0.02	-0.06*
Locus6-05	0.11**	0.11**	0.04*	-0.01	-0.03	-0.00
Locus6-06	0.00	0.00	0.01*	0.00	0.00*	-0.00
Locus6-07	0.08*	-0.05	0.07**	-0.06*	-0.01	-0.00
Locus6-08	-0.02	-0.03	-0.03	-0.02	0.04*	-0.01
Locus6-09	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01
Average	0.00	0.01	0.01	-0.01	-0.00	-0.01
<u>Rps34b</u> 2n = 238						
No. Pairs	702	702	702	702	702	702
Locus34-01	-0.01	-0.00	-0.02	-0.02	-0.01	-0.00
Locus34-02	-0.04	0.01	0.02	0.04	-0.04	-0.04
Locus34-03	-0.06	-0.01	0.02	0.00	-0.01	-0.06
Locus34-04	0.01	0.05*	-0.03	-0.03	0.01	-0.02
Locus34-05	0.06*	0.13**	0.04	-0.01	0.00	-0.08*
Average	-0.01	0.03	0.01	-0.00	-0.01	-0.04
<u>Rps39</u> 2n = 240						
No. Pairs	524	753	933	921	804	822
Locus39-01	-0.02	-0.01	-0.02	-0.01	-0.01	-0.00
Locus39-02	0.06	-0.04	-0.01	-0.09**	0.03	-0.02
Locus39-03	0.02	-0.02	-0.04	-0.05	0.09**	-0.04
Locus39-04	-0.01	-0.03	0.01	-0.05	-0.00	0.02
Locus39-05	-0.09*	0.00	0.02	0.01	-0.05	0.01
Average	-0.01	-0.02	-0.01	-0.04	0.01	-0.01

Table 14 continued

DistBound	75	85	95	105		
DistClass	7	8	9	10		
<u>Rps6</u>						
No. Pairs	780	678	472	453	p	q
Locus6-01	-0.01	-0.01	-0.00	-0.00	1.000	0.0042
Locus6-02	0.01	-0.01	-0.02	0.00	0.127	0.2083
Locus6-03	-0.00	0.02	-0.04	-0.02	1.000	0.6542
Locus6-04	-0.01	-0.03	-0.00	0.01	0.441	0.0167
Locus6-05	-0.14**	-0.05	-0.09*	-0.03	0.000	0.0417
Locus6-06	-0.00	-0.03**	-0.03*	-0.07**	0.009	0.0042
Locus6-07	-0.04	-0.04	-0.02	0.00	0.045	0.0583
Locus6-08	-0.01	-0.01	-0.00	0.01	0.151	0.0083
Locus6-09	-0.01	-0.01	-0.02	-0.05*	0.250	0.0042
Average	-0.02	-0.02	-0.03	-0.02		
<u>Rps34b</u>						
No. Pairs	702	702	702	703	p	q
Locus34-01	-0.01	-0.00	-0.01	-0.01	1.000	0.0042
Locus34-02	0.01	-0.02	-0.00	-0.02	0.887	0.2059
Locus34-03	0.05	0.02	-0.03	-0.00	0.534	0.7227
Locus34-04	-0.05	-0.00	-0.01	0.01	0.412	0.0168
Locus34-05	-0.05	-0.10**	-0.08*	-0.01	0.000	0.0504
Average	-0.01	-0.02	-0.03	-0.01		
<u>Rps39</u>						
No. Pairs	780	678	472	453	p	q
Locus39-01	-0.01	-0.01	-0.00	-0.00	1.000	0.0042
Locus39-02	0.05*	0.01	-0.07	0.03	0.050	0.6750
Locus39-03	-0.02	-0.01	-0.02	0.05	0.017	0.2625
Locus39-04	0.01	0.01	-0.04	-0.02	0.726	0.0292
Locus39-05	0.03	-0.01	-0.03	-0.02	0.184	0.0292
Average	0.01	-0.00	-0.03	0.01		

