Population genetic structure of red pine (*Pinus resinosa*) based on nuclear microsatellite markers

Jacquelyn Boys

A Thesis

in

the Department

of

Biology

Presented in Partial Fulfilment of the Requirements

for the degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

August 2004

© Jacquelyn Boys, 2004

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-94654-1
Our file Notre référence
ISBN: 0-612-94654-1

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Population genetic structure of red pine (*Pinus resinosa*)

based on nuclear microsatellite markers

Jacquelyn Boys

Red pine (*Pinus resinosa* Ait.) is an ecologically and economically important forest tree species of northeastern North America and is considered as one of the most genetically depauperate conifer species in the region. Traditional genetic markers have failed to identify significant genetic polymorphism in this species. In order to set conservation priorities and gain insight into the post-glacial dispersal history of red pine, highly sensitive markers such as nuclear microsatellites are valuable. I have isolated and characterized thirteen nuclear microsatellite loci by screening a partial genomic library with di-, tri-, and tetra-nucleotide repeat oligonucleotide probes. Analysis of 518 individuals representing 17 red pine populations from Manitoba through Newfoundland identified five polymorphic microsatellite loci with an average of 9 alleles per locus. The mean expected and observed heterozygosity values were 0.508 and 0.185 respectively. Private alleles were detected in six of the populations examined, and *F* coefficients showed significant departure from Hardy-Weinberg equilibrium with an excess of homozygosity indicating high levels of inbreeding in all populations studied. Populations were highly differentiated with approximately 31.5% of genetic variation among populations, and results assuming IAM and SMM were similar. Weak but significant

isolation by distance was detected, and genetic distances revealed regional genetic breaks and distinct populations of red pine. These findings suggest a complex and highly differentiated population genetic structure for red pine uncharacteristic of most conifers. The genetic patterns identified support a multiple refugia hypothesis for red pine, which should be taken into consideration when setting conservation plans for this species.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Selvadurai Dayanandan for giving me this opportunity and for supporting me throughout this project. His guidance and patience have been invaluable to me.

I would also like to thank my committee members Dr. Paul Widden and Dr. William Zerges for all of their helpful advice and kind support.

I am very grateful to Dr. L.A. Hermanutz (Memorial University), Dr. Daniel McLaughlin, and Ms. Cathy Calegoropolous (Concordia University), Mr. Normand Fleury (Montreal Botanical Gardens), Mr. Richard Jordan (The Guelph University Arboretum), Mr. Jim Bowen, and Mr. Tommy George (West Virginia Division of Forestry), Mr. Dale Simpson (Canadian Forest Service), Mr. Ron Wolanin (Massachusetts Audubon Society), and Dr. Rick Kesseli (University of Massachusetts) for their help in collecting samples. I would like to thank Mr. Jonathan Gray (Concordia University) for his help in preparing the map.

I would like to acknowledge the generosity of Mr. Steve D'eon, Petawawa Research Forest, who provided plant materials. His assistance and support are greatly appreciated.

I would like to thank Concordia University, the Natural Sciences and Engineering Research Council of Canada, Le Fonds Quebecois de la Recherche sur la Nature et les Technologies, and Canada Foundation for Innovation for the financial support, and Concordia Centre for Structural and Functional Genomics for laboratory resources.

My very special appreciation goes to my husband Jonathan and my mother Irma for

their love, encouragement, and patience.

Finally, I would like to thank all the members of the Forest and Evolutionary Genomic Laboratory for all of their support and friendship.

Table of Contents

List of Figures x

List of Tables xi

1. Introduction 1

    1.1 Red Pine 2

        1.1.1 Distribution and ecology of red pine 2

        1.1.2 History and genetics of red pine 3

    1.2 Nuclear Microsatellites 6

        1.2.1 General Characteristics of nuclear microsatellites 6

        1.2.2 Microsatellite mutation rates and mechanisms 7

        1.2.3 Microsatellite mutation models and patterns 8

        1.2.4 Population differentiation measures and genetic distances 10

        1.2.5 Potential problems 12

        1.2.6 Applications 13

    1.3 Study objectives 14

2. Materials and Methods 15

    2.1 Samples collected 15

    2.2 Isolation and characterization of microsatellites 16

    2.3 Assessment of genetic diversity and population genetic structure in red pine 19

3. Results 21

3.1 Recovery of microsatellite loci and detection of polymorphism..................21

3.2 Levels of genetic variation and population genetic structure........................ 22

4. Discussion..32

4.1 Characteristics of microsatellite loci in red pine.................................... 32

4.2 Patterns of genetic diversity 33

4.3 Inbreeding and evolution of mating system 34

4.4 Population differentiation and unique populations 38

4.5 Implications for glacial refugia and post-glacial colonization 41

4.6 Implications for conservation 43

5. Conclusions 45

Literature Cited.. 46

Appendix 1: Protocols.. 64

Protocol 1: DNA Extraction Procedure... 64

Protocol 2: Construction and screening of genomic library......................... 66

Protocol 3: 1% agarose gel recipe.. 69

Protocol 4: DNA sequencing reaction recipe... 70

Protocol 5: Sequencing reaction thermal cycling program 71

Protocol 6: Purification of products of sequencing reaction and sequencing procedure.. 72

Protocol 7: PCR thermal cycling program 73

Protocol 8: Genotyping procedure.. 74

Appendix 2: Hardy-Weinberg Equilibrium........75
Appendix 3: Test Statistics........76
Test 1: Expected heterozygosity........76
Test 2: Fisher's exact test for Hardy-Weinberg equilibrium and linkage disequilibrium........77
Test 3: Score test (U test) for heterozygote deficiency........79
Test 4: Hierarchical F coefficients........80
Test 5: Calculation of R_{ST}........82
Test 6: Test 6: Nei's standard genetic distance (D_S)........84
Test 7: Delta μ^2 genetic distance........85
Test 8: Self-fertilization rates........86
Test 9: Mantel test........87
Appendix 4: The Unweighted-Pair-Group Method (UPGMA)........88
Appendix 5: List of Abbreviations........91

List of Figures

Figure 1. Map showing the locations of the seventeen sampled populations of *Pinus resinosa*..16

Figure 2. UPGMA dendogram depicting standard genetic distances (Nei, 1978) between seventeen *Pinus resinosa* populations based on five polymorphic nuclear microsatellite loci..30

Figure 3. UPGMA dendogram depicting delta μ^2 genetic distances (Goldstein et al., 1995) between seventeen *Pinus resinosa* populations based on five polymorphic nuclear microsatellite loci...31

List of Tables

Table1: Population sizes (n) and locations including state or province, and longitude and latitude in decimal degrees....15

Table 2: Primer sequences and optimal annealing temperatures (T) for twenty nuclear microsatellite loci isolated from *Pinus resinosa*....18

Table 3: Repeat pattern, expected size, and number of alleles detected (A) for fourteen nuclear microsatellite loci isolated and optimized in *Pinus resinosa*....22

Table 4: Size and frequency of alleles and heterozygosity values for five polymorphic nuclear microsatellite loci in seventeen *Pinus resinosa* populations....24

Table 5: *F* coefficients calculated for five polymorphic *Pinus resinosa* nuclear microsatellite loci based on five hundred and seventeen individuals from seventeen different populations with bootstrap values at 95% confidence intervals....27

Table 6: F_{IS} and selfing rates (S) assuming $S=2F_{IS}/1+F_{IS}$ for seventeen *Pinus resinosa* populations....27

Table 7: Maternal parent and progeny genotypes for eight heterozygous parents.... 29

1. Introduction

The commercial, ecological and evolutionary significance of red pine make it a very important species to study. However, red pine is considered to be one of the most genetically uniform conifers, and little information has been obtained for this species using traditional genetic markers.

Molecular markers have become important tools for conservation, evolutionary and population genetics studies. Nuclear microsatellites are increasingly becoming the markers of choice for many of these studies due to the fact that they are codominant, highly variable, and easy to score. In order to apply these markers to population genetics analyses, it is necessary to have some understanding of their characteristics, such as mutation rates, patterns, and models.

The focus of this thesis is to develop polymorphic nuclear microsatellites for red pine and to use these markers to gain genetic information about this species.

1.1 Red pine

1.1.1 Distribution and ecology of red pine

The genus *Pinus*, the largest genus of conifers, comprises over 100 extant species, and is the most widespread genus of trees in the Northern Hemisphere (Price et al. 1998). It is subdivided into two major groups, subgenus *Strobus* (soft pines), and subgenus *Pinus* (hard pines). Red pine, *Pinus resinosa* Ait, belongs to the latter of the two groups.

Red pine is one of the most extensively planted trees in the northern United States and Canada and is widely used as a source of wood for poles, cabin logs and fuel wood. Red pine forests are also managed for recreation, wildlife habitat and watersheds. It is distributed throughout northeastern North America, ranging from southeastern Manitoba, eastward through the Great Lakes/St. Lawrence region to Newfoundland, and south to West Virginia (Rudolf, 1990). Red pine is native to areas with cool to warm summers, cold winters, and low to moderate precipitation. Its northern limit is related to the frost-free period and closely parallels the 2° C mean annual isotherm (Cherry, 2001). Red pine commonly grows on sandy soils, with silt-clay content of 10-40%, available water storage capacity of 6-23%, and pH of 4.5-6 (Rudolf, 1990). Red pine is monoecious and flowers between April and June. Seed production decreases with stand density and good seed crops occur at intervals of 3-7 years (Rudolf, 1990). Fire is considered to be an important agent that creates the conditions necessary for the establishment of red pine stands, by providing favorable seedbeds, eliminating competitors or pests, and opening up the overstory canopy for this relatively shade intolerant species (Rudolf, 1990).

Damage to red pine can occur through various agents although, in general, they have fewer pests than related species (Rudolf, 1990). The North American strain of canker (*Gremmeniella abietina*) is one of the pathogens that cause serious damage and mortality in red pine, particularly in plantations (Ylimartimo et al., 1997).

1.1.2 History and genetics of red pine

Despite its wide geographic distribution, red pine shows high morphological uniformity, and is considered to be one of the most genetically depauperate conifer species in North America (Fowler and Lester, 1970). Some studies however, have demonstrated that traits such as the timing of bud set and flowering are correlated with latitude (Fowler and Lester, 1970). The genetic uniformity of red pine may have resulted from passage through a genetic bottleneck during the last glacial maximum, about 18000 years before present (Fowler and Morris, 1977). The fossil data, dating from 16000 to 18000 years before present suggests that red pine populations persisted in refugia located in the southern Appalachian Mountains (Jackson et al., 2000). Red pine may also have survived during glacial maxima in refugia off the present coastline of the eastern seaboard, consisting of large exposed, non-glaciated islands and extensions of the mainland (Pielou, 1992). The highly fragmented population structure and self-compatible mating system (Fowler, 1964) of red pine may also have contributed to the loss of genetic variation through inbreeding and genetic drift during the post-glacial expansion from these refugia. Although the genetic uniformity of red pine contrasts with the high variability found in most conifer species (Hamrick and Godt, 1990), low genetic diversity

has also been reported in several conifer species such as torrey pine (*Pinus torreyana* Parry ex Carr) (Ledig and Conkle, 1983), red spruce (*Picea rubens* Sarg.) (Hawley and DeHayes, 1994), and western red cedar (*Thuja plicata* Donn. Ex D. Don) (Glaubitz et al., 2000).

The low genetic variability of red pine may compromise its ability to respond to selective pressures such as changes in environmental conditions, and may increase its risk of extinction due to threats such as pests and disease. It has been predicted that red pine will be extirpated from the U.S. within the next century due to increases in temperature during the growth season (Prasad and Iverson, 1999; Cherry, 2001). Further decline in natural populations of red pine has been caused by harvesting without regeneration, and suppression of fire that is necessary for seedling recruitment (Mosseler et al., 1992). As a mitigating measure against potential threats, it is necessary to identify and conserve genetically distinct populations of red pine to ensure long-term survival of the species. However, the traditional genetic markers are of limited utility for assessing patterns of genetic diversity in red pine. Isozyme studies have failed to detect genetic variation in red pine (Fowler and Morris, 1977; Allendorf et al., 1982; Simon et al., 1986; Mosseler et al., 1991), even when the most disjunct populations were included in the study (Mosseler et al., 1991). However isozyme variation reflects protein polymorphisms that may be under selective constraint, and molecular markers that examine variability at the DNA level, particularly non-coding regions may reveal more genetic variation (Mitton, 1994). Using randomly amplified polymorphic DNA (RAPD) analysis, Mosseler et al. (1992), compared the genetic variability of red pine to that of white spruce (*Picea glauca*) and black spruce (*Picea mariana*). Although high levels of variation were detected in both

spruce species, virtually none was detected in red pine. Interestingly, DeVerno and Mosseler (1997) reported that digestion of RAPD reaction products with restriction enzymes (RAPD-RFLP analysis) could detect low levels of polymorphism in red pine that was not apparent with simple RAPD analysis. Recent studies based on chloroplast simple sequence repeats (cpSSR) analysis have revealed some genetic polymorphism in red pine (Walter and Epperson, 2001; Echt et al., 1998). However, these markers are of limited utility due to low intra-population polymorphism and uniparental (paternal in pines) inheritance of the chloroplast genome (Mitton, 1994). It has been suggested that nuclear microsatellite markers would be more suitable to gain information about the population genetic structure of red pine (Echt et al., 1998).

1.2 Nuclear Microsatellites

1.2.1 General Characteristics of nuclear microsatellites

Microsatellites are short segments of DNA made up of a single sequence motif (1-6 bp) that is repeated numerous times. Longer tandemly repeated sequences with a motif of 10- 30 bp are called minisatellites (Jeffreys et al., 1985), while those of shorter motif length are most commonly referred to as "microsatellites", simple sequence repeats (SSRs) (Tautz, 1989), or short tandem repeats (STRs) (Edwards et al., 1991).

A repeat motif of 2 base pairs is referred to as a di-nucleotide repeat, and the terms tri-, tetra-, penta-, and hexa-nucleotide describe repeat motifs of three, four, five, and six base pairs, respectively. Microsatellites can further be divided into categories of perfect (CACACACACACA), interrupted (CACACA**G**CACACA) or compound (CACACATATATA) repeats.

Microsatellites are found in every eukaryotic organism examined so far. The most common di-nucleotide repeat in humans (Beckman and Weber, 1992) and mice (Stallings et al., 1991) is CA/GT, while the AT motif is most common in plants (Lagercrantz et al., 1993; Morgante et al., 2002). It is assumed that microsatellites are evenly distributed throughout the genome, and are rarely found in coding regions (Hancock, 1995) although some have been shown to affect expression in nearby genes (reviewed in Li et al., 2002). Most commonly, tri-nucleotide repeats have been associated with certain diseases including Huntington's disease, Kennedy's disease, myotonic dystrophy, and fragile X (reviewed in Kovtun et al., 2001). In most diseases where microsatellites are implicated,

repeat sequences reach a critical length after which they become unstable and undergo rapid expansion termed "dynamic expansion". A wide variety of cancers have also been associated with high levels of microsatellite instability (Halling et al., 1999).

1.2.2 Microsatellite mutation rates and mechanisms

Understanding the mutation model and rates of mutation underlying microsatellite evolution is of great importance when analyzing microsatellite data, especially since microsatellite loci are characterized by relatively high mutation rates. It is typically assumed that the mutation rate of microsatellites is approximately 10^{-3} in humans (Weber and Wong, 1993) although the rate varies among species from about 10^{-2} in *Escherichia coli* (Levinson and Gutman, 1987) to 6×10^{-6} in *Drosophila* (Schug et al., 1997). In plants, mutation rates of 10^{-2} and 3.9×10^{-3} have been found in chickpeas (Udupa and Baum, 2001), 2.4×10^{-4} in durum wheat (Thuillet et al., 2002), and 7.7×10^{-4} in maize (Vigouroux et al., 2002). It is also possible for mutation rates to vary among loci (Thuillet et al., 2002). One of the factors that influence the mutation rates is the type of repeat motif of the microsatellite (Ellegren et al., 1995; Chakraborty et al., 1997). For example di-nucleotides have the greatest mutation rates in *Drosophila* (Schug et al., 1998), humans (Chakraborty et al., 1997), and maize (Vigouroux et al., 2002). Gender may also influence mutation rates. In barn swallows the mutation rate was almost twice as high in males than in females (Primmer et al., 1998), and this bias has also been reported in humans (Ellegren et al., 2000a). There is also evidence that mutation rate increases with the size difference of alleles in a heterozygous individual (Amos et al., 1996). The

greatest effect on mutation rate however, appears to come from the microsatellite length (Schlotterer et al., 1998), and this has been demonstrated in humans (Weber and Wong, 1993) and plants (Thuillet et al., 2002; Vigouroux et al., 2002).

