Molecular Phylogeny of the Tree Genus *Populus* (Salicaceae)

based on

Chloroplast and Nuclear DNA Sequence Data

Mona Hamzeh

A Thesis

in

the Department

of

Biology

Presented in Partial Fulfilment of the Requirements

for the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

August 2003

Your file Votre référence
ISBN: 0-612-83848-X
Our file Notre référence
ISBN: 0-612-83848-X

Abstract

Molecular phylogeny of the tree genus *Populus* (Salicaceae)

based on chloroplast and nuclear DNA sequence data

Mona Hamzeh

Species of the genus *Populus*, collectively known as poplars, are widely distributed over the northern hemisphere, and well known for their ecological, economical, and evolutionary importance. The extensive inter-species hybridization and high morphological diversity in this group pose difficulties in identifying taxonomic units for comparative evolutionary studies and systematics. In order to understand the evolutionary relationships among poplars and to provide a framework for biosystematic classification, I reconstructed the phylogeny of the genus *Populus* based on nucleotide sequences of three non-coding regions of the chloroplast DNA (intron of *trnL*, and intergenic regions of *trnT-trnL* and *trnL-trnF*) and *ITS1* and *ITS2*of the nuclear *rDNA*. The resulting phylogenetic trees showed polyphyletic relationships among species in the sections *Tacamahaca* and *Aigeiros*. The chloroplast DNA sequence of *P. nigra* was similar to that of the species of section *Populus*, whereas nuclear DNA sequence data suggested a close affinity of *P. nigra* to species of the section *Aigeiros*, suggesting a possible hybrid origin for *P. nigra*. Similarly, the chloroplast DNA sequences of *P. tristis* and *P. szechuanica* were similar to that of the species of section *Aigeiros,* while the nuclear sequences revealed a close affinity to species of the section *Tacamahaca,* suggesting a hybrid origin for these two Asiatic balsam poplars. The incongruities

between nuclear- and chloroplast- DNA based phylogenetic trees suggest reticulate evolution in the genus *Populus*.

Acknowledgements

I would like to express my special thanks to my supervisor Dr. Selvadurai Dayanandan not only for giving me the opportunity to work in his laboratory, but also for all the help, support and patience through guidance and working with me, side by side.

I would also like to thank my committee members Dr. Paul Widden and Dr. William Zerges for their helpful comments and suggestions on the early version of this thesis.

I am very grateful to Dr. Barbara Thomas (Alpac, and the University of Alberta) and Mr. Alan Robertson (Alpac), Mr. Pierre Perinet (Le Ministere des Ressources Naturelles du Quebec), Mr. Normand Fleury and Ms. Edith Morin (Montreal Botanical Garden), Dr. Campbell Davidson (AAFC- Morden Research Station and the Arboretum, Manitoba), and Dr. Damase Khasa (Universite Laval) for their help in collecting leaf samples.

I would like to thank Concordia University, Natural Sciences and Engineering Research Council of Canada, Le Fonds Quebecois de la Recherche sur la Nature et les Technologies, and Canada Foundation for Innovation for their generous financial support that made this study possible.

My very special gratitude goes to my dear husband Reza for his never-ending love, patience, support and understanding.

I am deeply grateful to my wonderful parents and my sister Mahsa for their endless support and encouragements.

Finally I wish to thank all the crew of Forest and Evolutionary Genomics Lab for their great spirits that made lab working so much fun.

Table of Contents

List of Figures

List of Tables

1. Introduction

The family Salicaceae of the order Malpighiales (APG, 1998) comprise two genera, *Populus L.* with 30-40 species and *Salix L.* (willows) with 350-450 species. The species of genus *Populus* (aspen, cottonwood, and poplars) collectively known as poplars are one of the most widely used groups of forest trees, particularly in North America and Europe, where their ecological, economic and evolutionary importance is well known (Stettler et al., 1996). Poplars are widely distributed over the northern hemisphere, ranging from sub-tropical to boreal forests. They are most abundant in the temperate and sub-arctic regions of the northern hemisphere and have relatively few species in the southern hemisphere and tropics. To date no species of *Populus* native to Central and South America has been described (Argus, 1986; Eckenwalder, 1977 b).

Poplars are light- and moisture-loving colonizers of open and disturbed habitats and primarily grow in riparian or wet habitats on the fresh alluvial soils (Dickmann et al., 2001). All species of *Populus* are trees, although specific environmental conditions can result in localized stunting and shrub –like growth. Species of the genus *Populus* are all single-trunked, deciduous (or semi ever green) trees. All but one of the species (*P. heterophylla* of section *Leucoides* Spach.) is dioecious, with male (pollen-bearing) and female (ovule bearing) flowers occurring on separate trees (Dickmann & Stuart, 1983). Poplars show a high potential for long-range gene flow through wind-dispersed pollen and seeds. Poplar seeds are relatively small and embedded in a matrix of cotton-like fibres that provide floatation, enabling them to be carried long distances via wind and water (Strauss, 2002). Among temperate trees with female catkins, only the poplars and

willows (*Salix* spp.) have seeds with a cover of cottony hairs on partial placentas located in thin–walled capsules (Eckenwakder, 1996). Because of their short viability period, poplar seeds must reach sites with abundant moisture and sunlight shortly after dispersal or they will not survive (Strauss, 2002). Moreover, competition from herbaceous weeds soon after germination precludes or greatly reduces survival. For example, successful establishment of aspen seedlings in upland sites of the northern temperate and boreal regions usually requires fire or a comparable disturbance that exposes mineral soil and reduces competition (Dickmann et al., 2001). Similarly, cottonwoods in arid zones usually require riparian areas with newly deposited alluvial soils and little or no competition from herbaceous vegetation for their establishment (Strauss, 2002). Due to their high shade intolerance, poplars are unable to invade forest or herbaceous stands with a closed canopy (Strauss, 2002).

Because of the stringent conditions required for seed germination and seedling survival in poplars, vegetative reproduction is often more common than sexual reproduction, particularly for dispersion at the local level. These species can spread clonally by means of root-borne suckers (soboliferous), a feature uncommon among trees, yet useful in the management of poplar plantations (Stettler et al., 1996). All poplars tend to sprout vigorously from stumps after trees are cut or have fallen from natural causes (Dickmann et al., 2001). In some instances, poplars can proliferate and form large groves in areas where competition with other forest species is limited (e.g., in regions with recurrent flooding, high water tables, or fire; Smith, 1988). Thus, a given genotype can persist well beyond the lifespan of a single stem (approximately 50-300

years). The aspens (section *Populus*) in particular are vigorous root sprouters and can spread over a wide geographic area (Mitton & Grant, 1996).

