Evolutionary genetic studies of forest trees:

Genetic structure of the boreal forest tree *Pinus banksiana*,

and the

Molecular phylogeny of the tropical tree family Dipterocarpaceae

by

Sandra Lucy Chiovitti

A Thesis

In

the Department

of

Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

April 2006

Your file Votre référence
ISBN: 0-494-14220-0
Our file Notre référence
ISBN: 0-494-14220-0

ABSTRACT

Evolutionary genetic studies of forest trees:

Genetic structure of the boreal forest tree *Pinus banksiana*, and the molecular phylogeny of the tropical tree family Dipterocarpaceae

Sandra Lucy Chiovitti

In order to contribute to our understanding of the evolution of forest trees, the major structural life form of some of the most diverse ecosystems of the world, I conducted studies on a) the genetic structure of Jack Pine (*Pinus banksiana*) and b) the phylogeny of the tropical tree family Dipterocarpaceae.

Jack pine is one of the most widely distributed pine species in Canada and a valuable natural resource. The purpose of the present study was to examine the range-wide genetic diversity, population differentiation, inter-population migration, and genetic distances within Canada. The maritime populations were genetically distinct from the mainland populations. Genetic distance values for maritime populations ranged from 0.0504 to 0.0812 as compared to 0.0386 across all populations. The overall range-wide genetic structure of this species likely resulted from multiple refugia events.

The members of the pantropical tree family Dipterocarpaceae dominate Asian rainforests, and are considered an excellent model to study the evolution of high tree species diversity characteristic of tropical rainforests. The objective of my study was to decipher the evolutionary relationships among dipterocarps using the partial sequences from the gene that codes nuclear 18S ribosomal RNA (N18S) sequence to determine the phylogenetic relationships within the Dipterocarpaceae. Although the level of polymorphism detected was relatively low, the overall phylogenetic relationships inferred

from the N18S data were congruent with chromosome number of each group, where chromosome 7 is the derived state, and consistent with historical biogeographical events.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Selvadurai Dayanandan for giving me the opportunity to work in his laboratory, continuous guidance, and particularly for collecting and maintaining all plant material used in this study.

I would also like to express my gratitude to the members of the thesis committee; Dr. Patrick Gulick and Dr. William Zerges for their support. I would like to thank Dr. S. Dayanandan and Concordia University for their financial support that made this study possible.

I send special thanks to Dr. Dominique Anzellotti for her comments on earlier versions. I wish to thank all current and previous members of the Forest and Evolutionary Genomics laboratory for their cheerfulness and optimism that made my experience at Concordia intellectually stimulating and fun. Jonathon Gray for preparation of Figure 1: Distribution map indicating the locations of the seventeen *Pinus banksiana* populations sampled.

I would like to acknowledge Concordia University teaching assistantship and Dr. Dayanandan's research grants from 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 that made this study possible.

Finally I gratefully acknowledge my family members for their patience and endless encouragement. I am especially indebted to my parents, brother, and my best friend, Anthony, for their unwavering support and understanding.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

General Introduction

Despite long-standing debates the mechanisms underlying the evolution of the diversity of forest trees, the major structural life forms of some of the most diverse ecosystems of the world, remain poorly understood (Newstrom et al. 1994, Ledig 1988, Richardson et al. 2001, Webb 2000, Wen 1999). In order to contribute further to the understanding of the evolution of forest trees, I conducted studies on a) the genetic structure of jack pine (*Pinus banksiana*), a widely distributed coniferous tree in the boreal forests of North America, and b) the molecular phylogeny of the tropical tree family Dipterocarpaceae based on partial nucleotide sequences of a nuclear encoded 18S ribosomal RNA gene.

a) The genetic structure of jack pine (*Pinus banksiana*):

Jack pine is one of the widely distributed conifer species in North America. Knowledge of population genetic structure in current natural populations is crucial for its sustainable management and is an essential component required to understand how phenomena such as genetic drift or historical refugia events have influenced the evolutionary process. Genetic studies of jack pine based on traditional genetic markers have indicated relatively high levels of polymorphism coupled with low levels of genetic differentiation among populations suggesting an elevated level of gene flow between populations (Schoenike et al. 1976, Wagner et al. 1987, Dong et al. 1994, Saenz-Romero et al. 2001, Nkongolo et al. 2002, Ye et al. 2002). Most jack pine studies have restricted their sampling range to hybrid zones or to local sites. The purpose of this study is to examine range-wide genetic diversity, population differentiation, inter-population migration, and genetic distances among natural populations across North America.

b) Phylogeny of the Dipterocarpaceae:

The pan tropical tree family Dipterocarpaceae represents one of the most species rich tropical trees in the asiatic rainforests. Although chloroplast DNA based phylogenetic studies of the Dipterocarpaceae are available, comprehensive phylogenetic studies of the Dipterocarpaceae based on nuclear gene sequences are lacking. Thus, I conducted a phylogenetic study of the Dipterocarpaceae based on partial sequences of nuclear 18S rRNA gene and compare nuclear biparentally inherited data to chloroplast uniparentally inherited data to elucidate the classification of various taxa and examine post-glacial dispersal events.

I 1. Introduction

I 1.1 Jack Pine ecology and distribution

Jack pine, *Pinus banksiana,* is one of the most widely distributed pine species in Canada, and its current natural distribution range extends from the provinces of Prince Edward Island to Alberta. In the eastern part of its range, jack pine grows in a maritime climate but elsewhere it is found in diverse continental climates characterized by short and warm to cool summers to very cold winters with minimal rainfall. Jack pine can grow on a very dry sandy and calcareous soil (pH 8.2) provided a normal mycorrhizal association is present or on gravelly soils where other species can scarcely survive. However, it grows best on well-drained sandy soils (Rudolph et al. 1990). Due to its intolerance to shade, post-fire recruitment of jack pine is usually limited to a single even-aged cohort of individuals established within a few years after a fire (Gauthier et al. 1993).

Jack pine is a monoecious species and "flowers" between late June and August. Its ability to produce serotinous cones, requiring a temperature of 50°C to open, is a crucial adaptation for regeneration in fire prone boreal forests (Gauthier et al. 1992, Beland et al. 1993). Jack pine in the southern regions of its distribution range produces nonserotinous cones (Rudolph et al. 1990). It is considered a pioneer species and invades areas where mineral soil has been exposed to major disturbances such as fire (Rudolph et al. 1990). In the absence of fire or other catastrophes, jack pine is succeeded by more shade tolerant species.

The present distribution range of Jack pine was once covered by ice during the last glacial period of the Pleistocene era (Marsella et al. 2000, Dyke et al. 2002). Thus,

the distribution of the species is believed to have resulted from re-invasion and migration over a relatively short period of time (Rudolph et al. 1990). Additionally, the paleobotanical evidence suggests that jack pine may have survived the recent glacial maximum in the North Dakota region and migrated to the north and eastward through the Mississippi Valley into Ontario, and westward toward Manitoba and Saskatchewan regions (Rudolph et al. 1982). There is also the possibility that jack pine may have survived during the last glacial period in refugia in eastern North America (Rudolph et al. 1982). Studies on range-wide genetic structure may elucidate the historical distribution patterns of the species.

Although some studies have investigated the genetic and morphological diversity of jack pine in its entire distribution range (Schoenike et al. 1976, Wagner et al.1987), most studies were focused on hybrid zones or selected areas of the distribution range (Dong et al. 1994, Saenz-Romero et al. 2001, Nkongolo et al. 2002, Ye et al. 2002). A previous study, which included most of the populations across the natural distribution range of jack pine, found a correlation between the morphological traits and the geographic origins of the populations (Schoenike 1976). It also suggested that natural selection has a significant role in shaping morphological and genetic diversity (Schoenike 1976).

The importance of jack pine as a source of wood coupled with the concomitant reduction in genetic diversity resulting from various harvesting practices exerted on the boreal forests, have led to a growing interest in formulating effective management strategies for this species. Knowledge of population genetic structures, one of the components used to estimate biodiversity, is crucial for designing programs for

sustainable management of biodiversity (Lindenmayer et al. 2000). To prevent loss of biodiversity, protected populations are selected based upon genetic distinction, range disjunction, and/or biogeographic distinction (Green 2005). Therefore, insights into the genetic structure of a population are useful to identify regions for conservation (Bangert et al. 2005, Lindenmayer et al. 2000, Redford et al. 1999). The purpose of this study is to examine range-wide genetic diversity, population differentiation, inter-population migration, and genetic distances among natural populations.

I 1.2 Nuclear microsatellite markers

We used nuclear microsatellite markers to estimate the range-wide genetic diversity and population structure of jack pine. Microsatellites, also known as simple sequence repeats (SSR), are short segments of DNA composed of 10 to 50 copies of a unique repeat motif, typically between 1 to 6 base pairs (bp) in length. Microsatellites are further classified into three categories: perfect, interrupted, and compound repeats (Hartl et al. 2000). It is assumed that these markers are uniformly and randomly distributed throughout the eukaryotic genome at a high frequency, and rarely found in coding regions (Hancock 1995, Baker et al. 2000).

Microsatellites are characterized by relatively high mutation rates per generation and reported to vary between loci (Thuillet et al. 2002) and among organisms ranging from approximately 10^{-2} in *Escherichia coli* (Levinson et al. 1987), and 10^{-3} in humans (Weber et al. 1993), to $6x10^{-6}$ in *Drosophila* (Schug et al. 1997). The mutation rates in plants vary from $2.4x10^{-4}$ in durum wheat (Thuillet et al. 2002) to $7.7x10^{-4}$ in maize (Vigouroux et al. 2002). Mutation rates are influenced by several factors including: the type of repeat (Ellegren et al. 1995, Chakraborty et al. 1997), gender (Primmer et al. 1998, Ellengren 2000), and the length of the microsatellite (Schlotterer et al. 1998). The influence of microsatellite length on mutation rates has been reported in humans (Weber et al. 1993) and plants (Thuillet et al. 2002, Vigouroux et al. 2002). Although rare, some microsatellites have been shown to affect expression of nearby genes (Li et al. 2002). Most commonly, tri-nucleotide repeats have been associated with the onset of various human diseases such as fragile X-syndrome (CGC), myotic dystrophy (CTG), and Huntington's disease (CAG) (Klung et al. 2000, Kovtun et al. 2001). In most diseases

where microsatellites are implicated, repeat sequences reach a critical length after which they become unstable and undergo an increase termed "dynamic expansion" (Groenen et al. 1998, Smith et al. 1995).

There are two proposed mechanisms that may explain the relatively high mutation rates observed in microsatellite markers. The most commonly accepted theory involves DNA polymerase slippage during replication (Levinson et al. 1987, Hartl et al. 2000). During replication, the template strand momentarily dissociates from the newly synthesized strand whereby slippage occurs when and if the two strands re-associate out of phase with each other causing an expansion or compression of the microsatellite (Hartl et al. 2000). An alternate explanation for such elevated mutation rates involves unequal crossing over or gene conversion events (Smith 1976, Jeffreys et al. 1994).

The use of nuclear microsatellite markers may address the weaknesses of traditional genetic markers such as isozymes and RAPD that pose difficulties for fine-scale genetic analysis of natural populations due to their low levels of polymorphisms or limited reproducibility or inconsistent results (Jarne and Lagodo 1996). Microsatellite markers are reportedly effective in assessing genetic diversity at various scales, including quantifying seed and pollen-mediated gene dispersal as a function of geographical distances (Epperson et al. 1997, Richardson et al. 2002), as well as estimating the effects of harvesting practices on the genetic constitution of forest trees (Thomas et al., 1999) due to their high mutation rates. Thus far, microsatellite markers have been successfully used to assess genetic variability of plant species (Dayanandan et al 1998, Stacy et al 2001, Lemes et al. 2003), including *Pinus contorta* (Thomas et al. 1999), *Pinus radiata*

(Fisher et al 1998), *Pinus sylvestris* (Soranzo et al 1998), and *Pinus resinosa* (Boys et al 2005). However, no range-wide genetic studies were reported for *Pinus banksiana*.

I 1.3 Study objectives

The objectives of this study are (1) to isolate and characterize nuclear microsatellite markers for jack pine, and (2) to use these markers to analyze population genetic structure to gain insights into gene flow and post-glacial dispersal events.

I 2. Materials and Methods

I 2.1 Samples

A total of 495 samples (seeds or needle samples) from seventeen jack pine populations distributed throughout its distribution range within Canada were used for the present study. Approximately 25 to 30 individuals from each of the 17 representative natural populations were included in the analysis (Figure 1). The population NS1 is a group of half sib families collected from a single source tree. All other populations were collected from several unique trees in a given location. The geographical locations of the populations included in this study are given in Table 1 and in Figure 1.

One to two fresh needles, or 0.75g to 1.5g respectively, were obtained for DNA extraction for each individual. Two to three milliliters (ml) of extraction buffer (31.9 grams of Sorbitol (FW 182.2), 1.9 grams of Sodium bisulfite, 50 ml of 1 M Tris (pH 7.5), 5.0 ml of 0.5M EDTA and filled with distilled water to a final volume of 500ml) was added in a Ziploc plastic freezer bag containing the fresh samples. The needles were crushed with rollers and the extract was transferred into a labeled microcentrifuge tube and placed on ice until further processing.

Once a set of samples were crushed and transferred into the microcentrifuge tubes, the set of tubes were centrifuged at 6000 rpm for 5 minutes at room temperature. The supernatant was subsequently decanted and pellet(s) were re-suspended pellet with 600µl of Extraction/Nucleic Lysis (1:1) Buffer. To make 200 ml of nucleic lysis buffer we added 23.4 grams of NaCl, 4.0 grams of CTAB (FW 364.5), 40.0 ml of 1 M Tris (pH 7.5), 20.0 ml of 0.5M EDTA and filled with distilled water to a final volume of 200ml.

To the 600µl of Extraction/Nucleic Lysis (1:1) Buffer, we added 120µl of 5% Sarkosyl solution and mixed with a vortex mixer for several seconds. The tubes were incubated in a 65°C water bath for a minimum of 45 minutes and were mixed about half way through with a vortex mixer.

After the incubation, in a fumehood, we added 600µl of Chloroform/Isoamyl Alcohol (24:1) mixture and gently inverted the tubes to form an emulsion. The samples were subsequently centrifuged at 7000 rpm for 5 minutes. The upper phase was transferred to a clean microfuge tube containing 600µl of Isopropanol, gently inverted several times, and were left standing at room temperature for a minimum of 30 minutes to allow the DNA to precipitate out of the solution. The samples were then centrifuged at 14000 rpm for 10 minutes and the supernatant was decanted. We washed the pellets with 70% ethanol and centrifuged at 14000rpm for 10 minutes. The supernatant was decanted and the samples were left to dry overnight at room temperature. The next morning, to dissolve the dried pellets we added 150µl of RNase treated TE Buffer (2.0µl of RNase was added to every 1.5ml tube of TE buffer), mixed them on the vortex several times and left in the fridge for at least 4 hours.

I 2.2 Isolation and characterization of nuclear SSR markers

A partial genomic library of jack pine with DNA fragments ranging from 500 to 1000 base pairs was constructed and screened with $(TC)_{15}$, $(AC)_{15}$, $(AT)_{15}$, $(AC)_{15}$ synthetic oligonucleotide probes (Dayanandan et al. 1998). Aliquots of 50µl of DNA from each of six randomly selected jack pine individuals were pooled (for a total of 300µl). The pooled samples were digested with the restriction enzyme *Sau*3aI (Promega, Madison, Wisonsin) in a total volume of 600µl for 6 hours at 37°C. The solution was heated for 10 minutes at 65°C to deactivate the enzyme *Sau*3aI. The digested DNA was precipitated out with 1300µl of 95% ethanol, washed with 1000µl of 70% ethanol, and dissolved in 30µl 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 subsequently dephosphorylated by adding 0.5µl (1 unit/µl) of calf intestinal alkaline phosphatase (MBI Fermentas, Burlington, Ontario) and incubated for 1 hour at 37°C. The phosphatase was deactivated by adding 1.25µl of 0.1 M EDTA and heating to 75°C for 10 minutes.

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

The ligation of jack pine and the vector was performed in a total volume of 20µl, with 4µl of each of the gel-purified DNA samples (approximately 80 ng of each DNA sample) and 1 µl of T4 DNA ligase (Promega, Madison, Wisconsin), at 16°C for 16 hours. The ligated DNA was transfected into XL1 Blue MRF' competent bacterial cells

(Strategene, 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 nearly 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 into a plastic Tupperware containing 150ml of 2x SSC (diluted 10 fold from a solution of 20X SSC that consisted of 175.3g of NaCl, 88.2g of sodium citrate and filled with distilled water to 1 liter) and then transferred to a new container containing 350µl of the same solution. Membranes were then washed with 5x SSC at 50°C for 1 hour and pre-hybridized overnight at 50°C in a Tupperware containing 250 ml of hybridization medium. Following pre-hybridization, the hybridization medium in the Tupperware was replaced with 250 ml of fresh medium (consisting of 150 ml of 20x SSC, 5 ml of 100x Denhardts solution, 5 grams of BSA, 20ml of 10% SDS, and filled to 500 ml with distilled water), and then the labeled probe was added.

The oligonucleotides used as probes were $(TC)_{15}$, $(AC)_{15}$, $(AT)_{15}$, and $(AC)_{15}$. Each probe (20 pmol) was end lablled 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 50µl. The enzyme was inactivated by heating at 65°C for 20 minutes. The labelled probe was purified using Qiagen Nucleotide Removal Kit (Qiagen, Mississauga, Ontario) and denatured at 90°C before adding to the hybridization medium. The Tupperware was subsequently placed in a shaking water incubator at 50°C overnight (minimum of 3 hours). Hybridized blots were washed with 100 ml of 0.1% SDS/6x SSC

solution twice at room temperature, and once with 1000 ml of the sane solution at 50°C for 30 minutes. The blots were rinsed with 6x 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 150µl of overnight grown XL1-blueMRF' bacterial cells. Tubes were kept in a shaking incubator at 37°C for 6 hours. Positive clones that were not well isolated, due to high density of plaques per plate, were first re-plated and then picked and grown following the method just described. Once grown, a fraction of each cell culture was transferred to a 2 ml microfuge tube and spun at 14000 rpm for 5 minutes.

Positive clones were sequenced using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, California) from either single-stranded DNA, isolated from the Wizard M13 DNA purification system (Promega, Madison, Wisconsin), or double-stranded DNA, isolated from the Qiaprep Spin Miniprep Kit (Qiagen, Mississauga, Ontario).

The yield of isolated DNA was tested by electrophoresis on a 1% agarose gel with ethydium bromide (0.33µg/ml). To make a 1% agarose gel, in a 250 ml flask we added 1 gram of agarose powder to 20 ml 5 X TBE (Tris-borate) and 80 ml dH_2O. The mixture was heated in a microwave for 2 minutes to dissolve the powder and then cool down to55°C. We then added 3.3µl of ethydium bromide (10g/L) and swirled the flask several times. The solution was poured into a 100 ml gel container and left in a fumehood at room temperature to solidify.