There are two mechanisms that can explain the high mutation rates observed for microsatellite loci. The most commonly accepted is that of DNA polymerase slippage also referred to as slipped strand mispairing (Levinson and Gutman, 1987). During replication, the template strand dissociates from the newly synthesized strand temporarily and the two strands may re-anneal out of phase with each other. This creates a single-stranded loop on one of the strands. Depending on which strand contains the loop, an insertion or deletion will occur if the excision or filling in is done on the wrong strand. Additional explanations for the high mutation rates include recombination between DNA molecules by unequal crossing over or by gene conversion (Smith, 1976; Jeffreys et al., 1994).

1.2.3 Microsatellite mutation models and patterns

In order to estimate population differentiation measures and genetic distances using microsatellite data, several theoretical mutation models have been developed for use in population genetics. This study considers the two extreme mutation models. These models are the infinite alleles model (IAM) (Kimura and Crow, 1964), and the stepwise mutation model (SMM) (Kimura and Ohta, 1978), which was originally designed for allozyme data. The IAM assumes that each mutation creates a novel allele at the rate of mutation, μ. This model does not allow for homoplasy and therefore any identical alleles

are assumed to share the same ancestry and are identical by descent (IBD). However, the widely accepted slipped strand mispairing mechanisms (discussed above), will tend to cause small changes in the repeat number of microsatellite sequences such that alleles of similar size are more closely related. Under the SMM, microsatellites mutate by adding or deleting a repeat unit from the sequence, with the probability of a mutation in either direction being $\mu/2$. As a result the SMM should be more consistent with the mechanism of microsatellite mutation and a memory of allele size will exist under this mutation model (Slatkin, 1995). Population genetics estimators based on the SMM have been developed specifically for microsatellite data (discussed in the following section).

Evidence for the SMM has been tested in several organisms. It was demonstrated that the majority of mutations involve one or two repeat units in humans (Weber and Wong, 1993; Amos et al., 1996), and that 84% percent of mutations in barn swallows (Primmer et al., 1998), and 88% of mutations in pipefish (Jones et al., 1999) involved one repeat unit. Plants also conform largely to a SMM, with 100% of observed mutations in durum wheat, and 83% of observed mutation in maize resulting in changes of one repeat unit (Vigouroux et al., 2002). It should however be mentioned that significant amounts of non stepwise mutations has been documented in some organisms (Colson and Goldstein, 1999; Van Oppen et al., 2000) and that the frequency of non stepwise mutations varies between taxonomic groups ranging from 4 to 74% (reviewed in Ellegren 2000b). More recently, the two-phase model (TPM) (Valdès et al., 1993; Di Rienzo et al., 1994) has been developed, which is a modified SSM that allows for larger mutational events. However none of the mutation models developed to date, have been shown to perfectly conform to microsatellite evolution.

Although the stepwise mutation model appears to explain the majority of mutations at microsatellite loci in plants thus far, it has become increasingly evident that all the proposed models oversimplify the evolution of microsatellite loci. For example, there may be a bias for microsatellites to mutate to a larger rather than smaller size (Weber and Wong, 1993; Primmer et al., 1996; Ellegren, 2000a). In plants, 67% of mutations in durum wheat (Thuillet et al., 2002), and 79% of mutations in maize (Vigouroux et al., 2002) resulted in a larger allele. It has also been demonstrated that geographically derived groups of maize have larger average allele sizes relative to ancestral groups of maize, further corroborating this upward bias (Vigouroux et al., 2003). Although this bias could theoretically lead to microsatellites of infinite size, some reasons have been proposed for why this does not happen. One explanation is that there may be an equilibrium between mutations that cause alleles to increase in size and base substitutions that degrade the microsatellite sequence (Kruglyak et al., 1998). However, there is also evidence that the upwards mutational bias decreases as microsatellites increase in size in *Drosophila* (Harr and Schlotterer, 2000), humans (Weber and Wong, 1993), and plants (Vigouroux et al., 2002).

1.2.4 Population differentiation measures and genetic distances

The high mutation rates and complex patterns of microsatellite evolution should not be neglected when choosing appropriate population genetics measures. To assess levels of population differentiation measures using microsatellite data, many studies have continued to rely on estimators of Wright's (1951, 1965) F_{ST}, which assume the IAM,

and are used for predicting gene flow (Appendix 3, Test 4). The most commonly used unbiased estimator of F_{ST} is that of Weir and Cockerham (1984). There has been both support for and arguments against using F_{ST} as an indicator of gene flow (reviewed by Neigel, 2002) and when using highly polymorphic markers such as microsatellites, the relationship between F_{ST} and the number of migrants may not hold (Rousset, 1996). As discussed above, microsatellites may not conform to the IAM, and may in fact be closer to a SMM, a mutation model that generates homoplasy (where alleles are identical in state but not by descent). In this case, the probability of identity of two alleles decreases as the mutation rate increases (Rousset, 1996) and as a result F_{ST} may be underestimated (Hedrick, 1999; Balloux et al., 2000). Despite the limitations of F_{ST}, it is expected that it will continue to be a standard measure of population genetic structure (Neigel, 2002).

In order to accommodate microsatellite evolution models, Slatkin (1995) developed a population differentiation measure specifically for microsatellite data termed R_{ST}, an F_{ST} analogue that assumes the SMM and is independent of the mutation rate under this model (Slatkin, 1995; Balloux et al., 2000) (Appendix 3, Test 5). Although relatively insensitive to mutation rate, R_{ST} is more sensitive to the mutation model and tends to have a higher variance (Balloux et al., 2000). It has been suggested that even under a strict SMM, F_{ST} may outperform R_{ST} due to the high variance of R_{ST} (Gaggiotti et al., 1999), although it has also been predicted that R_{ST} will be superior when some memory of allele size exists (Slatkin, 1995). The relative performance of F_{ST} and R_{ST} may also depend on the level of population differentiation, where R_{ST} performs better for highly structured populations when the effect of mutation becomes more important than migration (Balloux et al., 2000). In general, R_{ST} increases accuracy as it conforms to the

expected mutation model of microsatellite loci but decreases precision due to its high variance.

In addition to population differentiation measures, many genetic distances have been developed although only a few remain in use (reviewed in Nei, 1987; Kalinowski, 2002), and several studies have compared the relative performance of these measures (Paetkau et al., 1997; Kalonowski, 2002). Nei's (1978) unbiased standard genetic distance D_S continues to be one of the most commonly used genetic distances assuming the IAM, while Goldstein et al.'s (1995) delta μ^2 distance developed for use with microsatellite data assumes the SMM (Appendix 3, Tests 6,7).

1.2.5 Potential problems

One of the main disadvantages of using microsatellite markers is the time and cost associated with their development. Various methods have been described in order to overcome these problems in isolating microsatellites (reviewed by Zane et al., 2002).

Another problem is constraint on allele size leading to homoplasy resulting in alleles that are identical in state but not by descent. There is evidence that microsatellite size is constrained, rarely surpassing 30 repeats (Balloux and Lugon-Moulin, 2002), and it is expected that homoplasy will increase with the time of divergence (Estoup and Cornuet, 1999). Population differentiation may be underestimated with increasing homoplasy (Balloux et al., 2000), and genetic distance measures may plateau depending on the degree of constraint, mutation rate and population size (Nauta and Weissing, 1996;

Feldman et al., 1997). The effects of homoplasy on population genetics studies are reviewed by Estoup et al. (2002).

Finally, null alleles, which are sequence variants at the primer sites that prevent primer annealing during PCR, may also present a problem when analyzing microsatellite data. When undetected these alleles may result in an overestimate of homozygotes.

1.2.6 Applications

Despite the potential problems listed above, the high degree of polymorphism, codominance, and the presence of multiple, easily scored loci have made microsatellites the marker of choice for genetic mapping, forensics, and evolutionary studies. Nuclear microsatellite markers have been useful in detecting genetic variation in animals, (Firestone et al., 2000; Feldheim et al., 2001; Bidlack and Cook, 2002) and plants, (Dayanandan et al., 1998; Rajora et al., 2001; Stacy et al. 2001; Bérubé et al., 2003), including pines such as *Pinus strobus* (Echt et al., 1996), *Pinus radiata* (Fisher et al., 1998), *Pinus sylvestris* (Soranzo et al., 1998), and *Pinus taeda* (Al-Rabab'ah and Williams, 2002). These markers have been used for examining the genetic effects of different harvesting practices of forest trees (Thomas et al., 1999), estimating population differentiation (Echt et al., 1996), and clonal fingerprinting (Dayanandan et al., 1998). To date, no such markers are available for red pine. Reported here are the results of a first genetic diversity study based on nuclear microsatellite loci of red pine identified through the screening of genomic libraries.

1.3 Study objectives

The objectives of my thesis were to (1) isolate and characterize nuclear microsatellite markers for red pine to detect previously undetected levels of genetic variation in this species, (2) use these markers to identify the areas of highest genetic diversity and distinct populations of red pine in order to set conservation priorities, (3) gain insight into the population genetic structure of red pine populations, in terms of mating system, isolation by distance, and post-glacial dispersal history, (4) determine if population differentiation and genetic distance measures based on different mutation models (IAM and SMM) yield similar results in this study.

2. Materials and methods

2.1 Samples collected

A total of 518 individuals were sampled from seventeen red pine populations. The locations of these are listed in Table 1 and illustrated in Figure 1. Samples from populations 1, 2, 3, 8, 13, and 14 were needles, while those from the remaining populations consisted of seeds that were germinated prior to DNA extraction.

Table 1. Population sizes (n) and locations including state or province, and longitude and latitude in decimal degrees of seventeen *Pinus resinosa* populations.

Population	Longitude	Latitude	n
1. Pendleton/ West Virginia	79.44	38.67	30
2. Hardy County/ West Virginia	79.01	38.88	33
3. High Ledges[1]/ Massachusetts	72.69	42.59	20
4. Beaver Lake/ Nova Scotia	65.20	44.14	35
5. Tracy1/ New Brunswick	66.39	45.38	35
6. Tracy2/ New Brunswick	66.39	45.38	36
7. Geary/ New Brunswick	66.32	45.42	31
8. Petawawa/ Ontario	77.83	46.00	31
9. Delta County/ Michigan	87.00	46.00	35
10. Beauceville/ Quebec	70.38	46.12	34
11. Cass Lake/ Minnesota	94.36	47.22	34
12. Isle Maligne/ Quebec	71.37	48.35	34
13. Abitibi/ Quebec	79.37	48.50	26
14. Terra Nova[2]/ Newfoundland	53.93	48.53	10
15. South Junction/ Manitoba	95.52	49.02	33
16. Rowsells Brook/ Newfoundland	56.22	49.14	32
17. Sioux Lookout/ Ontario	91.57	50.04	29

[1]High Ledges abbreviation for High Ledges Wildlife Sanctuary, [2]Terra Nova abbreviation for Terra Nova National Park)

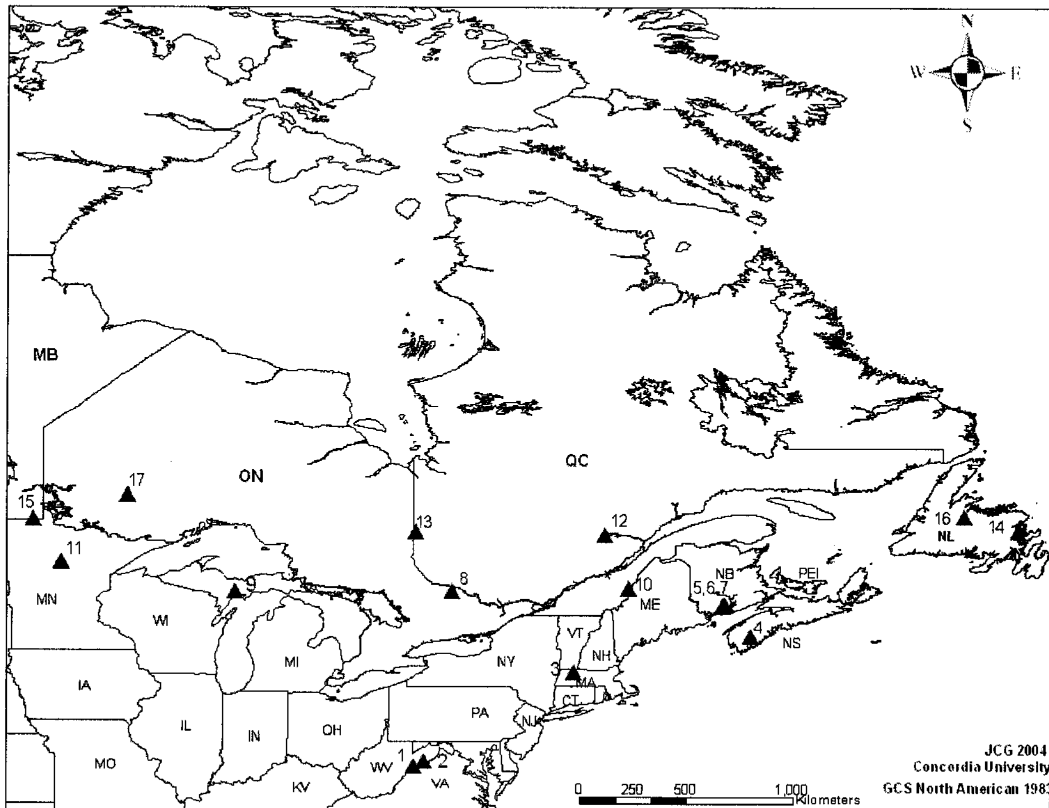


Figure 1. Map indicating the locations of the seventeen sampled populations of *Pinus resinosa.* Population numbers correspond to those in Table 1.

2.2 Isolation and characterization of nuclear microsatellite markers

DNA was extracted from each needle sample following the method described in Dayanandan et al. (1997) (Appendix 1, Protocol 1). A partial genomic library of red pine DNA fragments between 500 to 1000 bp was constructed and screened with $(TC)_{15}$, $(AC)_{15}$, $(AT)_{15}$, $(AAT)_{15}$, $(AAAT)_{15}$, $(TG)_{15}$, $(AG)_{15}$, $(AC)_{15}$, $(AGAT)_{15}$, and $(AAAG)_{15}$ following the protocol of Dayanandan et al. (1998) (Appendix 1, Protocol 2). Either

single-stranded DNA (isolated using the Wizard M13 DNA purification system; Promega, Madison, Wisconsin) or double-stranded DNA (isolated using the QIAprep Spin Miniprep Kit; Qiagen, Mississauga, Ontario), was sequenced using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, California) (Appendix 1, Protocol 4, 5, 6). The yield of isolated DNA was tested by electrophoresing on a 1% agarose gel with ethidium bromide (0.33μg/ml) (see Appendix 1, Protocol 3 for agarose gel recipe).

Oligonucleotide primers complimentary to regions flanking the repeat regions were designed and synthesized (Operon Technologies, Alameda, California). The polymerase chain reaction (PCR) was used to amplify microsatellites in the sampled individuals. Amplification reactions were performed in a total volume of 25 μl with 0.2 mM dNTP, 2.0 mM $MgCl_2$, 2.5 μl 10x buffer, and 2.5 pmol of each primer. The thermal cycling profile of PCR reactions consisted of an initial denaturation at 94 °C for 3 minutes, followed by 30 cycles of: 94 °C for 1 minute (denaturation), 1 minute annealing (temperature varies according to primers), and 72 °C for 1 minute (extension). This was followed by 72 °C for 4 minutes (final extension) (Appendix 1, Protocol 7). The optimal annealing temperature for each primer is given in Table 2. The sizes of amplified products were determined using an ABI 310 genetic analyzer and Genescan software (Applied Biosystems, Foster city, California) (Appendix 1, Protocol 8). Primers successfully optimized were tested for polymorphism based on 68 individuals from different populations.