Poplars play a significant ecological role as pioneer species in boreal forests and also as a dominant species in riparian forests that serve as rich wild life habitats and watersheds (Braatne et al., 1996). Poplars have also become one of the most economically important groups of trees, mainly due to their fast growth rates, profuse vegetative propagation (sucker development from root, sprouting from stump and stem, and rooting and sprouting from buried branches), rapid juvenile growth, adaptability to a variety of ecological sites, and multipurpose wood. Poplar wood is used for timber, paper pulp, veneer, fuel, pallets and furniture (Stettler et al., 1996). The importance of poplars is further increased because of their inherent biological characteristics. For example, poplars are relatively easy to breed through controlled pollination and there exist a number of natural and artificial hybrids with superior qualities that can be exploited in breeding programs. Therefore, poplars are a target for many breeding programs (Zsuffa, 1975) and are widely used in large-scale commercial plantations (Zsuffa et al., 1996). Since natural reservoirs of genetic variation have been found in several native *Populus* species, numerous attempts are underway to manipulate this variation through breeding, selection, biotechnology and genetic engineering (Dickmann et al., 2001; Stettlar et al., 1996). Poplars have also become useful models for the study of growth and development and tolerance to biotic and abiotic stresses in trees (Dickmann & Stuart, 1983).

Despite many recent advances in poplar biology (Stettler et al., 1996), relatively little is known about the evolutionary history of the genus *Populus*. Poplars are an ideal

system for a wide variety of evolutionary studies owing to their modest number of species in a morphologically diverse group, general abundance, rapid growth, ready vegetative propagation, easy cultivation, ability to flower and set seed on excised branches, obligate outcrossing, inter-fertility, cytological uniformity at the diploid level, extensive fossil record and rich array of pests and diseases (Eckenwalder, 1996).

A striking feature of poplars that has received considerable attention from many researchers is the occurrence of inter-species hybrids (Barnes et al., 1985; Eckenwalder, 1982). As in many other wind-pollinated, allogamous forest trees, natural hybridization is common in *Populus* and well documented (Eckenwalder, 1996; Whitham et al., 1996). Hybrids are regularly found in regions where species of sections *Aigeiros* and *Tacamahaca* are sympatric, such as in the contact zones of *P. angustifolia*, *P. trichocarpa*, and *P. balsamifera* (Brayshaw, 1965). Similarly, species of section *Populus* are known to hybridize naturally with other members of the section, such as *P. alba* with *P. tremula* (Stettler & Zuffa, 1996) and *P. grandidentata* with *P. tremuloides* (Barnes, 1961). However, to-date no natural hybrids between species of section *Populus* and species of other sections are known (Stettler & Zuffa, 1996). The readiness with which poplars and aspens reproduce asexually further facilitates gene flow among species through perpetuating hybrid plants and prolonging their role as agents of introgression (Smith, 1988). This may have been occurring for a long period of time, as there is evidence for ancient introgression between North American cottonwoods and sympatric balsam poplar species (Eckenwalder, 1984 a, b, c), as well as among North American aspens (Barnes, 1967). Moreover, RFLP analysis of cpDNA and rDNA (Smith, 1988) has raised the possibility of ancient hybridization by which, *P. nigra*

appeared to be an introgressant of the *P. alba* (cpDNA) lineage and some other presently unknown paternal lineage of section *Populus*. Similarly, *P. tristis* (a Central Asian member of section *Tacamahaca*) appeared to be an introgressant or a hybrid of the *P. nigra* (cpDNA) lineage and the lineage with Asian species of section *Tacamahaca*. Thus, it is not unreasonable to assume that there have been abundant opportunities for gene exchange among sympatric species, even between taxa of different sections (Stettler & Zsuffa, 1996). One of the interesting features of natural hybridization in hybrid zones is the backcrossing of hybrids with their parental species (Keim et al., 1989). Backcrossing is sometimes unidirectional (rather than bi-directional) as documented in hybrid swarms in the contact zone of *P. angustifolia* and *P. fremontii* in Utah (Keim et al. 1989). Restriction fragment length polymorphisms (RFLP) analysis showed that the hybrid population consisted of either F1 hybrids or backcrosses to *P. angustifolia*, but no trees attributable to crosses between F1s or between F1s and *P. fremontii* were found. The controlled backcrossing to *P. fremontii* led to early death of seedlings possibly due to developmental incongruity (Keim et al.,1989; Hogenboom, 1984). Spontaneous hybridization is also a common feature in regions where poplar cultivars are planted in the vicinity of native populations (Strauss, 2002). Hybrid breakdown and maladaptation are expected to limit the ability of hybrid progeny to invade established areas of wild poplar stands. However, when wild stands are relatively small compared to hybrid plantations, introgression may be detected even after a long period of time. For instance, introgression of introduced North American *P. deltoides* genes into wild stands of *P. nigra* in Europe has been documented (Heinze, 1997; Arens et al., 1998; Winfield et al.,

1998). This was detected soon after the introduction of *P. deltoides* into Europe in the 18th century (Houtzagers, 1937) and played a significant role in the spread of hybrid P. x *euramericana* cultivars that contain *P. deltoides* and *P. nigra* genomes. The widespread use of these P. x *euramericana* hybrids has subsequently compromised the genetic integrity of native *P. nigra* stands (Arens et al., 1998). The rapid shrinkage of those stands has further exacerbated the problem and called for vigorous gene conservation measures for *P. nigra* (Frison et al., 1995). Other examples of gene flow from introduced cultivars to local or native species include diverse hybrids generated between *P. alba* and *P. tremula* (*P.x canescens*) in Europe, and *P. tremuloides* and *P. grandidentata* in North America. Similarly, the occurrence of many hybrids in the vicinity of *P. nigra var. italica* (Lombardy poplar) plantings attests to the introgression of genes from an introduced male clone into various backgrounds (Stettler & Zsuffa, 1996). Controlled crosses among species of poplars have greatly augmented the hybrid combinations beyond those from spontaneous hybridization, especially to obtain combinations of desirable traits (Dickmann et al., 2001). For example, rootability of stem cuttings, which is an important trait in plantations culture, varies widely in *P. deltoides* and can be significantly improved in crosses with the well-rooting *Tacamahaca* poplars such as *P. balsamifera, P. maximowiczii, P. simonii*, and *P. trichocarpa* (Dickmann & Stuart, 1983). Another example is the individual level variation in poplar disease resistance based upon the species and the geographic origin of the individual. Resistance genes to *Marssonia*, *Septoria,* and *Dothichiza populea* are more likely to be found in *P. deltoides*, whereas those against *Xanthomonas populi* are found in *P. nigra*

(Stettler & Zsuffa, 1996). The acquisition of various resistance or tolerance genes is one of the desired trait combinations achieved through hybridization.

The extensive inter-species hybridization and the existence of high levels of morphological variation in poplars pose great difficulties in species delimitation for systematic and comparative evolutionary studies. The number of *Populus* species currently described in the literature ranges from 22 to 85 plus hundreds of hybrids, varieties, and cultivars (Dickmann & Stuart, 1983; Eckenwalder, 1996; Eckenwalder, 1977b). Discrepancies in the number of species could be attributed to the misinterpretation of some hybrids and to difficulties involved in delineating species boundaries. According to a recent classification (Eckenwalder, 1996), the genus *Populus* is classified into 29 species in six sections (*Abaso, Aigeiros, Leucoides, Populus, Tacamahaca, Turanga*; Appendix 1 Table 1). These sections are considered to be natural in most cases as they are delineated by the occurrence of major hybridization barriers (Eckenwalder, 1996; Zsuffa, 1975). However, the placement of several taxa within these sections remains controversial. For instance *P. nigra* of section *Aigeiros* shows genetic affinity to species of *Tacamahaca.* In *P. nigra*, cpDNA RFLP analysis showed similarity to species of the section *Populus*, but RFLP patterns of nuclear rDNA were distinct from the section *Populus*, suggesting a possible hybrid origin of *P. nigra* (Smith et al., 1990).