Table 14 continued

DistBound	15	25	35	45	55	65
DistClass	1	2	3	4	5	6
<u>Rps50</u> 2n = 240						
No. Pairs	524	753	933	921	804	822
Locus50-01	0.00	-0.01	-0.02	-0.03	-0.03	0.03
Locus50-04	-0.03	-0.02	-0.03	0.03	0.03	-0.03
Locus50-05	-0.01	-0.09*	0.01	0.04	0.00	-0.06
Locus50-06	0.02	-0.00	-0.03	0.02	0.01	-0.10*
Locus50-07	-0.02	0.09**	-0.02	0.02	-0.03	-0.04
Locus50-08	-0.02	-0.01	0.02	0.05*	-0.01	0.01
Locus50-09	-0.02	-0.02	0.01	-0.01	-0.01	-0.02
Locus50-10	-0.00	-0.05	0.01	0.02	-0.02	-0.05
Locus50-11	0.00	0.02	0.02	-0.01	0.04	-0.05
Locus50-12	0.01	-0.06	-0.06*	0.03	0.04*	-0.02
Locus50-13	0.04	-0.00	-0.01	-0.03	0.03	0.00
Locus50-14	-0.03	-0.01	0.03	-0.06*	0.01	-0.02
Average	-0.00	-0.01	-0.01	0.01	0.00	-0.03
<u>Rps84</u> 2n = 240						
No. Pairs	524	753	933	921	804	822
Locus84-01	0.01*	0.00	0.00*	0.01*	0.00*	0.00*
Locus84-02	0.00	0.06*	0.00	-0.05	-0.03	-0.03
Locus84-03	-0.01	-0.01	-0.01	-0.01	-0.02	0.04*
Locus84-04	0.01	0.09**	-0.05	-0.01	-0.03	0.01
Locus84-05	-0.03	0.02	-0.00	-0.04	-0.03	-0.00
Locus84-06	-0.03	-0.04	0.02	0.00	-0.03	0.00
Average	-0.01	0.02	-0.01	-0.02	-0.02	0.00
<u>Rps127</u> 2n = 232						
No. Pairs	496	720	897	887	757	753
Locus127-01	0.03	0.05*	0.05*	-0.04	0.03	-0.03
Locus127-02	0.03	0.05*	0.05*	-0.04	0.03	-0.03
Average	0.03	0.05*	0.05*	-0.04	0.03	-0.03

Table 14 continued

DistBound	75	85	95	105		
DistClass	7	8	9	10		
<u>Rps50</u>						
No. Pairs	780	678	472	453	p	q
Locus50-01	-0.03	-0.02	0.05	0.03	0.627	0.0125
Locus50-04	-0.03	-0.03	0.06*	-0.02	0.256	0.0125
Locus50-05	-0.03	0.05	0.05	-0.04	0.107	0.2292
Locus50-06	0.02	-0.02	0.03	-0.01	0.031	0.2750
Locus50-07	-0.04	0.02	-0.05	-0.03	0.021	0.1625
Locus50-08	-0.03	-0.05	-0.06	-0.04	0.276	0.0542
Locus50-09	-0.03	0.04	-0.01	-0.01	0.607	0.0167
Locus50-10	0.02	-0.02	0.03	-0.04	0.852	0.1458
Locus50-11	-0.02	-0.03	-0.02	-0.07	0.538	0.0208
Locus50-12	-0.02	0.02	-0.01	-0.00	0.257	0.0208
Locus50-13	-0.03	-0.02	-0.03	-0.04	0.628	0.0250
Locus50-14	-0.01	0.01	0.01	-0.00	0.378	0.0250
Average	-0.02	-0.01	0.00	-0.02		
<u>Rps84</u>						
No. Pairs	780	678	472	453	p	q
Locus84-01	-0.01	-0.02	-0.05**	-0.09**	0.001	0.0042
Locus84-02	-0.02	-0.02	0.03	0.00	0.299	0.6958
Locus84-03	-0.02	-0.03	-0.02	-0.01	0.195	0.0083
Locus84-04	-0.06	-0.00	-0.08*	0.05	0.022	0.1917
Locus84-05	-0.04	0.02	0.03	0.02	1.000	0.0417
Locus84-06	-0.00	0.00	-0.01	-0.01	1.000	0.0583
Average	-0.02	-0.01	-0.02	-0.01		
<u>Rps127</u>						
No. Pairs	719	608	426	407	p	q
Locus127-01	-0.06	-0.05	-0.14**	0.03	0.020	0.7457
Locus127-02	-0.06	-0.05	-0.14**	0.03	0.020	0.2543
Average	-0.06	-0.05	-0.14**	0.03	0.020	

LITERATURE CITED

LITERATURE CITED

- National data climate center (NCDC). Grayling weather station. June 2000. <<http://www.ncdc.noaa.gov>>.
- Adams, W. T. Gene dispersal within forest tree populations. *New Forests*. 1992; 6:217-240.
- Amos, Bill and Hoelzel, A. Rus. Applications of molecular genetic techniques to the conservation of small populations. *Biological Conservation*. 1992; 61:133-144.
- Beaulieu, J. and Simon, J.-P. Genetic structure and variability in *Pinus strobus* in Quebec. *Can. J. For Res.* 1994a; 24:1726-1733.
- . Inheritance and linkage relationships of allozymes in *Pinus strobus* L. *Silvae Genetica*. 1994b; 43(4):253-261.
- . Mating system in natural populations of eastern white pine in Quebec. *Can. J. For Res.* 1995; 25:1697-1703.
- Beaulieu, Jean and Simon, Jean-Pierre. Variation in cone morphology and seed characters in *Pinus strobus* in Quebec. *Can. J. Bot.* 1995; 73:262-271.
- Becher, S. A. and Griffiths, R. Genetic differentiation among local populations of the European hedgehog (*Erinaceus*) in mosaic habitats. *Molecular Ecology*. 1998; 7:1599-1604.
- Berg, Edward E. and Hamrick, James L. Fine-scale genetic structure of a turkey oak forest. *Evolution*. 1995; 49(1):110-120.