Table 2. Primer sequences and optimal annealing temperatures (T) for twenty nuclear microsatellite loci isolated from *Pinus resinosa.*

Locus	Primer sequences (5'-3')	T (°C)
PRE1	F: TAGAGTTAGTTTAGCAGTTAGAATTCA R: CGTTTGGAGTAGCGGGTCT	50
PRE2	F: ATCAAGCCGACTAGCGACTCA R: ATATTTCGCCGCACACCAGT	65
PRE3	F: ACTCGACCTCATTTGGATACTGT R: CCCAGAAGCAACTTCACAACT	60
PRE4	F: ATTATCAAGCCGACCTACAACT R: TGCAGTTACAGCTTCAAATTCA	60
PRE6	F: GTTGCAGCTTCAAATTCAACA R: AGGCTCAGAAGCCCAACAGT	65
PRE7	F: TTATGAAAACGGCTTCCGAAT R: CAGTTAAAGCTAACTTCGAGCCA	60
PRE8	F: CCCTTCTAGAGCTAGTCCACCAT R: GTGTGTTTACTTTCATCAAACTGGT	50
PRE10	F: CTGGTCTTGGCCTAAGAATATGAAG R: CATTGGGACGTAAACAACAATACCA	55
PRE11	F: AAAATACAAATATTTCATTATGAATC R: TAAAAACATATAACCTGTATAGGAG	50
PRE12	F: ATAAAAATAATTAAACAGTTATGTACGA R: TCATAACATTATTTTAAGGGACT	55
PRE13	F: GATGTGTCTTTAGGCTCGTTGC R: AGGGTTAGTAATCACGGCCTGT	60
PRE14	F: ATGATAAGTGTTATGCACAGTTATGA R: GCCAGCAAGCCTGGTTAGT	50
PRE15	F: AAAGCCTCTGTCAAGGCAAGA R: CATCTGCCTTGAGCGCCT	60
PRE16	F: TCCTGCGATGAGTCTCTTTGT R: TCCATTTTTTACTTTTGATAACTTTAC	50
PRE18	F: CATTTTTCAAGTCTTTAGCCA R: CTGAAATTGATGGTGTCTACTTG	55
PRE19	F: GACTAGGGAAATATGGTCCGA R: TTCCATAGCATAACCCTTAGCA	55
PRE22	F: CAGGCGAATTTATCGAACCGT R: TGAACTTTCGCGATTTAATACAG	55
PRE23	F: AAATGGATACCTACAATATCTAGTGA R: GGGGGGAGATAATGATAGAGT	50
PRE24	F: GTTTTTTAAATTGGGAAGGCG R: CGTGGGGGAGATAGTGATAGAGT	50
PRE25	F: TGGAGGGCCATTGTTCACAC R: CCACCCTGACGTGACGACATAG	50

2.3 Assessment of genetic diversity and population genetic structure in red pine

For each polymorphic locus, the total number of alleles per population was determined and the distribution of allele frequencies was calculated. The average number of alleles per locus (A) for each population over all loci, unbiased values of the expected proportion of heterozygotes (H_E), and the observed proportion of heterozygotes (H_O) were calculated using the Genetic Data Analysis (GDA) version 1.0 software (Lewis and Zaykin, 2001) (Appendix 3, Test 1). Private alleles were also computed for each population. Deviation from Hardy-Weinberg equilibrium (HWE) (Appendix 2) and linkage disequilibrium was examined using Fisher exact tests with 3200 reshufflings (Appendix 3, Test 2). Tests for heterozygote deficiency for each locus and population was determined by the method of Rousset and Raymond (1995) as implemented in Genepop 3.4 software (Raymond and Rousset, 1995) (Appendix 3, Test 3). Population genetic structure was examined by estimating Wright's (1951, 1965) hierarchical F coefficients (Appendix 3, Test 4) by the method of Weir and Cockerham (1984), as implemented in the program GDA, and the significance of the F coefficients was determined by constructing 95% confidence limits by bootstrapping with 10000 replications. Selfing rates (S) were calculated for each population using the formula $S=2F_{IS}/1+F_{IS}$ (Appendix 3, Test 8). Unbiased genetic distances were calculated (Nei, 1978), as implemented in GDA (Appendix 3, Test 6). Calculation of R_{ST} (Slatkin 1995) and delta μ^2 (Goldstein et al. 1995) was conducted using RSTCALC (Goodman 1997) (Appendix 3, Tests 5,7). F_{ST} values and Nei's standard genetic distance D_S assume an infinite alleles model (IAM), while R_{ST} and delta μ^2 assume a stepwise mutation model (SMM). Matrices of pairwise F_{ST} versus R_{ST} values, and Nei's genetic distances versus

delta μ^2 distances were compared using Mantel tests in order to determine whether the results obtained were dependent on the mutation model assumed for the analyses (Appendix 3, Test 9). Nei's genetic distance and delta μ^2 were used for constructing dendograms using UPGMA (unweighted pair group method with arithmetic average) (Appendix 4). The resulting dendograms were visualized with the TreeView program (Page, 1996). Isolation by distance was calculated using IBD (Isolation by distance) software (Bohonak, 2002). The correlation between genetic distances and geographic distances was tested using Mantel tests (Mantel, 1967). Isolation by distance was examined using Slatkin's (1993) measure of similarity $(1/F_{ST}-1)/4$ taking the logarithm of both geographic and genetic distance as suggested by the author, and Rousset's (1997) distance $(F_{ST}/1-F_{ST})$ taking the logarithm of geographic distance only as suggested by the author. Both of these measures are implemented in the IBD software.

Mendelian inheritance was tested for loci PRE10, PRE13, and PRE16 where progeny of heterozygous individuals were available.

3. Results

3.1 Recovery of microsatellite loci and detection of polymorphism

A partial genomic library consisting of ~13,000 clones was screened using di-, tri, and tetra-nucleotide repeat oligonucleotide probes. The sequencing of 35 positive clones identified through screening with TC/AG probes revealed 19 microsatellites (17 TC/AG, one AT, and one AT mixed repeat), and primers were designed for nine. Seven primers were designed based on 23 microsatellites (12 AC/TG, three AT, three TC/AG, three AC/TG and AT mixed repeats, one TCC/AGG repeat, and one 6-bp repeat of TCCACA) identified through the sequencing of 68 clones positive for AC/TG probes. The AT/AAT/AAAT probes identified 13 positive clones, while sequencing revealed nine microsatellites (AT/AAT/AAAT repeats). Primers were designed for all nine. A total of 25 oligonucleotide primer pairs complementary to flanking regions of microsatellites were designed. Optimization of PCR conditions resulted in successful amplification with 20 primer pairs (Table 3). Thirteen of these produced amplification products consistent with a single-locus pattern, and four of these loci showed polymorphism, one of which amplified two polymorphic loci that were easily distinguished and scored. Another seven primer pairs produced multiple amplification products, and amplification of the remaining five primer pairs was inconsistent.

Table 3. Repeat pattern, expected size, and number of alleles detected (A) for thirteen nuclear microsatellite loci isolated and optimized in *Pinus resinosa.*

Locus	repeat motif	size (bp)	A
PRE1	$(CA)_6C_3$	158	1
PRE2	$(TC)_2C(TC)_2A(TC)_7$	247	1
PRE3	$(TC)_8TA(TC)_2TG(TC)_2$	269	1
PRE4	$(TC)_7N_{16}(TCTCT)_2N_{16}(TC)_4$	252	1
PRE6	$(GA)_6C_3N_{40}(AG)_8$	222	1
PRE7	$(GA)_9$	345	1
PRE8	$(GA)_8C_{10}$	317	1
PRE10	$(AT)_{21}$	340	12
PRE11	mixed AT's	225	1
PRE13	$(AT)_{21}$	263	13
PRE16	$(AT)_{22}$	354	15
PRE19	$(CA)_{10}$	250	1
PRE24	$(AG)_{29}$	222	3
PRE24B		222	2

3.2 Levels of genetic variation and population genetic structure

From the 518 red pine individuals surveyed, 45 alleles were identified at five polymorphic microsatellite loci, with an average of 9 alleles per locus. The highest diversity was found at the $(AT)_{22}$ repeat (locus PRE16), with 15 alleles. The loci PRE10, PRE13, PRE24, and PRE24B showed thirteen, twelve, three, and two alleles, respectively.

Of the seventeen populations, Delta County (Michigan) demonstrated the highest allelic diversity with an average of 4.6 alleles per locus. The lowest allelic diversity was found in Hardy County (West Virginia) with an average of 1.2 alleles per locus. Private alleles were identified in six of the seventeen populations. One private allele was found in each of the Petawawa (Ontario), Terra Nova (Newfoundland), and Rowsells Brook (Newfoundland) populations, while two were found in Abitibi (Quebec) and South Junction (Manitoba), and three in Delta County (Michigan). The H_E values ranged from 0.242 at locus PRE24B to 0.789 at locus PRE16, while the H_O values ranged from 0.067 at locus PRE13 to 0.317 at locus PRE16. The H_E across all loci was highest in the Beauceville (Quebec), Tracy1 (New Brunswick), and the Geary (New Brunswick) populations with values of 0.489, 0.485, and 0.483 respectively. The lowest H_E values were found in Hardy County (West Virginia), High Ledges (Massachusetts), and Pendleton County (West Virginia) with values of 0.014, 0.142 and 0.211 respectively. Observed heterozygosity ranged from 0 in Hardy County (West Virginia) to 0.269 in Beaver Lake (Nova Scotia). Mean H_O and H_E across all loci were 0.185 and 0.508 respectively (Table 4).

Table 4. Size and frequency of alleles and heterozygosity values for five polymorphic nuclear microsatellite loci in seventeen *Pinus resinosa* populations.

Locus	Size (bp)	Frequency																	
		PET	SIO	DEL	CAS	SOU	BEC	ISL	ABI	HAR	TR1	TR2	GEA	BEV	TER	RO	HIG	PEN	ALL
PRE10																			
	325					.015													**.001**
	331					.015													**.001**
	335			.029	.029			.061	.060			.029	.016					.150	**.022**
	337	.758	.464	.429	.412	.758	.500	.636	.680	1.0	.387	.114	.403	.200	.100		.056	.617	**.442**
	339	.097	.089	.114	.059	.015	.067	.015	.160					.014				.016	**.038**
	341	.065	.071	.114	.221		.233	.136	.020		.161	.300	.468	.000	.200	.656	.944	.217	**.224**
	343	.048	.304	.214	.279	.167	.167	.136	.040		.452	.486	.113	.786	.700	.344			**.249**
	345		.054	.057		.030	.033	.015				.057							**.014**
	347	.032	.018	.014															**.004**
	351			.014															**.001**
	353			.014															**.001**
	355								.040			.014							**.003**
	HO	**.097**	**.429**	**.400**	**.353**	**.333**	**.333**	**.394**	**.040**	**.000**	**.194**	**.200**	**.323**	**.257**	**.000**	**.000**	**.000**	**.467**	**.242**
	HE	**.415**	**.688**	**.750**	**.710**	**.403**	**.673**	**.562**	**.515**	**.000**	**.630**	**.666**	**.616**	**.347**	**.484**	**.458**	**.108**	**.599**	**.683**
PRE13																			
	245								.019										**.001**
	247				.029									.029					**.003**
	249	.016																	**.001**
	253				.029								.017						**.003**
	257														.100				**.006**
	259								.019										**.001**
	261			.029				.030											**.003**
	263	.742	.769	.686	.824	.818	.786	.758	.558	1.0	.583	.514	.883	.429	.200	.806	.025	.286	**.627**
	265	.161	.231	.229	.118	.152	.190	.015	.327		.069	.071	.100	.543	.700	.194	.025	.714	**.226**
	267			.057		.030					.111								**.012**
	269	.048									.125	.114							**.017**
	275							.015	.038								.550		**.035**
	277	.032					.024	.182	.038		.111	.300					.400		**.064**
	HO	**.065**	**.000**	**.000**	**.000**	**.000**	**.048**	**.273**	**.115**	**.000**	**.111**	**.257**	**.033**	**.057**	**.000**	**.000**	**.100**	**.000**	**.067**
	HE	**.427**	**.362**	**.480**	**.311**	**.311**	**.354**	**.398**	**.590**	**.000**	**.623**	**.636**	**.213**	**.528**	**.484**	**.317**	**.550**	**.416**	**.526**
PRE16																			
	334	.000	.040						.019										**.003**
	337	.000					.045	.030								.379			**.027**
	339	.000					.015						.016						**.002**
	343	.000				.031						.029							**.004**
	345	.016					.015	.182			.057								**.016**
	347	.194	.140	.086	.086	.125	.273	.485	.115	.034	.614		.516	.514		.466	.026		**.216**

		PET	SIO	DEL	CAS	SOU	BEC	ISL	ABI	HAR	TR1	TR2	GEA	BEV	TER	RO	HIG	PEN	ALL
PRE16	349	.129	.040	.129	.155		.121	.061	.096		.057	.029	.032	.029	.150				**.060**
	350	.355	.420	.386	.259	.188	.364	.167	.615	.966	.100	.771	.371	.057		.017		.960	**.353**
	352	.048	.040	.057	.172	.156	.091	.045			.129	.143	.032				.974	.040	**.113**
	354	.000				.047		.030											**.005**
	356	.000		.014	.017	.063					.014		.032						**.008**
	358	.226	.260	.243	.259	.313	.061		.077		.029	.029		.371					**.110**
	360	.032	.040	.071	.017	.078	.015		.077					.029	.850	.138			**.079**
	362	.000	.020		.034														**.003**
	369	.000		.014															**.001**
	HO	**.419**	**.440**	**.514**	**.483**	**.469**	**.273**	**.273**	**.385**	**.000**	**.286**	**.229**	**.258**	**.543**	**.268**	**.345**	**.053**	**.000**	**.317**
	HE	**.778**	**.744**	**.771**	**.817**	**.827**	**.776**	**.707**	**.598**	**.068**	**.597**	**.388**	**.602**	**.601**	**.100**	**.631**	**.053**	**.078**	**.789**
PRE24																			
	227	.790	1.0	1.0	.912	1.0	.765	.971	1.0	1.0	.819	.900	.613	.729	.100	.094	1.0	1.0	**.805**
	229	.210			.088		.235	.029	.000		.181	.100	.387	.271	.900	.859			**.192**
	231															.047			**.003**
	HO	**.290**	**.000**	**.000**	**.000**	**.000**	**.235**	**.059**	**.000**	**.000**	**.306**	**.143**	**.194**	**.143**	**.000**	**.250**	**.000**	**.000**	**.105**
	HE	**.337**	**.000**	**.000**	**.163**	**.000**	**.365**	**.058**	**.000**	**.000**	**.300**	**.183**	**.482**	**.401**	**.189**	**.254**	**.000**	**.000**	**.302**
PRE24B																			
	233	.817	1.0	1.0	.912	1.0	.838	.971	1.0	1.0	.833	.914	.452	.829	.500	.359	1.0	1.0	**.849**
	235	.183			.088		.162	.029			.167	.086	.548	.171	.500	.641			**.151**
	HO	**.367**	**.000**	**.000**	**.000**	**.000**	**.324**	**.059**	**.000**	**.000**	**.333**	**.171**	**.387**	**.343**	**1.0**	**.719**	**.000**	**.000**	**.192**
	HE	**.305**	**.000**	**.000**	**.163**	**.000**	**.275**	**.058**	**.000**	**.000**	**.282**	**.159**	**.503**	**.288**	**.526**	**.468**	**.000**	**.000**	**.242**
All loci	***HO***	**.248**	**.174**	**.183**	**.167**	**.160**	**.243**	**.211**	**.108**	**.000**	**.246**	**.200**	**.239**	**.269**	**.220**	**.263**	**.031**	**.093**	**.185**
	HE	**.452**	**.359**	**.400**	**.433**	**.308**	**.489**	**.357**	**.341**	**.014**	**.486**	**.406**	**.483**	**.433**	**.391**	**.426**	**.142**	**.211**	**.508**

Table abbreviations: PET = Petawawa, Ontario SIO = Sioux Lookout, Ontario DEL = Delta County, West Virginia CAS = Cass Lake, Minnesota SOU = South Junction, Manitoba BEC = Beauceville, Quebec ISL = Isle Maligne, Quebec ABI = Abitibi, Quebec HAR = Hardy Count, West Virginia TR1 = Tracy 1, New Brunswick TR2 = Tracy 2, New Brunswick GEA = Geary, New Brunswick BE V= Beaver Lake, Nova Scotia TER = Terra Nova National Park, Newfoundland RO = Rowsells Brook, Newfoundland HIG = High Ledges Wildlife Sanctuary, Massachusetts, PEN = Pendleton County, West Virginia

A significant departure from the Hardy Weinberg equilibrium was observed for most loci in most populations and tests for heterozygote deficiency revealed significant heterozygote deficits in all populations and for all loci except for PRE24B. No linkage disequilibrium was observed, indicating that all five loci were behaving independently of each other.