Recent phylogenetic analyses of the family *Salicaceae* using DNA sequence data from chloroplast *rbcL* (Azuma et al., 2000) and *ITS* of nuclear rDNA (Leskinen et al., 1999) strongly suggest that *Populus* is a monophyletic group sister to *Salix*. However, phylogenetic relationships within the genus remain controversial. Based on restriction fragment length polymorphism analysis, the section *Tacamahaca* has been suggested to

be polyphyletic and the section *Populus* is considered as the terminal clade (Smith, 1988). On the other hand, a phylogenetic tree based on DNA sequences from the *ITS* region of the nuclear rDNA of four species of *Populus* showed an opposite trend with *P. alba* of section *Populus* as basal followed by *P. lasiocarpa* of the section *Leucoides* and species of the sections *Aigeiros* and *Tacamahaca* as the terminal clade (Leskinen et al., 1999). Mitochondrial and chloroplast restriction site analysis of four *Populus* species suggests polyphyletic relationships for species in the section *Aigeiros* (Barrett et al., 1993; Rajora et al., 1995). Therefore, a detailed phylogenetic study involving a large set of taxa and carefully chosen data, such as DNA sequence data is needed to better understand the evolutionary history and classification of poplars.

The data based upon DNA sequences are particularly well suited for systematic studies owing to near independence from selective pressures and minimal bias toward convergence (Judd et al., 2002). Among the three genomes available for systematic studies in plants (nuclear, mitochondrial, and chloroplast), chloroplast DNA (cpDNA) is particularly well suited as a source of data for phylogenetic analysis (Judd et al., 2002). The linear order and arrangement of genes, DNA sequence, and overall size are mostly conserved in known chloroplast genomes (Palmer, 1985a, 1986). The relatively conserved chloroplast genome is particularly useful as a tool for phylogeny reconstruction as mutations occur relatively infrequently, persist over an evolutionary time scale and serve as heritable markers to identify taxa of shared ancestry (Smith, 1988). Moreover, the cpDNA lacks a strong transition-transversion bias, making convergence events less likely than in other forms of DNA (Curtis and Clegg, 1984). Furthermore, cpDNA occurs in high copy numbers and is relatively easy to isolate from

green plant tissues. The chloroplast genome is maternally inherited in most angiosperms with few instances of biparental inheritance (Palmer, 1988; Birky, 1995). However, no evidence for plastid fusion and intermolecular recombination exist (Palmer, 1985 a, b; Birky, 1995). As a result, trait "leakage" between lineages is minimal, and therefore phylogenies derived from cpDNA depict maternal lineages (Smith, 1988).

Despite the many advantages of cpDNA as a tool for phylogenetic studies, the lack of information corresponding paternal lineages limits the use of cpDNA data to gain a comprehensive understanding of the evolutionary relationships for a given group of organisms. This drawback is critical in studies of groups in which hybridization between lineages has played a substantial role in their evolution (Smith, 1988). Therefore, analysis of nuclear genomic data is invaluable to study the evolutionary history of both maternal and paternal lineages of a group of taxa. Thus, the combined use of cpDNA and nuclear DNA data is crucial to gain a comprehensive understanding of the evolutionary history of a given group of plant taxa. As a source of nuclear genomic data, nucleotide sequences of the rDNA array, which encodes for ribosomal RNA (rRNA), part of the cytosolic ribosome, have been used frequently in phylogeny reconstruction (Page & Holmes, 1998). Because of the crucial role of ribosomes in protein synthesis, rRNA is ubiquitous in all organisms, and often present in large quantities. The rDNA arrays found in the nuclear genomes of eukaryotes, produce three types of rRNA: *18S*, one component of the small subunit (~1800 bp), *28S* (>4000bp) and *5.8S* rRNA (~160bp) part of the large subunit (where S stands for Svedberg unit, a measure of the sedimentation rate). The rDNA array also contains an external transcribed spacer (ETS) and two internal transcribed spacers (*ITS1 and ITS2*) that contain signals needed to process the pre-rRNA.

The highly variable nucleotide sequences of spacers and less variable (conserved) coding regions make rDNA arrays suitable to reconstruct the phylogenetic relationships of both distantly- and closely-related species (Page & Holmes, 1998).

The DNA sequence based phylogenies may not be congruent with "true phylogeny" or "species" phylogenies due to a number of inherent biological phenomena (Sytsma, 1990) including gene conversion between multi-copy genes, interspecific hybridization (Raven, 1980) and subsequent backcrossing with a parental species (Anderson, 1953). It is likely that many *Populus* species may have undergone complex reticulate evolution (i.e. all the lineages have not been evolved independently but, some of them could be the result of hybridization between different lineages that has given rise to a new lineage; Eckenwalder, 1996). In a hybrid lineage resulting from reticulate evolution, parental nuclear genes may get fixed for one of the parental types through homogenization by concerted evolution of gene families (Page & Holmes, 1998), segregation during sexual reproduction or lineage sorting (Doyle, 1992). These events will generate discrepancies among phylogenetic trees based on uniparentally-inherited organeller sequences (Sears, 1980, Neal et al., 1986) and biparentally-inherited nuclear sequences or their phenotypic expression in morphology. Therefore, data from both nuclear and chloroplast genes are required to parse out reticulate evolutionary events (Page, 2000), and reconstruct robust phylogenetic trees.

The objective of my thesis is to reconstruct the phylogeny of genus *Populus* based on DNA sequences of chloroplast and nuclear genomes to improve understanding of the evolutionary history of the genus and to provide a framework for taxonomy and classification of *Populus*. The specific questions to be addressed in this thesis are:

a) What are the phylogenetic relationships among species of genus *Populus* as recognized in major taxonomic treatments?

b) Are there any discrepancies between phylogenetic trees based on nuclear and chloroplast DNA sequence data suggesting reticulate evolutionary events in *Populus*?

c) Is there any relationship between known interspecies hybridization and the evolutionary relatedness of corresponding taxa?