- Blanquer-Maumont, Anne and Crouau-Roy, Brigitte.
Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. *J Mol Evol.* 1995; 41:492-497.
- Blouin, M S; Parsons, M; Lacaille, V, and Lotz, S. Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* 1996; 5, 393-401.
- Botstein. David; White, Raymond L.; Skolnick, Mark, and Davis, Ronald W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Human. Genet.* 1980; 32:314-331.
- Brookfield, J. F. Y. A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology.* 1996; 5:453-455.
- Brym, Petr and Eckert, Robert T. Within-stand clustering of eastern white pine genotypes. in. *Proceedings of the 30th NE forest tree improvement conference.* Orono, ME: School of Forestry Resources, U of ME; 1986 Jul 22-1986 Jul 24; pp. 126-138.
- Buchert, George P.; Rajora, Om P.; Hood, James V., and Dancik, Bruce P. Effects of harvesting on genetic diversity in old-growth eastern white pine in Ontario, Canada. *Conservation Biology.* 1997; 11(3):747-758.
- Bunville, Julie; Krakowski, Karen, and Seto, Donald. Cerium oxide removes background fluorescent contaminants, "green smears", from glass plates used for automated fluorescence-based DNA sequencing. *BioTechniques.* 1997; 22(2):236-238.
- Callen, David F.; Thompson, Andrew D.; Shen, Yang; Phillips, Hilary A.; Richards, Robert I.; Mulley, John C., and Sutherland, Grant R. Incidence and origin of "null" alleles in the (AC)_n microsatellite markers. *Am. J. Human. Genet.* 1993; 52:922-927.

Chagala, E. M. Inheritance and linkage of allozymes in *Pinus strobus* L. *Silvae Genetica*. 1996; 45(4):181-187.

Chung, M. G., and Epperson, B. Spatial genetic structure of clonal and sexual reproduction in populations of *Andenophora grandiflora* (Campanulaceae). *Evolution*. 1999; 53(4):1068-1078.

Clark, James M. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research*. 1988; 16(20):9676-9687.

Cliff, A. D. and Ord, J. F. Spatial process-methods and applications. Pion, London; 1981.

Cornelius, Jonathan. Heritabilities and additive genetic coefficients of variation in forest trees. *Can. J. For. Res.* 1994; 24:372-379.

Crawford, T. J. The estimation of neighbourhood parameters for plant populations. *Heredity*. 1984; 52(2):273-283.

Dallimore, W. and Jackson, A. Bruce. A handbook of coniferae including ginkgoaceae. London: Edward Arnold & Co.; 1948.

Davis, Margaret B. and Shaw, Ruth G. Range shifts and adaptive responses to quaternary climate change. *Science*. 2001; 292:673-679.

Dewey, Sarah E. and Heywood, John S. Spatial genetic structure in a population of *Psychotria Nervosa*. I. Distribution of genotypes. *Evolution*. 1988; 42(4):834-838.

Don, R. H.; Cox, P. T.; Wainwright, B. J.; Baker, K., and Mattick, J. S. "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*. 1991; 19(14):4008.

- Dong, Jinsheng and Wagner, David B. Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and test of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics*. 1994; 136:1187-1194.
- Echt, C. S.; May-Marquardt, P.; Hseih, M., and Zahorchak, R. Characterization of microsatellite markers in eastern white pine. *Genome*. 1996; 39(6):1102-1108.
- Echt, C. S. and Nelson, C. D. Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). *Theor Appl Genet*. 1997; 94:1031-1037.
- Echt, C. S.; Vendramin, G. G.; Nelson, C. D., and Marquardt, P. Microsatellite DNA as shared genetic markers among conifer species. *Can. J. For Res*. 1999; 29:365-371.
- Eckert, R. T.; Joly, R. J., and Neale, D. B. Genetics of isozyme variants and linkage relationships among allozyme loci in 35 eastern white pine clones. *Can. J. For Res*. 1981; 11:573-579.
- Eckert, Robert T. and Ryu, Jang Bal. Foliar isozyme variation in twenty seven provenances of *Pinus strobus* L.: relationship of three foliar isozyme alleles with height growth and multivariate patterns of range-wide variation. in. *Proceedings of the seventh North America biotechnology workshop on physiology and genetics of intensive culture*. Lexington, Kentucky: University of Kentucky; 1982 Jul 26-1982 Jul 28: 179-187.
- El-Kassaby, Y. A.; Meagher, M. D., and Davidson, R. Temporal variation in the outcrossing rate in a natural stand of western white pine. *Silvae Genetica*. 1993; 42(2-3):131-135.