F_{IS}, F_{IT}, and F_{ST} values ranged from –0.133-0.840, 0.221-1, and 0.211-0.455 respectively, and all overall F values were significantly higher than 0 with bootstrap values at 95% confidence levels (Table 5). The overall F_{ST} value was 0.280 while the overall R_{ST} value was 0.350. F_{IS} values ranged from 0.384 to 1 across all populations and were significantly higher than 0 for all populations except for the two Newfoundland populations. Estimates of selfing rates (S) assuming $S=2F_{IS}/1+F_{IS}$ ranged from 0.555 in the Beaver Lake (Nova Scotia) population to 1 in the in the Hardy County (West Virginia) (Table 6).

The genetic distances ranged from 0 between Sioux Lookout (Ontario) and Delta County (Michigan) to 1.487 between Terra Nova (Newfoundland) and High Ledges (Massachusetts) for Nei's standard genetic distances. They ranged from 0.010 between Sioux Lookout (Ontario) and Delta County (Michigan), to 3.887 between Rowsells Brook (Newfoundland) and High Ledges (Massachusetts) for delta μ^2 distances.

Pairwise estimates of F_{ST} (IAM) and R_{ST} (SMM) were found to be highly correlated ($r=0.91$, $p=0.0001$), as were both genetic distance measures ($r=0.94$, $p=0.0001$) based on Mantel tests.

Table 5. *F* coefficients calculated for five polymorphic *Pinus resinosa* nuclear microsatellite loci based on five hundred and seventeen individuals from seventeen different populations with bootstrap values at 95% confidence intervals.

Locus	F_{IS}	F_{IT}	F_{ST}
PRE10	0.535	0.651	0.250
PRE13	0.840	0.874	0.211
PRE16	0.456	0.605	0.273
PRE24	0.381	0.663	0.455
PRE24B	-0.133	0.221	0.312
Overall	*0.504*	*0.643*	*0.280*
Bootstrap upper	0.690	0.770	0.361
Bootstrap lower	0.278	0.492	0.236

Table 6. F_{IS} and selfing rates (S) assuming $S=2F_{IS}/1+F_{IS}$ for seventeen *Pinus resinosa* populations.

Population	F_{IS}	S
Beaver Lake/ Nova Scotia	0.384	0.555
Rowsells Brook/ Newfoundland	0.387	0.558
Isle Maligne/ Quebec	0.411	0.583
Terra Nova/ Newfoundland	0.450	0.621
Petawawa/ Ontario	0.457	0.627
South Junction/ Manitoba	0.483	0.651
Tracy1/ New Brunswick	0.498	0.665
Beauceville/ Quebec	0.508	0.674
Geary/ New Brunswick	0.510	0.675
Tracy2/ New Brunswick	0.511	0.676
Sioux Lookout/ Ontario	0.521	0.685
Delta County/ Michigan	0.547	0.707
Pendleton/ West Virginia	0.561	0.719
Cass Lake/ Minnesota	0.618	0.764
Abitibi/ Quebec	0.687	0.814
High Ledges/ Massachusetts	0.790	0.883
Hardy County/ West Virginia	1.000	1.000

Both UPGMA dendograms (Figs. 2 and 3) revealed a genetically distinct group that included both Newfoundland populations and the Geary (New Brunswick) population. The remaining populations formed a separate group with the High Ledges (Massachusetts) population most divergent within this group in both trees. The only major difference between the topologies of the two trees was the placement of Beauceville (Quebec) which grouped with New Brunswick and the Nova Scotia populations in the tree based on Nei's standard genetic distances, but closer to Petawawa (Ontario) in the tree based on delta μ^2 distances.

A weak but significant positive correlation was found between geographic distance and genetic distances. For Nei's standard genetic distances (D_S) r=0.3812 (p=0.0072), delta μ^2 distances r=0.3323 (p=0.0144), and Rousset's distance (Fst/1-Fst) r=0.1320 (p=0.0291). A weak but significant negative correlation was also found between geographic distance and Slatkin's measure of similarity ((1/Fst-1)/4), r= –0.2500 (p=0.0214).

Mendelian inheritance was demonstrated for the three loci tested (PRE10, PRE13, PRE16), as all progeny tested had at least one allele in common with the maternal parent. (Table 7).

Table 7. Maternal parent and progeny genotypes for eight heterozygous parents.

Ind.	Locus	Parent Genotype[1]	Progeny genotypes[1] P1	P2	P3	P4	P5	P6
21	PRE10	337/347	337/337	337/337	337/337	337/337	337/337	----
18	PRE13	249/263	263/263	263/263	263/263	263/263	263/263	263/277
71	PRE13	363/269	263/269	263/277	263/263	263/263	263/269	263/263
33	PRE16	350/358	350/358	358/358	352/358	350/358	352/358	350/350
38	PRE16	347/358	347/350	347/350	358/358	347/347	350/358	347/347
45	PRE16	347/358	347/347	347/352	347/350	358/358	347/358	----
49	PRE16	347/352	352/352	347/350	347/352	347/358	347/347	352/360
57	PRE16	347/358	347/347	347/347	347/350	347/358	350/358	349/358

[1]Genotypes are given in terms of allele sizes in base pairs where "/ " separates the two alleles

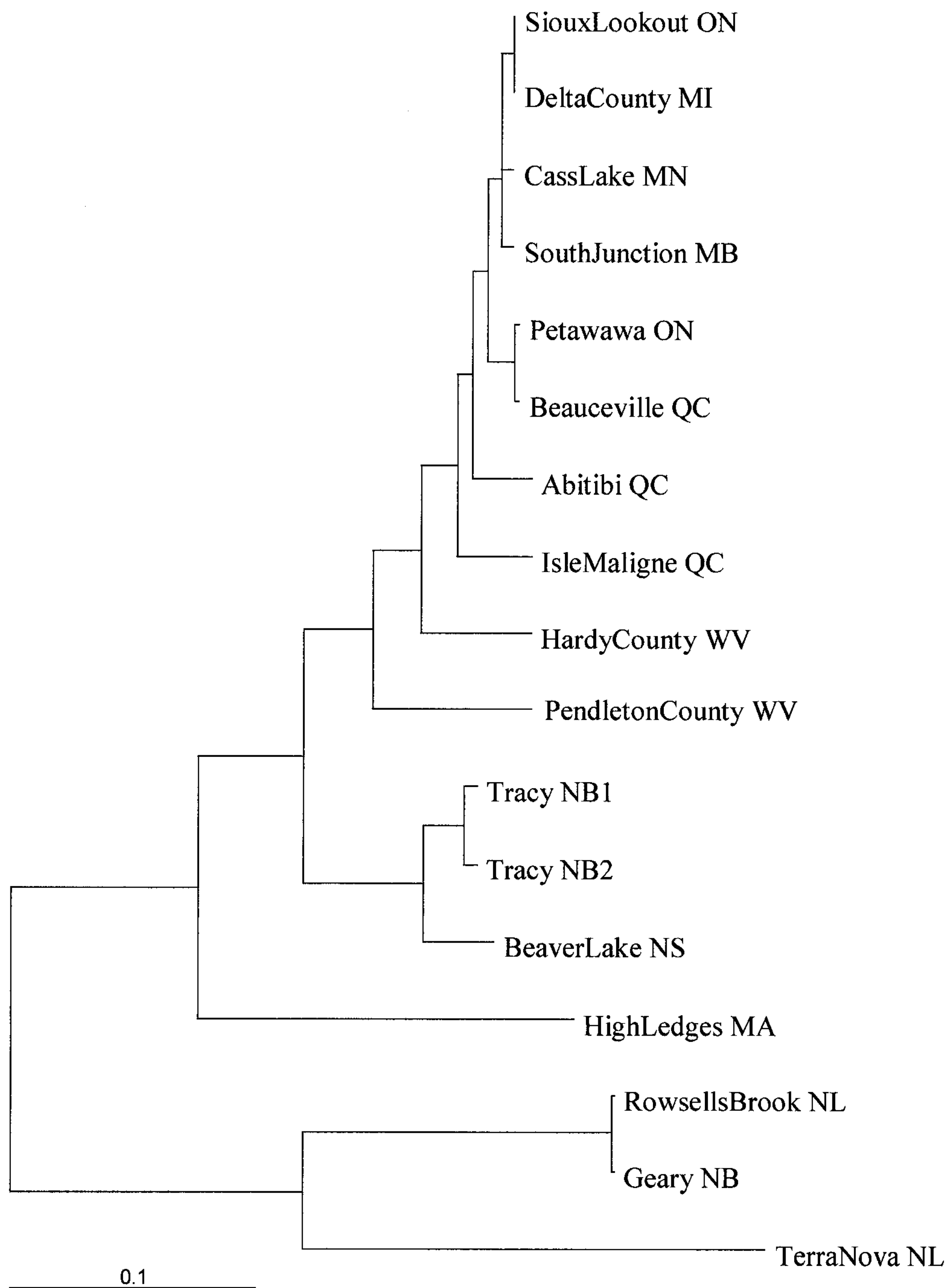


Figure 2. UPGMA dendogram depicting standard genetic distances (Nei, 1978) between seventeen *Pinus resinosa* populations based on five polymorphic nuclear microsatellite loci.

Figure abbreviation: ON=Ontario, MI=Michigan, MN=Minnesota, MB=Manitoba, QC=Quebec, WV=West Virginia, NB=New Brunswick, NL=Newfoundland, NS=Nova Scotia, MA=Massachusetts.

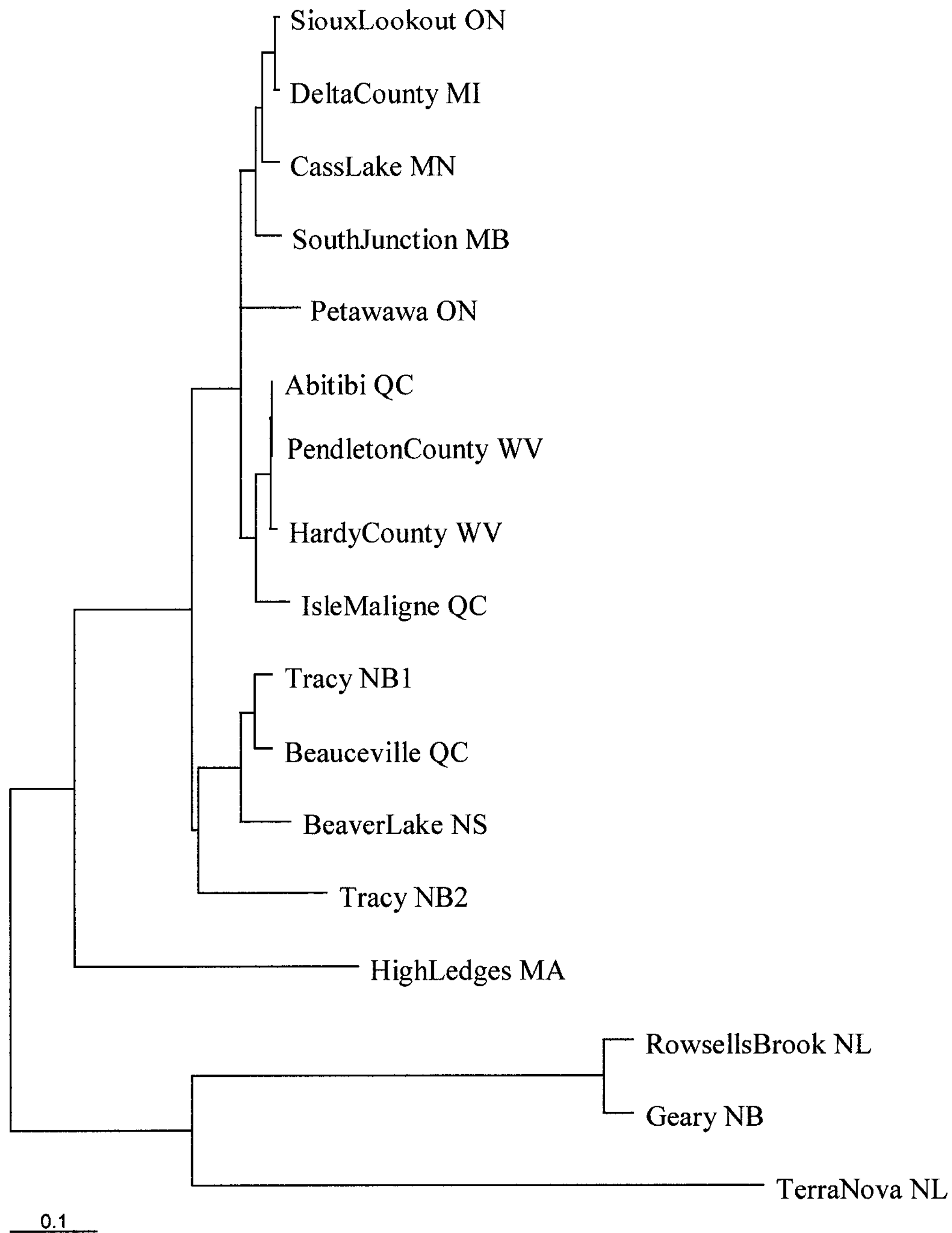


Figure 3. UPGMA dendogram depicting delta μ^2 genetic distances (Goldstein et al., 1995) between seventeen *Pinus resinosa* populations based on five polymorphic nuclear microsatellite loci.

Figure abbreviation: ON=Ontario, MI=Michigan, MN=Minnesota, MB=Manitoba, QC=Quebec, WV=West Virginia, NB=New Brunswick, NL=Newfoundland, NS=Nova Scotia, MA=Massachusetts.

4. Discussion

4.1 Characteristics of microsatellite loci in red pine

The low recovery of single locus polymorphic microsatellite loci in this study is typical of organisms with large genome sizes such as conifers (Fischer and Bachman, 1998), and is consistent with a number of studies (Smith and Devey, 1994; Kostia et al., 1995; Pfeiffer et al., 1997; Fisher et al., 1998; Berube et al., 2003). The reason for this could be a higher proportion of duplicated regions in a larger genome or an increased probability of non-specific binding sites for the designed primers (Garner, 2002). In the first case, the probability of amplifying multiple loci would be increased rendering scoring difficult, while in the second case the probability of isolating single genomic regions would be decreased.

The greater number of TC/AG repeats relative to AC/TG repeats found in red pine is consistent with findings for several tree species, including *Populus tremuloides* (Dayanandan et al., 1998), *Pinus radiata* (Fisher et al., 1998), and *Picea glauca* (Rajora et al., 2001). An exception is white pine (*Pinus strobus*) in which AC/TG repeats were found to be more abundant than TC/AG (Echt et al., 1996). Although microsatellites containing longer repeats are expected to demonstrate higher levels of polymorphism (Weber, 1990), the greatest polymorphism in red pine was detected with locus PRE16, $(AT)_{22}$, with 15 alleles, while the longest repeat $(AG)_{29}$ had the least polymorphism with only three alleles at locus PRE24, and two alleles at locus PRE24B. This pattern may not be unusual for forest trees, however, as other studies have indicated that longer repeats do

not necessarily show higher levels of polymorphism (Echt et al., 1996; Dayanandan et al., 1998). The microsatellite locus based on the shortest cloned repeat demonstrating polymorphism in red pine was $(AT)_{21}$, while all primer pairs based on clones with repeats shorter than 21 were monomorphic.