Oligonucleotide primers complimentary to the regions flanking the microsatellites were designed and synthesized (Operon Technologies, Alameda, California). The polymerase chain reaction (PCR) conditions were optimized for each oligonucleotide primer pair and used for amplifying the target regions in the sampled individuals. PCR reactions were performed in a total volume of 25 µl with 0.2mM dNTP, 2.0 mM $MgCl_2$, 2.5 µl 10X buffer (1 ml of 10X buffer consisted of 200.0 µl of 1 M Tris (pH 9.5), 62.5 µl of 4 M KCL, 5.0 µl of Tween, 1.0 mg of BSA, and filled to a final volume of 1 ml with distilled water), 0.3 units of *Taq* Polymerase, 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 35 to 40 cycles of: 94°C for 30 seconds (denaturation), 15 seconds at the optimal annealing temperature as given in Table 2, and 72°C for 30 seconds (extension). These parameters were followed by a 5 minute final extension period at 72°C.

The amplified products were visualized using the ABI 310 genetic analyzer and size of each fragment was determined using Genescan and Genotyper software packages (Applied Biosystems, Foster city, California). Since multiple DNA fragments per primer pair were observed (Figure 4), possibly due to duplicated regions of the pine genome, we were unable to establish co-dominant segregation patterns of alleles for each locus. Although non-specific primer binding could also have contributed to amplification of additional fragments, because the signal intensity of the fragments were consistent among each SSR marker, the multiple fragments are likely resulting from additional SSR sites than random regions of the genome. Thus, as in another study that faced the same situation (Bockelmann et al. 2003), we scored the presence (1) versus absence (0) of

DNA fragments and each fragment was considered as a unique site. A minimum ratio of 3:1 signal intensity to background noise was used as a threshold for scoring the presence of a given band.

I 2.3 Assessment of genetic diversity

The population genetic analysis software POPGENE (Yeh et al. 1997) was used for analyzing the data. A pair-wise genetic distance analysis between populations (Nei 1972) was conducted to construct a dendrogram depicting genetic relationships among all populations using the neighbor-joining algorithm of the PHYLIP software (Felsenstein 1989). One hundred trees were constructed, based on 100 distance matrices (SEQBOOT), and a final unrooted consensus tree was subsequently constructed using the computer program CONSENSE (Felsenstein 1989). Genetic distance results indicated a distinction between mainland and maritime populations; therefore mainland and maritime populations were treated as two distinct groups for some of the analyses.

The measure of genetic diversity demonstrates how much or how little variability and in which population(s) genetic variation exists. The genetic diversity statistics include the percentage of polymorphic loci, the effective number of alleles per locus, and the expected heterozygosities for each population including the pooled mainland population. These analyses were conducted under the assumption of Hardy-Weinberg equilibrium (HWE).

Population genetic structure analyses include the estimation of F_{st} among populations located in the mainland and amongst all populations. The F-statistic (F_{st}) provides a comparable numerical value representing the degree of genetic differentiation, how different one population or groups of populations are from one another. This parameter (F_{st}) is used for estimating the number of migrants N_m among the populations of interest [$N_m = (1/F_{st} - 1)/4$]. This project compared 3 methods of estimating the number of migrants; Wright's (1969), Hudson's (1998), and Yeh and Hu's (2005) methods. Yeh

and Hu (2005) provided a detailed comparison of each of these methods and are discussed below. N_m estimates the number of migrants successfully entering a population per generation; where N represents the effective population size and m the migration rate.

These three estimates of N_m rely on the estimates of heterozygosity values. As we were unable to elucidate the mode of inheritance of the amplified DNA fragments, we scored the fragments as a binary code (presence or absence) and consequently analyzed the data as information obtained from a dominant marker. This approach is similar to microsatellite data analyzed by Bockelmann et al. (2003), and many RAPD marker based population genetic studies. Dominant markers are unable to differentiate between two states; homozygous dominant (in which 2 fragments of the same size are amplified from a given region of the genome) and the heterozygous state (in which 1 fragment amplifies from one chromosome but the equivalent region on the opposing chromosome fails to amplify). In both states a single band or fragment would be amplified and scored as 1 for presence. However, a third state exists (in which neither fragments on homologous regions of the chromosome amplify – perhaps due to a mutation in the primer binding region) referred to as the homozygous recessive state. This third state would be scored as 0 for absence. The relative measure of the homozygous recessive state is used in Hardy-Weinberg equations to estimate the q^2 value; where q^2 represents the frequency of the recessive genotypic sate. After taking the square root of q^2, we used the equations from Hardy-Weinberg, $1-q = p$, so we can get our p value. The p-value corresponds to the events when a band is amplified or the mean frequency of the dominant allelic state; either homozygous dominant or heterozygous states. Finally, the heterozygosity value of populations is obtained by multiplying 2pq (from Hardy-Weinberg equation). In each of

the three methods, Wright's (1969), Hudson's (1998), and Yeh and Hu's (2005), the heterozygosity values are averaged among all pairwise comparisons for each fragment analyzed (Yeh et al. 2005).

The spatial pattern of genetic variation amongst all populations and among the populations in mainland was tested for evidence of isolation by distance. The estimates of the ratios of $F_{st}/(1-F_{st})$ were regressed against the natural log of geographical distance, i.e. $F_{st}/(1-F_{st})$ =a+b ln(distance) where a and b are the intercept and regression coefficients respectively (Rousset 1997). According to Wright's formula, the ratio of $F_{st}/(1-F_{st})$ refers to the inverse of the number of migrants. The geographical distance here is two-dimension Euclidian distance based on latitude and longitude coordinates between pairwise populations.

Due to different assumptions used in various models of population structure and gene flow analysis (Yeh and Hu, 2005), three distinct methods were applied to estimate the number of migrants from the mainland to the maritime regions: Wright's F_{st} (Wright 1969), Hudson's method (Hudson 1998), and the Yeh and Hu's method (Yeh and Hu, 2005). The latter two methods are based on unidirectional gene flow assumption (from mainland to maritime populations). Hudson's method may overestimate gene flow results, while Yeh and Hu's method proves to be more realistic since it is based on a more general condition of allele frequency distribution (Yeh and Hu 2005). With the Yeh and Hu's method the mutation effect is assumed negligible in the maritime populations and effective population size is taken into consideration. The effective population size in each population is assumed to be about 1000 individuals (Govindaraju 1988). The effective population size of the pooled mainland population could be approximately

calculated by Wright's formula (Wright 1943) of $LN/(1-F_{st})$ where L is the number of populations investigated in mainland and N is the effective size of each population in the mainland.

I 3. Results

I 3.1 Characterization of nuclear microsatellite markers

The partial genomic library constructed consisted of approximately 12000 clones. The sequencing of 90 positive clones revealed 38 unique microsatellites (10 AC/TG; 4 AT/AAT/AAAT; and 24 AG/TC). A total of 18 oligonucleotide primer pairs complementary to the flanking regions of the microsatellites were designed (Table 2). Fifteen of these primer pairs produced consistent PCR amplification products. Eight (PBA13, PBA15, PBA21, PBA22, PBA24, PBA26, PBA27, and PBA28) of those showed polymorphism while the remaining six were monomorphic.

The five primer pairs that showed consistent PCR amplification with the most polymorphism were selected for genetic analysis of population, and yielded a total of 40 polymorphic fragments or sites. The microsatellite marker PBA13 uncovered 8 unique fragments, PBA15 uncovered 5 fragments, PBA22 uncovered 10 unique fragments, PBA24 uncovered 6 unique fragments, and PBA28 uncovered 11 unique fragments (Figure 5).

I 3.2 Genetic distances

The average genetic distance within mainland populations (0.0217±0.0012) was smaller than those among all populations (0.0386±0.0020) (Table 5). An unrooted tree indicated that these maritime populations were genetically well separated from the mainland populations (Figure 3). The average genetic distance across all populations was 0.0386±0.0020, with the range from 0.0059 to 0.1205 (Table 5). The genetic distances between mainland and maritime populations varied with the highest distance between mainland and the Black brook NS1 population (0.0812±0.0063 and ranged from 0.0444 to 0.1205), and the lowest between mainland and the Grand lake NB1 population (0.0504±0.0047 and ranged from 0.0265 to 0.0825). The distance between the mainland and the Coal branch NB2 population (0.0504±0.0047) is comparable to that between mainland and the Foxley river (PE1) population (0.0486±0.0033). The distance between mainland and another Black brook population (NS2) remained relatively large, 0.0593±0.0043.

I 3.3 Genetic diversity

None of the populations showed polymorphism at all of the 40 fragments (Table 1). The number of polymorphic loci (n_p) was variable from population to population, with an average of 27 in the 12 mainland populations and 21 in the 5 maritime populations. The percentage of polymorphic loci also showed differences among populations, with an average value of 67.5% in the 12 mainland populations and 61.5% in the 5 maritime populations. Similar results were observed among populations in the number of observed alleles per locus (n_a), the expected number of alleles per locus (n_e), and the average gene diversity (h). Generally the values of these genetic statistics demonstrate a tendency of greater diversity among the populations in mainland than in those of the maritime (Table 1).

The Ewens-Watterson test for neutrality indicated that all estimates were within the lower and upper 95% confidence limits suggesting that all 40 loci fits the hypothesis of neutrality. No significant linkage disequilibria were observed among all pair-wise comparison of loci.

I 3.4 Population structure

Genetic differentiation varied with different loci, with F_{st} values ranging from 0.0164 to 0.4667 among all 17 populations, and from 0.0125 to 0.3057 among the 12 mainland populations (Table 3). In general, differentiation among all populations was greater than that among mainland populations, with 16.62% variation among populations for the former and 10.22% for the latter.

Regression analysis showed that the allele frequencies and expected heterozygosities at a majority of loci, with few exceptions, did not show a correlation with latitude and longitude (Table 4). The allele frequencies of three markers (PBA24-3, PBA13-7, and PBA15-4) have significant correlation with both latitude and longitude. The allele frequencies of the marker PBA28-1 showed a significant correlation with longitude. The expected heterozygosities of the markers PBA13-7 and PBA15-1 showed a significant correlation with latitude, while the expected heterozygosities of the marker PBA28-3 showed a significant correlation with longitude (Table 4). No correlation between the percentage of polymorphic loci and geographic location was observed. There was weak, but not significant isolation by distance effect was noted among all 17 populations, $F_{st}/(1-F_{st}) = 0.0629 + 0.0104\ln(\text{distance})$ (P-value=0.0699; $R^2 = 0.19$; Figure 2a). Also, there was no significant isolation by distance among all 12 mainland populations, $F_{st}/(1-F_{st}) = 0.0389 + 0.0057\ln(\text{distance})$ (P-value =0.1508; $R^2 = 0.22$; Figure 2b).

I 3.5 Inter-population migration

Estimation of the number of migrants with Wright's F_{st} (migration from "migrant pool"; Wright 1969) indicated more than one individual per generation was exchanged between the pooled mainland population and individual maritime populations (Table 6). The number of exchanged migrants per generation was large for the Coal branch NB2 (N_m=3.21) and Grand lake NB1 populations (N_m =3.20), but small for the Black brook NS1 (N_m =1.39) and NS2 (N_m =1.74) populations. The number of migrants between mainland and Foxley river (PE1) showed an intermediate value (N_m =2.64).

The results based on Hudson (1998) method indicated that the numbers of migrants from pooled mainland to Black brook NS1 (=0.45) and NS2 (=0.77) populations were small compared to the numbers of migrants to the rest three maritime populations (Table 7). The numbers of migrants to the Coal branch NB2 (=2.78) and Grand lake NB1 (=1.56) populations were high and the number of migrants to Foxley river PE1 was about 3.16 per generation (Table 7).

Compared with the estimates based on Wright's and Hudson's methods, the estimates of the number of migrants based on Yeh and Hu's method were small (less than 1.0 individual per generation), but the overall pattern was similar. The numbers of migrants were small to Black brook NS1 (=0.52) and NS2 (=0.55) populations, but large to the Coal branch NB2 and Grand lake NB1 populations (Table 7).

I 4. Discussion

I 4.1 Genetic Diversity

The low recovery rate of microsatellites (sequencing of 90 positive clones uncovered 38 unique microsatellites) reported in this study is typical of conifers and other species with large genomes (Kostia et al. 1995, Pfeiffer et al. 1997, Fischer and Bachmann 1998, Bérubé et al. 2003, Boys et al. 2005). It has been suggested that due to genome duplication, a phenomenon common in species with large genomes, there are multiple sites for primer binding and an increased probability of nonspecific primer binding sites resulting in the observed low microsatellite recovery rate (Garner 2002). With the exception of white pine (Echt et al. 1996), the greater number of TC/AG repeats relative to AC/TG repeats in jack pine is consistent with many other tree species, including *Populus tremuloides* (Dayanandan et al. 1998), *Pinus radiata* (Fisher et al. 1998), and *Pinus resinosa* (Boys et al. 2005).

Genetic diversity indices among mainland populations were greater than those among maritime populations. In conjunction with the relatively low levels of gene flow (discussed below), the maritime populations are genetically distinct from each other, as well as from mainland populations (Figure 3).

I 4.2 Population structure

The genetic differentiation values (F_{st}) obtained in our study ranging from 0.0164 to 0.4667 among all 17 populations (Table 3), is congruent with parameters obtained from previous allozyme studies reporting low differentiation in western populations (Dancik and Yeh 1983, Saenz-Romero et al. 2001, Ye et al. 2002). Population differentiation values derived from allozyme markers ranged between 0.021 and 0.070 from Alberta (Dancik and Yeh 1983, Ye et al. 2002), Manitoba (Ross and Hawkins 1986), Ontario (Danzmann and Buchert 1983), Quebec (Gauthier et al. 1992), and Wisconsin (Saenz-Romero et al. 2001). More recently, population differentiation values obtained from minisatellite mtDNA analysis, representing several populations across jack pine distribution within Canada, ranged from 0.265 to 0.597 (Godbout et al. 2005). The higher values reported from Godbout et al. (2005), as compared to the other allozyme studies mentioned, could be attributed to the eastern-most populations. When the maritime populations were excluded from our analysis, the genetic differentiation parameters decrease to values ranging from 0.0125 to 0.3057. When maritime populations are included in the data set, significantly higher genetic differentiation is observed in pine species (Godbout et al. 2005, Boys et al. 2005). These observations suggest that the maritime populations are relatively more genetically distinct from one another as compared to the mainland populations. Similar observations were reported for black spruce, *Picea mariana* Mill., (Jaramillo-Correa et al. 2004). Based on four polymorphic mitochondrial DNA loci, a significant subdivision of population genetic diversity was detected, suggesting low levels of gene flow between populations (Jaramillo-Correa et al. 2004).

The lack of a significant correlation between geographic distance and genetic differentiation, F_{st} values, among all 17 jack pine populations, as well as among the mainland populations, suggests that multiple factors have contributed to the population genetic structure of this species. Although reforestation efforts can influence the genetic structure of populations, our samples were obtained from natural stands and therefore the effects are unlikely attributed to human involvement. Multiple refugia events during the last glacial maxima have been reported to influence the genetic structure of pine species across their natural range (Boys et al. 2005, Godbout et al. 2005). Jack pine studies have reported the possibility that three such events occurred during the Pleistocene era (Godbout et al. 2005). Each of these events could have influenced the genetic constitution of each region uniquely and consequently eliminated a geographical pattern across the species' entire distributional range.

I 4.3 Inter-population migration

F_{st} is a measure of genetic differentiation among subpopulations. Differentiation levels between populations provide an indication of the degree of evolutionary divergence. This measure is influenced by opposing effects of migration, such as forces that tend to homogenize populations, while mutation or genetic drift can lead to differentiation. The relatively high values of F_{st} estimated in certain populations of this study indicate that gene flow is limited in maritime jack pine populations as compared to the mainland populations. As in Godbout et al. (2005), our overall values are still high for all populations suggesting jack pine populations are relatively highly differentiated. This is an unusual characteristic for most conifer species as long-distance seed and pollen dispersal ability is facilitated by wind.

Nei's genetic distance assumes an infinite alleles model with all loci having the same rate of neutral mutation. F_{st} estimates the degree of genetic differentiation among all subpopulations and Nei's genetic distance values estimate the degree of genetic differentiation among all pairs of populations. The same trend was observed in both measures; maritime populations are relatively more genetically differentiated from each other than are the mainland populations (Figure 3).

Extensive gene flow among contiguous jack pine populations may have contributed to the lower population differentiation among mainland populations. In contrast, low gene flow due to discontinuous populations may have contributed to high genetic distances among maritime populations. As maritime populations are separated by physical barriers, gene flow is more restricted as compared to the mainland populations; which are not physically separated. Similarly, Godbout et al. (2005) reported low levels

of gene flow with N_m = 0.379 migrant per population for jack pine. The low level of gene flow, in particular for the Nova Scotia populations, NS1 and NS2, is below that which is necessary to counteract the effects of genetic drift, which should to be considered when developing conservation programs. (However, the NS1 population is a group of half sib families collected from a single source trees, therefore gene flow is expected to be low).

I 4.4 Geological history

Patterns of genetic variation reported in this study could be attributed to postglacial colonization of individuals from multiple refugia. Several authors have suggested the likelihood of multiple, at least three southern and one northeastern, colonization paths for various conifers within Canada (Boys et al. 2005, Godbout et al. 2005, Jaramillo-Correa et al. 2004). The postglacial recolonization into the present distribution range of several North American trees has been constructed from fossil and pollen records (Davis 1983, Webb 1988, Jackson et al. 1997). During the last glacial maximum, the *Pinus*-dominated vegetation (mainly *P. banksiana* with local *P. resinosa* and *P. strobus*) in the east occurred extensively to 34°N and possibly as far south as 30°N and from the Atlantic coastal to the interior highland regions (Jackson et al. 2000, Rudolph and Yeatman 1982). From 14000 to 10000 B.P., northern hard pines are reported to have spread northwards and westwards from the central and western refugia respectively (Davis 1983, Webb 1988, Jackson et al. 1997). In the midcontinent, *Pinus* pollen in palynological records declined by 12000 B.P., but they remained high along the Atlantic coast (Jackson et al. 1997). The distinct eastern coastal refuge reported for *Pinus resinosa* (Walter and Epperson 2001, 2005, Boys et al 2005) and for black spruce (Jaramillo-Correa et al. 2004), suggests that glacial populations would have initially expanded inland from the coastal line establishing itself in Prince Edward Island, Nova Scotia, and New Brunswick. Subsequently, based on Godbout et al. (2005), jack pine would have been followed by a second wave of migrants from the southeastern glacial populations of the Appalachian Mountains.