- Epperson, B. K. Spatial patterns of genetic variation within plant populations. in: Brown, Anthony H. D.; Clegg, Michael T.; Kahler, Alex L., and Weir, Bruce S., editors. Plant population genetics, breeding, and genetic resources. Sunderland, Massachusetts: Sinauer Associates; 1989; pp. 229-253.
- . Spatial autocorrelation of genotypes under directional selection. *Genetics*. 1990; 124:757-771.
- Epperson, B. K. and Allard, R. W. Allozyme analysis of the mating system in lodgepole pine populations. *The Journal of Heredity*. 1984; 75:212-214.
- . Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics* . 1989; 121:369-377.
- Epperson, B. K.; Huang, Z., and Li, T.-Q. Measures of spatial structure in samples of genotypes for multiallelic loci. *Genet. Res. Camb.* 1999; 72:251-261.
- Epperson, Bryan (epperson@msu.edu). Low frequency alleles may reduce autocorrelation. 2001.
- Epperson, Bryan K. and Chung, Myong Gi. Spatial genetic structure of allozyme polymorphisms within populations of *Pinus strobus* (Pinaceae). *American Journal of Botany*. 2001.
- Epperson, Bryan K. and Li, TianQuan. Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *Proc. Natl. Acad. Sci. USA*. 1996; 93:10528-10532.
- Erpelding, John (mayje@ars-grin.gov). Sucrose improves amplification of weak PCR reactions. 2001.

- Estoup, Arnaud; Tailiez, Christian; Cornuet, Jean-Marie, and Solignac, Michel. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol. Biol. Evol.* 1995; 12(6):1074-1084.
- Etterson, Julie R. and Shaw, Ruth G. Constraint to adaptive evolution in response to global warming. *Science*. 2001; 294:151-154.
- Farrar, John Laird. Trees of the Northern United States and Canada. Ames, Iowa: Iowa State University Press; 1995.
- Fields, R. L. and Scribner, K. T. Isolation and characterization of novel waterfowl microsatellite loci: cross-species comparisons and research applications. *Molecular Ecology*. 1997; 6:199-202.
- Fowler, D. P. Effects of inbreeding in red pine, *Pinus resinosa* Ait. IV. Comparison with other Northeastern *Pinus* species. *Silvae Genetica*. 1965; 14:76-81.
- Fowler, D. P. and Heimbürger, C. Geographic variation in eastern white pine, 7-year results in Ontario. *Silvae Genetica*. 1969; 18(4):97-144.
- Garza, John Carlos; Slatkin, Montgomery, and Freimer, Nelson B. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* 1995; 12(4):594-603.
- Goldstein, David B.; Linares, Andres Ruiz; Cavalli-Sforza, Luigi Luca., and Feldman, Marcus W. An evaluation of genetic distance for use with microsatellite loci. *Genetics*. 1995; 139:463-471.
- Gonick, Larry and Smith, Woollcott. The cartoon guide to statistics. New York: HarperCollins Publishers, Inc.; 1993.

Goodman, S. J. Rst Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology*. 1997; 6:881-885.

Grimaldi, Marie-Claude and Crouau-Roy, Brigitte.
Microsatellite allelic homoplasy due to variable flanking sequences. *J Mol Evol*. 1997; 44:336-340.

Gullberg, Annica; Tegelström, Håkan, and Olsson, Mats.
Microsatellites in the sand lizard (*Lacerta agilis*): description, variation, inheritance, and applicability. *Biochemical Genetics*. 1997; 35(7/8):281-295.

Guries, Raymond P. and Ledig, F. Thomas. Genetic structure of populations and differentiation in forest trees. in. *Proceedings of the symposium on isozymes of North American forest trees and forest insects*. General technical report PSW-48. Berkeley, CA: US Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station; 1979 Jul 27; pp. 42-47.

---. Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). *Evolution*. 1982; 36(2):387-402.

Hamrick, J. L. and Godt, M. J. W. Allozyme diversity in plant species. in: Brown, A. H. D.; Clegg, M. T.; Kahler, A. L., and Weir, B. S., editors. *Plant population genetics, breeding and genetic resources*. Sunderland, Massachusetts: Sinauer Associates; 1989; pp. 43-63.

Hamrick, J. L.; Godt, Mary Jo W., and Sherman-Broyles, Susan L. Factors influencing levels of genetic diversity in woody plant species. *New Forests*. 1992; 6:95-124.