4.2 Patterns of genetic diversity

The observed heterozygosity (H_O) values of 0.067-0.317 with a mean of 0.185 reported here for red pine are lower than those reported for other pine species using microsatellite markers. For example, H_O values for *P. radiata* ranged from 0-0.85 with a mean of 0.625 (Smith and Devey, 1994), and those for *P. strobus* ranged from 0.125 to 0.812 with a mean of 0.515 (Echt et al., 1996). The observed heterozygosity values for red pine are much lower than expected heterozygosity (H_E) values (Table 4), and all populations as well as four of the five loci showed significant heterozygote deficiency. Despite the lower levels of observed heterozygosity and relatively few polymorphic loci, the average of 9 alleles per locus across the five polymorphic loci characterized for red pine was not lower than the 6 alleles per locus observed for *P. radiata* (Smith and Devey, 1994), 5.4 alleles per locus for *P. strobus* (Echt et al., 1996), and 6.7 alleles per locus for *P. sylvestris* (Soranzo et al., 1998). A similar finding of relatively high genetic diversity and low heterozygosity has been reported in *Elymus alaskanus* (Sun and Salomon, 2003), and the authors propose high levels of inbreeding in this species as the primary reason for this pattern.

Genetic variation across red pine populations was highly variable with expected heterozygosity values ranging from 0.489 in Beauceville (Quebec) to as low as 0.014 in Hardy County (West Virginia). Delta County (Michigan) showed the greatest allelic diversity with 4.6 alleles per locus, while Hardy County (West Virginia) was the least diverse with 1.2 alleles per locus. Unique or private alleles can be seen as a measure of genetic distinctiveness and were detected in six of the populations sampled (three for Delta County, two for Abitibi, two for South Junction, and one each in Petawawa, Terra Nova, and Rowsells Brook). This demonstrates that private alleles are not concentrated in one area but are distributed across several populations spanning both western and eastern extremes of red pine's range. Interestingly both Newfoundland populations had private alleles despite the relatively low allelic diversity in these two populations (2.4-2.6 alleles per locus). The lower number of alleles detected in the Terra Nova population however, may simply be due to the lower number of individuals examined. The distribution of alleles, the presence of unique alleles in several populations examined, and the distribution of genetic variation in red pine demonstrates the importance of protecting as many distinct populations as possible throughout its range in addition to those that demonstrate higher levels of genetic variation.

4.3 Inbreeding and evolution of mating system

F_{IS} and F_{IT} values measure the deviation from Hardy-Weinberg equilibrium within subpopulations and in the total population, respectively. Overall F_{IS} and F_{IT} values of 0.504 and 0.643, respectively, were significantly higher than zero at 95% confidence

levels suggesting a departure from Hardy-Weinberg equilibrium, with a heterozygote deficiency at both levels. In addition, F_{IS} values were significantly different from 0 in all populations except for the two Newfoundland populations and selfing rates across all populations (assuming inbreeding equilibrium) were high, ranging from 0.555 in the Beaver Lake (Nova Scotia) population to 1 in the Hardy County (West Virginia) population.

Increases in levels of homozygosity and the subsequent departure from Hardy-Weinberg equilibrium demonstrated here in red pine can result from a mating system with a high level of inbreeding by self-fertilization (Fowler, 1964) or mating between closely related individuals. The results reported here further support a self-compatible mating system in red pine. Other factors such as the presence of null alleles due to sequence variation at the primer sites that prevent amplification during PCR, and selection against heterozygotes, can also result in heterozygote deficiency. However, the pattern detected here was consistent across four of the five loci, and it is unlikely that the high levels of observed homozygosity are due to null alleles at all four loci.

Inbreeding is a form of non-random mating where the frequency of self-fertilization or mating between close relatives is greater than expected under Hardy-Weinberg equilibrium, and will act to increase the frequency of homozygotes in the population. As discussed above, evidence from Hardy-Weinberg tests, F coefficents and corresponding selfing rates all indicate a strong tendency towards inbreeding in red pine populations. Loss of fitness as a result of this type of mating relative to outcrossing is termed inbreeding depression. Inbreeding depression is common among all organisms, and has been quantified by numerous studies (Lande and Schemske, 1985; Husband and

Schemske, 1996). It can act to offset the advantage conferred by self-fertilization (Fisher, 1941), and explains the evolution of cross-fertilization and maintenance of self-incompatibility in plant mating systems (Lande and Schemske, 1985; Glemin et al., 2004). Most pines are predominantly outcrossing (Hamrick et al., 1992) and evidence of inbreeding depression has been noted in most conifers (reviewed by Williams and Savolainen, 1996). The extent of inbreeding depression varies across species, among populations within species (Karkkainen et al., 1996; Hedrick et al., 1999; Krakowski et al., 2003), and may even vary temporally within the same individual (El-Kassaby et al., 1993), or be expressed at different stages of the life cycle (Husband and Schemske, 1996). It is known that considerable self-sterility exists in most pine species with red pine as the exception (reviewed in Ledig, 1998), and that in some pine species the level of inbreeding depression is high enough to ensure that no progeny produced by selfing survive (Lande et al., 1994). However, low inbreeding depression relative to other conifers may also be found in *Pinus radiata* (Wu et al., 1998). In a review by Williams and Savolainen (1996) it was found that estimates of average numbers of lethal equivalents per diploid zygote in conifers ranges from 0.1 in red pine (Fowler, 1965) to 10.8 in *Larix laricina* (Park and Fowler, 1982). Durel et al. (1996) also summarized inbreeding depression values in a number of conifer species. They found that inbreeding resulted in a reduction in height from 9% in *Pinus radiata* by age 4 (Wilcox, 1983) to 61% in *Picea abies* by age 19 (Eriksson et al., 1973), relative to outcrossed species. By age 11, *Pinus sylvestris* showed a decrease in volume of up to 66% (Lunkvist et al., 1987) as a result of inbreeding. In addition the number of cones produced for first and second generation selfed progeny was reduced by 53% and 89% respectively (Durel et al., 1996).

Lande and Schemske (1985) propose an association between the levels of inbreeding depression and the evolution of self-fertilization. The inbreeding coefficient, F_{IS}, may reflect inbreeding in a population that arises for reasons other than mating system such as a population bottleneck (Husband and Schemske, 1996). Husband and Schemske (1996) suggest that inbreeding in red pine is a consequence of reduction in population size increasing the probability of self-fertilization or mating between closely related individuals. Low levels of inbreeding will expose homozygotes (for deleterious alleles) to natural selection and hence increase the rate of purging of deleterious alleles (Lande and Schemske, 1985), a process which reduces inbreeding depression and hence increases selfing rates in a population. It should however be mentioned that some highly inbred populations still exhibit inbreeding depression (Husband and Schemske, 1996) and that purging may not occur if the genomic mutation to lethals is sufficiently high (Lande et al., 1994). Thus, there is stronger selection in favor of reducing the mutation rate in asexual or selfing organisms relative to those that are outcrossing (Drake et. al., 1998), since the cost of a high mutation rate is greater in inbreeding populations. As a consequence, highly selfing populations may evolve lower mutation rates, and if this is the case for red pine it may help to explain the lower levels of genetic variation based on previous studies relative to most other conifers. High rates of self-fertilization would also act to reduce the effective population size relative to the actual population size, hence increasing the loss of diversity during bottlenecks. Interestingly, western red cedar (*Thuja plicata*) also demonstrates high selfing rates along with low genetic diversity relative to other conifers (Glaubitz et al., 2000), which is consistent with the pattern seen in red

pine. A selfing rate of 68% has been estimated in a *T. plicata* seed orchard, one of the highest reported values for a conifer (El-Kassaby et al., 1994).

High levels of selfing in red pine may have evolved in small refugial populations, through purging of deleterious alleles and decreasing the magnitude of inbreeding depression. This in turn would further facilitate the maintenance of high selfing rates. The low levels of inbreeding depression relative to other conifers may indicate that many lethals and semi-lethals have been eliminated in red pine populations through purging.

4.4 Population differentiation and unique populations

F_{ST} is a measure of genetic differentiation among subpopulations. Differentiation between populations indicates the level of evolutionary divergence between populations and is influenced by the opposing effects of migration, that tends to homogenize populations and genetic drift or mutation, which differentiates them. The overall F_{ST} value of 0.280 and overall R_{ST} value of 0.350 indicate that on average about 31.5 % of the genetic variation can be accounted for among subpopulations. This indicates that red pine populations are highly differentiated. This is generally uncommon for conifers due to long-distance pollen flow facilitated by wind pollination. For example, allozyme studies demonstrated low differentiation among northern European Scots pine populations ($F_{ST} \leq 0.02$) (Gullberg et al., 1985), and the proportion of genetic diversity detected among populations of widespread conifer species in usually less than 10% (Ledig, 1998). The level of population differentiation determined by the F_{ST} and R_{ST} values is consistent with the high selfing rates of this species since studies show lower

among population differentiation in outcrossing plant species relative to selfers (Loveless and Hamrick, 1984; Hamrick and Godt, 1990; Awadalla and Ritland, 1997). Higher rates of self-fertilization in red pine populations, may result in the exclusion of pollen from outside sources, further promoting the greater among population differentiation relative to other conifers.

Microsatellites are biparentally inherited, codominant, highly variable markers occurring throughout the genome, containing short repeat units which vary in the number of repeats between individuals (Tautz, 1989). There has been much debate concerning the mutation model of microsatellites and hence the appropriate statistics for calculating genetic distances and population differentiation using microsatellite data. Traditionally estimators of Wright's (1951, 1965) F-statistics and Nei's (1978) standard genetic distance have been used, both of which assume an infinite alleles model (IAM) where every new mutation creates a novel allele. R_{ST} (Slatkin, 1995) an analogue to F_{ST}, and delta μ^2 (Goldstein et al., 1995), a measure of genetic distance, both assume the stepwise mutation model (SMM) and were specifically developed for use with microsatellite data. The SMM model is based on differences in allele length as opposed to differences in allele frequencies alone. However F_{ST} is predicted to perform better with smaller sample sizes or fewer loci (Gaggiotti et al., 1999). Studies where both methods are used and where similar results are obtained under both mutation models (e.g. Firestone et al., 2000; Feldheim et al., 2001) are more robust. The fact that pairwise R_{ST} and F_{ST} values, as well as the two distance values, were highly correlated in this study confirms that the genetic patterns detected are not dependent on the mutation model assumed in the analyses.

The Newfoundland populations used in this study represent the northeastern extreme of the geographical range of red pine. Newfoundland red pines have been isolated from those on the mainland by the Gulf of St. Lawrence since the last glacial retreat. Previous studies have attempted to distinguish these populations from those of the mainland using isozymes (Mosseler et al., 1991), RAPD markers (Mosseler et al., 1992), and RAPD-RFLP analysis (Deverno and Mosseler, 1997) with limited success, although some unique banding patterns were noted in Newfoundland populations (Deverno and Mosseler, 1997). The striking differences in allele frequencies and greater genetic distances noted between Newfoundland and most other populations may indicate a lack of gene flow between Newfoundland and mainland populations, possibly dating back to the last glacial retreat, as well as genetic drift experienced by the Newfoundland population since the time of its isolation. These results indicate that the Newfoundland populations are genetically distinct from most other populations studied, and that special attention should be given to the conservation of Newfoundland populations of red pine.

Within the group containing the remaining populations, High Ledges (Massachusetts) is shown as highly divergent. There is also evidence for a genetic break around the border of New Brunswick and Quebec, and the major difference between the two dendograms is the placement of Beauceville (Quebec), which lies near this border (see Figs. 2 & 3). This population most likely experiences some gene flow from populations to both the west and east of this genetic break.

4.5 Implications for glacial refugia and post-glacial colonization

Pleistocene glaciations have greatly impacted the genetic structure and ecology of North American species (Pielou, 1992) and, as with many tree species, the history, current population structure, and genetic variation of red pine are intimately linked to these events. Repeated glacial cycles would have caused severe reductions in available habitat during the Pleistocene during which time surviving populations were displaced and isolated into one or several refugia. Fossil data suggests that red pine populations persisted in refugia located in the southern Appalachian Mountains (Jackson et al., 2000). Among the red pine populations examined in this study, the area of highest diversity based on expected heterozygosity values is centered in the New Brunswick/Nova Scotia region extending into Quebec. This is consistent with evidence from cpSSR data (Walter and Epperson, 2001; Echt et al., 1998). Although areas of high genetic diversity are often an indication of the presence of a glacial refugium, this area was completely ice covered during the last glaciation and therefore cannot correspond to a refugial population. However, it is believed that some ice-free areas existed off the current northeastern seaboard during the last glacial maximum (Pielou, 1992). Walter and Epperson (2001) suggest that the greater genetic diversity in the New Brunswick red pine populations could have arisen from post-glacial colonization by individuals that persisted in two different refugia, one in the southern Appalachians and another in the northeast. This could result in an admixed population due to secondary contact, and this population(s) would be located in an area intermediate between the hypothesized refugial areas. It is also interesting that both the highest allelic diversity and greatest number of private

alleles were found in the Michigan population, a pattern consistent with the presence of a glacial refugium, although there is no direct evidence that red pine persisted in this area during the last glacial maximum. However macrofossils of other conifers (*Picea sp.*) have been found in areas south of Michigan, in both Indiana and Ohio (Jackson et al., 2000). Therefore, based on the genetic evidence presented here, the possibility exists that the current range of red pine was colonized from individuals originating from as many as three glacial refugia.

Higher genetic variation is often found in southern populations relative to the northern populations, as demonstrated for several European taxa that survived in much reduced refugial populations during Pleistocene glaciations (Hewitt, 2001). A similar trend has also been demonstrated for plants in western (Soltis et al., 1997), and eastern (Lewis and Crawford, 1995) North America. This pattern may hold when the majority of genetic diversity is maintained in southern refugial populations during glaciations resulting in lower levels of diversity in the north as it is colonized. In contrast, for red pine the lowest diversity was found in the three most southern populations, a pattern inconsistent with post-glacial colonization from a single southern glacial refugium. This further supports the existence of a second refugium in the north, most likely off the current northeastern seaboard. It should also be noted however, that genetic drift and lack of gene flow might have contributed to the reduced genetic diversity in the smaller and more isolated populations in West Virginia. It has been shown that genetic drift may be 2 to 30 times greater near the edge of a species' range (Vucetich and Waite, 2003) due to the smaller effective population sizes. The lack of reduced genetic diversity in glaciated versus ice-free areas in the current range of *Trillium grandiflorum* has been explained by

the larger effective population sizes in the north (Griffin and Barrett, 2004). Another factor contributing to lower diversity in the south may be the events of the Hypsithermal period that began around 8000 years ago and lasted several thousand years. At this time warmer conditions existed than those of the present day and the tree line was higher. This may have given an advantage to species adapted to warmer conditions at the expense of cold-tolerant species surviving in southern refugia (Pielou, 1992) such as red pine. This could help to explain lower genetic diversity of southern red pine populations as they may have experienced additional bottlenecks during this period.

Despite this, there is no reason to discount the collective genetic evidence pointing to the persistence of a second red pine refugium off the current northeastern seaboard in ice-free areas. In addition, the high concentration of endemics and fossils of conifer consuming mastodons found in this region indicate that plant communities may have persisted in the region during the last Pleistocene glaciation (Pielou, 1992).

4.6 Implications for conservation

In setting conservation priorities, protecting genetic diversity and hence the evolutionary potential of a species should be considered. Loss of genetic diversity may affect populations by decreasing their ability to adapt to changing environments and increase the probability of extinction. It appears that the population genetic structure of red pine is quite complex and that areas of high diversity and/or genetic distinctiveness can be found throughout the range of this species. The presence of private alleles and distinctive microsatellite frequencies resulting from past and continuing differentiation of

red pine populations indicate the evolutionary significance of these populations. It is therefore advisable that representative populations of this species across its range be considered for conservation measures.