Eastern and central glacial refugia may have contributed to the observed division between mainland and maritime populations in our study. The central and westward glacial refugia hypotheses are congruent with our data, which show a low genetic differentiation and higher number of migrants per generation. Furthermore, the suggestion that a third glacial refugia originating from the Atlantic coast is highly probable due to the elevated genetic differentiation parameters reported from the maritime to mainland populations. It has been proposed that the northern part of the Appalachian Mountains and Adirondacks in the United States may have created a long-lasting physical barrier limiting gene flow between eastern and western glacial populations of the mountain range (Godbout et al. 2005). Mitochondrial DNA minisatellite analysis in jack pine also reports three distinct refugial populations corresponding to the western, central-eastern and eastern regions of Canada, demonstrating the significant role of the northern part of the US Appalachian Mountains as a factor of vicariance, a mode of speciation which occurs as a result of the separation and subsequent isolation of fragments of an original unified population, during the ice age (Godbout et al. 2005). Similarly, in red pine, coalescence-based analysis suggests that northeastern and main populations likely became isolated during the most recent Pleistocene glacial period, and the pattern of genetic structure reported could have arisen from postglacial colonization by individuals from multiple refugia, including southern Appalachian and northeastern refugia, forming an admixed population through secondary contact (Walter and Epperson 2001, 2005, Boys et al 2005). The patterns revealed in this study are in agreement with Godbout et al. (2005) and suggest that jack pine shared its

geological history with other forest-associated North American species (Walter and Epperson 2001, 2005, Boys et al. 2005, Jaramillo-Correa et al. 2004).

I 5. Conclusion

The polymorphic microsatellite markers developed have been valuable in assessing the genetic diversity of jack pine, one of the commercially exploited and widely distributed boreal tree species. The use of microsatellite markers in elucidating fine-scale genetic variability, as in other studies (Boys et al. 2005), likely due to their high mutation rate is evident in our project. The microsatellite markers have been able to provide further evidence that 1) the current population genetic structure of jack pine likely resulted from range expansion from at least two glacial refugia, and 2) the maritime populations are genetically distinct from the mainland populations. The information gained from this study contributes to the understanding the evolution of forest trees in relation to forest refugia events during Pleistocene era, as well as for planning and implementation of genetically sound conservation and management practices for jack pine particularly for the maritime populations; due to their higher levels of genetic differentiation and lower levels of gene flow relative to the mainland populations.

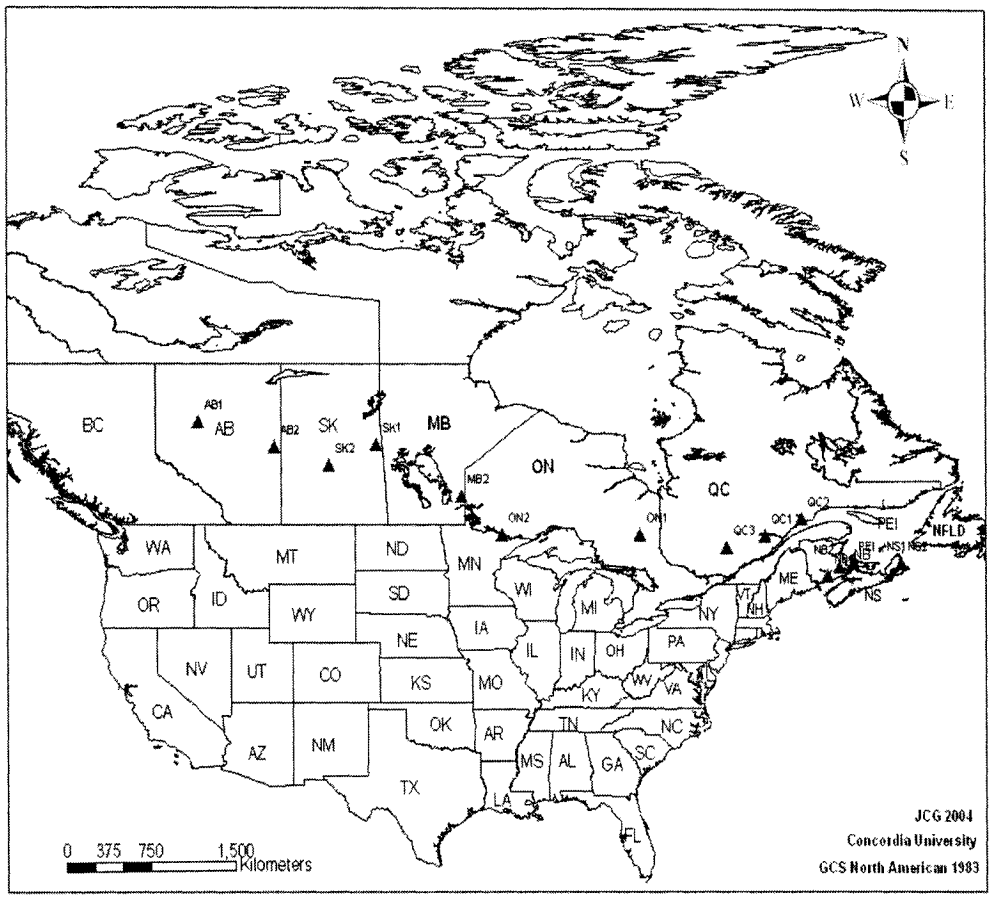


Figure 1 Distribution map indicating the locations of the seventeen *Pinus banksiana* populations sampled. Abbreviations correspond to those described in Table 1.

Figure 2a Isolation by distance analysis among all 17 populations;
$F_{st}/(1-F_{st}) = 0.0629 + 0.0104\ln(\text{geographicdistance})$

P-value=0.0699, R^2=0.19, not significant

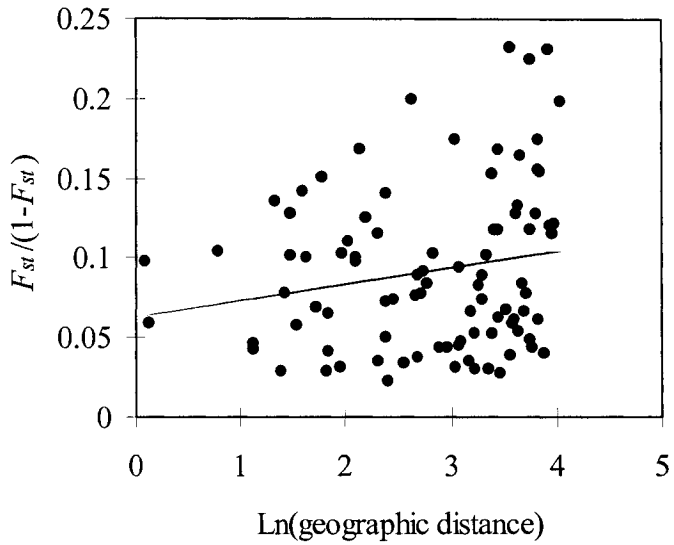


Figure 2b Isolation by distance analysis among the 12 mainland populations;
$F_{st}/(1-F_{st}) = 0.0389 + 0.0057\ln(\text{geographicdistance})$

P-value=0.1508, R^2=0.22, not significant.

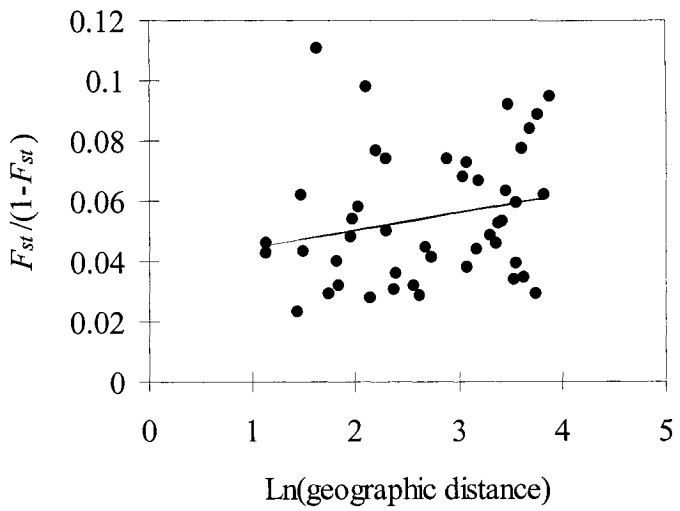


Figure 3 An unrooted consensus Neighbor-Joining tree based on 40 polymorphic sites.

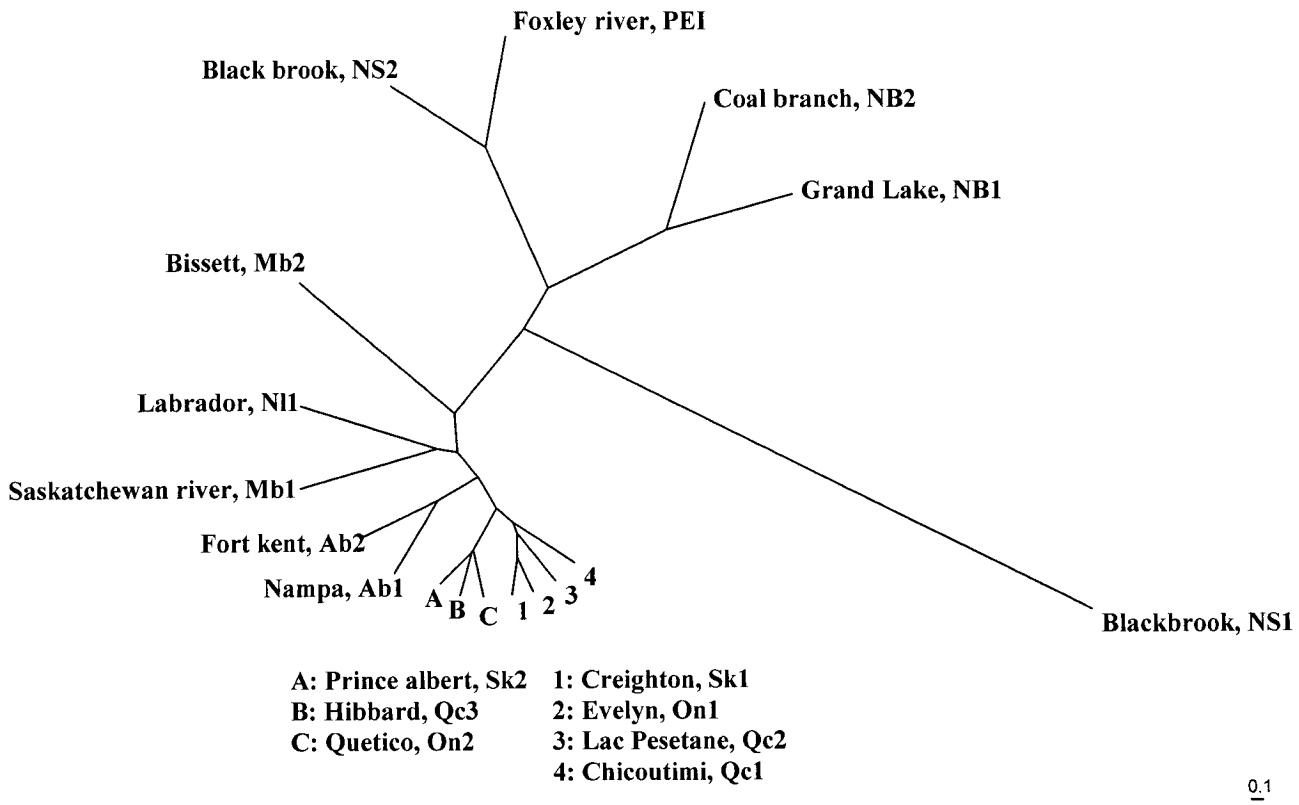


Figure 4 Multiple DNA fragments per primer pair. The figure depicts the output file from the ABI 310 Genetic Analyzer after genotyping a representative individual from one of the 17 populations. The green trace outlines the number and sizes of each of the fragments amplified with one SSR marker. The y-axis represents the signal intensity of each peak (fragment) while the x-axis represents the sizes of the fragments in base pairs. The program also provides a summary table indicating the sizes of each fragment to the nearest tenth (not shown). This individual would have been scored for the presence (1) of 4 fragments and absence (0) for all other fragments found in other individuals. This process was repeated with each of the 5 SSR markers included in the analysis and for all of the individuals.

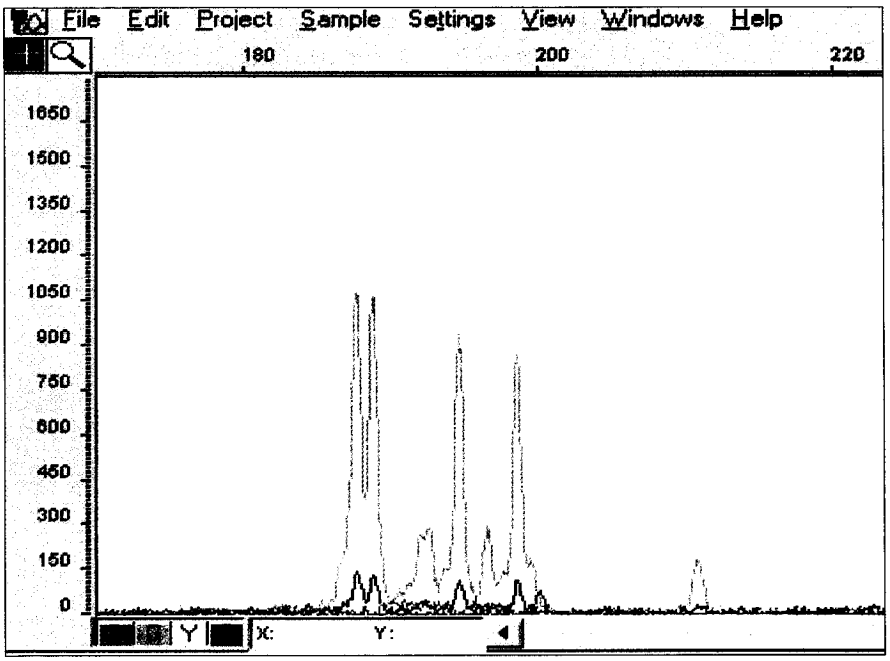


Figure 5 The binary matrix results of one SSR marker analysis from one population (Chicoutimi, QC). The figure represents the binary matrix created for one SSR marker (PBA22) from the analysis of one population. The first heading row indicates the name of the SSR marker used to amplify the fragments. The second row indicates the fragment number. In this case PBA22 uncovered 10 unique fragments from all the individuals genotyped. The third row indicates the sizes of each fragment in base pairs. Each column represents a unique fragment while each row, beginning at the forth, represents one individual. A matrix was constructed for each of the 5 markers and each of the 17 populations in the same way. The matrices were subsequently compiled to create the complete matrix (40 fragments by 495 individuals).

PBA 22									
Fragment1	Fragment2	Fragment3	Fragment4	Fragment5	Fragment6	Fragment7	Fragment8	Fragment9	Fragment10
168	**181**	**195**	**200**	**207**	**216**	**226**	**232**	**238**	**248**
0	1	0	1	0	0	0	0	1	1
1	0	0	0	0	0	0	0	0	1
0	0	0	1	1	0	1	0	0	1
1	1	0	0	0	0	1	1	0	0
1	1	0	0	0	0	1	1	0	0
1	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	1	1	0	1
0	0	1	0	0	0	1	0	0	1
0	1	0	0	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	0	0	0	1
1	0	0	0	0	1	0	0	0	1
1	1	0	0	0	0	1	1	0	1
0	0	0	1	0	1	0	1	0	1
0	1	1	0	0	0	1	1	0	0
0	0	0	1	0	0	0	0	0	1
0	0	0	1	0	1	0	1	0	1
1	1	0	0	0	0	1	0	0	0
1	0	0	0	1	0	1	1	1	0
0	0	0	0	1	0	1	1	0	1
1	0	0	0	1	0	1	1	0	1
1	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	1	1	0	0
0	1	0	0	0	0	0	1	0	1
0	0	0	0	0	1	1	0	0	1
0	1	0	0	0	0	1	0	0	1

Table 1 Geographic locations and genetic diversity in the 17 natural jack pine populations studied†

Population	°N	°E	N	n_p	P(%)	n_a	n_e	h
Mainland								
Chicoutimi (QC1)	48.30	71.00	27	27	67.5	1.675 ± 0.474	1.300 ± 0.355	0.183 ± 0.190
Lac Pesetane (QC2)	49.45	68.13	29	27	67.5	1.675 ± 0.474	1.232 ± 0.281	0.154 ± 0.159
Hibbard (QC3)	47.53	74.00	29	24	60.0	1.600 ± 0.496	1.244 ± 0.313	0.154 ± 0.174
Saskatchewan river (MB1)	---	---	30	22	55.0	1.550 ± 0.504	1.253 ± 0.321	0.158 ± 0.179
Bissett (MB2)	51.02	95.41	28	28	70.0	1.700 ± 0.464	1.267 ± 0.290	0.173 ± 0.169
Creighton (SK1)	54.51	102.25	29	22	55.0	1.550 ± 0.504	1.258 ± 0.327	0.160 ± 0.181
Prince albert (SK2)	53.14	106.03	27	26	65.0	1.650 ± 0.483	1.277 ± 0.312	0.176 ± 0.174
Labrador (NL1)	---	---	30	29	72.5	1.725 ± 0.452	1.231 ± 0.286	0.153 ± 0.157
Nampa (AB1)	56.07	116.55	26	30	75.0	1.750 ± 0.439	1.286 ± 0.313	0.182 ± 0.172
Fort kent (AB2)	54.38	110.51	30	29	72.5	1.725 ± 0.452	1.256 ± 0.282	0.170 ± 0.160
Evelyn (ON1)	48.40	81.00	29	30	75.0	1.750 ± 0.439	1.304 ± 0.321	0.192 ± 0.176
Quetico (ON2)	48.39	92.14	30	30	75.0	1.750 ± 0.439	1.261 ± 0.315	0.166 ± 0.170
Maritime								
Black brook (NS1)	46.46	60.19	29	20	50.0	1.500 ± 0.506	1.217 ± 0.346	0.128 ± 0.185
Black brook (NS2)	46.46	60.19	29	20	50.0	1.500 ± 0.506	1.187 ± 0.299	0.118 ± 0.165
Coal branch (NB2)	46.20	65.10	30	31	77.5	1.775 ± 0.423	1.321 ± 0.358	0.194 ± 0.188
Grand lake (NB1)	45.57	66.04	29	27	67.5	1.675 ± 0.474	1.256 ± 0.339	0.158 ± 0.178
Foxley river (PE1)	46.43	64.02	30	25	62.5	1.625 ± 0.490	1.266 ± 0.337	0.164 ± 0.184

† °N: latitude; °E: longitude; N: sample size; n_p : number of polymorphic "loci" sites; P(%): percentage of polymorphic "loci" sites is estimated from dividing the n_p by the total number of sites recorded (which equals 40) ; n_a : observed number of alleles; n_e : effective number of alleles estimates the reciprocal of homozygosity according to Kimura et al. (1964); h: Nei's (1973) gene diversity.