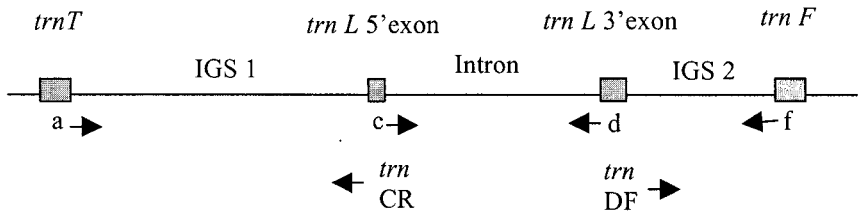
2. Material And Methods

2.1 Taxon sampling, DNA extraction, PCR amplification, and sequencing

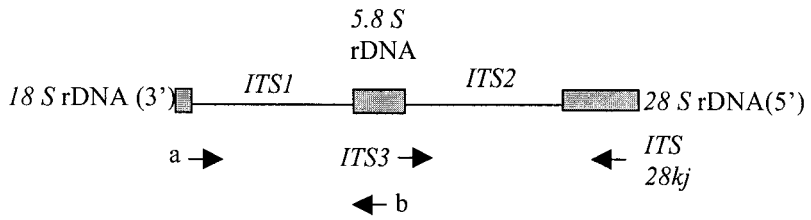
Taxon sampling was based on the recent classification of genus *Populus* proposed by Eckenwalder (Eckenwalder, 1977a,b; Appendix 1, Table 1). Fresh leaves were obtained from 21 species representing three sections of *Populus* (sections *Tacamahaca* Spach, *Aigeiros* Duby, and *Populus* (Leuce) Duby. This sampling includes two varieties of *P. deltoides* and two putative species not included in Eckenwalder's (1977 a, b) taxonomic treatment. DNA sequences from two species of *Salix* were used as outgroups. The list of species used in this study and the accession numbers are given in Appendix 1 and Table 2. The collected leaf samples were stored at –80 °C prior to extraction of DNA. Total genomic DNA was extracted from frozen leaf tissue using the methods of Doyle and Doyle (1987) and Dayanandan et al., (1997; Appendix 2, Protocol 1).

The three non-coding regions of cpDNA: *trnT-trnL* intergenic spacer (IGS1), *trnL intron*, and *TrnL –trnF* intergenic spacer (IGS2) (hereafter referred to as cpDNA) and *ITS1, ITS2* and part of the *5.8S* and the 5' region of the *28S* sub-unit of the rDNA (hereafter referred as rDNA) were amplified by the polymerase chain reaction (PCR). The oligonucleotide primers a, c, d, and f designed by Taberlet et al., (1991) for cpDNA, primers "a" and "b" modified from Leskinen (1999) for *ITS1*, and primers *ITS3* (Becerra, 1999) and *ITS 28kj* modified from Culling (1992) for *ITS2* were used for PCR amplification and sequencing. Two additional primers, *trn* CR and *trn* DF were designed

for sequencing to cover some gaps in the cpDNA sequences (Figure1: a, b; Appendix 1, Table 3).



(1a) — 100 bp



(1b) —— 100 bp

Figure 1: Relative positions and directions of primers used to amplify and sequence three non-coding regions of chloroplast DNA (1a), and *ITS* regions (1b).

Amplification reactions contained 230μM dNTP, 2.5mM $MgCl_2$, 5μM of each primer, 1 unit of Taq DNA polymerase, and 2.5μL buffer (0.2M Tris PH.9.5; 0.25 M KCl; 1mg/ml BSA, 5μl/ml tween 20) in a total volume of 25 μL (Appendix 2, Protocol 2). PCR amplification was performed in an Eppendorf Mastercycler gradient thermal cycler at 94°C for 60 seconds, 55°C for 30seconds, and 72°C for 60 seconds for 35 cycles (Appendix 2, Protocol 3). Amplified DNA was purified using a QIAGEN PCR purification kit, electrophoresed on 1% agarose gel with ethidium bromide (0.33μg/mL) at 3.5 v/cm for 90 minutes (see Appendix 2, Protocol 4 for agarose gel recipe). Amplified fragments were visualized and documented using a GeneSnap 4.00-Gene Genius Bio Imaging System (Syngene; A division of Synoptics LTD.). The digital image files were analyzed using Gene Tools software from Syngene. The quantity of DNA was estimated using a Mass Ruler DNA Ladder Mix (Fermentas). The purified, amplified DNA was directly sequenced using ABI Big Dye™ Terminator v. 3.0 and 3.1 Cycle Sequencing Ready Reaction kit (diluted 1: 4) and an ABI310 automated genetic analyser (Applied Biosystems, Foster city CA). Each region was sequenced between 2-7 times. Same primers were used for both PCR amplification and sequencing. The thermal cycling profile of sequencing reactions were: 96 °C for 10 sec., 50 °C for 5 sec., and 60 °C for 4 min. for 25 cycles (Appendix 2, Protocol 5, 6, 7). The chromatograms of DNA sequencing results were processed and analysed using the Staden software package (Staden group, MRC Laboratory of Molecular Biology, Cambridge, UK.). The assembled contigs of cpDNA and rDNA of each species were aligned using the ClustalW (Thompson et al., 1994) multiple sequence alignment software. Aligned DNA sequences were imported to MacClade 4.0 software (Maddison & Maddison, 2001) for verification

and manual editing of the sequence alignments. The 5'/3'*trnL* exon (recognized by comparing with the *Nicotiana* chloroplast sequence; Genebank accession number = NC_001879; gi = 11465934) was excluded from the matrix. The final data set included three non-coding regions of the *trnT-F* of cpDNA (intergenic region of *trnT-trnL*, *trnL intron*, and intergenic region of *TrnL –trnF*) and *ITS I*, partial *5.8S* rRNA, *ITS 2*, and part of *28S* subunit of the rDNA. The two data sets (cpDNA and rDNA) were analysed separately.

2.2 Phylogenetic analysis

The task of a molecular phylogeneticist is to infer the evolutionary relatedness of taxa through the reconstruction of phylogenetic trees based on molecular (DNA, RNA, and Protein) sequences. Three major methods, namely maximum parsimony, distance and maximum likelihood (Felsenstein, 1988), which differ from each other in their assumptions and algorithms of character state optimization are frequently used to reconstruct phylogenetic trees (Huelsenback et al., 1993; Appendix 4).

The phylogenetic analyses 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 investigated by chi-square test of homogeneity of base distribution across sequences (Appendix 3, Test1)

Parsimony analysis:

Maximum parsimonious (MP) phylogenetic trees were reconstructed through heuristic search with equal character weights, gaps treated as missing, multi state taxa interpreted as uncertainty, starting tree obtained via stepwise addition and sequence addition was as-is for the cpDNA data set (random addition was not possible due to the complexity of the data matrix) and random addition of sequences with 1000 replicates for rDNA data set. Tree-bisection reconstruction (TBR) was used as the branch-swapping algorithm. Strict and 50% majority rule consensus trees were obtained. The phylogenetic trees were rooted using *Salix* species as outgroups. Bootstrap analysis with

fast-heuristic search based on 1000 replicates was performed to assess the robustness of branches.

Maximum likelihood analysis:

The software program Modeltest v. 3.06 (Posada, 1998) was used to find the best nucleotide substitution model that fits the data set using hierarchical likelihood ratio test (LRT; Appendix 3, Test 2). The maximum likelihood (ML) analyses of both cpDNA and rDNA were performed through heuristic search with TBR branch swapping, addition of sequences as-is and the Tamura-Nei +I model (Tamura & Nei, 1993) as the nucleotide substitution model. The bootstrap analysis with fast-heuristic search based on 100 replicates was performed to assess the robustness of branches.

Distance method:

The corrected distances based on Modeltest results were used to build the trees. Robustness of branches was tested via bootstrapping of 1000 replicates with the neighbor-joining (NJ) search algorithm.