5. Conclusions

In summary, these polymorphic microsatellite markers have been highly valuable in assessing genetic diversity and identifying genetically unique populations of red pine, one of the ecologically and commercially important, but genetically depauperate tree species in North America. The high levels of population differentiation maintained in red pine may be due to high levels of inbreeding that have evolved in response to small effective population sizes during population bottlenecks caused by Pleistocene glaciations. The geographical distribution and genetic structuring of red pine populations strongly suggests dispersal from at least two different refugia with an intermediate admixture zone, as well as the possibility of a third more western refugium. These results demonstrate that the use of a small number of polymorphic loci can successfully determine the geographical patterns of red pine, and that these markers have been able to identify levels of genetic diversity previously undetected by other marker systems. The information obtained in this study will be valuable not only for the theoretical advancement of our knowledge on the plant genetic diversity in relation to Pleistocene refugia, but also valuable for implementing genetically sound conservation and management programs for red pine.

Literature Cited

ALLENDORF, F.W., KNUDSEN, K.L., AND G..M. BLAKE. 1982. Frequencies of null alleles at enzyme loci in natural populations of ponderosa and red pine. *Genetics* 100: 497-504.

AL-RABAB'AH, M.A., AND C.G. WILLIAMS. 2002. Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. *Forest Ecology and Management* 163: 263-271.

AMOS W., SAWCER S.J., FEAKES R.W., AND D.C. RUBINSZTEIN. 1996. Microsatellites show mutational bias and heterozygote instability. *Nature Genetics* 13: 390-391.

AWADALLA, P., AND K. RITLAND. 1997. Microsatellite variation and evolution in the *Mimulus guttatus* species complex with contrasting mating systems. *Molecular Biology and Evolution* 14: 1023-1034.

BALLOUX, F., AND H. BRUNNER. 2000. Microsatellites can be misleading: An empirical and simulation study. *Evolution* 54: 1414-1422.

BALLOUX F., AND N. LUGON-MOULIN. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11: 155-165.

BECKMANN J.S., AND J.L. WEBER. 1992. Survey of human and rat microsatellites. *Genomics* 12: 627-631.

BERUBE, Y., RITLAND, C., AND K. RITLAND. 2003. Isolation, characterization, and cross-species utility of microsatellites in yellow cedar (*Chamaecyparis nootkatensis*). *Genome* 46: 353-361.

BIDLACK, A.L., AND J. COOK. 2002. A nuclear perspective on endemism in northern flying squirrels (*Glaucomys sabrinus*) of the Alexander Archipelago, Alaska. *Conservation Genetics* 3: 247-259.

BOHONAK, A.J. 2002. IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity* 93: 153-154.

CALLEN, D.F., THOMPSON, A.D., SHEN, Y., PHILLIPS, H.A., RICHARDS, R.I., MULLEY, J.C., AND G.R. SUTHERLAND. 1993. Incidence and origin of 'null' alleles in the $(AC)_n$ microsatellite marker. *American Journal of Human Genetics* 52: 922-927.

CHAKRABORTY R., KIMMEL M., STIVERS D.N., DAVISON J., AND R. DEKA. 1997. Relative mutation rates at di-, tri-, and teranucleotide microsatellite loci. *Proceedings of the National Academy of Science USA* 94: 1041-1046.

CHERRY, M. 2001. Options for allocating afforestation stock in Ontario with anticipated climate change. Ontario Ministry of Natural Resources, Ontario Forest Research Institute, Forest research paper no. 148: p.29.

COLSON, I., AND D.B. GOLDSTEIN. 1999. Evidence for complex mutations at microsatellite loci in *Drosophila. Genetics* 152: 617-627.

DAYANANDAN, S., BAWA, K.S., AND R. KESSELI. 1997. Conservation of microsatellites among tropical trees (Leguminosae). *American Journal of Botany* 84: 1658-1663.

DAYANANDAN, S., RAJORA, O.P., AND K.S. BAWA. 1998. Isolation and characterization of microsatellites in trembling aspen (*Populus tremuloides*). *Theoretical and Applied Genetics* 96: 950-956.

DEVERNO, L.L., AND A. MOSSELER 1997. Genetic variation in red pine (*Pinus resinosa*) revealed by RAPD and RAPD-RFLP analysis. *Canadian Journal of Forest Research* 27: 1316-1320.

DIRIENZO, A., PETERSON, A.C., GARZA, J.C., VALDES, A.M., SLATKIN, M. AND N.B. FREIMER. 1994. Mutational -processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Science USA* 91: 3166-3170.

DRAKE, J.W., CHARLESWORTH, B., CHARLESWORTH, D., AND J.F. CROW. 1998. Rates of spontaneous mutation. *Genetics* 148: 1667-1686.

DUREL C.E., BERTIN P., AND A. KREMER. 1996. Relationship between inbreeding depression and inbreeding coefficient in maritime pine (*Pinus pinaster*). *Theoretical and Applied Genetics* 92: 347-356.

ECHT, C.S., MAY-MARQUARDT, P., HSEIH, M., AND R. ZAHORCHAK. 1996. Characterization of microsatellite markers in eastern white pine. *Genome* 39: 1102-1108.

ECHT C.S., DEVERNO L.L., ANZIDEI M., AND G.G. VENDRAMIN. 1998. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology* 7: 307-316.

ECHT, C.S., VENDRAMIN, G.G., NELSON, C.D., AND P. MARQUARDT. 1999. Microsatellite DNA as shared genetic markers among conifer species. *Canadian Journal of Forest Research* 29: 365-371.

EDWARDS A., CIVITELLO A., HAMMOND H.A., AND C.T. CASKEY. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* 49: 749-756.

EL-KASSABY, Y.A., MEAGHER, M.D, AND R. DAVIDSON. 1993. Temporal variation in the outcrossing rate in a natural stand of western white pine. *Silvae Genetica* 42: 131-135.

EL-KASSABY, Y.A., RUSSEL, J., AND K. RITLAND. 1994. Mixed mating in an experimental population of western red cedar, *Thuja plicata*. *Journal of Heredity* 85: 227-231.

ELLEGREN, H., PRIMMER, C.R., AND B.C. SHELDON. 1995. Microsatellite evolution: directionality or bias in locus selection. *Nature Genetics* 11: 360-362.

ELLEGREN, H. .2000a. Heterogenous mutation processes in human microsatellite DNA sequences. *Nature Genetics* 24: 400-402.

ELLEGREN, H. .2000b. Microsatellite mutations in the germline; implications for evolutionary inference. *Trends in Genetics* 16: 551-558.

ERIKSSON, G., SCHELANDER, B., AND V. ÅKEBRAND. 1973. Inbreeding depression in an old experimental plantation of *Picea abies*. *Heriditas* 73: 185-194.

ESTOUP, A., AND J.M. CORNUET. 1999. Microsatellite evolution: inferences from population data. *In*: D.B. Goldstein, and C. Schlötterer [eds]. Microsatellite evolution and applications. Oxford University Press, New York, USA, p. 49-65.

FELDHEIM. K.A., GRUBER, S.H., AND M.V. ASHLEY. 2001. Population genetic structure of the lemon shark (*Negaprion brevirostris*) in the western Atlantic: DNA microsatellite variation. *Molecular Ecology* 10: 295-303.

FELDMAN M., BERGMAN A., POLLOCK D.D., AND D.B. GOLDSTEIN. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* 145: 207-216.

FIRESTONE, K.B., HOULDEN, B.A., SHERWIN, W.B, AND E. GEFFEN. 2000. Variability and differentiation of microsatellites in the genus *Dasyrus* and conservation implications for the large Australian carnivorous marsupials. *Conservation Genetics* 1: 115-133.

FISCHER, D., AND K. BACHMANN. 1998. Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *Biotechniques* 24: 796–798.

FISHER, R.A. 1941. Average excess and average effect of a gene substitution. *Annals of Eugenetics* 11: 53-63.

FISHER, P.J., RICHARDSON, T.E., AND R.C. GARDNER. 1998. Characteristics of single and multi-copy microsatellites from *Pinus radiata. Theoretical and Applied Genetics* 96: 969-979.

FOWLER, D.P. 1964. Effects of inbreeding in red pine, *Pinus resinosa* Ait. *Silvae Genetica* 13: 170-177.

FOWLER, D.P. 1965. Effects of inbreeding in red pine, *Pinus resinosa* Ait. *Silvae Genetica* 12: 12-23.

FOWLER, D.P., AND D.T. LESTER. 1970. Genetics of red pine. USDA Forest Service Research Paper WO-8.

FOWLER, D.P., AND R.W. MORRIS. 1977. Genetic diversity in red pine: evidence for low genetic heterozygosity. *Canadian Journal of Forest Research* 7: 343-347.

GAGGIOTTI, O.E, LANGE, 0., RASSMAN, K., AND C. GLIDDON. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, 8: 1513-1520.

GARNER, T.W.J. 2002. Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome* 45: 212-215.

GLAUBITZ J.C., EL-KASSABY Y.A., AND J.E. CARLSON. 2000. Nuclear restriction fragment length polymorphism analysis of genetic diversity in western red cedar. *Canadian Journal of Forest Research* 30: 379-389.

GLEMIN, S., BATAILLON, T., RONFORT, J., MIGNOT, A., AND I. OLIVIERI. 2001. Inbreeding depression in small populations of self-incompatible plants. *Genetics* 159: 1217-1229.

GOLDSTEIN, D.B., LINARES, A.R., CAVALLI-SFORZA, L.L., AND M.W. FELDMAN.1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics*. 139:463–471.

GOODMAN, S. J. 1997. R_{ST} Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 6: 881-885.

GRIFFIN, S.R., AND S.C.H. BARRETT. 2004. Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: inferences from phylogeography. *American Journal of Botany* 91: 465-473.

GULLBERG, U., YAZANI, R., RUDIN, D., AND N. RYMAN. 1985. Allozyme variation in Scots pine (*Pinus sylvestris* L.) in Sweden. *Silvae Genetica* 34: 193-200.

HALLING, K.C., HARPER, J., MOSKALUK, C.A., THIBODEAU, S.N., PETRONI, G.R., YUSTEIN, A.S., TOSI, P., MINACCI, C., ROVIELLO, F., PIVA, P., HAMILTON, S.R., JACKSON, C.E., AND S.M. POWELL. 1999. Origin of Microsatellite Instability in Gastric Cancer. *American Journal of Pathology* 155: 205 - 211.

HAMRICK J.L., AND M.J.W.GODT. 1990. Allozyme diversity in plant species. *In* A.H.D. Brown, M.T. Clegg, A.L. Kahler, and B.S. Weir c[eds.], Plant population genetics, breeding and genetic resources. Sinauer Associates Inc, Sunderland, Massachusetts, USA, p. 43-63.

HAMRICK, J.L., GODT, M.J.W., AND S.L. SHERMANBROYLES. 1992. Factors affecting levels of genetic diversity in woody plant species. *New Forests* 6: 96-124.

HANCOCK, J.M. 1995. The contribution of slippage-like processes to genome evolution. *Journal of Molecular Evolution* 41: 1038-1047.

HARR, B., AND C. SCHLOTTERER. 2000. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* 155: 1213-1220.

HAWLEY, G.J., AND D.H. DEHAYES. 1994. Genetic diversity and population structure of red spruce (*Picea rubens*). *Canadian Journal Botany* 72: 1778-1786.

HEDRICK, P.W., SAVOLAINEN, O., AND K. KÄRKKÄINEN. 1999. Factors influencing the extent of inbreeding depression: an example from Scots pine *Heredity* 82: 441-450.

HEDRICK, P.W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53: 313-318.

HEDRICK, P.W. 2000 Genetics of Populations (Second Edition), 185-194. Jones and Barlett, Boston, Massachusetts, USA. p. 179-193.

HEWITT, G.M. 2001. Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology* 10: 537-549.

HUSBAND, B. C., AND D. W. SCHEMSKE. 1996. Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* 50: 54-70.

JACKSON, S.T., WEBB, R.S., ANDERSON, K.H., OVERPECK, J.T., WEBB III, T., WILLIAMS, J.W., AND B.C.S. HANSEN. 2000. Vegetation and environment in Eastern North America during the last glacial maximum. *Quaternary Science reviews* 19: 489-508.

JEFFREYS, A. J., WILSON, V., AND S.L. THEIN. 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* 314: 67-73.

JEFFREYS, A.J., TAMAKI, K., MACLEOD, A., MONCKTON, D.G., NEIL, D.L. AND J.A.L. ARMOUR.1994. Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics* 6: 136-145.

JONES, A.G., ROSENQVIST, G.,BERGLUND, A., AND J.C. AVISE. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics* 152: 1057-1063.

KALINOWSKI, T.S. 2002. Evolutionary and statistical properties of three genetic distances. *Molecular Ecology* 11: 1263-1273.

KÄRKKÄINEN, K., KOSKI, K., AND O. SAVOLAINEN. 1996. Geographical variation in the inbreeding depression of Scots pine. *Evolution* 50: 111-119.

KIMURA, M. AND T. OHTA. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Science USA* 75: 2868-2872.

KOSTIA, S., VARVIO, S.-L., VAKKARI, P., AND P. PULKKINEN. 1995. Microsatellite sequences in *Pinus sylvestris*. *Genome* 38: 1244-1248.

KOVTUN, I.V., AND C.T. MCMURRAY. 2001 Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genetics*. 27: 407–411.

KRAKOWSKI, J., AITKEN, S.N., AND Y.A EL-KASSABY. 2003. Inbreeding and conservation genetics in whitebark pine. *Conservation Genetics* 4: 581-593.

KRUGLYAK, S.,DURRETT, R.T, SCHUG, M.D., AND C. F. AQUADRO. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceeding of the National Academy Science USA* 95: 10774-10778.

LAGERCRANTZ, U., ELLEGREN, H., AND L. ANDERSSON. 1993. The abundance of various polymorphic microsatellite motif differs between plants and vertebrates. *Nucleic Acids Research* 21: 1111-1115.

LANDE , R., SCHEMSKE, D.W., AND S.T. SCHULTZ. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. *Evolution* 48: 965-978.

LANDE, R., AND D.W. SCHEMSKE. 1985. The evolution of self-fertilization and Inbreeding depression in plants. I. Genetic models. *Evolution* 39:24-40.

LEDIG, F.T., AND M.T. CONKLE. 1983. Gene diversity and genetic structure in a narrow endemic, Torrey pine (*Pinus torreyana* Parry ex Carr). *Evolution* 37: 79-85.

LEDIG, F.T. 1998. Genetic variation in *Pinus*. *In* D.M. Richardson [ed.], Ecology and Biogeography of *Pinus*. Cambridge University Press, Cambridge, UK, p. 251-273.

LEVINSON, G., AND G.A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203-221.

LEWIS, P.O., AND D. J. CRAWFORD. 1995. Pleistocene refugium endemics exhibit greater allozymic diversity than widespread congeners in the genus *Polygonella* (Polygonaceae). *American Journal of Botany* 82: 141-149.

LEWIS, P.O., AND D. ZAYKIN. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html.

LI, Y.-C., KOROL, A.B. FAHIMA, T., BEILES, A., AND E. NEVO. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms. *Molecular Ecology* 11: 2453-2465.

LOVELESS, M.D., AND J.L. HAMRICK. 1984. Ecological determinants of genetic structure of plant populations. *Annual Review of Ecology and Systsematics* 15: 65-95.

LUNDKVIST, K., ERIKSSON, G., NORREL, L., AND I. EKBERG. 1987. Inbreeding depression in two field trials of young *Pinus sylvestris* (L.). *Scandinavian Journal of Forest Research* 2: 281-290.

MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27:209-220.

MITTON, J.B. 1994. Molecular approaches to population biology. *Annual Review of Ecology and Systematics* 25: 45-69.

MORGANTE, M., HANAFEY, M., AND W. POWELL. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* 30: 174-200.

MOSSELER A., INNES, D.J., AND B.A. ROBERTS. 1991. Lack of allozyme variation found in disjunct Newfoundland populations of red pine (*Pinus resinosa*). *Canadian Journal of Forest Research* 21: 525-528.