Table 3 Average estimates of F_{st} between pair-wise comparisons at individual polymorphic sites among all *P. banksiana* populations and among the *P. banksiana* populations within the mainland[†].

	All populations			Mainland populations		
Loci	H_t	H_s	F_{st}	H_t	H_s	F_{st}
PBA24-1	0.4438	0.3478	0.2161	0.4739	0.3921	0.1725
PBA24-2	0.4842	0.4409	0.0895	0.4927	0.4510	0.0846
PBA24-3	0.4992	0.4585	0.0815	0.5000	0.4639	0.0722
PBA24-4	0.1952	0.1855	0.0501	0.1707	0.1654	0.0310
PBA24-5	0.0489	0.0426	0.1285	0.0540	0.0454	0.1595
PBA24-6	0.0145	0.0128	0.1180	0.0205	0.0182	0.1152
PBA13-1	0.3315	0.2457	0.2588	0.3681	0.2809	0.2370
PBA13-2	0.0876	0.0820	0.0642	0.1189	0.1130	0.0489
PBA13-3	0.3826	0.3541	0.0745	0.3379	0.3315	0.0190
PBA13-4	0.2118	0.2039	0.0373	0.2468	0.2418	0.0202
PBA13-5	0.0646	0.0623	0.0360	0.0767	0.0739	0.0364
PBA13-6	0.1490	0.1441	0.0329	0.166	0.1614	0.0272
PBA13-7	0.1866	0.1768	0.0523	0.2205	0.2107	0.0446
PBA13-8	0.1367	0.1286	0.0591	0.0993	0.0981	0.0125
PBA15-1	0.4616	0.3184	0.3103	0.3471	0.3016	0.1312
PBA15-2	0.4652	0.3967	0.1472	0.4655	0.3801	0.1836
PBA15-3	0.2033	0.1855	0.0875	0.1921	0.1784	0.0711
PBA15-4	0.4776	0.2547	0.4667	0.3202	0.2637	0.1765
PBA15-5	0.2883	0.2696	0.0648	0.2616	0.2511	0.0401
PBA28-1	0.2004	0.1882	0.0612	0.2527	0.2440	0.0346
PBA28-2	0.4304	0.3329	0.2265	0.4447	0.3702	0.1675
PBA28-3	0.3040	0.2203	0.2751	0.2035	0.1863	0.0848
PBA28-4	0.0102	0.0094	0.0824	0	0	-----
PBA28-5	0.3555	0.2496	0.2979	0.3412	0.2369	0.3057

PBA28-6	0.0143	0.0135	0.0603	0.0031	0.0031	0.0172
PBA28-7	0.0683	0.0659	0.0347	0.0849	0.0825	0.0290
PBA28-8	0.0020	0.0020	0.0164	0.0029	0.0028	0.0160
PBA28-9	0.1228	0.0859	0.3005	0.0639	0.0530	0.1711
PBA18-10	0.0412	0.0370	0.1017	0.0286	0.0275	0.0404
PBA28-11	0.0050	0.0048	0.0402	0.0071	0.0068	0.0392
PBA22-1	0.0128	0.0124	0.0305	0.0151	0.0146	0.0331
PBA22-2	0.0021	0.0021	0.0170	0	0	----
PBA22-3	0.0044	0.0043	0.0166	0.0062	0.0061	0.0157
PBA22-4	0.0352	0.0332	0.0578	0.0200	0.0195	0.0271
PBA22-5	0.3379	0.3213	0.0491	0.3369	0.3307	0.0184
PBA22-6	0.3727	0.2758	0.2600	0.2929	0.2851	0.0264
PBA22-7	0.1445	0.1385	0.0412	0.1636	0.1571	0.0397
PBA22-8	0.1873	0.1733	0.0751	0.2332	0.2196	0.0579
PBA22-9	0.0486	0.0473	0.0274	0.054	0.0525	0.0277
PBA22-10	0.0165	0.0160	0.0280	0.0143	0.0140	0.0169

Table 4 Expected heterozygosities in mainland and maritime jack pine populations. Allele frequencies of markers in bold and underlined were found to have a significant correlation with latitude and longitude. Markers in bold were found to have a correlation with latitude, while those only underlined were found to have a correlation with longitude.

	Mainland	NS1	NS2	NB2	NB1	PE1
PBA24-1	0.4926	0.3065	0.0000	0.4056	0.0000	0.4954
PBA24-2	0.4891	0.4972	0.3095	0.3492	0.4935	0.4329
PBA24-3	0.4997	0.5000	0.2794	0.4944	0.4589	0.4944
PBA24-4	0.1685	0.3924	0.0753	0.3492	0.2530	0.0973
PBA24-5	0.0499	0.0702	0.0753	0.0000	0.0342	0.0000
PBA24-6	0.0219	0.0000	0.0000	0.0000	0.0000	0.0000
PBA13-1	0.4355	0.0000	0.0000	0.4959	0.3109	0.0000
PBA13-2	0.1170	0.0000	0.0000	0.0000	0.0369	0.0000
PBA13-3	0.3331	0.4910	0.4994	0.2815	0.2703	0.4994
PBA13-4	0.2454	0.0000	0.0330	0.1949	0.2398	0.0973
PBA13-5	0.0770	0.0000	0.0000	0.0342	0.0726	0.0655
PBA13-6	0.1667	0.0354	0.1590	0.1006	0.0000	0.2174
PBA13-7	0.2147	0.1698	0.0655	0.1006	0.1421	0.0330
PBA13-8	0.1000	0.3441	0.3492	0.0342	0.0367	0.2461
PBA15-1	0.3334	0.0000	0.4998	0.3249	0.4856	0.4381
PBA15-2	0.4416	0.4812	0.3544	0.5000	0.4987	0.3492
PBA15-3	0.1861	0.3690	0.0000	0.4142	0.1643	0.0655
PBA15-4	0.3869	0.2014	0.2997	0.1590	0.1328	0.3722
PBA15-5	0.2556	0.0702	0.2703	0.3311	0.4848	0.4142
PBA28-1	0.2458	0.0342	0.0655	0.0973	0.0330	0.0413
PBA28-2	0.4726	0.4241	0.0000	0.4944	0.2985	0.0000
PBA28-3	0.1972	0.1006	0.4324	0.3492	0.2178	0.4107
PBA28-4	0.0000	0.0000	0.0000	0.1590	0.0000	0.0000
PBA28-5	0.4274	0.4848	0.0000	0.4831	0.4324	0.0000
PBA28-6	0.0030	0.0000	0.0000	0.0330	0.0000	0.1590

PBA28-7	0.0849	0.0000	0.0973	0.0330	0.0000	0.0000
PBA28-8	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000
PBA28-9	0.1428	0.0000	0.0000	0.4995	0.0000	0.3249
PBA18-10	0.0268	0.0000	0.0000	0.0000	0.2997	0.0000
PBA28-11	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
PBA22-1	0.0153	0.0000	0.0000	0.0354	0.0000	0.0000
PBA22-2	0.0000	0.0000	0.0000	0.0000	0.0354	0.0000
PBA22-3	0.0062	0.0000	0.0000	0.0000	0.0000	0.0000
PBA22-4	0.0216	0.0354	0.0000	0.0702	0.0000	0.2247
PBA22-5	0.3322	0.0354	0.1949	0.4142	0.4910	0.3589
PBA22-6	0.2900	0.0000	0.4856	0.3203	0.1041	0.3589
PBA22-7	0.1562	0.0702	0.1006	0.1041	0.0000	0.1949
PBA22-8	0.2321	0.0000	0.1006	0.1041	0.0702	0.0677
PBA22-9	0.0548	0.0000	0.0000	0.0354	0.1041	0.0342
PBA22-10	0.0153	0.0000	0.0000	0.0000	0.1041	0.0000

Table 5 Unbiased Nei's genetic distances among the populations within the jack pine mainland, and among maritime and mainland jack pine populations.

	Average ± Standard deviation	Range
Among all populations	0.0386±0.0020	0.0059~0.1205
Maritimes — mainland		
NS1	0.0812±0.0063	0.0444~0.1205
NS2	0.0593±0.0043	0.0379~0.0858
NB2	0.0504±0.0047	0.0265~0.0825
NB1	0.0416±0.0043	0.0233~0.0712
PE1	0.0486±0.0033	0.0350~0.0714
Within mainland	0.0217±0.0012	0.0059~0.0445

Table 6 Estimates of the number of migrants between mainland and maritime populations using the multiple loci estimates of F_{st} (Nm=(1/F_{st}-1)/4).

Maritimes-mainland	F_{st}	Nm
NS1	0.1521	1.39
NS2	0.1258	1.74
NB2	0.0723	3.21
NB1	0.0724	3.20
PE1	0.0865	2.64

Table 7 Estimates of the number of migrants from mainland to maritime populations. "—" stands for the estimates that could not be obtained either due to monomorphism or the reduced heterozygosity in mainland populations compared with that in maritime populations. The estimates were calculated using Yeh and Hu (2005) method at each locus in each maritime population, and using the Hudson (1998) method (value in parenthesis).

	NS1		NS2		NB2		NB1		PE1	
PBA24-1	0.46	(0.41)	—		0.57	(1.17)	—		—	
PBA24-2	—		0.47	(0.43)	0.51	(0.62)	—		0.61	(1.93)
PBA24-3	—		0.44	(0.32)	0.70	(23.32)	0.64	(2.81)	0.71	(23.32)
PBA24-4	—		0.65	(0.20)	—		—		0.69	(0.34)
PBA24-5	—		—		—		—	(0.54)	—	
PBA13-1	—		—		—		0.50	(0.62)	—	
PBA13-2	—		—		—		1.31	(0.12)	—	
PBA13-3	—		—		0.55	(1.36)	0.54	(1.08)	—	
PBA13-4	—		0.42	(0.04)	0.58	(0.96)	0.64	(10.71)	0.47	(0.16)
PBA13-5	—		—		—	(0.20)	—	(4.13)	—	(1.42)
PBA13-6	0.60	(0.07)	0.85	(5.16)	0.71	(0.38)	—		—	
PBA13-7	0.62	(0.95)	0.49	(0.11)	0.53	(0.22)	0.58	(0.49)	0.46	(0.04)
PBA13-8	—		—		3.64	(0.13)	3.78	(0.15)	—	
PBA15-1	—		—		0.61	(9.55)	—		—	
PBA15-2	—		0.54	(1.02)	—		—		0.53	(0.94)
PBA15-3	—		—		—		0.72	(1.88)	0.56	(0.14)
PBA15-4	0.43	(0.27)	0.52	(0.86)	0.40	(0.17)	0.39	(0.13)	0.61	(6.32)
PBA15-5	0.44	(0.09)	—		—		—		—	
PBA28-1	0.42	(0.04)	0.44	(0.09)	0.47	(0.16)	0.42	(0.04)	0.43	(0.05)
PBA28-2	0.61	(2.18)	—		—		0.47	(0.43)	—	
PBA28-3	0.57	(0.26)	—		—		—		—	

PBA22-5	0.35	(0.03)	0.46	(0.35)	—		—		—	
PBA22-6	—		—		—		0.42	(0.13)	—	
PBA22-7	0.72	(0.20)	0.79	(0.45)	0.80	(0.49)	—		—	
PBA22-8	—		0.49	(0.19)	0.50	(0.20)	0.47	(0.11)	0.46	(0.10)
Average	**0.52**	**(0.45)**	**0.55**	**(0.77)**	**0.81**	**(2.78)**	**0.84**	**(1.56)**	**0.55**	**(3.16)**

II 1. Introduction

II 1.1 Tropical rainforests

The tropical rainforests are one of the most complex biomes in terms of structure and species diversity. These forests are defined by their location in the tropics and the presence of significant rainfall. Presently occupying approximately seven percent of the world-wide land area, they are believed to contain nearly fifty percent of the world's species (Whitmore 1998). Although rainforests originally exhibited a much more widespread distribution, they have been receding due to past and present threats - particularly those from humans. There are four types of woody vegetation existing in the tropics; they include rain forests, then progressively drier and more seasonal, monsoon forests, savana forests, and thorn forests (Whitmore 1998).

Most species co-existing in rain forests are rare and endemic to particular regions. For example the species *Shorea cordifiolia*, belonging to the genus *Shorea* section *Doona*, is a locally abundant subcanopy species that can be found only in Sri Lanka's rainforests. About eighty percent of all green plants are angiosperms, most of which can be found in tropical rainforests. Some scientists estimate that a given 2.5 acres of land in these forests has the potential to contain over 750 types of trees and 1500 species of higher plants (Whitmore 1998). A vast number of medications, approximately one forth of those available today, as well as daily provisions, such as avocado, banana, or black pepper, originated in the tropical rainforests (Whitmore 1998). There are several components that contribute to the species richness of these biomes: interactions with animals acting mainly as pollinators or dispersers, a forest canopy providing large numbers of spatial and temporal niches, and a stable climatic history (Whitmore 1998).

Despite the long-standing debates (Ashton 1969), the origin of high biological diversity in tropical rainforests is still poorly understood. High rates of speciation or low rates of extinction under optimal environmental conditions are two components that may be responsible for the present day biodiversity in tropical rainforests.

The Malaysian forests have been categorized into 16 types with special reference to the dipterocarps; represented by lowland, hill, and upper dipterocarp forests (Symington 1943). The lowland dipterocarp forests are composed of approximately 130 species of dipterocarps that form a high proportion of the emergent and dominant strata. The hill dipterocarp forests have representatives of all the main dipterocarp groups, except *Dryobalanops*. The main difference between the hill and upper forest types is in the specific composition of the constituents of the upper strata. *Shorea curtisii* is predominant among these dipterocarps, and is a useful indicator of a hill dipterocarp forest. The upper dipterocarp forests usually contain few species; the typical predominant ones are *Dipterocarpus costatus* and *D. retusus* (Symington 1943, Pooma 2003).

Dipterocarps are confined to tropical climates with a mean annual rainfall exceeding 1,000 mm, and/or a dry season of less than six months (Ashton 1982). Mean annual rainfall and varying abiotic factors within different plots, are important factors in explaining differences in floristic composition between locations (Slik et al. 2003). The dipterocarps of South Asia are confined to tropical, moist, deciduous, and tropical evergreen forests in Bangladesh, moist tropical forest in India, lowland wet evergreen and highland wet evergreen forests in Sri Lanka, and tropical evergreen, tropical semi-evergreen, and dry deciduous forests in Myanmar (FAO 1985). The *Shorea* spp. found on the border between evergreen and the deciduous forest, attending the former state in very

moist fertile localities and the latter state in less moist or dry situations (Pooma 2003). Studies have identified five main floristic regions within the lowland dipterocarp rain forests of Borneo, each of which had its own set of characteristic genera (Slik et al. 2003).

Phylogenetic studies of selected groups of co-occurring congeneric species are extremely valuable to our understanding of the mode of speciation in rainforest inhabitants. They can be used to visualize networks that represent historical interactions between taxa resulting in the presently observed biodiversity. Phylogenetic analysis can provide a means to describe biodiversity, reveal population processes and consequently define conservation units in order to maintain lineages that represent the breadth of evolutionary diversity (Moritz 1995).

Phylogenetic inferences that are consistent with other data, be it comparisons among genetic information from various organelles or between molecular, temporal, or ecological sources, provide greater value in evaluating their relative importance than phylogenetic inferences alone. The significance of assessing the evolutionary history of an organism in a temporal scale, using molecular dating analysis, was elegantly highlighted using the Dipterocarpaceae family (Gunasekara 2004). Our approach to the study of the biogeography of dipterocarps consisted of obtaining nuclear gene based molecular data to reveal evolutionary relationships, comparing them to phylogenies based on data from other organelles, and subsequently uncover possible explanations for the presently observed disjunct distribution.

II 1.2 Phylogeny

Phylogenetic trees illustrate evolutionary relationships among genes or a given segment of a genome, and species. These trees provide a graphical representation depicting the degree of relatedness among a group of entities that are believed to share a common ancestor. Although phylogenetic trees can be constructed based on morphological characteristics, today they are most often constructed on the basis of molecular data (DNA, RNA, or protein). Therefore, the associations from one entity to another also depict the amount of genetic variation between them. Closely related taxa would have less genetic variation between them as compared to more distantly related ones, and would therefore cluster together. Since the data set is assumed to share a common ancestor, evolutionary patterns can be observed based on the topology of the tree.

There are three main methods of constructing phylogenetic trees, which differ from one another in their assumptions and algorithms of character state optimization: maximum parsimony, maximum likelihood, and distance-based methods such as neighbour-joining (Felsenstein 1988, Huelsenback 1993). Each method has its own advantages and limitations. Parsimony methods are commonly used in phylogenetic analysis but can generate inconsistent results. Such scenarios can occur when the rates of evolution among sequences show considerable variation or when the sequences are highly divergent (Page and Holmes 1998) resulting in long branch attraction; a tendency of species at the ends of long branches to be made artificially closer to each other than their evolutionary relatedness. Compared to parsimony models, the maximum likelihood model generally constructs consistent results, as parameters being estimated (branch

lengths) do not change as more nucleotide sites are considered, and permits additional statistical tests of evolutionary hypotheses. However, it is computationally intensive and consequently limits its applications to relatively small data sets. The distance-based methods are computationally highly efficient. Although their application to various data sets is not limited, this strategy does not guarantee to produce a minimum evolution tree.

The parsimony analysis principle suggests that one seeks the topology that minimizes the amount of evolutionary change, or the number of changes from one character sate into another, required to explain the data (Kluge and Farris 1969, Fitch 1971). The maximum parsimonious (MP) tree is the one with the minimum number of evolutionary changes (equivalent to the tree length) and the total number of changes between any given two sequences is attributed to respective branch lengths (Felsenstein 1983, Page and Holmes 1998). Hence, these assumptions dictate that the tree that minimizes change is probably the best estimate of the actual phylogeny. Parsimony informative sites are identified when a given site provides information that distinguishes different topologies from one another. When the observed substitution rate or transition-transversion bias is low in the DNA sequence matrix, all characters are given equal costs and equally weighed parsimony settings are used for analyses.

The maximum likelihood (ML) method infers phylogenetic relationships based on the parameters, such as transition/transversion ratio, base frequencies, and rate of variation among sites, of an evolutionary model. In other words, the evolutionary associations are constructed based on finding the highest likelihood or probability ratio that the proposed model and the hypothesized history would give rise to the observed data set. Similarly to the MP method, ML analysis generates the tree with the highest

likelihood through estimating branch lengths (number of changes among informative sites between sequences), given a model of evolution. Therefore the tree that best fits the observed data set as the most likely evolutionary outcome is regarded as the maximum likelihood estimate of phylogeny.