2.3 Comparison of cpDNA with rDNA tree

The two consensus trees obtained through maximum parsimony analysis of two data sets were compared using non-parametric Templeton (Wilcoxon signed-ranks) and winning-sites tests (Templeton, 1983; Appendix 3, Test 3). In order to compare two maximum likelihood trees, I used the Kishino-Hasegawa test using bootstrap with full optimization, two-tailed test (Kishino & Hasegawa, 1989; Appendix 3, Test 4) and also the Shimodaira-Hasegawa test using bootstrap with full optimization, one-tailed test (Shimodaira &Hasegawa, 1999; Appendix 3, Test 4). Each bootstrap analysis was performed with 1000 replicates.

3. Results

Direct sequencing of purified PCR products of the *trnT-trnF* chloroplast region and rDNA followed by ClustalW alignment of all sequences resulted in a data matrix with the length of 1414 characters (nucleotides) and 23 taxa with 71 parsimony informative ones for the cpDNA, and 791 characters and 23 taxa with 43 parsimony informative ones for the rDNA genomic region respectively. The corresponding DNA sequences of multiple individuals of the same species were identical.

The frequency distribution of the length of 10^6 random trees yielded g1 = -1.725565, g2 = 4.396228 for cpDNA and g1 = -1.313342, g2 = 2.71817 for the rDNA suggesting a strong phylogenetic signal in these data matrices. The Chi-square test of homogeneity of base frequencies across taxa showed p = 1.0 in both cases, and the null hypothesis of homogeneous base distribution across sequences was accepted.

The length of ITS 1 in all *Populus* species included in my study was 224 bp except for *P. fremontii*, which was 225 bp. The length of ITS 2 was 212 bp for all *Populus* species examined except for *P davidiana* and *P.tremula* (214 bp) and *P.maximowiczii* and *P.simonii* (203bp) which shared a deletion of 11 bp.

The cpDNA data matrix comprised IGS1 (473 bp), trn T-L intron (628 bp) and IGS 2 (312 bp).

3.1 cpDNA Trees

3.1.1 Maximum Parsimony analysis (MP)

Maximum parsimony searches of chloroplast DNA yielded 30939 equally parsimonious trees (tree length 118; CI= 0.924;RI= 0.927; RC= 0.857). The 50% majority rule consensus tree (Figure 2) and strict consensus tree differed only in the placement of *P. tremula* and *P. tremuloides* (with 53% occurrence). They appeared as sister taxa in a basal position to the other species of section *Populus* in the 50% majority rule tree, while in the strict consensus tree they grouped with other members of the section with an unresolved polytomy. The bootstrap value also did not support a basal position for these two aspens.

Based on MP analysis, all *Populus* species formed a strongly supported monophyletic group comprising three major clades. One clade comprised *P. simonii*, *P. maximowiczii, P. laurifolia*, and *P. songarica*. The other clade comprised species of section *Populus* (*P. grandidentata, P. alba, P. davidiana, P. tremula,* and *P. tremuloides*) and *P. nigra* of section *Aigeiros*. The phylogenetic relationships among taxa within this clade remained unresolved. The remaining clade comprised the remainder of the species. Within this clade, *P. angustifolia* clustered with *P. cathayana*, whereas *P. trichocarpa* clustered with *P. balsamifera. P. szechuanica* occupied a position basal to a group of species comprising *P. tristis, P. fremontii, P. roegneriana*, *P. angulata, P. deltoides*, and *P. sargentii*. The phylogenetic relationships among species of this group remained unresolved.

Besides *P. nigra*, which did not group with the other *Aigeiros* group members, the cpDNA data of *P. tristis* and *P. szechuanica* (Asiatic balsam poplars) suggested that these were more related to North American cottonwoods of the *Aigeiros* group than to species of section *Tacamahaca*. MP analysis placed the lineage consisting of North American cottonwoods (*P. fremontii*, *P. deltoides*, and its two varieties) and *P. tristis* of Asian balsam poplar (section *Tacamahaca*) in a terminal clade. However, based on the MP tree, the basal lineage could not be determined. Sections *Populus* was monophyletic and sections *Tacamahaca* and *Aigeiros* were polyphyletic groups (Figure 2).

3.1.2 Maximum likelihood analysis (ML)

The results of the Modeltest analysis showed that TrN +I (Tamura & Nei, 1993: equal rate for all transversions and different transition rates with unequal base frequencies) nucleotide substitution model was the most suitable model for the cpDNA. The parameters of the model were:

a) Base frequencies: A= 0.4175, C= 0.1336,G= 0.1467, T= 0.3023

b) Rate matrix: (A-C) = (A-T) = (C-G) = (G-T) =1.0; (A-G) =1.7645; (C-T) =2.5848

c) Among-site rate variation: proportion of invariable sites I =0.6953; equal rates of substitution for all variable sites.

The maximum likelihood analysis of cpDNA with the above model parameters retained a single tree (Figure 3) with a topology identical to the 50% majority rule consensus tree obtained from parsimony analysis. Moreover, the bootstrap values for branch robustness under the maximum likelihood criterion were similar to the values obtained from the same test under the parsimony criterion.

As with the MP analysis, the ML analysis also showed the monophyletic origin of section *Populus*, and the polyphyletic origin of sections *Tacamahaca a*nd *Aigeiros*. *Populus nigra* clustered with members of the section *Populus*. *Populus tristis* and *P. szechuanica* grouped with the lineage comprising the American cottonwoods of section *Aigeiros*.

3.1.3 Distance method: Neighbor-joining analysis (NJ)

Distance analysis of cpDNA data using the neighbor-joining method yielded a single tree (Figure 4). Although the tree showed a complete resolution, after collapsing branches with low bootstrap consensus values (<50%), the resulting tree topology was similar to the MP and ML trees with few exceptions. Further resolution of phylogenetic relationships was observed in the clade of *P. sargentii, P. deltoides* and *P. tristis*, as well as in the clade of *P. maximowiczii*, *P. simonii, P. laurifolia* and *P. songarica.*

As with the MP and ML trees, the NJ tree suggests a monophyletic origin of sections *Populus* and a polyphyletic origin of sections *Aigeiros* and *Tacamahaca.* Chloroplasts of *P. nigra* share the same recent ancestor as the members of section *Populus*. The genetic distance (based on nucleotide sequences) between *P. nigra* and either *P. davidiana*, or *P. tremula* was less than the genetic distance between *P. nigra* and any other species. Moreover the maternal lineage of *P. tristis* and *P. szechuanica* showed a close affinity with the North American cottonwoods of section *Aigeiros*. Based on distance measures, *P. tristis* and *P. szechuanica* have chloroplasts that are genetically closer to those of *P. deltoides* and *P. fremontii* than to those of members of the section *Tacamahaca.*

In the NJ tree (after collapsing branches with the low bootstrap values) *P. angustifolia* and *P. cathayana* are basal to the other poplars and, due to a higher resolution compared with MP or ML trees, *P. deltoides* and *P. tristis,* grouped as sister taxa at the terminal position of the tree.