- Hamrick, J. L.; Mitton, J. B., and Linhart, Y. B. Levels of genetic variation trees: influence of life history characteristics. in. Proceedings of the symposium on isozymes of North American forest trees and forest insects. General technical report PSW-48. Berkeley, CA: US Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station; 1979 Jul 27; pp. 35-41.
- Hancock, John M. Microsatellites and other simple sequences: genomic context and mutational mechanisms. in: Goldstein, David B. and Schlötterer, Christian, editors. Microsatellites: evolution and applications. Oxford: Oxford University Press; 1999; pp. 1-9.
- Hartl, Daniel L. and Clark, Andrew G. Principles of population genetics. Third ed. Sunderland, Massachusetts: Sinauer Associates; 1997.
- Hecker, Karl H. and Roux, Kenneth H. High and low annealing temperatures increase both specificity and yield in touchdown and stepdown pcr. Biotechniques. 1996; 20(3):478-485.
- Herms, Daniel A.; Mattson, William J.; Karowe, David N.; Coleman, Mark D.; Trier, Terry M.; Birr, Bruce A., and Isebrands, J. G. Variable performance of outbreak defoliators on aspen clones exposed to elevated CO₂ and O₃. in: Hom, John; Birdsey, Richard, and O'Brian, Kelly, editor. Proceedings, 1995 meeting of the northern global change program ; Pittsburgh, PA. PA. Gen. Tech. Rep. NE-214: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station; 1996.
- Hubby, J. L. and Lewontin, R. C. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics. 1966; 54:577-594.

- Isebrands, J. G.; McDonald, E. P.; Kruger, E.; Hendrey, G.; Percy, K.; Prietzer K.; Sober, J., and Karnosky D.F. Growth responses of *Populus tremuloides* clones to interacting elevated carbon dioxide and tropospheric ozone. *Environmental Pollution*. 2001; 115:359-371.
- Jarne, Philippe and Lagoda, Pierre J. L. Microsatellites, from molecules to populations and back. *TREE*. 1996; 11(10):424-429.
- Jin, Li; Macaubas, Claudia; Hallmayer, Joachim; Kimura, Akinori, and Mignot, Emmanuel. Mutation rate varies among alleles at a microsatellite focus: Phylogenetic evidence. *Proc Natl Acad Sci*. 1996; 93:15285-15228.
- Johnson, L. P. V. Reduced vigour, chlorophyll deficiency, and other effects of self-fertilization in *Pinus*. *Canadian Journal of Research*. 1945; 23, Sec. C.(5):145-149.
- Karhu, A.; Dieterich, J.-H., and Savolainen, O. Rapid expansion of microsatellite sequences in pines. *Mol. Biol. Evol.* 2000; 17(2):259-265.
- Kebelmann-Betzing, C.; Seeger, K.; Dragon, S.; Schmitt, G.; Möricke A.; Schild T.A.; Henze G., and Beyermann B. Advantages of a new *Taq* DNA polymerase in multiplex PCR and time-release PCR. *BioTechniques*. 1998; 24(1):154-158.
- Kim, Z. S.; Lee, S. W., and Hwang, J. W. Genetic diversity and structure of natural populations of *Pinus thunbergii* in Korea. *Silvae Genetica*. 1997; 46(2-3):120-124.
- Kimura, Motoo and Crow, James F. The number of alleles that can be maintained in a finite population. *Genetics*. 1964; 49(4):725-738.

- Knowles, P. Spatial genetic structure within two natural stands of black spruce (*Picea mariana* (Mill.) B.S.P.). *Silvae Genetica*. 1991; 40(1):13-19.
- Kubiske, Mark E. and Pregitzer, Kurt S. Effects of elevated CO₂ and light availability on the photosynthetic light response of trees of contrasting shade tolerance. *Tree Physiology*. 1996; 16:351-358.
- Kubiske, Mark E.; Pregitzer, Kurt S.; Mikan, Carl J.; Zak, Donald R.; Maziasz, Jennifer L., and Teeri, James A. *Populus tremuloides* photosynthesis and crown architecture in response to elevated CO₂ and soil N availability. *Oecologia*. 1997; 110:328-336.
- Kuehl, Robert. Statistical principles of research design and analysis. California: International Thompson Publishing; 1994.
- Latta, Robert G. and Mitton, Jeffry B. A comparison of population differentiation across four classes of gene markers in limber pine (*Pinus flexilis* James). *Genetics*. 1997; 146:1153-1163.
- Le Corre, Valérie; Machon, Nathalie; Petit, Rémy J., and Kremer, Antoine. Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genet. Res. Camb.* 1997; 69:117-125.
- Ledig, Thomas K. Genetic variation in *Pinus*. in: Richardson, David M., in. *Ecology and biogeography of Pinus*. New York: Cambridge University Press; 1998; pp. 251-280.
- Legendre, Pierre. Spatial autocorrelation: trouble or new paradigm? *Ecology*. 1993; 74(6):1659-1673.
- Legendre, Pierre and Fortin, Marie-Josée. Spatial pattern and ecological analysis. *Vegetatio*. 1989; 80:107-138.