MOSSELER, A., EGGER, K.N., AND G.A. HUGHES. 1992. Low levels of genetic diversity in red pine confirmed by random amplified polymorphic DNA markers. *Canadian Journal of Forest Research* 22: 1332-1337.

NAUTA, M.J., AND F.J. WEISSING. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* 143: 1021-1032.

NEI, M. 1972. Genetic distance between populations. *American Naturalist* 106: 283-292.

NEI, M., AND A.K. ROYCHOUDHURY. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* 76: 379-390.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 147: 1943-1957.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEIGEL, J. 2002. Is Fst obsolete? *Conservation Genetics* 3: 167-173.

PAGE, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357-358.

PARK, Y.S., AND D.F. FOWLER. 1982. Effects of inbreeding and genetic variation in a natural population of tamarack (*Larix laricina* (Du Roi) K. Koch) in eastern Canada. *Silvae Genetica*: 31:21-26.

PAETKAU, D., WAITS, L.P., CLARKSON, P.L., CRAIGHEAD, L., AND C. STROBECK. 1997. An empirical evaluation of genetic distance statistics using microsatellites data from bear (Ursidae) populations. *Genetics* 147: 1943-1957.

PEMBERTON, J.M., SLATE, J., BANCROFT, D.R., AND J.A. BARRETT. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology* 4: 249-252.

PFEIFFER, A., OLIVIER, A.M., AND M. MORGANTE. 1997. Identification and characterization of microsatellites in Norway Spruce (*Picea Abies* (L.) Karst.) *Genome* 40: 411-419.

PIELOU, E.C. 1992. After the ice age: the return of life to glaciated North America. University of Chicago Press, Chicago, Illinois, USA.

PRASAD, A.M., AND L.R. IVERSON. 1999. A climate change atlas for 80 forest tree species of the eastern united states. http://www.fs.fed.us/ne/delaware/atlas/index.html, Northeastern Research Station, USDA Forest Service, Delaware, Ohio.

PRICE R.A., LISTON A., AND STRAUSS S.H. 1998. Phylogeny and systematics of *Pinus*. *In*: D.M. Richardson [ed] Ecology and Biogeography of *Pinus.* Cambridge University Press, Cambridge, UK, p 49-68.

PRIMMER C.R., ELLEGREN H., SAINO, N., AND A.P. MØLLER. 1996. Directional evolution in germline microsatellite mutations. *Nature Genetics* 13: 391-393.

PRIMMER C.R., SAINO N., MOLLER A.P., AND H. ELLEGREN. 1998. Unraveling the processes of microsatellite evolution through analysis of germline mutations in barn swallows *Hirundo rustica*. *Molecular Biology and Evolution* 15: 1047-1054.

RAJORA, O.P., RAHMAN, M.H., AND S. DAYANANDAN. 2001. Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulness in other spruce species. *Molecular General Genetics* 264: 871-882.

RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86:248-249.

ROUSSET, F., AND M.RAYMOND. 1995. Testing heterozygote excess and deficiency. *Genetics* 140: 1413-1419.

ROUSSET, F .1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* 142: 1357-1362.

ROUSSET, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219-1228.

RUDOLF, R.O. 1990. Pinus resinosa Ait. Red pine. In Burns, R.M. and B.H. Honkala (Eds.) Silvics of North America, Vol. 1, Conifers. Washington DC: U.S.D.A. Forest Service Agriculture Handbook 654: 442-455.

SCHEMSKE, D.W., AND R. LANDE. 1985. The evolution of self-fertilization and inbreeding depression in plants. II. Empirical observations. *Evolution* 39:41-52.

SCHLOTTERER, C., RITTER, R., HARR, B., AND G. BREM. 1998. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution* 15:1269–1274.

SCHUG, M.D., MACKAY, T.F.C., AND C.F. AQUADRO. 1997. Low mutation rates of microsatellite loci in *Drosophila melanogaster. Nature Genetics* 15: 99-102.

SCHUG, M.D., WETTERSTRAND, K.A., GAUDETTE, M.S., LIM, R.H., HUTTER, C.M., AND C.F. AQUADRO. 1998. The distribution and frequency of microsatellite loci in *Drosophila melanogaster. Molecular Ecology* 7: 57–70.

SIMON, J.-P., BERGERON, Y., AND D. GAGNON. 1986. Isozyme uniformity in populations of red pine (*Pinus resinosa*) in the Abitibi region, Quebec. *Canadian Journal of Forest Research* 16: 1133-1135.

SLATKIN, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47: 264-279.

SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

SMITH, G.P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* 191: 528-535.

SMITH, D.N., AND M.E. DEVEY. 1994. Occurrence and inheritance of microsatellites in *Pinus radiata. Genome* 37: 977-983.

SOKAL, R.R., AND P.H.A. SNEATH. 1963. Principles of Numerical Taxonomy. W.H. Freeman, San Francisco.

SOLTIS, D.E., GITZENDANNER, M.A., STRENGE, D.D., AND P.S. SOLTIS .1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution* 206: 353-373.

SORANZO, N., PROVAN, J., AND W. POWELL. 1998. Characterization of microsatellite loci in *Pinus sylvestris* L. *Molecular Ecology* 7: 1247-1263

STACY, E.A., DAYANANDAN, S., DANCIK, B.P., AND P.D. KHASA. 2001. Microsatellite DNA markers for the Sri Lankan rainforest tree species, *Shorea cordifolia* (Dipterocapaceae), and cross-species amplification in *S. megistophylla. Molecular Ecology Notes* 1: 53-54.

STALLINGS, R., FORD, A., NELSON, D., TORNEY, D., HILEBRAND, C. AND R. MOYZIS. 1991. Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* 10: 807-815.

SUN, G.-L., AND B. SALOMON. 2003. Microstaellite variability and heterozygote deficiency in the arctic-alpine Alaskan wheatgrass (*Elymus alaskanus*) complex. *Genome* 46: 729-737.

TAUTZ, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463–6471.

THOMAS, B.R., MACDONALD, S.E., HICKS, M., ADAMS, D.L., AND R.B. HODGETTS. 1999. Effects of reforestation methods on genetic diversity of lodgepole pine: an assessment using microsatellite and randomly amplified polymorphic DNA markers. *Theoretical and Applied Genetics* 98: 793-801.

THUILLET, A.C., BRU, D., DAVID, J.L., ROUMET, P., SANTONI, S., SOURDILLE, P., AND T. BATAILLON. 2002. Direct estimation of mutation rate for ten microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. Ssp durum desf. *Molecular Biology and Evolution* 19: 122-125.

UDUPA, S.M., AND M.. BAUM. 2001. High mutation rate and mutational bias at (TAA) (n) microsatellite loci in chickpea (*Cicer arietinum* L.). *Molecular Genetics and Genomics* 265: 1097–1103.

VALDES, A.M., SLATKIN, M., AND N.B. FREINER. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revised. *Genetics* 133: 737-749.

VAN OPPEN, M.J.H., RICO, C., TURNER, G.F., AND G.M. HEWITT. 2000. Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in lake Malawi cichlids. *Molecular Biology and Evolution* 17: 489-498.

VIGOUROUX, Y., JAQUETH, J.S., MATSUOKA, Y., SMITH, O.S., BEAVIS, W.D., SMITH, J.S., AND J. DOEBLEY. 2002. Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution* 19:1251–60.

VIGOUROUX, Y., MATSUOKA, Y, AND J. DOEBLEY. 2003. Directional evolution for microsatellite size in maize. *Molecular Biology and Evolution* 20: 1480-1483.

VUCETICH, J.A., AND T.A. WAITE. 2003. Spatial patterns of demography and genetic processes across the species' range: Null hypotheses for landscape conservation genetics. *Conservation Genetics* 4: 639-645.

WALTER, R., AND K. EPPERSON. 2001. Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology* 10: 103-111.

WEBER, J.L.1990. Informativeness of human $(dC\text{-}dA)_n$: $(dG\text{-}dT)_n$ polymorphisms. *Genomics* 7: 524-530.

WEBER, J.L. AND C. WONG. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics* 2: 1123-1128.

WEIR, B.S., AND C.C. COCKERHAM. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

WEIR, B.S. 1996. Genetic Data Analysis II. Sinauer, Sunderland MA.

WILCOX, M.D. 1983. Inbreeding depression and genetic variances estimated from self- and cross-pollinated families of *Pinus radiata*. *Silvae Genetica*. 32: 89–96.

WILLIAMS, C.G., AND O. SALVOLAINEN. 1996. Inbreeding depression in conifers: implications for breeding strategy. *Forest Science* 42: 102-117.

WRIGHT, S. 1943. Isolation by distance. *Genetics* 28: 114–138.

WRIGHT, S. 1951. The genetical structure of populations. *Annals of Eugenetic* 15: 323–354.

WRIGHT, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19: 395-420.

WU, H.X., MATHESON, A.C., AND D. SPENCER. 1998. Inbreeding in *Pinus radiata*. The effect of inbreeding on growth, survival and variance. *Theoretical and Applied Genetics* 97: 1256-1268.

YLIMARTIMO, A., LAFLAMME, G., SIMARD, M., AND D. RIOUX. 1997. Ultrastructure and cytochemistry of early stages of colonization by *Gremmeniella abietina* in *Pinus resinosa* seedlings. *Canadian Journal of Botany* 75: 1119-1132.

ZANE, L., BARGELLONI, L., AND T. PATARNELLO. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1-16.

Appendix 1: Protocols

Protocol 1: DNA Extraction Procedure.

1) Obtain clean and preferably fresh needle samples (1 to 2 needles or 0.75 to 1.5 g). Place in a labeled plastic bag with 2 to 3 ml Extraction Buffer. Crush with rollers. Transfer extract into a microcentrifuge tube, label, and place on ice.
2) Centrifuge the sample(s) at 6000 rpm (3000 G) for 5 minutes at room temperature. Decant the supernatant. Pellet(s) may be frozen at –80 °C for future extraction, if necessary.
3) Re-suspend the pellet with 600 µl of Extraction/Nucleic Lysis (1:1) Buffer. Add RNAse, if required. Add 120 µl of 5% Sarkosyl solution. Mix with a vortex mixer.
4) Heat to 65 °C for a minimum of 45 minutes mixing the tube(s) about halfway through the heating procedure with a vortex mixer.
5) In the fume-hood, add 600 µl of Chloroform/Isoamyl Alcohol (24:1) mixture. Invert tube(s) to form an emulsion (gently). Centrifuge at 7000 rpm (5000 G) for 5 minutes. Transfer upper phase to a clean microcentrifuge tube containing 600 µl of Isopropanol. Gently invert tube(s) and allow the DNA to precipitate out of solution at room temperature for a minimum of 30 minutes.
6) Centrifuge tube(s) at 14000 rpm (20000 G) for 10 minutes. Decant the supernatant.
7) Wash the pellet(s) with 70% ethanol. Centrifuge at 14000 rpm (20000 G) for 10 minutes. Decant the supernatant and dry overnight.
8) Dissolve dried pellet (s) in 150 µl of TE Buffer.

Extraction Buffer (500 ml):

Sorbitol (MW 182.2)	31.9 g
Sodium bisulfite	1.9 g
1 M Tris (pH 7.5)	50 ml
0.5 M EDTA	5 ml

Add distilled water to volume.

Nucleic Lysis Buffer (200ml)

NaCl	23.4 g
CTAB (MW 364.5)	4.0 g
1 M Tris (pH 7.5)	40 ml
0.5 M EDTA	20 ml

Add distilled water to volume. It may be necessary to heat this solution in order to aid in dissolving.

Protocol 2: Construction and screening of genomic library.

Samples of 50 μl of DNA, from each of six red pine individuals were pooled (for a total of 300 μl). The pooled sample was digested with the restriction enzyme *Sau*3aI (Promega, Madison, Wisconsin) in a total volume of 600 μl for 6 hours at 37 °C. The *Sau*3aI enzyme was deactivated by heating the solution for 10 minutes at 65 °C. The digested DNA was precipitated by adding 1300 μl of 95% ethanol, washed with 1000 μl of 70% ethanol, and dissolved in 30 μl of TE buffer (pH 7.4). One microgram of the M13 vector DNA (M13mp19RF1, MBI Fermentas, Burlington, Ontario) was digested with the restriction enzyme *Bam*HI for 2 hours at 37 °C and dephosphorylated by adding 0.5 μl (1unit/μl) of calf intestinal alkaline phosphatase (MBI Fermentas, Burlington, Ontario) and incubating for 1 hour at 37 °C. The phosphatase was deactivated by adding 1.25 μl of 0.1M EDTA and heating to 75 °C for 10 minutes.

Both digested DNA samples were electrophoresed on a 1% agarose gel. Red pine DNA fragments between 500 and 1000 bp were excised from the gel along with the M13 DNA in its entirety. The red pine and the M13 DNA were then purified with QIAquick gel extraction kit (Qiagen, Mississauga, Ontario).

The ligation of red pine and the vector DNA was performed in a total volume of 20 μl, with 4 μl of each of the gel-purified DNA samples (approximately 80 ng of each DNA) and 1 μl of T4 DNA ligase (Promega, Madison, Wisconsin), at 16 °C for 16 hours. Ligated DNA was transfected into XL1 BlueMRF'- competent bacterial cells (Stratgene, California) following the manufacturer's recommendations. The transformation product was then plated on 137 mm culture plates with LB/tetracycline agar, Xgal, and IPTG. A

library of about 13000 clones with a density of about 500-1700 plaques per plate was prepared.

Plates were blotted with nylon membranes (Hybond N+, Amersham) for 1 minute. Membranes were baked at 80 °C for 2 hours and then placed in a plastic tupperware containing 150 ml of 2 x SSC and transferred to 350 ml of the same solution. Membranes were then washed with 5 x SSC at 50 °C for 1 hour and prehybridized overnight at 50 °C in a tupperware containing 250 ml of hybridization medium. Following prehybridization, the hybridization medium in the tupperware was replaced with 250 ml of fresh medium, and the labeled probe was added (see below for preparation of probe). The tupperware was then placed in a shaking incubator at 50 °C overnight (minimum of 3 hours suggested). Hybridized blots were washed with 100 ml of 0.1% SDS/6 x SSC solution twice at room temperature, and once with 1000 ml of the same solution at 50 °C for 30 minutes. The blots were rinsed with 6 x SSC and exposed overnight to autoradiographic film. Films were developed and positive clones were identified and picked using a sterile pipette tip. Clones that were well isolated were grown in 15 ml culture tubes containing 2.5 ml of LB/Tetracycline and 150 μl of overnight-grown XL1 BlueMRF' bacterial cells. Tubes were incubated at 37 °C for 6 hours with shaking. Because of the high density of plaques per plate, positive clones that were not well isolated were first replated and then picked and grown following the method described above. Once grown, a portion of each culture was transferred to 2 ml microfuge tubes and spun at 14000 rpm (20000 G) for 5 minutes. Single-stranded DNA was isolated using the Wizard M13 DNA purification system (Promega, Madison,

Wisconsin) and double-stranded DNA (for reverse sequencing) was isolated using the QIAprep Spin Miniprep Kit (Qiagen, Mississauga, Ontario).

Probe labelling:

The oligonucleotides used as probes were $(TC)_{15}$, $(AC)_{15}$, $(AT)_{15}$, $(AAT)_{15}$, $(AAAT)_{15}$, $(TG)_{15}$, $(AG)_{15}$, $(AC)_{15}$, $(AGAT)_{15}$, and $(AAAG)_{15}$. Each probe (20 pmol) was endlabeled with 1.11×10^6 Bq of γ-AT(^{32}P) (Amerisham) using T4 Polynucleotide Kinase (Fermentas, Burlington, Ontario) at 37 °C for 1 hour in a total volume of 50 µl. The enzyme was inactivated by heating at 65 °C for 20 minutes. The labeled probe was purified using Qiagen Nucleotide Removal Kit (Qiagen, Mississauga, Ontario) and denatured at 90 °C before adding to the hybridization medium.