In distance based methods, the goal is to create a distance matrix based on the number of substitutions that have occurred between all pairs of taxa by applying a specific evolutionary model that makes assumptions about the nature of evolutionary changes. The neighbor-joining (NJ) method constructs phylogenetic trees based on a matrix of pair-wise distances between taxa. It clusters pairs of taxa in an additive method while minimizing the total branch length of each stage of clustering (Saitou and Nei 1987). For nucleotide sequence data, branch lengths are calculated from the fraction of sites that differ between two sequences. Therefore the output that minimizes the total tree length based on the calculated distance matrix, given an evolutionary model, is the neighbor joining evolutionary reconstruction.

II 1.3 Literature review

The tropical tree family Dipterocarpaceae, more commonly known as dipterocarps, dominates the Asian rainforests. This family is composed of over 500 species with co-occurring congeneric species groups. It is one of the region's most important resources due to their extreme species richness, economical value, and has been used as a source of timber and non-timber forest products (Appanah 1998, FRA 2000). This particular family of the south East Asian lowland evergreen forest has been of particular interest to investigators for their noticeably distinct groups (Ashton 1982, Ashton 1988, Whitmore 1989, Whitmore 1998, Maury and Curtet 1998). The classification of the three subfamilies, Dipterocarpoideae, Monotoideae, and Pakaraimoideae, within Dipterocarpaceae has been accepted based on unique floral characteristics, reproductive biology and anatomical features (Maguire and Ashton 1977, Maury and Curtet 1998).

However, classification among genera within dipterocarps has been controversial. The classical way of inferring phylogenetic relationships has been by comparing morphological characteristics. Before estimating any historical relationship between individuals, an understanding of the biological mechanism of change influencing the species can be insightful. Fedorov (1966) discussed the role of genetic drift in dipterocarp biology, and concluded that genetic drift, or the random change in allelic frequency occurring by chance in small and isolated populations play a more prominent role in dipterocarp evolution than natural selection. Since individuals of the same species do not tend to aggregate together, they would each adapt to their environment and eventually become distinct from one another. Even if pollinators or wind dispersal mechanisms are

effective, mating will not occur because the two individuals will most likely develop different flowering periods and thereby prevent cross fertilization (Fedorov 1966).

Ashton (1969) noted that tropical tree species are sometimes distributed in clusters while other groups are more dispersed. Hence, dipterocarps may not necessarily be as isolated as suggested by Federov (1966). Furthermore, Ashton (1969) pointed out that flowering does occur at the same time between individuals thereby permitting cross-pollination at great distances through long distance pollination by animal vectors. Ashton (1969) suggested that the theory of natural selection influenced Dipterocarpaceae biology more than genetic drift.

The classification of dipterocarps has been established based on morphological data, fossil records and pollen analysis (Ashton 1982, Maury and Curtet 1998) and only few molecular based approaches have been reported. Studies by Chase et al. (1993) and Dayanandan (1996) placed the Dipterocarpaceae family in the order Malvales based on the *rbc*L chloroplast gene, which codes for the large subunit of RuBisCo (enzyme used to incorporate CO_2). Alverson (1998) supported this classification. Alverson's study was more comprehensive and included a large number of taxa which supported the Malvalean affinity of the Dipterocarpaceae. Molecular tools, in particular sequence data, as four nucleotide character states, are discreet and less homoplastic in contrast to morphological features, and provide higher resolution power to decipher phylogenetic relationships (Alverson 1998).

Based on *rbc*L sequence and morphological data, Dayanandan (1995) pointed out that the African genus, *Monotes*, is more closely related to Asiatic dipterocarps than to the Tiliaceae, and the genera *Shorea* and *Hopea* are poylphyletic. Furthermore, he

suggested that certain groups of species are a result of recent diversification as evident by relatively low level of *rbc*L sequence variation in *Shorea* section *Doona*.

Using restriction fragment length polymorphism (RFLP) methods on chloroplast DNA, Tsumura et al (1996) examined genetic variation among 30 species from 10 genera. They obtained a strict consensus tree using the maximum parsimony method and a similar neighbour-joining tree. The position of *S. multifolia* and slight differences in positions of *Parashorea* and some *Shorea* species varied between the two tree topologies. They were able to clearly differentiate between tribe Dipterocarpeae, chromosome number = 11, and tribe Shoreae, chromosome number = 7. In addition to chromosome number, the tribe Dipterocarpeae is characterized by valvate arrangement of the fruit sepals and consists of more than 150 species in 8 genera (*Dipterocarpus, Upuna, Cotylelobuim, Stemenoporus, Anisoptera, Vatica, Vateriopsis, and Vateria*). In contrast, the tribe *Shoreae* is distinguished based on the overlap arrangement of the fruit sepals as well as chromosome number and is composed of over 300 species classified within 6 genera (*Hopea, Shorea, Balanocarpus, Neobalanocarpus, Parashorea* and *Dryobalanops*). Tribe Shoreae is shown to be monophyletic. Due to the observed phylogenetic topology of the tree, Tsumura et al. (1996) concluded that the change in chromosome number (x) must have occurred at the base of this taxon and from $x = 11$ to $x = 7$. They were also able to show that there is a close affinity between *Parashorea lucida*, *Shorea*, *Hopea*, and *Neobalanocarpus* species, and that *Cotylelobium* and *Vatica* are closely related (Tsumura 1996). The authors found that from the 11 genes studied, 4 chloroplast genes, *rpo*C, *rbc*L, *pet*B, and *trn*K, were the most suitable for molecular

phylogenetic studies. Yet even those four were not variable enough to infer infrageneric relationships (Tsumura 1996).

Kajita *et al.* (1998) identified two potential problems in the study by Tsumura *et al.* (1996). The outgroup selection, *Upuna*, was unreliable and unjustified. Secondly, no confidence levels were reported for the potential phylogenetic topologies provided (Kajita 1998). Using many of the same species, Kajita *et al.* (1998) sequenced the *mat*K, *trn*L/F, and *trn*L intron regions for 17 species among 11 genera. Maximum parsimony analysis yielded 2 equally parsimonius trees, one tree was obtained from neighbor-joining analysis, and bootstrap analysis provided confidence levels for each tree respectively (Kajita 1998). The rate of evolution of *mat*K was found to be higher than *trn*, due to increased number of nucleotide substitutions; it has also been suggested that the *mat*K 5' end is more variable for lower taxonomic levels compared to its' 3' end (Kajita 1998, Hilu 1996). The phylogenetic relationship between *Parashorea* and other *Shorea* species were resolved, whereas Tsumura *et al.* (1996) failed to produce a clear topology. Both studies agreed that $\chi = 7$ was the derived state.

However inconsistent with previous classification based on chromosome number (Tsumura et al., 1996), the species of *Dipterocarpus*, from the Dipterocarpeae tribe, was considered a genus monophyletic with the tribe Shoreae (Kajita 1998, Dayanandan 1999). The genus *Dipterocarpus* has been reported as basal taxa of the sub-family Dipterocarpoideae and the genus *Dryobalanops* has been recognized as intermediate between the two sub-families (Meijer 1979, Maury and Curtet 1998). Dayanandan et al. (1999), using the chloroplast gene *rbc*L, deciphered relationships among 35 species representing 20 genera belonging to 5 families. This study had difficulties with

establishing the species *Dipterocarpus* and *Dryobalanops* topologies. However, they were able to resolve the relationships between the two tribes Dipterocarpeae and Shoreae. Some infrageneric relationships were resolved but more studies involving representative taxa from *Hopea*, *Shorea*, *Dryobalanops*, and *Dipterocarpus* are needed to further clarify the phylogenetic relationships within Dipterocarpaceae. In light of these obscurities the classification of Dipterocarpaceae should be reconsidered.

Recently, phylogenetic studies of the Ditperocarpaceae family have included the use of nuclear genes (Kamiya et al. 2005), as well as molecular dating analysis (Gunasekara 2004). Kamiya et al. (2005) reconstructed the phylogeny of dipterocarps using the nuclear *PgiC* gene. The number of taxon used in this study was limited, but the phylogenetic tree topology was consistent with cpDNA trees.

II 1.4 Objectives

It should be noted that all the molecular studies, with the exception of the nuclear gene *PgiC* based study (Kamiya et al. 2005), have examined data from the chloroplast genome. Since chloroplast DNA is maternally inherited, the reported topologies might not reflect the complete picture as other sources of genetic information can be obtained from the other organelles representing paternal and biparentally inherited data. In comparing the topologies gathered from chloroplast, mitochondrial, and nuclear DNA can a comprehensive mode of evolution be proposed. The *PgiC* gene based phylogeny (Kamiya et al. 2005) had limited taxa representation, therefore the objective of the present study is to reconstruct a comprehensive phylogeny of Dipterocarpaceae based on a nuclear 18S rRNA gene sequence data including representative taxa from all genera.

Materials and Methods

II 2.1 Samples

Leaf samples were collected and dried for long-term storage. We followed the classification of Ashton (1982). One to two individuals representing each genus or section were used in the present study (Table 8). Total genomic DNA was extracted from dried leaf samples using a modified version of the protocol outlined by Dayanandan et al. 1997.

We obtained dried leaf samples (either 1 to 2 leaves or 0.75g to 1.5g respectively) and placed them in a clean mortar with a pestle. We added enough liquid nitrogen to cover the leaves and the tip of the pestle. Once the liquid nitrogen had evaporated, we vigorously crushed the leaves into powder. Then we added 0.03g to 0.04g of dried leaf powder to a 1.5ml centrifuge tube containing 600µl Extraction/Nucleic Lysis (1:1) buffer and 120µl 5% Sarkosyl. We inverted the tubes several times and placed all the samples in 65°C water bath for 3 hours and inverted the tubes intermittently. Following the incubation period, we centrifuged the samples at 7000 rpm for 5 minutes and decanted supernatant into a new 1.5ml centrifuge tube. In the fume-hood, we added 600µl Chloroform/Isoamly (24:1) solution. The tubes were left at room temperature for 10 minutes while gently inverting tubes to form an emulsion intermittently.

The tubes were then centrifuged at 7000 rpm for 5 minutes and the upper phase was transferred to a clean 1.5ml centrifuge tube containing 600µl Isopropanol. The tubes were gently inverted and the DNA was precipitated out of the solution at room temperature for at least 2 hours. Then the tubes were placed in -20°C freezer overnight.

The following day, they were centrifuged at 14,000 rpm for 10 minutes and the supernatant was decanted. The pellets were washed with 250µl 70% ethanol and centrifuged at 14,000 rpm for 10 minutes. The supernatant was subsequently decanted and the wash was repeated is the pellets were darkly colored. The pellets were dried in a vacuum spinner for 30 minutes at medium temperature. Finally, the dried pellets were dissolved in 100µl TE buffer. Using the Qiagen DNeasy Kit the samples were purified according to the protocol outlined in the manufacturer provided manual.

We amplified and sequenced a section of the nuclear 18S rRNA gene (hereafter referred as the N18S segment) using the primers N18S1 and N18C18H (Bult et al. 1992). The 18S rRNA segment has been used for phylogeny reconstruction among eukaryotes, including most major groups of plants (Kuzoff et al. 1998) such as green algae (Chapman and Buchheim 1991) and angiosperms (Soltis et al. 1997). PCR amplification reactions consisted of 230 µM dNTP, 2.5 mM $MgCl_2$, 10 µM each primer, 1 unit of Taq DNA polymerase, and 2.5 µL buffer (0.2M Tris PH 9.5; 0.25 M KCl; 1 mg/ml BSA, 5 µl/ml tween 20) in a total volume of 25 µl (Appendix 1, Protocol 5). The thermal cycling profile of PCR reactions, consisted of an initial denaturation step at 94°C for 3 minutes, followed by 40 cycles of: 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 2 minutes. These parameters were followed by a final elongation step at 72°C for 5 minutes. Amplified products were electrophoresed on a 1% agarose gel with ethium bromide (0.33 µg/ml) at 5.5 v/cm for 35 minutes (Appendix 1, Protocol 4 for agarose gel recipe). PCR products were then purified and sequenced using the same primers as those used in the PCR reactions. Each individual was sequenced twice, at both forward and reverse directions. The chromatograms of the DNA sequencing results were processed

and analyzed using the Staden software package (Staden group, MRC Laboratory of Molecular Biology, Cambridge, UK). The assembled contigs were aligned using ClustalW (Thomson et al. 1994) multiple sequence alignment software. Alignments were subsequently imported into MacClade 4.0 software (Maddison & Maddison 2001) for verification and manual editing of the sequence alignments. The final data set included 27 taxa.

II 2.2 Phylogenetic Analysis

The phylogenetic analyses of the present study were conducted using PAUP 4.0 beta 8 (Swofford 2001). For each data set 1000000 random trees were analyzed to obtain the frequency distribution of tree lengths to assess the phylogenetic signal of the data matrix by calculating the skewness (g1) and kurtosis (g2). The nucleotide frequency distribution was examined by the chi-square test of homogeneity of base distribution across all sequences.

Maximum parsimony phylogenetic trees were reconstructed through conducting a heuristic search with equal character weights, gaps treated as missing data, multi state taxa were interpreted as uncertainty, starting tree was obtained by stepwise addition, and random sequence addition was used for the data set. Tree-bisection reconstruction (TBR) was used as the branch swapping algorithm, and the consensus tree was obtained through strict and 50% rules. The trees were rooted using *Durio zibathinus* species, from the family tropical tree family Bombacaceae. To examine the confidence levels of the branches, bootstrap analysis with fast-heuristic search based on 1000 replicates was conducted to complete the analysis.

Modeltest v3.06 (Posada 1998) was used to find the optimal nucleotide substitution model that fits the data set using hierarchical likelihood ratio test. The ML analysis was performed through heuristic search with TBR branch swapping, addition of sequences as-is and the Modeltest v3.06 (Posada 1998) parameters as nucleotide substitution model. To examine the confidence levels of the branches, bootstrap analysis with full-heuristic search based on 1000 replicates was conducted to complete the analysis.

We have selected the evolutionary model for our data set based on results from Modeltest v.3.06 (Posada 1998) and applied the parameters to reconstruct phylogenetic trees. Robustness of branches was tested by bootstrapping 1000 replicates with the neighbor-joining (NJ) search algorithm. When more than one tree was generated from any of the algorithms used, MP, ML, or NJ, two methods were used to obtain the consensus trees. The strict consensus method means that only clades that show up in all of the most parsimonious trees are recognized and retained. Whereas the majority-rule approach only retains branches that occur more than 50% of the time in all of the most parsimonious trees generated. When relationships are not maintained among all optimal trees, the majority-rule analysis will reveal a more resolved tree; in the strict consensus the branches are collapsed if they are not found in all of the trees.

II 3 Results

II 3.1 Preliminary data analysis

Direct sequencing of purified PCR products of the N18S rRNA segment was followed by ClustalW alignment of all sequences and resulted in a 1030 character (nucleotide) long data matrix. The data matrix was composed of 27 taxa with 27 parsimony informative sites. The frequency distribution calculated from 10^6 random trees, yielded g1 = - 0.437036, and g2 = 0.203389 suggesting a significant phylogenetic signal in the data matrix. The chi-square test of homogeneity of base distribution across all sequences was used to check if the observed nucleotide frequencies differed significantly from the expected frequencies. This test revealed p = 1.0 allowing us to accept the null hypothesis; the nucleotide frequency distribution is homogeneous and consistent with the expected values.

II.3.2 Maximum parsimony (MP) analysis

Maximum parsimony analysis generated 277 equally parsimonious trees. The 50% majority rule consensus tree and the strict consensus tree differed in several relationships. Most internal nodes were not supported by bootstrap analysis (Figure 4). Although *Monotes* and *Marquesia* are grouped together, they are placed within the Dipterocarpoideae subfamily as opposed to other studies, which have previously classified them within the subfamily Monotoideae (Dayanandan et al. 1999, Ducousso et al. 2004).

Distinction between tribes Ditperocarpeae and Shoreae based on chromosome number is maintained. The associations of the tribe Shoreae in the strict consensus tree are collapsed but maintained in the majority rule. Furthermore, the phylogenetic relationships among taxa within this clade remain unresolved. Relationships between *Hopea odorata*, *Hopea brevipetiolaris*, *Hopea dryobalanoides*, and *Neobalanocarpus heimii* remain unresolved. Although they all belong to the same tribe, Shoreae, their reported topology is intertwined. *Hopea odorata*, *Hopea brevipetiolaris*, and *Hopea dryobalanoides* are considered different species belonging to the same genus, yet the placement of *Hopea dryobalanoides* occurred as a sister taxa to *Neobalanocarpus heimii*, genus *Neobalanocarpus*, and separated from the other *Hopea* species. The position of *Neobalanocarpus heimii* in our study, Figure 6, suggests evidence of reticulate evolution; this specie is likely a result of hybridization between *Shorea* and *Hopea* taxa.

Within the tribe Dipterocarpeae, relationships between species of the same genus are maintained; *V. venulosa* and *V. bantamensis*, *D. zeylanicus* and *D. glandulosus*, in addition to *Stemenoporus gilimalensis* and *Stemenoporus sp.* (collected from

Sooriyakanda) group together within their genera *Vatica*, *Dipterocarpus*, and *Stemenoporus* respectively. Two major clades are depicted within this tribe. The first consists of the genera *Dipterocarpus* and *Vateria*. The second is composed of *Upuna*, *Stemenoporus*, *Vateriopsis*, *Cotylelobium*, and *Vatica*.

II.3.3 Maximum likelihood (ML) analysis

Modeltest analysis suggested that the SYM+I+G (Zharkikh 1994) nucleotide substitution model, with the following parameters, was the most suitable given our data:

a. Equal base frequencies
b. Rate matrix: (A-C)=0.2940; (A-G)=3.0287; (A-T)=1.4838; (C-G)=0.2326; (C-T)=1.4278; (G-T)=1.0000
c. Among-site rate variation: proportion of invariable sites = 0.8372
d. Gamma distribution shape parameter = 0.5734

The maximum likelihood analysis was conducted with the above parameters and yielded one tree. Similar relationships are maintained between ML and MP topologies. With the exception of the genus *Shorea*, the ML tree is fully resolved. Bootstrap values did not support most of the internal nodes (Figure 7) and are very similar to those obtained under the parsimony criterion.

With the exception of the placement of the species *S. cordifolia*, and the species from the subfamily Monotes, topologies within the tribe Shorea are identical to those depicted under the parsimony criterion. The species *S. cordifolia* is placed basal in this tribe in contrast to the parsimony topology where it nestled within the group. In the parsimony analysis, the species from the subfamily Monotes is basal to the tribe Dipterocarpeae in contrast to the tribe Shoreae and the species from the subfamily Marquesia.