- Levinson, Gene and Gutman, George. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 1987; 4(3):203-221.
- Li, C. C. and Horvitz, D. G. Some methods of estimating the inbreeding coefficient. *Am. J. Human. Genet.* 1953; 5(2):107-117.
- Li, Peng; Beaulieu, Jean; Daoust, Gaétan, and Plourde, Ariane. Patterns of adaptive genetic variation in eastern white pine (*Pinus strobus*) from Quebec. *Can. J. For Res.* 1997; 27:199-206.
- Linhart, Y. B.; Mitton, J. B.; Sturgeon, K. B., and Davis, M. L. Genetic variation in space and time in a population of ponderosa pine. *Heredity.* 1981; 46(3):407-426.
- Litt, Michael and Luty, Jeffrey. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Human. Genet.* 1989; 44:397-401.
- Lodge, David M. Biological invasions: lessons for ecology. *Tree.* 1993; 8:133-137.
- Mariette, Stephanie; David Chagne; Stephane Decroocq; Giuseppe Giovanni Vendramin; Celine Lalanne; Delphine Madur, and Christophe Plomion. Microsatellite markers for *Pinus pinaster* Ait. *Ann. For. Sci.* 2001; 58(2001):203-206.
- Marshall, D. R. and Brown, A. H. D. Optimum sampling strategies in genetic conservation. in: Frankel, O. H. and Hawkes, J. G., editors. *Crop genetic resources for today and tomorrow*. London: Cambridge University Press; 1975; pp. 53-80.

- Mattson, William J. Exotic insects in North American forests: ecological systems forever altered. in: Britton, K. O., editor. Exotic pests of eastern forest; Nashville, TN. S. For. Res. Sta., P.O. Box 2680, Asheville, NC 28802: USD For. Serv & Tn Exotic Pest Plant Council; 1998: 187-194.
- Mergen, François. Ecotypic variation in *Pinus strobus* L. Ecology. 1963; 44(4):716-726.
- Messler, Walter; Li, Shou-Hsien, and Stewart, Caro-Beth. The birth of microsatellites. Nature. 1996; 381:483.
- Moore, S. S; Sargeant, L. L.; King, T. J.; Mattick, J. S.; Georges, M., and Hetzel, D. J. S. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics. 1991; 10:654-660.
- Moran, P. A. P. Notes on continuous stochastic phenomena. Biometrika. 1950; 37:17-23.
- Moretti, Tamara; Koons, Barbara, and Budowle, Bruce. Enhancement of PCR amplification yield and specificity using AmpliTaq Gold™ DNA Polymerase. BioTechniques. 1998; 25(4):716-722.
- Mullis, Kary B. and Faloona, Fred A. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods in Enzymology. 1987; 155:335-350.
- Nauta, Maarten J. and Weissing, Franz J. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics. 1996; 143:1021-1032.
- Nei, Masatoshi. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA. 1973; 70(12):3321-3323.

- . F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum Genet.*, Lond. 1977; 41:225-233.
- . Genes in populations. in. *Molecular evolutionary genetics*. New York: Columbia University Press; 1987; pp. 149-175.
- Niemela, Pekka and Mattson, William. Invasion of North American forests by European phytophagous insects. *Bioscience*. 1996; 46(10):741-753.
- Paetkau, D and Strobeck, C. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology*. 1995; 4:519-520.
- Panetsos, K. P.; Aravanopoulos, F. A., and Scaltsoyiannes, A. Genetic variation of *Pinus brutia* from islands of the northeastern Aegean sea. *Silvae Genetica*. 1998; 47(2-3):115-120.
- Pauley, Scott; Spurr, Stephen H., and Whitmore, Frank W. Seed source trials of eastern white pine. *Forest Science*. 1955; 1(3):244-256.
- PE Applied Biosystems. Genotyper analysis ver. 2.0 software. User's manual. Foster City, CA; 1994.
- . Genescan analysis ver. 2.1 software. User's manual. Foster City, CA; 1996.
- Pemberton, J. M.; J. Slate; D.R. Bancroft, and J.A. Barrett. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*. 1995; 4:249-252.