Hybribization medium (500 ml):

20 x SSC	150 ml
100x Denhardts solution	5 ml
BSA	5 g
10% SDS	20 ml

Add distilled water to volume.

20X SSC (1 L):

NaCl	175.3 g
Sodium citrate	88.2 g

Add distilled water to volume.

Protocol 3: 1% agarose gel recipe.

For 100 ml gel:

Agarose powder	1 g
5 X TBE (Tris-borate)	20 ml
dH_2O	80 ml

Heat in microwave to dissolve (approximately 1minute). Cool to 55 °C, and add 3.3 μl of ethidium bromide (10g/L).

5 X TBE (Tris-Borate):

Tris Base (MW 121.14)	54 g
Boric Acid (MW 61.83)	27.5 g
0.5 M EDTA (pH 8.0)	20 ml

Add distilled water to a final volume of 1 L.

Protocol 4: DNA sequencing reaction recipe.

5X buffer (Applied Biosystems)	3 μl
Forward or reverse primer (3.2 pmol/μl)	1 μl
Purified DNA	10 ng/ 100 bp (14 μl max.)
Reaction mix	2 μl

Reaction mix:

ABI Big Dye^{TM} Terminator v.3.0 and 3.1 Cycle Sequencing Ready Reaction kit (diluted 1:4).

Protocol 5: Sequencing reaction thermal cycling program.

Lid preheated to 105 °C

1) Denaturation	94 °C for 10 seconds
2) Annealing	50 °C for 5 seconds
(R= 1 °C/S)	
3) Elongation	60 °C for 4 minutes
4) Go to step 1	Repeat 25 times
5) Cooling and storage	4 °C

Protocol 6: Purification of products of sequencing reaction and sequencing procedure.

1) For each 20 µl sequencing reaction product, prepare the 80 µl of mixture below:

3 M NaOAC	3 µl
95% non-denatured ethanol	62.5 µl
dH_2O	14.5 µl

2) Mix the products and the cocktail with a vortex mixer and leave the solution at room temperature for 15 minutes.
3) Centrifuge the solution at 14000 (20000 G) at room temperature for 20 minutes.
4) Aspirate the supernatant with a pipette carefully.
5) Wash the pellet with 250 µl of 70% ethanol, and centrifuge at 14000 (20000 G) for 10 minutes.
6) Aspirate the supernatant and repeat the wash
7) Dry the pellet overnight and proceed to sequencing. The dried pellet may be stored at -20˚C if necessary.
8) Dissolve the pellet in 25 µl of Hi-Di formamide (deionized formamide).
9) Denature the sample at 94 ˚C for 4 minutes and place on ice for a minimum of 5 minutes.
10) Sequence using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, California).

Protocol 7: PCR thermal cycling program.

Lid preheated to 105 °C

1) Denaturation	94 °C for 3 minutes
2) Annealing	*Optimum temperature for 30 seconds
(R= 3 °C/S)	
3) Elongation	72 °C for 45 seconds
4) Go to step 1	Repeat 30 times
5) Final elongation	72 °C for 4 minutes
6) Cooling and storage	4 °C

*Optimum temperatures for each primer pair are listed in Table 2.

Protocol 8: Genotyping procedure.

1) Mix 15 µl of Hi-Di formamide (deionized formamide) with 0.1 µl of Tamara GeneScan-500^{TM} size standard (Applied Biosystems).
2) Add 1µl of undiluted PCR products for PRE10, PRE13, and PRE16, or 1 µl of 1:10 dilution of PCR product for PRE24 to mixture.
3) Denature at 94 °C for 4 minutes and place on ice for a minimum of 5 minutes.
4) Proceed to genotype with the ABI 310 genetic analyzer and Genescan software (Applied Biosystems, Foster city, California).

Appendix 2: Hardy-Weinberg Equilibrium

The Hardy-Weinberg model describes and predicts genotype and allele frequencies in a non-evolving population. The model assumes, the population is large (i.e., there is no genetic drift), there is no gene flow between populations, mutations are negligible, individuals are mating randomly, and natural selection is not operating on the population. Given these assumptions, a population's genotype and allele frequencies will remain unchanged over successive generations, and the population is said to be in Hardy-Weinberg equilibrium (HWE). This model allows us to compare a population's actual genetic structure with what we would expect if the population is in HWE. If the observed genotype frequencies differ from those expected under HWE, one or more of the assumptions has been violated. The HWE can be represented by a simple equation.

In a diploid organism with two alleles *A* and *a* at a given locus, there are three possible genotypes: *AA*, *Aa*, and *aa*. We can use p to represent the frequency of *A* and q to represent the frequency of *a*, so that $p + q = 1$. The genotype frequencies can be written as p^2 for *AA*, q^2 for *aa*, and $2pq$ for *Aa*. The equation for genotype frequencies is

$$p^2 + 2pq + q^2 = 1$$

(This is the simplest form of the equation but it can easily be extended to cases where 3 or more alleles are present).

Appendix 3: Test Statistics

Test 1: Expected heterozygosity.

One of the most commonly used measures of genetic variation in a population is the amount of heterozygosity, also referred to as gene diversity. Assuming Hardy-Weinberg Equilibrium, heterozygosity for a particular locus with *n* alleles can be calculated as

$$H_E = 1 - \sum_{i=1}^{n} p_i^2 \qquad \text{(Nei, 1987)}$$

which is one minus the Hardy-Weinberg homozygosity, while an unbiased estimate of the expected heterozygosity at a locus for *N* individuals is

$$H_E = 2N/2N\text{-}1 \left(1 - \sum_{i=1}^{n} p_i^2\right) \qquad \text{(Nei and Roychoudhury, 1974)}$$

Test 2: Fisher's exact test for Hardy-Weinberg equilibrium and linkage disequilibrium.

<u>Null hypotheses:</u>

Hardy Weinberg Equilibrium (HWE)

Null hypothesis: Ho = random union of gametes

Linkage disequilibrium

Null hypothesis: Ho = loci associating independently

These tests can be performed using a chi-squared test of independence using the observed and expected (under HWE) genotype frequencies to construct a contingency table. However, when dealing with multiple alleles at each locus, several cells in the table may have very small numbers of observations. It is then recommended to apply Fisher's exact test as described by Weir (1996). This test is calculated by generating all tables that are more extreme than the table given by the data. The p-values of these tables are added up, including the p-value of the table itself. The p-value of the test corresponds to the probability that gametic frequencies are as extreme as those observed. Therefore, assuming the 95% rule, a p-value (or exact probability) of less than 0.05 indicates that there is significant deviation from HWE. In the same way, all possible pairs of loci can be tested for statistically significant amounts of linkage disequilibrium.

A brief example follows.

Example:

The following table represents a data set where locus A and B are characterized by two alleles each.

	A_1	A_2	*Total*
B_1	n_{11}	n_{21}	$n_{.1}$
B_2	n_{12}	n_{22}	$n_{.2}$
Total	$\mathbf{n_{1.}}$	$\mathbf{n_{2.}}$	**n**

(Where n_{11} is the number of individuals with gametes A_1 B_1, n_{12} with gametes A_1 B_2, etc)

The probability of a table containing such an array given the marginal numbers is given by:

$$Pr = n_{1.}!\ n_{2.}!\ n_{.1}!\ n_{.2}!\ /\ n_{11}!\ n_{12}!\ n_{21}!\ n_{22}!$$

All possible samples with the same marginal totals are generated and their probabilities calculated. They are then ranked in ascending order, and the probability of obtaining the observed sample by chance is the cumulative probability up to and including the observed sample.

Test 3: Score test (U test) for heterozygote deficiency.

Null hypothesis:

Ho = random union of gametes (HWE)

Alternative hypothesis:

H1 = deviation from HWE with heterozygote deficiency

Although exact tests can examine deviation from HWE, when the alternative hypothesis (H1) of interest is heterozygote excess or deficiency, more powerful tests such as the score test (U test) can be used.

The test statistic for a locus with k alleles where p_i is the frequency of the ith allele, and n_{ii} is the number of individuals homozygous for the ith allele, is given by:

$$U = \sum_{i=1}^{k} n_{ii}/p_i - N$$ (Rousset and Raymond, 1995)

Assuming the 95% rule, a corresponding p-value of less than 0.05 indicates significant heterozygote deficiency.

Test 4: Hierarchical F coefficients.

Wright's (1951, 1965) F coefficients can be used to determine levels of population genetic differentiation and examine population substructure under the assumption of the IAM.

For subdivided populations and over multiple loci the following estimates can be made:

Deviation from HWE within subpopulations:	$F_{IS} = (H_S - H_O)/ H_S$
Deviation from HWE in total population:	$F_{IT} = (H_T - H_O)/ H_T$
Genetic differentiation among subpopulations:	$F_{ST} = (H_T - H_S)/ H_T$

Where

H_O = average observed heterozygosity within a subpopulation over all loci

H_S = average expected heterozygosity within subpopulations over all loci

H_T = the average of the expected heterozygosity

Various estimators of F coefficients have been developed. One of the most commonly used is Weir and Cockerham's (1984) theta (θ).

Theta (θ) is an unbiased estimator of F_{ST} that corrects for error associated with differences in population sizes. See Weir and Cockerham (1984) for a full derivation.

F_{ST} is often related to the number of migrants per generation by the formula

$$Nm = 1 - F_{ST} / 4 F_{ST}$$

where N is the local population size and m is the migration rate among populations.

Interpreting F_{ST}:

Values can range from 0 (no genetic differentiation) to 1 (fixation of alternative alleles).

Wright's Guidelines:

0 - 0.05, little differentiation

0.05 – 0.15, moderate differentiation

0.15 – 0.25, high differentiation

> 0.25, very high differentiation

Test 5: Calculation of R_{ST}.

R_{ST} is an analogue to F_{ST} that explicitly accounts for mutation rates at microsatellite loci assuming the SMM and is defined as

$$R_{ST} = (ST - SW)/ ST \qquad \text{(Slatkin, 1995)}$$

where ST is twice the estimated variance in allele size across populations, and SW is twice the estimated variance in allele size within each population

This derivation assumes populations of equal sample size and that all loci have equivalent variances. However this is rarely the case for most natural populations. As a result unbiased estimators of Slatkin's R_{ST} have been developed. Those implemented in the RSTCALC program (Goodman, 1997) presented here.

For loci of differing variances the data set is transformed before the R_{ST} calculations are carried out, by globally standardising the data set so alleles are expressed in terms of standard deviations from the global mean rather than repeat unit number. Each locus in the transformed data set will then have a global mean allele size of zero and a standard deviation of one.

Each allele in each locus is standardised as follows:

$$Ys=(Y-GM)/\text{std dev}$$

Where Ys is the standardised value of allele Y, Y is allele (n) at locus (l), GM is the mean allele size in repeat units for locus (l) over the whole data set. The std dev is the standard deviation in allele size in terms of repeat units for locus (l) over the whole data set.

The effects of differences in sample size between populations can be taken into account by obtaining variance components using conventional statistical approaches to calculate an unbiased estimator of R_{ST}, termed Rho. Rho is defined as

$$Rho = Sb/ (Sb+SW)$$

where Sb is the component of variance that is between populations.

Test 6: Nei's standard genetic distance (D_S).

The standard genetic distance of Nei (1972, 1978) assumes the IAM and is defined as

$$DS = -\ln\left(J_{XY} / (J_{XX} J_{YY})^{1/2}\right)$$

where

$$J_{XY} = \sum_{i=1}^{m} \sum_{j=1}^{r} x_{ij}\, y_{ij} \,/\, r$$

$$J_{XX} = \sum_{i=1}^{m} \sum_{j=1}^{r} x_{ij}^{2} \,/\, r$$

$$J_{YY} = \sum_{i=1}^{m} \sum_{j=1}^{r} y_{ij}^{2} \,/\, r$$

(x_{ij} is the frequency of ith allele at the jth locus in population X, and y_{ij} is the frequency of the ith ith allele at the jth locus in population Y).

Test 7: Delta μ^2 genetic distance.

Delta μ^2 (Goldstein et al., 1995), a genetic distance that explicitly accounts for mutation rates at microsatellite loci assuming the SMM, is defined as

$$\text{Delta } \mu^2 = (\mu A - \mu B)^2$$

where μA and μB are the mean allele sizes in populations A and B respectively. The μA and μB values are calculated by finding the average allele size at each locus in each population and by then squaring the difference in mean allele size, and averaging over loci.

Test 8: Self-fertilization rates.

In cases where partial-selfing is taking place, and assuming inbreeding equilibrium, selfing rates (S) can be defined as

$$S = 2F_{IS}/ 1+F_{IS}$$

For a complete derivation see Hedrick (2000).

Test 9: Mantel test.

If the extent and/or characteristics of the geographical range of a species imposes limits on the ability for populations to exchange migrants, nonrandom spatial patterns of genetic variation may develop. When the rates of genetic exchange among populations depend on the on the distances separating them, a correlation between geographical and genetic distance, termed isolation by distance (Wright, 1943) will develop.
Significance in the isolation by distance relationship can be tested statistically using the Mantel test (Mantel, 1967). As not all points in the isolation by distance relationship are independent, the Mantel test is appropriate because it considers the unit of replication to be a population and not a pairwise contrast. The test assesses whether the pairwise genetic distance matrix is correlated with the pairwise geographic distance matrix. To test the association of one n by n matrix Y of measures of genetic distance among pairs on n populations with a second matrix X of measures of geographical distance among populations

The test statistic is given by

$$M = \sum_{i \neq j} X_{ij} Y_{ij}$$

Null Hypothesis:

Ho = Spatially random distribution of populations

Appendix 4: The Unweighted-Pair-Group Method (UPGMA)

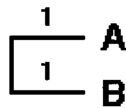
The UPGMA is the simplest method of tree construction and is an example of a distance matrix method as it uses distance information. It was originally developed by Sokal and Sneath (1963) for analyzing morphological data. The main assumption of this approach is that the rates of evolution are approximately constant among the different lineages. UPGMA employs a sequential clustering algorithm, in which local topological relationships are identified in order of similarity, and the phylogenetic tree is build in a stepwise manner. We first identify from among all the OTUs (operational taxonomic units) the two OTUs that are separated by the shortest genetic distance, and then treat these as a new single OTU. Such a OTU is referred to as a composite OTU. We then calculate estimates of genetic distance between this composite OTU and the remaining OTUs by taking the arithmetic average of their distances. This is repeated at each step until we are left with only two OTUs.

Consider the following distance matrix consisting of 4 OTUs representing 4 red pine populations:

	A	**B**	**C**	**D**
A	0			
B	2	0		
C	4	4	0	
D	6	6	6	0

The two most genetically similar populations appear to be populations A and B since they are separated by the smallest genetic distance of 2.

The subtree should therefore be constructed as follows:

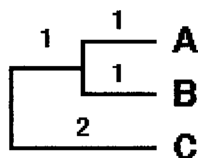


The next step is to consider cluster A and B as a single composite OTU (A,B) and calculate the distances between (A,B) and each of the remaining populations:

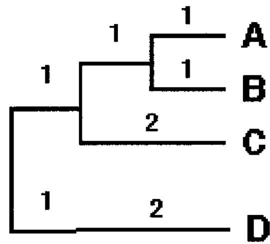
Distance (A,B) to C = (distance A-C + distance B-C) / 2 = 4

Distance (A,B) to D = (distance A-D + distance B-D) / 2 = 6

Since population C is separated from (A,B) by a smaller genetic distance than is population D, the tree should now look like this:



Finally using the genetic distance (A,B,C) to D of 6, we have:



Appendix 5: List of Abbreviations

A – alleles per locus

bp – base pairs

cpSSR – chloroplast simple sequence repeat

DNA – deoxyribonucleic acid

H_E – expected heterozygosity

H_O – observed heterozygosity

HWE – Hardy-Weinberg equilibrium

IAM – infinite alleles model

IBD – identical by descent

PCR – polymerase chain reaction

RAPD – randomly amplified polymorphic DNA

RFLP – restriction fragment length polymorphism

S – selfing rates

SMM – stepwise mutation model

SSR – simple sequence repeat

STR – short tandem repeat

TPM – two-phase model

UPGMA – unweighted-pair-group method with arithmetic mean