Within the major clade Dipterocarpeae, *Upuna* is placed basal to two subsequent groups. The first contains the genera *Dipterocarpus*, *Vatica* and *Vateria*. The second

group is composed of the following genera: *Stemenoporus*, *Vateriopsis*, and *Cotylelobium*.

II.3.4 Neighbor-joining (NJ) analysis

Using the same parameters suggested by Modeltest for ML analysis, the NJ algorithm retained only one tree. This tree does depict fully resolved relationships, and is similar to those obtained from MP and ML analysis. Bootstrap values did support several internal nodes (Figure 8) in comparison to the previous methods of analysis.

The subfamilies Monotoideae, *Monotes sp.* and *Marquesia sp.*, are found basal to the Dipterocarpoideae subfamily. Within the Dipterocarpoideae subfamily the same two major clades, consistent with chromosome number distinction, exist as in the previous methods of analysis. Basal to the first clade, tribe Shoreae, is the species *Shorea siamensis*. Similar relationships are maintained within the group as the MP and ML methods.

Relationships within the tribe Dipterocarpeae are more resolved but overall topologies are very similar to MP and ML analysis. The genus *Upuna* is depicted to be ancestral to the genera *Dipterocarpus*, *Vatica* and *Vateria*. The taxa *Stemenoporus*, *Vateriopsis*, and *Cotylelobium* are depicted as sister taxa. In cases where greater than one species was sequenced per genus, such taxa cluster together; for example *Dipterocarpus glandulosus* and *Dipterocarpus zeylanicus*, the two *Stemenoporus* individuals, and the three *Vatica* individuals.

II.4 Discussion

II.4.1 Comparisons between MP, ML, and NJ topologies

DNA sequences and other molecular data compared among organisms may contain phylogenetic signal, or they may be randomized with respect to phylogenetic history. Some method is needed to distinguish phylogenetic signal from random noise to prevent analyzing data that has been randomized with respect to the historical relationships among the taxa being compared. Distributions of tree lengths with a left skew indicate that relatively few solutions exist near the optimal solution compared to elsewhere in the distribution (Hillis and Huelsenbeck 1992). This, in turn, is an indication of correlation among characters beyond that expected at random. Comparison between g1 critical values (Hillis and Huelsenbeck 1992) and our result suggest that the performance of analyses in finding the true phylogeny has greater than 99% chance of occurring.

Estimating the robustness of branches through statistical measures, such as bootstrap, has been controversial (Brocchieri 2001, Henderson 2005). The bootstrap method analyzes a set of nucleotide sites randomly, with replacement, and constructs a new tree (Felsenstein 1985). This process is repeated, in our case 1000 times, and the frequency of appearance of a particular node among the bootstrap trees is viewed as a support or confidence value for deciding on the significance of that node. Recall that only the informative sites provide information relevant to constructing phylogenetic trees. If during resampling these sites are not selected, due to randomization effects, the same tree will evidently not be re-produced. As our data set has a low proportion of informative sites, our low bootstrap values can be expected.

ML and NJ methods use explicit statistical models of nucleotide evolution to estimate phylogenetic trees. The objective of model selection is to find the model that will allow one to most accurately estimate unknown phylogenetic parameters while avoiding bias and excessive variance (Bos and Posada 2005). Selection of the most appropriate model for a given data set depends on several criteria; substitution rates (rates of evolution between pairs of nucleotides), overall among-site rate variation at nucleotide positions (gamma distribution), and nucleotide frequencies (Bos and Posada 2005). Applying a specific model of nucleotide sequence evolution prior to constructing phylogenetic trees, such as in NJ and ML analysis, will increase the chances of generating a consistent tree; whereas both weighted and un-weighted parsimony methods are generally less efficient than the NJ and ML methods even in the case where the MP method gives a consistent tree (Tateno et al. 1993). When all the assumptions of the ML method are satisfied, this method is slightly more efficient than the NJ method (Tateno et al. 1993). However, when the assumptions are not satisfied, the NJ method with incorporated optimal gamma distances is slightly more efficient in generating the correct topology than is the ML method (Tateno et al. 1993).

Many phylogenetic relationships were not fully resolved in the MP analysis. With the exception of *Balanocarpus utilis*, belonging to the tribe Dipterocarpeae, the distinction between the two tribes of Dipterocarpaceae is maintained as is the distinction in chromosome number in the majority rule topology. In the strict consensus version, the clade corresponding to the tribe Dipterocarpeae is collapsed, but suggests that chromosome 7 was derived. Genera *Upuna*, *Vatica*, *Vateriopsis*, *Cotylelobium*, and *Stemenoporus* form a terminal clade. The grouping is consistent with other studies

(Tsumura et al. 1996, Kajita et al. 1998, Dayanandan et al. 1999), but internal nodes differ within the clade. *Vatica* and *Cotylelobium* have described as sister groups (Tsumura et al. 1996, Kajita et al. 1998). In our analysis, *Vatica* is a sister taxa to *Upuna* and gives rise to *Cotylelobium*. The relationship depicted in the MP majority rule analysis between *Stemenoporus* and *Vateriopsis*, as well as those within the tribe Shoreae, are consistent with other studies based on chloroplast DNA analysis (Tsumura et al. 1996, Kajita et al. 1998, Dayanandan et al. 1999). However, the placement of *Neobalanocarpus hemii* is controversial. MP analysis has placed it basal to the chromosome 7 clade that consists of the genera *Shorea* and *Hopea*, whereas other studies place species from the genus *Shorea* basal to the chromosome 7 clade. MP methods are effective when site-to-site heterogeneity is moderate and the matrix contains enough parsimony informative sites. As our matrix only generated 27 informative sites from over 1400 characters, it likely impeded production of a consistent phylogeny.

While maintaining the same basal relationships in MP analysis, associations within each major clade are more resolved in ML analysis. Inconsistencies such as the division of *Monotes* and *Marquesia*, and *V. bantanmensis* and *V. venulosa* whom each pair belong to a unique genus and whose associations are supported by morphological and molecular features (Ashton 1982, Dayanandan et al. 1999), suggest that the proposed topology does not reflect the most accurate evolutionary relationships (Gunasekara 2004).

Most phylogenetic relationships are resolved in the NJ tree. NJ analysis had the highest probability of generating the most accurate topology. When the gamma correction and other model parameters defined by Modeltest were incorporated in the NJ analysis, it

became more efficient than the ML method in constructing the most accurate phylogenetic tree since not all the assumptions were met (Tateno et al. 1993). Assumptions were violated in the ML analysis due to limitations in PAUP*4; the nucleotide substitution model defined by Modeltest to construct the tree was not available.

NJ analysis divides the monophyletic tribe Dipterocarpeae into two major clades. The first includes *Stemenoporus*, *Vateriopsis* and *Cotylelobium* while the other includes *Dipterocarpus*, *Upuna*, *Vateria*, and *Vatica*. This division is consistent with the notion that *Stemenoporus*, and *Vateriopsis* are the most recent ancestors within the tribe Dipterocarpeae (Gunasekara 2004) and consistent with other studies suggesting that 7 chromosomes is the derived or most evolutionary recent state (Tsumura et al. 1996, Kajita et al. 1998). Relationships within the monophyletic tribe Shoreae depict a close affinity among *Hopea*, *Neobalanocarpus*, and *Shorea*.

4.2 Phylogenetic relationships among dipterocarps

Disagreements concerning topology of *Dipterocarpus* and *Dryobalanops* remain to be addressed (Kajita et al. 1998, Dayanandan et al. 1999). The differentiation between chromosome numbers, that was once sufficient to classify the two tribes as monophyletic, is no longer fully accepted. Classification of *Dipterocarpus* in our study is consistent with chromosome number division (Tsumura et al., 1996). Although the phylogenetic tree reconstruction results, particularly the classification of *Dipterocarpus* and *Dryobalanops*, is consistent with the theory that dipterocarps originated in Gondawana and migrated eastwards toward Southeast Asia, the relationships remain unresolved (Dayanandan et al. 1999). In light of this new evidence further studies, using several species from each taxa and sequencing of other genomic regions, are required to elucidate the position of these genera.

Gathering and examination of mtDNA or nuclear DNA in comparison with the cpDNA information published will enhance our understanding and resolve the phylogeny of dipterocarps. Although mtDNA is also uniparentally inherited, from the mother, when the analysis is coupled with cpDNA and nuclear DNA, we would then be able to describe the phylogeny of dipterocarps more precisely. The placement of *Neobalanocarpus* has been debated. Our results place this species nested within *Hopea* and *Shorea* species. This arrangement is congruent with the phylogeny based on *PgiC* (nuclear) and *matK* sequence data analysis (Gunasekara 2004, Kamiya et al. 2005); providing further evidence that *Neobalanocarpus* is derived via hybridization between *Shorea* and *Hopea*.

As can be seen in all the methods of analysis, terminal lineages within the genus *Shorea* remain unresolved. Developing a method to resolve the relationships of such

terminal lineages would be valuable for the understanding of how the species evolved in their unique environment. Our attempts to decipher the relationships of such terminal lineages, such as *Shorea* section *Doona* and *Stemenoporus*, demonstrated the close degree of relatedness with little or no polymorphism suggesting a recent diversification of these species groups. Phylogenetic studies in conjunction with ecological, biogeographical and edaphic relationships among these species will be valuable to understand the mode of speciation prevalent among tropical forest trees.

4.3 Biogeography of Dipterocarpaceae

Plant geography has been described as the study of the distribution and biological processes of plants in attempt to understand the geographical range of particular taxa or floras as a means to elucidate their origin, evolution and dispersion (Tivy 1993). Contemporary disjunct distributions of various taxonomic units reflect the historical influence of geological and climatic events (Gunasekara 2004). It is widely held that angiosperms evolved in west Gondwanaland and radiated into east Gondwanaland (Australia) and through South America and Africa into Laurasia (Audley-Charles 1987, Morley 1998, 1999, Pooma 2003).

The tropical tree family Dipterocarpaceae presently dominates the west Malesian and Indian/Sri Lankan rainforests. Dipterocarp pollen first appeared in the fossil record of west Malesia about 30 million years ago (Adams 1994, Pooma 2003). The timing of this event follows the collision of the India-Ceylon plate with southern Asia during the early Tertiary, specifically in the middle Eocene (Adams 1994, Pooma 2003). Many African plant species dispersed into India, and many of their descendants subsequently dispersed into Southeast Asia following the collision. This can explain the present distribution of Dipterocarpaceae, originating from Africa, or possibly South American plate with subsequent invasion from India (Morley 2001). The ages estimated as the point of origin of dipterocarps are consistent with those of the origin of angiosperms based on *rbcL* and *18S* rRNA sequences (Sanderson and Doyle 2001, Wikstrom et al. 2001, Gunasekara 2004). Evidence from dipterocarp pollen dated to the late oligocene and early miocene, suggests that at this time dipterocarps were common in Southeast Asia.

The split between the subfamilies Pakaraimoideae and Monotoideae, documented to have occurred 119 Mya, and current distribution, support the existence of direct land connections between the South American and African plates (Gunasekara 2004). Pakaraimoideae is considered the most archaic subfamily due to its simplified pollen structure as compared to the specialized walls and exine structures of the Monotoideae and Dipterocarpoideae subfamilies (Gunasekara 2004). Fossil records from East Africa suggest the presence of the Dipterocarpoideae subfamily in African forests until the middle Miocene, approximately 88 Mya (Bancroft 1935, Gunasekara 2004). Further evidence supporting the notion that ancestors of Dipterocarpoideae existed in Africa results from the chronological sequence of interplate movement; the Indian plate separated from the Southern Gondwanaland (113-108 Mya) and drifted northward until it was in close proximity to Madagasgar (74 Mya) thereby facilitating subsequent plant dispersal from the African to Deccan plate (later renamed the Indian plate) (Morley 2003, Gunasekara 2004).

Warm climates remained during the Miocene era and early Pliocene, while cooler climates only occurred intermittently (Pooma 2003). These environmental effects promoted the expansion of the Malaysian rain forests to extend to northern India, and the flora of Southeast Asia began to take on its present form with Dipterocarpaceae constituting a major component within moist rain forests (Pooma 2003). Consistent with other molecular studies (Gunasekara 2004) our results (Figure 8) suggest parallel evolution among *Vateriopsis* and *Stemenoporus*, as well as among *Vateria* and *Upuna*. Divergence times reported for these taxa support the theory that they shared a common ancestor during the Eocene era (Gunasekara 2004). The division between the two groups

may be due to their independent evolution incurred during adaptation to their respective microhabitats and microclimatic conditions during the historical range reduction (Gunasekara 2004).

The most ancestral lineage of the genus *Shorea* is reported to be section *Doona* based on morphological as well molecular data (Ahston 1982, Gunasekara 2004). As seen in Figure 8, *Shorea cordifolia* of section *Doona*, is found basal to the Shorea group and its' position is congruent with other studies; with the exception of *Shorea siamensis*, section *Pentacme*. Studies have reported that the estimated time of divergence indicate that the common ancestor of *Shorea* section *Doona* existed in the Deccan plate prior to its collision with Eurasia (Gunasekara 2004).

II.5 Conclusions

Although previous studies have investigated the relationships within Dipterocarpaceae, we present a comprehensive investigation of the nuclear 18S (N18S) ribosomal segment based phylogenetic relationships within Dipterocarpaceae by including, with the exception of genera *Dryobalonops* and *Anisoptera*, at least one representative individual per genus. Phylogenetic analysis revealed low levels of polymorphism for the N18S segment, which can explain discrepencies such as the classification of *Monotes sp.* and *Marquesia sp.*, as compared to other studies. However, the overall segregation of taxa based on chromosome number, with chromosome 7 as the derived state, remains similar to other investigations. Furthermore groupings of taxa are consistent with consistent with historical biogeographical events. Dipterocarpaceae diverged into its genera and major sections by the end of the Miocene era and taxa, such as *Vateriopsis* and *Stemenoporus* as well as *Vateria* and *Upuna,* reportedly share common ancestors during the Eocene era suggesting parallel evolution.

Table 8: Abbreviations and classifications of species included in the phylogeny of the Dipterocarpaceae family using partial rRNA sequences. Abbreviations represent those seen in phylogenetic trees (Figures 6, 7, and 8).

Subfamily	Tribe	Genus	Section	Subsection	Species Name	Abbreviations
		Durio			*Durio zibathinus*	*Durio zibathinus*
Dipterocarpoideae	Dipterocarpeae	*Dipterocarpus*			*Dipterocarpus zeylanicus*	*D. zeylanicus*
		Dipterocarpus			*Dipterocarpus glandulosus*	*D. glandulosus*
		Vatica	*Vatica*		*Vatica venulosa*	*V. venulosa*
		Vatica	*Sunaptea*		*Vatica bantamensis*	*V. bantamensis*
		Upuna			*Upuna borneensis*	*U. borneensis*
		Cotylelobium			*Cotylelobium lanceolatum*	*C. lanceolatum*
		Balanocarpus			*Balanocarpus utilis*	*B. utilis*
		Vateriopsis			*Vateriopsis seychellarum*	*V. seychellarum*
		Vateria			*Vateria copallifera*	*V. copallifera*
		Stemonoporus			*Stemonoporus gilimalensis*	*St. gilimalensis*
		Stemonoporus			*Stemonoporus sp.*	*St. sp.*
	Shoreae	*Shorea*	*Richetioides*	*Richetioides*	*Shorea multiflora*	*S. multiflora*
		Shorea	*Brachypterae*	*Brachypterae*	*Shorea pauciflora*	*S. pauciflora*
		Shorea	*Brachypterae*	*Smithiana*	*Shorea smithiana*	*S. smithiana*
		Shorea	*Ovales*		*Shorea ovalis*	*S. ovalis*
		Shorea	*Pentacme*		*Shorea siamensis*	*S. siamensis*
		Shorea	*Rubella*		*Shorea rubella*	*S. rubella*
		Shorea	*Anthoshorea*		*Shorea stipularis*	*S. stipularis*
		Shorea	*Doona*		*Shorea cordifolia*	*S. cordifolia*
		Neobalanocarpus			*Neobalanocarpus heimii*	*N. heimii*
		Hopea	*Hopea*	*Hopea*	*Hopea odorata*	*H. odorata*
		Hopea	*Hopea*		*Hopea brevipetiolaris*	*H.brevipetiolaris*
		Hopea	*Dryobalanoides*	*Dryobalanoides*	*Hopea dryobalanoides*	*H.dryobalanoides*
Monotoideae		*Monotes*			*Monotes sp.*	*Monotes sp.*
Monotoideae		*Marquesia*			*Marquesia sp.*	*Marquesia sp.*
Pakaraimoideae		*Pakaraimaea*			*Pakaraimaea dipterocarpacea*	*P. dipterocarpacea*

Figure 6: The maximum parsimony majority rule consensus tree of 277 equally parsimonious trees based on the partial rRNA sequences. Numbers above branches depict frequency of occurrence in 50% majority rule consensus, while numbers below the branches indicate bootstrap % values. Numbers in brackets show branch lengths (the number of nucleotide substitution).