- Rajora, O. P.; Rahman, M. H.; Buchert, G. P., and Dancik, B. P. Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Molecular Ecology*. 2000; 9:339-348.
- Rajora, Om P.; DeVerno, Linda; Mosseler, Alex, and Innes, David J. Genetic diversity and population structure of disjunct Newfoundland and central Ontario populations of eastern white pine (*Pinus strobus*). *Can. J. Bot.* 1998; 76:500-508.
- Robertson, G. Philip and Tiedje, James M. Denitrification and nitrous oxide production in successional old-growth Michigan forests. *Soil Sci. Soc. A. J.* 1984; 48:383-389.
- Rose, Willard M. Biomass, net primary production and successional dynamics of a virgin white pine (*Pinus strobus*) stand in northern Michigan. East Lansing, Michigan: Michigan State University; 1984; p. 95. PhD dissertation.
- Routman, E. and Cheverud J. A rapid method of scoring simple sequence repeat polymorphisms with agarose gel electrophoresis. *Mammalian Genome*. 1994; 5:187-188.
- Rundel, Philip W. and Yoder, Barbara J. Ecophysiology of *Pinus*. in : Richardson, David M., in. *Ecology and biogeography of Pinus*. New York: Cambridge University Press; 1998; pp. 296-323.
- Salvador, L.; Alía, R. Agúndez, and Gil L. Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait) in the Iberian peninsula. *Theor Appl Genet*. 2000; 100:89-95.
- Schlötterer, Christian; Amos, Bill, and Tautz, Diethard. Conservation of polymorphic simple sequence loci in cetácean species. *Nature*. 1991; 354:63-65.

- Schlötterer, Christian and Tautz, Diehard. Slippage synthesis of simple sequence DNA. 1992; 20, (2): 211-215.
- Schnabel, Andrew and Hamrick, J. L. Organization of genetic diversity within and among populations of *Gleditsia triacanthos* (Leguminosae). Amer. J. Bot. 1990; 77(8):1060-1069.
- Schnabel, Andrew; Laushman, Roger H., and Hamrick, J. L. Comparative genetic structure of two co-occurring tree species, *Maclura pomifera* (Moraceae) and *Gleditsia triacanthos* (Leguminosae). Heredity. 1991; 67:357-364.
- Scribner, Kim T.; Gust, Judy R., and Fields, Raymond L. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic applications. Can. J. Fish. Aquat. Sci. 1996; 53:833-841.
- Slatkin, Montgomery. Gene flow in natural populations. Ann. Rev. Ecol. Syst. 1985; 16:393-430.
- . A measure of population subdivision based on microsatellite allele frequencies. Genetics. 1995; 139:457-462.
- Smith, R. H.; Peloquin, R. L., and Passof, P. C. Local and regional variation in the monoterpenes of ponderosa pine wood oleoresin. USDA Forest Service Research Paper PSW-56. 1969.
- Sokal, Robert R. Ecological parameters inferred from spatial correlograms. in: Patil, Ganapati P. and Rosenzweig, Michael L., editors. Contemporary quantitative ecology and related econometrics. Fairland Maryland: International Co-operative Publishing House; 1979; pp. 167-196.

Sokal, Robert R.; Jacquez, Geoffrey M., and Wooten, Michael C. Spatial autocorrelation analysis of migration and selection. *Genetics*. 1989; 121:845-855.

Sokal, Robert R. and Oden, Neal L. Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society*. 1978a; 10:199-228.

---. Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. *Biological Journal of the Linnean Society*. 1978b; 10:229-249.

Sommers, Lawrence M.; Darden, Joe T.; Harman, Jay R., and Sommers, Laurie K. Michigan: a Geography. Boulder, CO: Westview Press, Inc.; 1984.

Steel, Robert G. D. and Torrie, James H. Principles and procedures of statistics. New York: McGraw-Hill Book Company; 1980.

Stine, M. and Keathley, D. E. Paternal inheritance of plastids in Englemann spruce x blue spruce hybrids. *The Journal of Heredity*. 1990; 81(6):443-446.

Strahler, Arthur. Introduction to Physical Geography. New York: John Wiley and Sons, Inc.; 1965.

Streiff, Réjane; Labbe, Thierry; Bacilieri, Roberto; Steinkellner, Herta; Glössl, Joseph, and Kremer, Antoine. Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*. 1998; 7:317-328.

Takeuchi, Y.; Kubiske, M. E.; Isebrands, J. G.; Pregtizer, K. S.; Hendrey, G., and Karnosky, D. F. Photosynthesis, light and nitrogen relationships in a young deciduous canopy under open-air CO₂ enrichment. *Plant, Cell and Environment*. 2001; 24:1257-1268.