3.2 rDNA trees

3.2.1 Maximum Parsimony analysis (MP)

The maximum parsimony search based on nuclear rDNA yielded 497 equally parsimonious trees (tree length 94; CI= 0.851; RI= 0.888; RC= 0.756). The 50% majority rule consensus tree (Figure 5) and the strict consensus tree differed at:

(i) the placement of two American aspens; *P. tremuloides* and *P. grandidentata*. In the 50% majority rule consensus tree, these two species grouped as sister taxa to the lineage consisting of other *Populus* species, but their placement in the strict consensus tree is not resolved.

(ii) the relatively resolved relationships among groups of *Tacamahaca* species with the 50% majority rule consensus, and the unresolved polytomy in the strict consensus tree. In the 50% majority rule consensus, a group of balsam poplars consisting of *P. angustifolia*, *P. cathayana*, *P. trichocarpa*, *P. balsamifera*, *P. tristis*, and *P. szechuanica* clustered as a group sister to the lineage comprising members of section *Aigeiros*. Although the branch representing these lineages occurred in 75% and 100% of 497 most parsimonious trees respectively, they were not supported by bootstrap analysis. Moreover, in the 50% majority rule consensus tree, within the balsam poplar lineage *P. trichocarpa*, *P. balsamifera*, *P. tristis*, and *P. szechuanica* clustered together as a group sister to *P. angustifolia* and *p. cathayana*. However none of these internal nodes and relationships was supported by the bootstrap analysis, even though they occurred with a high percentage in the 497 retained equally parsimonious trees (Figure5). Moreover

sister relationship of *P. maximowiczii* and *P. laurifolia* to the clade comprising two lineages of the *Aigeiros* and group of *Tacamahaca* species (as mentioned earlier) was very weakly supported. The basal position of *P. simonii* was not supported by the bootstrap values, even though they all had 100% occurrence in the MP trees.

In the MP analysis of rDNA with respect to the results of the bootstrap analysis, all *Populus* species studied formed a strongly supported monophyletic group comprising two major clades. One clade comprised all species of the section *Populus* with *P. tremuloides* and *P. grandidentata* occupying a position sister to the clade comprising Eurasian species of section *Populus* (*P. tremula, P. alba,* and *P. davidiana*), suggesting a monophyletic origin for this section. The other major clade included all of the remaining species studied.

The relationships among species within section *Tacamahaca* were unresolved and they grouped as a polytomy. However this section could be divided into two distinct groups of taxa: *P. maximowiczii*, *P. simonii*, and *P. laurifolia* in one group and the other members of the section in another group. The relationships among species within section *Aigeiros* were resolved and the MP analysis of rDNA suggested a monophyletic origin for this section. However, this was not supported by bootstrap analysis (<50%). A close relationship between *P. nigra* and *P. deltoides*, *var. angulata, and P. roegneriana* was evident. In contrast to the cpDNA MP tree, *P. nigra* nuclear DNA did not show a close affinity to members of section *Populus*. *Populus tristis* and P. *szechuanica* clustered as an unresolved polytomy with the remaining species of sections *Tacamahaca* and *Aigeiros*.

3.2.2 Maximum likelihood analysis (ML)

The result of the Modeltest analysis showed that Tamura-Nei +I (Tamura, Nei, 1993: equal transversion rate with variable transition substitution rate and unequal base frequencies) was the most suitable model for the rDNA. The parameters of the model were:

a) Base frequencies: A= 0.1841, C= 0.3205,G= 0.3242, T= 0.1711

b) Rate matrix: (A-C) = (A-T) = (C-G) = (G-T) =1.0; (A-G) =3.5870; (C-T) =5.6386

c) Among-site rate variation: proportion of invariable sites I =0.8085; equal rates of substitution for all variable sites.

The ML analysis of rDNA with the above model parameters yielded three trees. The 50% majority rule consensus tree (Figure 6) and the strict consensus tree differed only at the placement of *P. deltoides, var. angulata*. In the 50% majority rule consensus it occupied the basal position to the clade comprising *P. nigra* and *P. roegneriana* whereas in strict consensus they all were sister to each other.

The 50% majority rule consensus ML tree showed a topology similar to the 50% majority rule consensus MP tree except (i) in the placement of *P. deltoides, var. angulata* as mentiond before and (ii) in clustering *P. catahyana* with *P. trichocarpa*, *P. balsamifera*, *P. tristis*, and *P. szechuanica*. Likewise, bootstrap analysis of rDNA under the maximum likelihood criterion resulted in a tree with a topology compatible with the MP consensus tree after collapsing branches with low bootstrap values (<50%), however the clade comprising *P.balsamifera, P.trichocarpa, P. cathayana, P.angustofolia*, *P.szechuanica*, and *P. tristis* still was supported by bootstrap analysis (Figures 5 and 6).

Distance method: Neighbor-joining tree (NJ)

Distance analysis of rDNA data with the neighbor-joining method resulted in a single tree (Figure 7). Although the tree showed a complete resolution, after collapsing branches with negative or very short lengths and low bootstrap consensus values (<50%), the resulting tree topology was similar to the MP and ML trees. The NJ tree showed more resolution than the MP, or ML trees. However some of the internal branches showed low support and branch lengths were very short. For example the branch that placed *P. fremontii* in a basal position to *P. nigra lineage* is supported by a bootstrap value of only 56% and the length was 0.001.

According to bootstrap analysis, the rDNA NJ tree (Figure7) suggested a monophyletic origin for section *Populus* and polyphyletic origin for section *Tacamahaca* with at least two distinct lineages: one (similar to the ML tree) including *P. angustifolia, P. cathayana, P. trichocarpa, P. balsamifera, P. szechuanica*, and *P. tristis* and the second consisting of *P. maximowiczii, P. simonii*, and *P. laurifolia*. The nuclear genome of black poplar (*P.nigra*) is genetically very close to that of *P.deltoides, var. angulata*. Also, *P. tristis* and *P. szechuanica* have the shortest genetic distance with two North American balsam poplars, *P. balsamifera* and *P. trichocarpa* than the other Asian balsam poplar species.

3.3 Topology Comparison Test Results

The Templeton (Wilcoxon signed-ranks) test statistic Z = -5.9244 was smaller than the absolute values of the rank sums. The Winning-site test with $p<0.0001$ indicates a significant difference at $p<0.05$ under the null hypothesis of no difference between two trees.

Both the Kishino-Hasegawa and Shimodaira-Hasegawa tests with P = 0.000 rejected the null hypothesis that there was no difference between the two trees.

Since there were incongruences between the cpDNA and rDNA data sets and a lack of concordance between the phylogenetic trees based on the two genomic regions, it is likely that species in the genus *Populus* may have undergone a complex reticulate evolution.

4. Discussion

4.1 Phylogenetic relationships of *Populus* and hybridization

Comparison of the trees derived from three major methods of analysis showed that all trees were compatible with the tree topology obtained through maximum parsimony analysis. Therefore, in order to discuss the evolutionary patterns in *Populus*, I will consider the tree topologies similar to the maximum parsimony trees to be the most plausible and conservative hypotheses for the phylogenetic relationships within the genus *Populus* based upon cpDNA and nuclear rDNA sequence data (Figures 2 and 5).