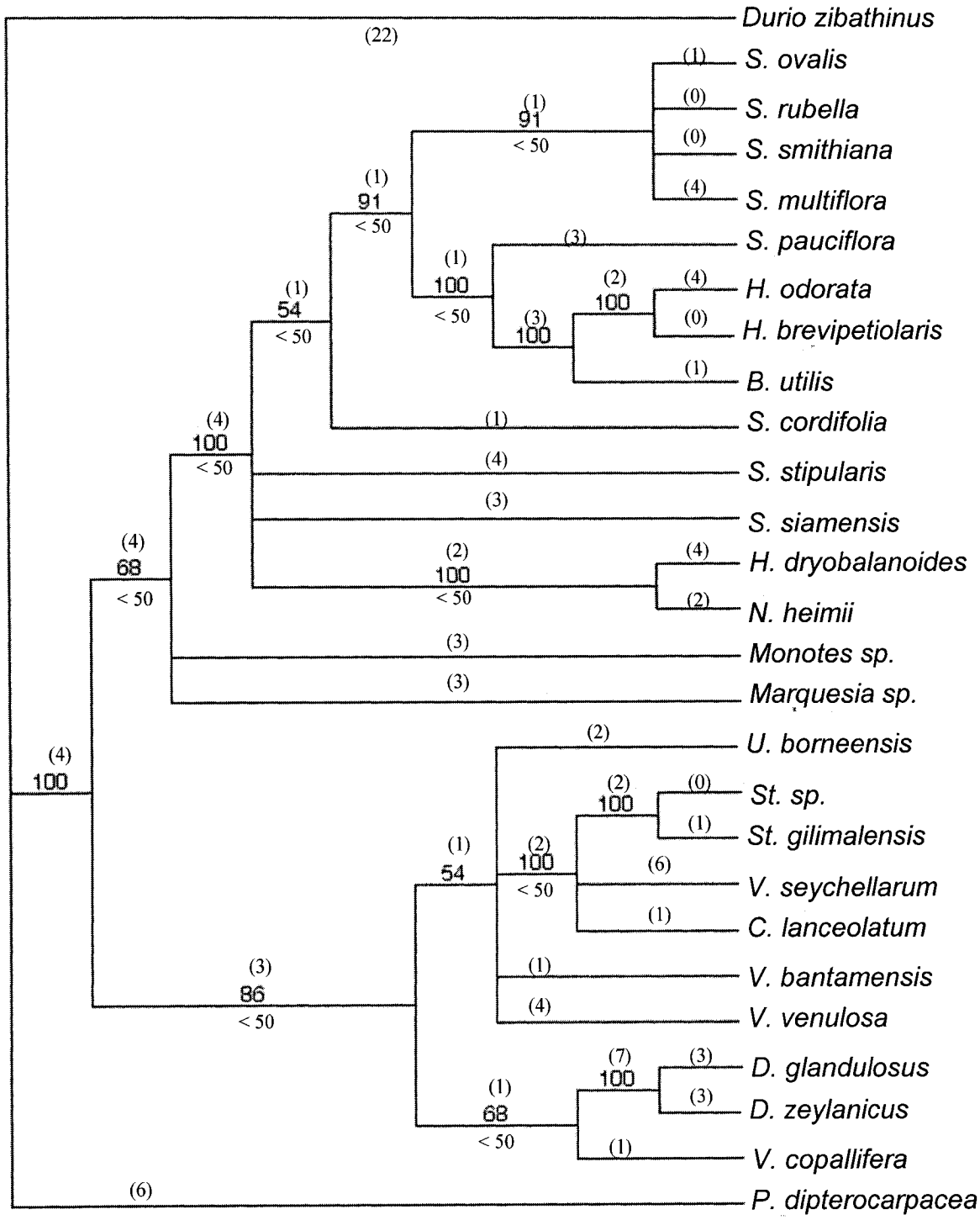


Figure 7: The maximum likelihood phylogenetic tree based on the partial rRNA sequences. Numbers above branches depict frequency of occurrence in 50% majority rule consensus, while those below branches depict bootstrap percentage values. Numbers in brackets show branch lengths (the number of nucleotide substitution).

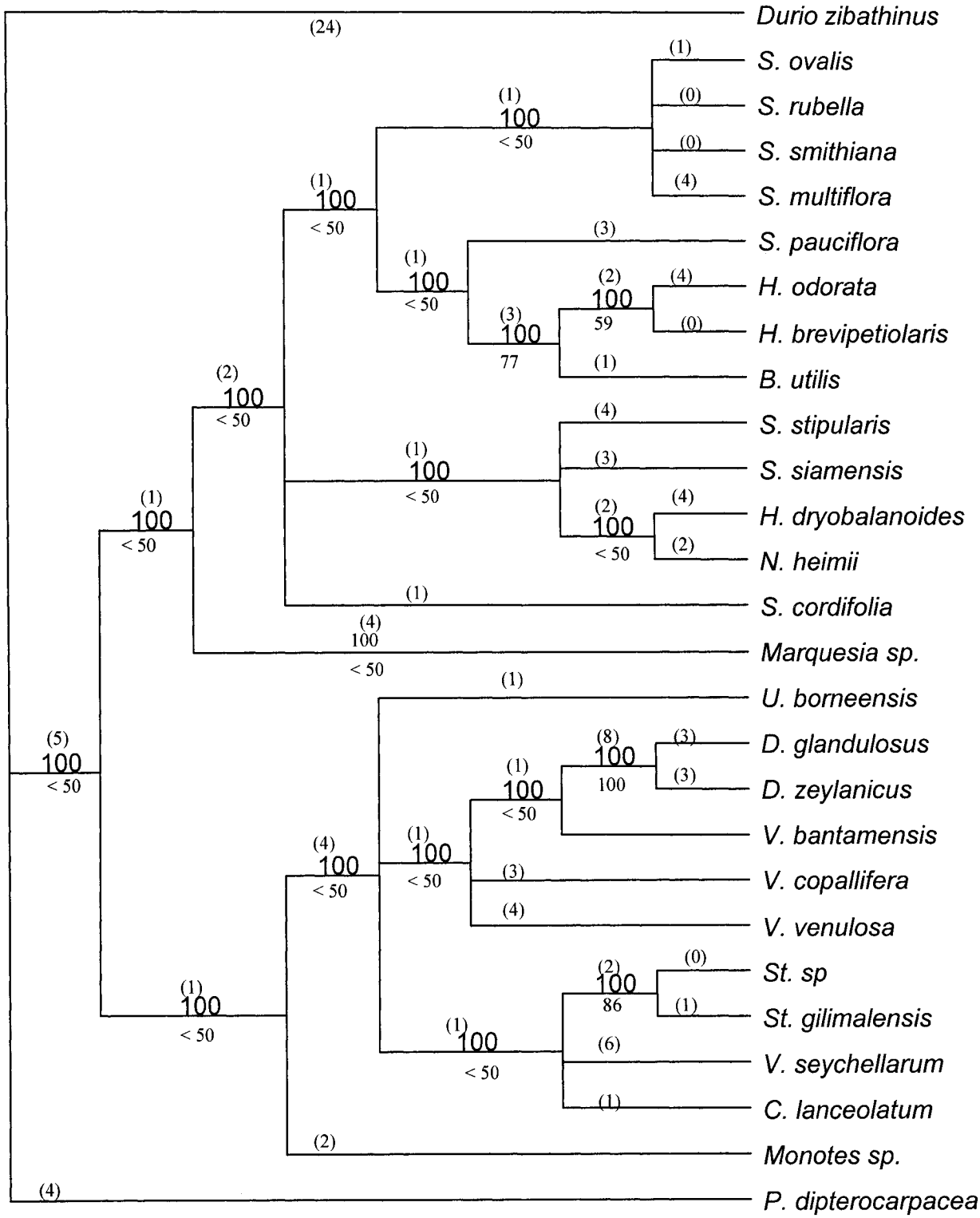
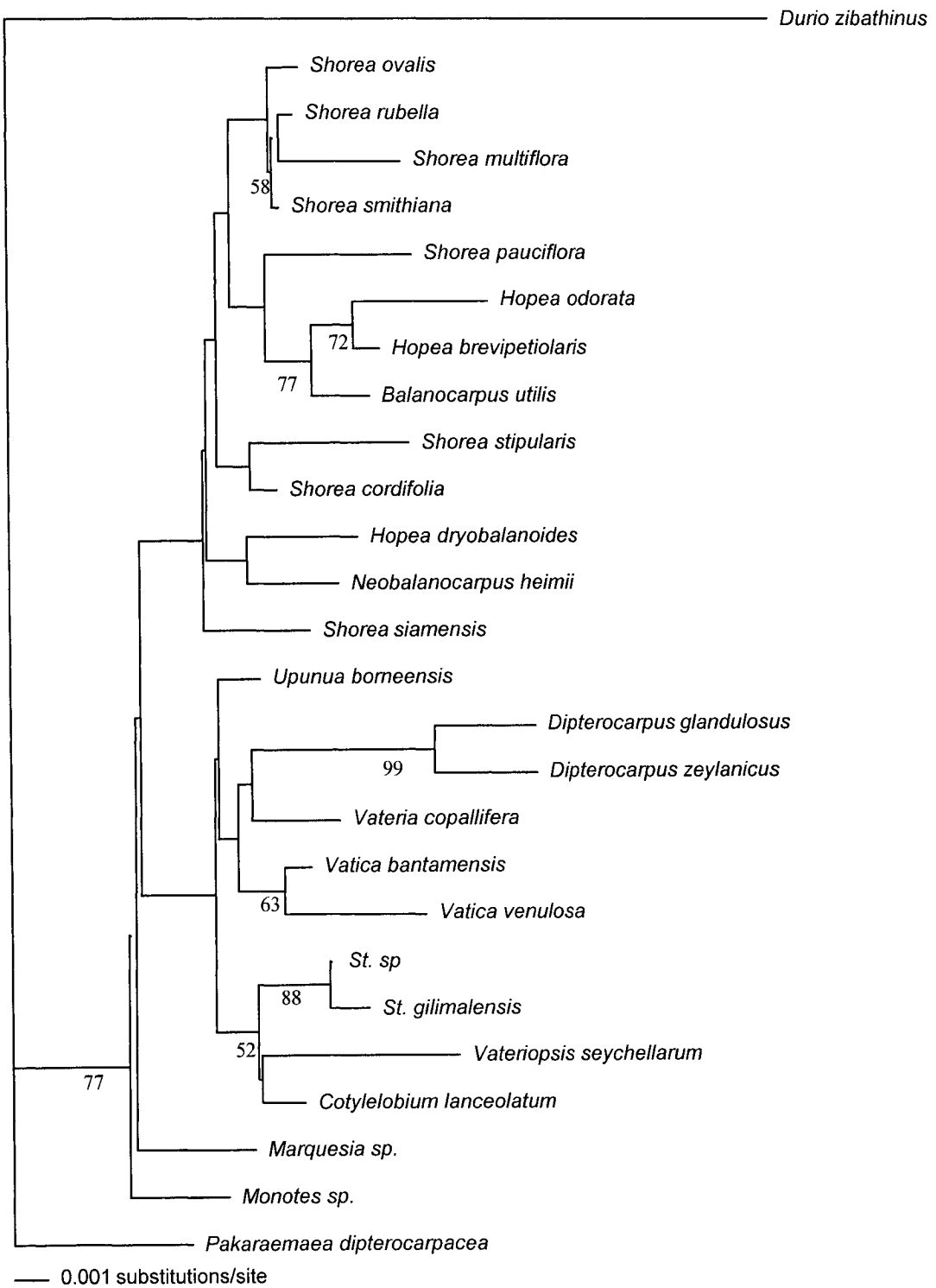


Figure 8: Ditperocarpaceae neighbor-joining (NJ) tree based on the partial rRNA sequences. Numbers indicate bootstrap percentage values; unless otherwise indicated values are less than 50%. Branch lengths (based on the distance matrix) are proportional to the number of substitutions.



General conclusions

The population genetic study of jack pine, *Pinus banksiana*, has revealed that the maritime populations are genetically distinct from the mainland populations and current gene flow values indicate that maritime populations have a significant potential to become genetically depauparate. This information is crucial for its sustainable management. Furthermore, detailed genetic structure parameters have elucidated the effects of various historical refugia events that have influenced the evolutionary process of this species. At least 2 glacial refugia events have influenced the genetic structure of jack pine.

The phylogenetic study of the Dipterocarpaceae based on partial sequences of nuclear rRNA is among the first of its kind. Although these segments revealed low levels of polymorphism overall segregation of taxa and divisions based on chromosome number remained consistent with other research studies. Groupings of sequences are generally as expected, with members of the same genus grouping together, and are consistent with historical biogeographical events.

Literature cited

Adams, J. (1994). The Distribution and Variety of Equatorial Rain Forest. Encyclopedia Catalana. Barcelona, Spain. Available Source http://www.esd.ornl.gov/projects/qen/rainfo.html, January 10, 2006.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Alverson, W. S., K. G. Karol, D. A. Baum, M. W. Chase, S. M. Swensen, R. McCourt, and K. J. Sytsma. (1998). Circumscription of the Malvales and relationships to other Rosidae: Evidence from rbcL sequence data. *American Journal of Botany*. 85: 876-887.

Appanah, S. (1998). Management of natural forests. *In* A review of Dipterocarps, taxanomy, ecology and silviculture (*eds.* S. Appanah and J. M. Turnbull). Center for forest research institute, Malesia. Pp. 133-1493.

Ashton, P.S. (1969). Speciation among tropical forest trees: some deductions in the light of recent evidence. *Biological Journal of the Linnaean Society*. 14: 155-196.

Ashton, P. S. (1982). Dipterocarpaceae, in Flora Malesiana, Series 1, Spermatophyta (*ed.* C. G. G. J. Van Steenis), Martinus Nijhoff Publishers, The Hague. 1: Pp. 237-289.

Ashton, P. S., Givnish, T. J. and Appanah, S. (1988). Staggered flowering in the Dipterocarpaceae. New sights into floral induction and the evolution of mast fruiting in the aseasonal tropics. *The American Naturalist*. 132: 44-66.

Audley-Charles, M.G. (1987). Dispersal of Gondawanaland: Relevance to Evolution of the Angiosperms. In T.C. Whitmore, ed. Biogeographical Evolution of the Malay Archipelago. Oxford University Press, New York.

Baker, A. 2000. *Molecular methods in Ecology*. Alden Press, Oxford and Northampton.

Bancroft, H. (1935). Some fossil dicotylodonous woods from Mont Elgon, East Africa. *American Journal of Botany*. 22: 505-519.

Bangert, R. K., Turek, Richard J., Martinsen, G. D., Wimp, G. M., Bailey, J. K. & Whitham, T. G. (2005). Benefits of Conservation of Plant Genetic Diversity to Arthropod Diversity.*Conservation Biology*. **19:** 379-390.

Beland, M., Bergeron, Y. (1993). Ecological factors affecting abundance of advanced growth in jack pine (*Pinus banksiana* Lamb.) stands of the boreal forest of northwertern Quebec. *The Forestry Chronicle*. 69: 561-568.

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

BOs, H.D., Posada, D. (2005). Using models of nucleotide evolution to buil phylogenetic trees. *Developmental & Comparative Immunology*. 29: 211-227.

Bockelmann, A.C., Reusch T.B.H., Bijlsma, R., and Bakker, J.P. (2003). Habitat differentiation vs. isolation-by-distance: the genetic population structure of *Elymusathericus* in European salt marshes. *Molecular Ecology*. 12: 505-515.

Boys, J., Cherry, M., Dayanandan, S. (2005). Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). *American Journal of Botany*. 92: 833-841.

Brocchieri, L. (2001). Phylogenetic inferences from molecular sequences: review and critique. *Theoretical population biology*. 59: 27-40.

Bult, C. Kallersjo,M., Suh, Y. (1992). Amplification and sequencing of 16/18S rDNA from gel-purified total plant DNA. *Plant molecular biology reporter*. 10: 273-284.

Chakraborty, R., Kimmel, M., Stivers, D.N., Davison, J., Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Science USA*. 94: 1041-1046.

Chapman, R. L., Buchheim, M. A. (1991). Ribosomal RNA sequence analysis and its significance in the phylogeny and taxonomy of green algae. *Critical Reviews in Plant Sciences*. 10:343-368.

Chase, M W, ET AL. (1993). Phylogenetics of seed plants: analysis of nucleotide sequences from the plastid gene *rbc*L. *Ann. Missouri. Bot. Gard.*, 80: 528-580.

Crow, J. F. and Kimura, M. (1972). Introduction to population genetics theory. Harper & Row, New York.

Dancik, B. P., Yeh F. C. (1983). Allozyme variability and evolution of lodegepole pine (*Pinus contorta* var*latifolia*) and jack pine (*P. banksiana*) in Alberta. *Canadian Journal of Genetics and Cytology*. 25: 57-64.

Danzmann, R. G., Buchert, G. P. (1983). Isozyme variability in central Ontario jack pine. In: *Procedings of the 28th Northeastern Forest Tree Improvement Conference* (ed. Ekert, R. T.) Institute of Natural and Environmental Resources, University of New Hampshire, Durham, New Hampshire. pp. 232-248.

Davis, M. B. (1983). Quaternary history of deciduous forests of eastesn North America and Europe. *Annals of the Missouri Botanical Garden*. 70: 550-563.

Dayanandan, S. (1995). Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast rbcL gene and morphology. Ph.D. thesis, Boston University, Boston, Massachusetts.

Dayanandan, S., Bawa, K. S., Kesseli, R. V. (1997). Conservation of Microsatellites among tropical trees (Leguminosae). *American Journal of Botany*. 84: 1658 – 1663.

Dayanandan, S., Rajora, O. P., Bawa, K. S. (1998). Isolation and characterization of microsatellites in trembling aspen (*Populus tremuloides*). *Theoretical and applied genetics*. 96: 950-956.

Dayanandan, S., Asthon, P. S., Williams, S. M., Primack, R. B. (1999). Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of chloroplast *rbc*L gene. *American journal of botany*, 86: 1182-1190.

Dong J, Wagner DB. (1994). Paternally inherited chloroplast polymorphism in Pinus: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics.* 136:1187-94.

Ducousso M., Béna G., Bourgeois C., Buyck B., Eyssartier G., Vincelette M., Rabevohitra R., Randrihasipara L., Dreyfus B., Prin Y.. (2004). The last common ancestor of Sarcolaenaceae and Asian dipterocarp trees was ectomycorrhizal before the India–Madagascar separation, about 88 million years ago. *Molecular Ecology*. 13: 231-236.

Dyke, A.S., Andrews, J.T., Clark, P.U., England, J.H., Miller, G.H., Shaw, J., Veillette, J.J., (2002). The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary Science Reviews*. 21: 9–31.

Echt, C. S., Deverno, L. L., Anzidel, M. Zahorchak, R. (1996). Characterization of microsatellite markers in eastern white pine. *Genome*. 39: 1102-1108.

Ellengren, H., Primmer, C.R., Sheldon, B.C. (1995). Microsatellite evolution: directionality or bias in locus selection. *Nature Genetics*. 11: 360-362.

Ellengren, H. (2000). Heterogenous mutation in the germline; implications for evolutionary inference. *Trends in Genetics*. 16: 551-558.

Epperson, B.K., Li, T.Q. (1997). Gene dispersal and spatial genetic structure. *Evolution*. 51: 672-681.

FAO. (1985). Mangrove management in Thailand, Malaysia and Indonesia. Food and Agriculture Organization (FAO) of the United Nations. Publishing and multimedia service, Information division, FAO, Rome.

Fedorov, A. (1966). The structure of the tropical rain forest and speciation in the humid tropics. *Journal of Ecology*. 54: 1-11.

Felsenstein, J. (1983). Parsimony in systematics: biological and statistical issues. *Annual Review of Ecology and Systematics*. 14: 313-333.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 39: 783-791.

Felsenstein J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics*. 22: 521-5265.

Felsenstein, J. (1989). PHYLIP-Phylogeny inference package (version 3.2). *Cladistics*. 5: 164-166.

Fisher, R. A. (1930). The genetical theory of natural selection. Clarendon Press, Oxford.

Fischer, D., Bachmann, K. (1998). Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *Biotechniques*. 24: 796-798.

Fisher, P.J., Richardson, T. E., Gardner, R. C. (1998). Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. *Theoretical and applied genetics*. 96: 969-979.