- Taylor, John S.; Durkin, Jonathan M. H., and Breden, Felix. The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. *Mol. Biol. Evol.* 1999; 16(4):567-572.
- Valdes, Ana Maria; Slatkin, Montgomery, and Freimer, Nelson B. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*. 1993; 133:737-749.
- Van Treuren, R. Van. Estimating null allele frequencies at a microsatellite locus in the oystercatcher (*Haematopus ostralegus*). *Molecular Ecology*. 1998; 7:1413-1417.
- Vendramin, G. G.; Lelli, L.; Rossi, P., and Morgante, M. A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology*. 1996; 5:595-598.
- Voelker, Robert A.; Schaffer, Henry E., and Mukai, Terumi. Spontaneous allozyme mutations in *Drosophila melanogaster*: rate of occurrence and nature of the mutants. *Genetics*. 1980; 94:961-968.
- Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; Theo Van de Lee, M.; Hornes, M.; Frijters, A.; Pot, J.; Peleman, J.; Kuiper, M., and Zabeau, M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*. 1995; 23 :4407-4414.
- Wackerman, A. E. Growth of the "Grayling Pine". *Journal of Forestry*. 1924; 22:796-797.
- Wagner, David B. Nuclear, chloroplast, and mitochondrial DNA polymorphisms as biochemical markers in population genetic analyses of forest trees. *New Forests*. 1992; 6:373-390.

- Wang, Chi-Wu; Perry, Thomas O., and Johnson, Albert G. Pollen dispersion of slash pine (*Pinus elliottii* Engelm.) with special reference to seed orchard management. *Silvae Genetica*. 1960; 9:78-86.
- Wartenberg, Daniel. SAAP spatial autocorrelation software. Version 4.3. 1989.
- Weber, James L. Informativeness of human (dC-dA)_n • (dG-dT)_n polymorphisms. *Genomics*. 1990; 7:524-530.
- Weber, James L. and May, Paula E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Human. Genet.* 1989; 44:388-396.
- Weber, James L. and Wong, Carmen. Mutation of human short tandem repeats. *Human Molecular Genetics*. 1993; 2(8):1123-1128.
- Wheeler, C. F. A sketch of the original distribution of white pine in the lower peninsula of michigan. *Mich. Agr. Exp. Sta. Bull.* 162. 1898:5-6.
- Whitney, Gordon G. Relation of Michigan's presettlement pine forests to substrate and disturbance history. *Ecology*. 1986; 67(6):1548-1559.
- Williams, Clarie G. and Savolainen, Outi. Inbreeding depression in conifers: implications for breeding strategy. *Forest Science*. 1996; 42(1):102-116.
- Williams, J. G. K.; Kubelik, A. R.; Livak, K. J. ; Rafalski, J. A., and Tingey, S. V. DNA polymorphisms amplified by arbitraty primers are useful as genetic markers. *Nucleic Acids Research*. 1990; 18(22):6531-6535.

- Workman, P. L. and Niswander, J. D. Population studies on southwestern indian tribes. II. Local genetic differentiation in the Papago. *Am. J. Human. Genet.* 1970; 22:24-49.
- Wright, Jonathan W. Pollen dispersion of some forest trees. USDA, Northeastern Forest Experiment Station Paper 46. 1952.
- . Genetics of eastern white pine. USDA For. Ser. Res. Pap. WO-9. 1970.
- Wright, Sewall. Systems of mating. I. The biometric relations between parent and offspring. *Genetics.* 1921; 6:111-178.
- . Coefficients of inbreeding and relationship. *The American Naturalist.* 1922; 56:330-338.
- . Isolation by distance. *Genetics.* 1943; 28:114-138.
- . Isolation by distance under diverse systems of mating. *Genetics.* 1946; 31:39-59.
- . The genetical structure of populations. *Annals of Eugenics.* 1951; 15:323-354.
- . The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution.* 1965; 19:395-420.
- . Evolution and the genetics of populations. II. The theory of gene frequencies. Chicago: The University of Chicago Press; 1969.
- Wu, H. X.; Matheson, A. C., and Spencer, D. Inbreeding in *Pinus radiata*. 1. The effect of inbreeding on growth, survival and variance. *Theor Appl Genet.* 1998; 97:1256-1268.

Wu, Junyuan; Krutovskii, Konstantin V., and Strauss, Steven H. Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. *Genetics*. 1998; 150(1605-1614):1605-1614.

Yazdani, R.; Muona, O.; Rudin D., and Szmidt A.E. Genetic structure of a *Pinus sylvestris* L. seed-tree stand and naturally regenerated understory. *Forest Science*. 1985; 31(2):430-436.

Yeh, Francis C.; Yang, Rong-cai, and Boyle, Timothy. *Popgene*. Version 1.31. 1999.

Zavarin, Eugene; Critchfield, William B., and Snajberk, Karel. Turpentine composition of *Pinus contorta* x *Pinus banksiana* hybrids and hybrid derivatives. *Canadian Journal of Botany*. 1969:1443-1453.

MICHIGAN STATE LIBRARIES



3 1293 02320 6778