Compared to the willows, there are relatively few species of poplars, and they fall into a number of morphologically and ecologically distinct groups, traditionally recognized as sections. With few exceptions, there is a reasonable agreement in the literature on the characteristics and species composition of these sections and major barriers to hybridization within the genus lie between the sections (Zsuffa, 1975). The most interesting question of sectional affiliation concerns the relationships between sections and among species within each section and also the placement of some controversial species or those of purported hybrid origin.

The species of section *Populus* are monophyletic with respect to their maternal lineage (cpDNA), with the exception of *P. nigra* of section *Aigeiros,* which clusters with species in section *Populus*. In terms of rDNA (nuclear) lineage section *Populus* is monophyletic, and *P. nigra* clusters with the other species of section *Aigeiros*. The phylogenetic relationships among the species within section *Populus* based on cpDNA

and rDNA sequence data are not in agreement with each other. Based on maternal chloroplast sequence data, European aspen (*P. tremula)* and the American trembling aspen (*P. tremuloides*) cluster together as sister taxa, and occupy the terminal position in the *Populus* clade. However, this relationship is weak and has less than 50% bootstrap support). In the rDNA based phylogenetic tree, white poplar (*P. alba*) and Korean aspen (*P. davidiana*) cluster together as sister species and occupy the terminal position. These two species along with European aspen (*P. tremula*) form a monophyletic group within section *Populus* with sister relationships to the North American aspens, *P. tremuloides* and *P. grandidentata.* Therefore, the phylogenetic trees based on these two genomic regions does not support the division of section *Populus* into two subsections namely *Albidae* (white poplars) and *Trepidae* (aspens) as proposed by Spach in1841 (Smith, 1988).

The relationships between sections *Aigeiros* and *Tacamahaca* have been controversial. The species of these two sections are known to be freely inter-fertile (Zsuffa, 1975; Eckenwalder, 1984a). Both cpDNA and rDNA sequence data suggest a polyphyletic origin of section *Tacamahaca.* The polyphyly of section *Tacamahaca* was also suggested by morphology-based phylogenenetic analysis (Eckenwalder1996). The rDNA based phylogenetic tree suggests a monophyletic origin of section *Aigeiros*. Similarly the cpDNA based phylogeny, in contrast, suggests a monophyletic origin, with the exception of *P.nigra* which shares its chloroplast ancestor with the section *Populus* and *P. tristis* from section *Tacamahaca* which clustered with North American cottonwoods of section *Aigeiros*. In other words, North American cottonwoods of

section *Aigeiros* are monophyletic with respect to their chloroplast lineage with the exception of *P.tristis*, which clusters within this group.

The phylogenetic relationships among species within sections, based on both cpDNA and rDNA are not well resolved. However, a few congruencies between the phylogenetic trees based on cpDNA and rDNA (with a low confidence level in the case of rDNA) can be recognized. The three Asian balsam poplars *P. simonii, P. maxomowiczii*, and *P. laurifolia* form a clade distinct from other members of the section *Tacamahaca*. Section *Aigeiros* show closer affinity to one of the lineages of balsam poplars that include *P. tristis* and *P. szechuanica* (Figures 2 and 5). Interspecific hybridization within *Populus* is generally limited to intrasectional crosses, or intersectional crosses between species of *Aigeiros* and *Tacamahaca* (Ronald, 1982). The crossing relationships between members of a given pair of poplar species may range from complete compatibility to complete incompatibility through the combined action of multiple barriers (Stettler & Zsuffa, 1996). Although intrasectional crosses tend to be fully fertile (Zsuffa 1975), the direction of the cross may affect the fertility. For instance, *P. deltoides* and *P. nigra* hybridize well if *P. nigra* is used as the male parent. The reciprocal cross is hampered by postzygotic barriers (Villar et al., 1996), possibly due to nuclear/cytoplasmic incompatibility. As discussed above, cpDNA of *P. nigra* shows close affinity to cpDNA of the species of section *Populus* whereas nuclear rDNA shows a closer affinity to species of section *Aigeiros*. As the cytoplasmic chloroplast is usually transmitted through maternal lineage in angiosperms, the hybrid with *P. nigra* as a maternal parent may render the cytoplasm a dysfunctional host for the hybrid nucleus through cytonuclear disequilibrium (Arnold, 1993). However, the existence of other

hybrids such as *P. nigra* x *P. trichocarpa* and *P. nigra* x *P. laurifolia* (Dickmann & Stuart, 1983) suggests that *P.nigra* chloroplasts are capable of functioning with nuclear genomes from certain other species. Intersectional crosses between *Tacamahaca* and *Aigeiros* are generally compatible in most combinations (Zsuffa, 1975) and have given rise to many vigorous clones used in plantations (Dickmann & Stuart, 1983).

The polyphyletic origin of section *Tacamahaca* and *Aigeiros* and the unresolved relationships among species within and between sections suggest relatively close evolutionary relationships among the species. This finding is consistent with observations of spontaneous hybridization among these species in nature and in cultivation. For instance, spontaneous hybridization between European *P. nigra* and introduced North American *P. deltoides* resulted in a dominant and widespread hybrid *P.x euramericana* (Lefevre et al., 2001).

Inter- and intra-sectional hybridization is common in regions of sympatry (Eckenwalder, 1996; Figure 8). Only sect. *Populus* is strongly reproductively isolated from the others and exhibits strong intersectional incompatibility barriers. However, the cellular and molecular bases of this have not yet been characterized (Villar et al., 1986). Nevertheless, this partial but significant reproductive isolation reflects the evolutionary divergence of species of section *Populus* from those in other sections including the species of two closest sections, *Aigeiros* and *Tacamahaca* (Eckenwalder, 1996; Figures 2 and 5). This is consistent with the manipulation required to succeed in the artificial crossing of aspens with other poplars (Guriers & Stettler, 1976; Ronald, 1982). A variety of artificial manipulations have been developed recently which successfully bypass these barriers and greatly expand the number of possible crosses. These efforts include

applications of mentor pollen, compatible pollen extract, and solvent washings of incompatible pollen and non-receptive stigmas (Guriers & Stettler, 1976).

It is generally assumed in the taxonomic literature that the relative inter-fertility of poplars (especially where these crosses are spontaneous) is a reflection of their genetic similarity (Stettler et al., 1996; Dickmann et al., 2001) and therefore, can form a basis on which to infer taxonomic affinities.

According to Eckenwalder (1977 b), *P. trichocarpa* and *P. balsamifera* of section *Tacamahaca* show clear signs of introgression in their regions of overlap (e.g., the Rocky Mountains), probably due to the affinity of their genomes, which permits natural hybridization and repeated back crossing with parental species. Sections *Tacamahaca* and *Aigeiros* are broadly sympatric in North America (Little, 1971), with overlapping ecological preferences (Fowells, 1965). The North American representatives of sections *Aigeiros* and *Tacamahaca* show such strong similarities in floral traits that they can hardly be distinguished on the basis of floral morphology (Eckenwalder, 1977b, 1984b). In fact, based on his extensive study of New World poplars, Eckenwalder (1977b) suggests the eventual merger of these two sections. The sister relationship of section *Aigeiros* (with North American cottonwoods in the basal position) with the rDNA based clade containing three North American balsam poplars of section *Tacamahaca* (*P. balsamifera, P. trichocarpa* and *P. angustifolia;* Figure 5), and the close affinities of *P. deltoides* and *P. fremontii* with North American balsam poplars in the cpDNA based phylogenetic tree (Figure 2) suggests a close evolutionary relatedness between these species and supports the merger proposed by Eckenwalder (1977b).