Fitch, W. M. (1971). Toward defininf the course of evolution: Minimum change for a specific tree topology. *Systematic Zoology*. 27: 401-410.

FRA (2000). Global forest resources assessment 2000. FAO Forestry paper 140 (Main Report). Food and Agriculture Organization (FAO) of the United Nations. Publishing and multimedia service, Information division, *FAO*, Rome.

Garner, T. W. J. (2002). Genome size and microsatellites: the effects of nuclear size on amplification potential. *Genome*. 45: 212-215.

Gauthier, S., Simon, J. P., Bergeron, Y. (1992). Genetic structure and variability in jack pine populations: effects of insularity. *Canadian Journal of Forest Research*. 22: 1958-1965.

Gauthier, S., Gagnon, J., Bergeron, Y. (1993). Population age structure of *Pinus banksiana* at the edge of the Canadian boreal forest. *Journal of Vegetative Science*. 4: 783-790.

Gillespie, J. (1998). Population genetics: a concise guide. Johns Hopkins Press, Baltimore, Md.

Godbout, J., Jaramillo-Correa, J. P., Beaulier, J., Bousquet, J. (2005). A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molecular Ecology*. 14: 3497-3512.

Govindarju, D.R. (1988). Life histories, neighbourhood sizes, and variance structure in some North America conifers. Biol. *J. Linn Soc.* 35: 69-78.

Green, D. M. (2005). Designatable Units for Status Assessment of Endangered Species. *Conservation Biology.* 19: 1813-1820.

Groenen P., Wieringa B. (1998). Expanding complexity in myotonic dystrophy. *Bioessays.* 20: 901-12.

Gunasekara, N. (2004). Phylogenetic and molecular dating analyses of the tropical tree family Dipterocarpaceae based on chloroplast *matK* nucleotide sequence data. Ph.D. dissertation, Concordia University, Montreal, Qc.

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

Harding, R. (1996). New phylogenies: an introductory look at the coalescent. In *New Uses for New Phylogenies* (Harvey, P.H. *et al.*, eds), Oxford University Press.

Hartl, D. Jones, E. (2000). Genetic analysis of genes and genomes Fifth edition. Jones and Bartlett publishers. Sudburry, Massachusetts.

Harvey, P. H. (1996). Phylogenies for ecologists. *J. Anim. Ecol.* 65: 255-263.

Henderson, A,R. (2005). The bootstrap: a technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clinica chimica acta; international journal of clinical chemistry.* 359: 1-26.

Hillis, D.M., Huelsenbeck, J.P. (1992). Signal, noise, and reliability in molecular phylogenetic analyses. *Journal of Heredity.* 83: 189-95.

Hilu, K. W. (1996). The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany.* 84: 819-839.

Hudson, R.R. (1998). Island models and the coalescent process. *Mol Ecol.* 7: 413-418.

Huelsenback J. P., Hillis D. M. (1993). Success of phylogenetic methods in the four-taxon case. *Systematic Biology*. 42: 247-264.

Jackson, S. T., Webb, R. S., Anderson, K. H. (2000). Vegetiation and environment in eastern North America during the last glacial maximum. *Quaternary science reviews*. 19: 489-508.

Jackson, S. T., Overpeck, J. T., Webb, R. S., Keattch, S. E., Anderson, K. H. (1997). Mapped plant-macrofossils and pollen records of the late Quaternary vegetation change in eastern North America. *Quaternary Science Reviews*. 19: 489-508.

Jaramillo-Correa, J. P., Beaulieu, J., Bousquet, J. (2004). Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (Picea mariana), a transcontinental North American conifer. *Molecular Ecology*. 13: 2735-2747.

Jarne, P., Lagoda, P. J. L. (1996). Microsatellites, from molecules to populations and back. *Tree*. 11: 424-429.

Jeffreys, A.J., Wilson, V., Thein, S.L. (1985). Hypervariable "minisatelllite" regions in human DNA. *Nature*. 314: 67-73.

Jeffreys, A.J., Tamaki, K., Macleod,A., Monckton, D.G., Neil, D.L., Armour, J.A.L. (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics*. 6: 136-145.

Kajita, T., Kaniya, K., Nakamura, K., Tachida, H., Wickneswari, R., Tsumura, Y., Yoshimaru, H., Yamazaki, T. (1998). Molecular Phylogeny of Dipterocarpaceae in Southeast Asia based on nucleotide sequences of *mat*K, *trn*L intron, and *trn*L-*trn*F

intergenic spacer region in chloroplast DNA. *Molecular phylogenetics and evolution*, 10: 202-209.

Kamiya, K., Harada, K., Tachida, H., Ashton, P. S.. (2005). Phylogeny of PgiC gene in Shorea and its closely related genera (Dipterocarpaceae), the dominant trees in Southeast Asian tropical rain forests. *Am. J. Bot.* 92: 775-788.

Kimura, M., and Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*. 49: 725-738.

Kluge, A. G. and Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Systamatic Zoology*. 18: 1-32.

Klung, W.S., Cummings, M.R. (2000). Concepts of Genetics 6th edition. Jones and Bartlett publishers. Sudburry, Massachusetts.

Kovtun, I.V., McMurray, C.T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genetics*. 4: 581-593.

Kostia, S., Varvio, S. L., Vakkari, P., Pulkkinen, P. (1995). Microsatellite sequences in *Pinus sylvestris*. *Genome*. 38:1244-1248.

Kuzoff, R.K., Sweere, J.A., Soltis, D.E., Soltis, P.F., Zimmer, E.A. (1998). The Phylogenetic potential of entire 26S rDNA sequences in plants. *Mol. Biol. Evol.* 15: 251–263.

Ledig, F.T. (1988). The Conservation of Diversity in Forest Trees: Why and How Should Genes Be Conserved? *BioScience*. 38: 471-479.

Lemes M, Gribel R, Proctor J & Grattapaglia D. (2003) Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon,

based on variation at microsatellite loci: implications for conservation. *Mol Ecol.* 12: 2875–2883.

Levinson, G., Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution.* 4: 203-221.

Li, Y.C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E. (2002). Microsatellites: genomic distribution, putative functions, and mutational mechanisms. *Molecular Ecology.* 11: 2453-2465.

Lindenmayer, David B., Margules, Chris R. & Botkin, Daniel B. (2000). Indicators of Biodiversity for Ecologically Sustainable Forest Management. *Conservation Biology.* 14: 941-950.

Maddison, W. P., Maddison, D. R. (2001). *MacClade 4.02.* Analyses of phylogeny and character evolution. Sinauer Associates, Sunderland.

Maguire, B. P. C. and Ashton, P. S. (1977). Pakaraimoideae, Dipterocarpaceae of the western hemisphere II. Systematic, geographic and phyletic consideration. *Taxon.* 26: 341-285.

Marsella, K.A., Bierman, P.R., Davis, P.T., and Caffee, M.W. (2000). Cosmogenic 10Be and 26Al ages for the last glacial maximum, eastern Baffin Island, arctic Canada. *Geological Society of America Bulletin.* 112: 1296–1312.

Maury-Lechon, G. and Curtet, L. (1998). Biogeography and evolutionary systematics of family Dipterocarpaceae. *In* A review id Dipterocarps, taxonomy, ecology, and silviculture (*eds* S. Appanah and J. M. Turnbull). Center for forest research institute, Malesia. Pp. 5-44.

Meijer, W. (1979). Taxonomic studies in the genus *Dipterocarpus*. *In* Dipterocarpacees: Taxonomia-Phylogenie-Ecologie, Memoires du Museum National d'Histoire Naturelle, (*ed.* G. Maury-Lechon). Editions du Museum, Paris. Serie B, Botanique. 26: 50-56.

Morgante, M & Olivieri, AM (1993). PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal*. 3:175-182.

Morley, R.J. (1998). Palynological Evidence for Tertiary Plant Dispersals in the SE Asian Region in Relation to Plate Tectonics and Climates. *In* R. Hall and J.D. Holloway, eds. Biogeography and Grological Evolution of SE Asia. Backbuys Publishers, Leiden. Pp. 211-234.

Morley, R.J. (1999). Origin and evolution of tropical rain forests. John Wiley and Sons, Chicester, U.K.

Morley, R.J. (2001). Tertiary History of the Malesian Flora: A Palynological Perspective. *In* L.G. Saw, L.S.C. Chua and K.C. Khoo, eds. Taxonomy: The Cornerstone of Biodiversity. Proceeding of the Fourth Internatioanl Flora Malesiana Symposium 1998, The Research Institute Malaysia, Kepong, Kualaumpur. Pp. 197-210.

Morley, R.J. (2003). Interplate dispersal paths for megathermal angiosperms. Perspective in *Plant Ecology, Evolution and systematics*. 6: 5-20.

Moritz, C. (1995). Uses of molecular phylogenies for conservation. *Philosophical Transactions: Biological Sciences*. 349: 113-118.

Nei, M. (1972). Genetic distance between populations. *Am Nat*. 106: 283-292.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA*. 70: 3321-3323.

Newstrom L.E., Frankie G. W., Baker H. G. (1994). A New Classification for Plant Phenology Based on Flowering Patterns in Lowland Tropical Rain Forest Trees at La Selva, Costa Rica. *Biotropica*. 26: 141-159.

Nkongolo KK, Michael P, Gratton WS. (2002). Identification and characterization of RAPD markers inferring genetic relationships among Pine species. *Genome*. 45: 51-8.

Page, R. D. M., Holmes, E. C. (1998). Molecular evolution: A phylogenetic approach. Blackwell publishers, Oxford. U.K.

Parcy, F. (2005). Flowering: a time for integration. *Int. J. Dev. Biol*. 49: 585-593.

Pfeiffer, A., Olivieri, A.M., Morgante, M. (1997). Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome*. 40: 411-419.

Pooma, R. (2003). Dipterocarpaceae in Thailand: Taxonomic and Biogeographical Analysis. Ph. D. dissertation, Kasetsart University, Chatuchak, Bangkok.

Posada, D. and Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics*. 14: 817-818.

Primmer, C.R., Saino, N., Moller, A.P, Ellengren, H. (1998). Unravelling the processes of microsatellite evolution htrough analysis of germline mutations in barn swallows *Hirundo rustica*. *Molecular Biology and Evolution*. 15: 1047-1054.

Redford, Kent H. & Richter, Brian D. (1999). Conservation of biodiversity in a world of use. *Conservation Biology*. 13: 1246-1256.

Richardson, B. A., Brunsfeld, S. J., Klopfenstein, N. B. (2002). DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (Pinus albicaulis). *Molecular Ecology*. 11: 215-227.

Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM.(2001). Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science.* 293: 2242-5.

Ross, H. A., Hawkins, J. L. (1986). Genetic variation among local population of jack pine (*Pinus banksiana*). *Canadian Journal of Genetics and Cytology.* 28: 453-458.

Rudolph, T.D., and Yeatman, C.W. (1982). Genetics of jack pine. U.S. For. Serv. Wash. Off. Res. Pap. WO-38.

Rudolph, T. D., Laidly, P. R.. (1990). Pinus banksiana Lamb. Jack 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: 555-586.

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

Saenz-Romero, C., Guries, R. P., Monk, A. I. (2001). Landscape genetic structure of *Pinus banksiana*: allozyme variation. *Canadian Journal of Botany.* 79: 871-878.

Saitou, N., Nei M. (1987). The neighbour joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution.* 4: 406-425.

Sanderson, M.J, Doyle, J.A. (2001). Sources of error and confidence intervals in estimating the age of Angiosperms form *rbcL* and *18S* rDNA data. *American Journal of Botany.* 88: 1499-1516.

Schoenike, Roland E. (1976). Geographic variation in jack pine. University of Minnesota Agriculture Experiment Station, Technical Bulletin. 304: 49.

Schlotterer, C., Ritter, R., Harr, B., Brem, G. (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., Aquadro, C.F. (1997). Low mutation rates of microsatellite loci in *Drosophila melanogaster. Nature Genetics.* 15: 99-102.

Sellick GS, Longman C, Tolmie J, Newbury-Ecob R, Geenhalgh L, Hughes S, Whiteford M, Garrett C, Houlston RS. (2004). Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. *Nucleic Acids Res.* 32:e164.

Slik, J.W.F., Poulsen, A.D., Ashton, P.S., Cannon, C.H., Eichhorn, K.A.O., Kartawinata, K., Lanniari, I., Nagamasu, H., Nakagawa, M., Van Nieuwstadt, M.G.L., Payne, J., Purwaningsih, J., Saridan, A., Sidiyasa, K., Verburg, R.W., Webb, C.O., Wilkie, P. (2003). A floristic analysis of the lowland dipterocarp forests of Borneo. *Journal of Biogeography.* 30. 1517-31.

Sltakin, M. (1987). Gene flow and the geographic structure of natural populations. *Science.* 236: 787-792.

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

Smith, D.N., Devey, M.E. (1994). Occurrence and inheritance of microsatellites in Pinus radiate. *Genome.* 37: 977-983.

Smith, G,K., Jie, J., Fox, G.E., Gao, X. (1995). DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes. *Nucleic Acids Res.* 23: 4303-11.

Soltis, D. E., Soltis, P. S., Nickrent, D. L., Johnson, L. A., Hahn, W. J., Hoot, S. B., Sweere, J. A., Kuzoff, R. K., Kron, K. A., Chase, M. W., Swenson, S. M., Zimmer, E. A., Chaw, S., Gillespie, L. J., Kress, W. J., Sytsma, K. J.. (1997). Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Missouri Bot. Gard.* 84: 1-49.

Soranzo N, Provan J, Powell W. (1998). Characterization of microsatellite loci in *Pinus sylvestris* L. *Molecular Ecology.* 7:1260-1.

Sowfford D. L. (2001). PAUP*: Phylogenetic analysis unsing parsimony (* and other methods). Version 4.0b8 for Macintosh, and 4.0b4 and 4.0b5 for UNIX. Sinauer Associate, Sunderland, Massachusetts, USA.

Stacy, E. A., Dayanandan, S., Dancik, B. P., Khasa, P. D.. (2001). Microsatellite DNA markers for the Sri Lankan rainforest tree species, *Shorea cordifolia* (Dipterocarpaceae), and cross-species amplification in *S. megistophylla. Molecular Ecology Notes.* 1: 53-54.

Sturtevant, A. H. (2001). A history of genetics. Cold Spring Harbor Laboratory Press and Electronic Scholarly Publishing Project. pp. 107-116.

Symington, C.F. (1943). Forester's Manual of Dipterocarps. *Malayan Forest Records.* 16: 1-244.

Tamura, K. Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution.* 10: 512-526.

Tateno, Y, Takezaki, N., Nei,M..(1993). Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Molecular Biology and Evolution.* 11: 261-277.

Thomas, B.R., MacDonald, S.E., Hicks, M., Adams, D.L., Hpdgetts, RB. (1999). Effects of reforestation methods on genetic diversity of lodgepole pone: an assessement using microsatellite and randomly amplified polymorphic DNA markers. *Theoretical and Applied Genetics.* 98: 793-801.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* 22: 4673-4680.

Thuillet, A.C., Bru, D., David, J.L., Roumet, P., Santoni, S., Sourdille, P., Bataillon, T. (2002). Direct estimation of mutation rate for ten microsatellite loci in drum wheat, *Triticum turgidum* (L.) Thell. Ssp durum desf. *Molecular Biology and Evolution.* 19: 122-125.

Tivy, J. (1993). Biogeography: A Study of Plants in the Ecosphere. Longman Scientific and Technical. Essex.

Tsumura, Y., Kawahara, T., Wickneswari, R. (1996). Molecular phylogeny of Dipterocarpaceae in Southeast Asia using RFLP of PCR-amplified chloroplast genes. *Theor Appl Genet.* 93: 22-29.

Vigouroux, Y., Jaqueth, J.S., Matsuoka, Y., Smith, O.S., Beavis, W.D., Smith, J.S., Doebley, J. (2002). Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution.* 19: 1251-1260.

Wagner, D. B., Furnier, G. R., Saghai-Maroof, M. A.,Williams, S. M.,Dancik, B. P., Allard, R.W. (1987). Chloroplast DNA polymorphisms in Lodgepole and Jack pines and their hybrids. *PNAS.* 84: 2097-2100.

Walter, R., Epperson, B. K. (2005). Geographic pattern of genetic diversity in *Pinus resinosa*: contact zone between descendants of glacial refugia. *American Journal of Botany*. 92: 92-100.

Walter, R,. Epperson, B. K. (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.

Webb, C. O. (2000). Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees. *The American Naturalist.* 156: 145–155.

Webb, T. III (1988). Glacial and Holocene vegetation history: eastern North Amrica. In: *Vegetation history* (eds. Huntley, B, Webb T. III), pp. 385-414. Kluwer Academic, Dordrecht, The Netherlands.

Webb, C. O., Ackerly, D. D., McPeek, M. A., Donoghue, M. J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**: 475–505.

Weber, J.L., Wong, C. (1993). Mutation of human short tandem repeats. Human Molecular *Genetics*. 2: 1123-1128.

Wen, J. (1999). Evolution of Eastern Asian and Eastern North American Disjunct Distributions in Flowering Plants. Annual Review of Ecology and Systematics. 30: 421-455.

Whitmore, T. C. (1989). South East Asian tropical forests. *In* Tropical rain forest ecosystems: biogeographical and ecological studies (*eds.* H. Lieth and M. J. A. Werger). Elsevier Science Publishers, B. V. Amsterdam, The Netherlands.

Whitmore, T. C. (1998). An introduction to tropical rain forests. Oxford university press.

Wikipedia contributors (2005). Molecular evolution. *Wikipedia, The Free Encyclopedia.* Retrieved 16:45, January 9, 2006 from http://en.wikipedia.org/w/index.php?title=Molecular_evolution&oldid=32991453.

Wikstrom, N., Savolanien, V., Chase, M.W. (2001). Evolution of angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London.* 268: 2211-2220.

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

Wright, S. (1969). *Evolutionary and the Genetics of Populations. Vol. 2. The Theory of Gene Frequencies.* The University of Chicago Press, Chicago.

Ye, T., Yang, R.C., Yeh F. C. (2002). Population structure of a lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*) complex as revealed by random amplified polymorphic DNA. *Genome.* 45: 530-540.

Yeh, FC, Hu, XS. (2005). Genetic structure and migration from mainland to island population in *Abies procera* Rehd. *Genome* 48: 461-473.

Yeh, F.C., Yang, R.C., Boyle, T.J.B., Ye, Z.H., and Miao, J.X. (1997). POPGENE, the user friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, Univ. of Alberta, Edmonton, Canada.

Zharkikh, A. (1994). Estimation of evolutionary distances between nucleotide sequences. *Journal of Molecular Evolution.* 39: 315-329.

Zuckerkandl E, Pauling L. (1965). Molecules as documents of evolutionary history. *J Theor Biol.* 8: 357-66.