This trend of extensive hybridization in *Populus* is a major cause of disagreements on the total number of poplar species and their classification. In fact, some poplars such as *P. Balm-of-Gilead* (a variety of *P. x jackii*) and *P. x tomentosa*, that were originally described as species and cultivated for centuries have later been identified as natural or near natural hybrids (Eckenwalder, 1996). Although the long-term commercial benefit of hybridization in *Populus* includes exploitation for increased variability and novel gene combinations, hybridization may have played a significant role in the evolution of sections and their rapid allopatric speciation (Eckenwalder, 1996).

4.2 Discrepancies between phylogenetic trees based on chloroplast and nuclear nucleotide sequence data

The major discrepancies between the nuclear rDNA and cpDNA phylogenetic trees involve the placement of *P. nigra*, *P. tristis*, and *P. szechuanica,* and may suggest an ancient hybridization event between ancestors of paternal and maternal lineages of these extant species.

Similar to the results of chloroplast RFLP analysis (Smith 1990), the cpDNA sequence data indicates that *P. nigra* has a chloroplast genome derived from the clade of section *Populus* and divergent from the American cottonwoods of section A*igeiros*. However, based on the number of nucleotide substitutions mapped on the maximum parsimonious tree (Figure 2), *P. nigra, P. tremula*, and *P. davidiana,* with a single substitution from the common ancestral node, are more closely related to each other than to *P. alba,* which has four base substitutions. This indicates that the chloroplasts of *P. nigra* may have originated from ancestors of *P. tremula* or *P. davidiana* ancestors rather than from *P. alba,* as suggested by Smith et al., (1990). Alternatively we can conclude that either an extinct ancestor of section *Populus* or any extant species within this section but not included in this study may have played the maternal role in this hybridization event. Moreover, similarity between data from rDNA sequences and rDNA restriction site variation (Smith et al., 1990) indicates that the nuclear genome of black poplar (*P. nigra*) is distinct from species in section *Populus* and, contrary to Smith et al, (1990), very closely related to *P. deltoides* variety *angulata* from section *Aigeiros* (Figure 5). It also shows sister relationships with *P. fremontii*, and *P. deltoides* and *P. deltoides* var.

sargentii occupied a basal position. However, these relationships were not strongly supported by bootstrap analysis (<50%). Thus, the extant *P. nigra* may have derived from an ancient hybridization event involving an ancestor or extant species of section *Populus* as the maternal (cpDNA) donor, and the ancestor of the cottonwoods (probably the immediate ancestor of *P. deltoides*) of section *Aigeiros* as the paternal (rDNA) donor. The possible geographic location of this ancient hybridization event and its likely geographic isolation from the paternal species are discussed below.

The evidence from classical morphological, chemical, crossing, and pathogenic studies do not consistently favour the placement of black poplar in either section *Aigeiros* or section *Populus* (Smith, 1988) both of which have been proposed as the origin of *P. nigra*. Black poplar (*P. nigra*) and white poplar (*P. alba*) have been taxonomically distinct since the first classification of the genus by Duby in 1828 (Smith, 1988), in which *P. alba* and *P. tremula* (aspen) were included in section *Populus* and *P. nigra* and *P. deltoides* (cottonwood) were included in section *Aigeiros*. Duby based this separation only on stamen number (8 vs. 12-30) and bud characteristics (tomentose and dry vs. glabrous and sticky, respectively). Indeed, both black poplars and aspens have glabrous, sticky buds (as do other poplars; Smith, 1988). Stamen number also seems to be a questionable criterion by which to align *P. nigra* with *P. deltoides*. Section *Populus* has 5-12 (rarely 20) stamens (Eckenwalder, 1977 b) and *P. nigra* has 15-30 stamens. Considering that cottonwoods have 30-60 stamens (Eckenwalder, 1977 b), *P. nigra* would appear to be more similar to section *Populus*. Additional morphological traits also lead to this conclusion. *Populus nigra* has 4-8 ovules per carpel and thus resembles section *Populus* (2-6) more than the cottonwoods (7-25 ovules; Eckenwalder, 1984 c).

Populus nigra has red winter buds like those in section *Populus* and unlike the tan buds of the cottonwoods (Eckenwalder, 1977b, 1982). *Populus nigra* like all members of section *Populus* has two carpels, whereas the cottonwoods have 3-4 (Eckenwalder, 1977 b). Flattened petioles are common to aspens as well as to black poplars and cottonwoods (Smith, 1988). Cross-sections of the vigorous long shoots of *P. nigra* and members of section *Populus* are terete, while those of the cottonwoods and other poplars are angular (Smith, 1988). Eckenwalder's (1977 b) chromatographic analysis of flavonoids also links *P. nigra* to *P. alba* (but not to the morphologically more similar aspens). However, the relationship between *P. nigra* and section *Populus* is not supported by other morphological and crossing studies (Smith, 1988). The morphological characters of the fertile parts (stigma dilation and lobing, disk shape, capsule shape) and of vegetative parts (bracts of piscillate aments, compression of petioles, viscidness of overwintering terminal buds) provide no evidence of an affinity of *P. nigra* to section *Populus* but rather to the cottonwoods (Smith, 1990). Furthermore, in the cladistic analysis of 76 morphological characters of buds, leaves, inflorescences, male and female flowers and fruits performed by Eckenwalder (1996), *P. nigra* showed a sister relationship with the clade comprising *P. fremontii* and *P. deltoides* from section *Aigeiros*. Thus it seems that the placement of *P. nigra* within section *Populus* would necessitate invoking a considerable amount of morphological convergence of *P. nigra* to the cottonwoods.

In addition, of all possible crosses involving *P. nigra* as one parent (Dickmann et al., 2001), the most problematic crosses are among members of its own maternal lineage, section *Populus*. Successful crosses of section *Populus* with *P. nigra* apparently are only rarely achieved. Willing and Pryor (1976) reported successful *P. tremuloides* x *P. nigra*

and *P. alba* x *P. nigra* crosses using solvent washes (Smith, 1988). Moreover, section *Populus* is also distinguished from *P. nigra* by its relative immunity to certain leaf rust organisms. *Melampsora medusae* (American) and *Melampsora populina* (European) infect members of section *Aigeiros* including *P. nigra* and *Tacamahaca*, but rarely attack the aspens and white poplars of section *Populus* (Lefevre et al., 2001). Based on this evidence, the placement of *P. nigra* in either section *Populus* or section *Aigeiros* remains controversial.

Another major incongruence between our data sets is the status of *P. tristis*. This small tree, which except for its bright red petioles and forked bole, is very similar to American balsam poplar (*P. balsamifera*; Dickmann & Stuart, 1983), has chloroplast affinity to section *Aigeiros* (cottonwoods) and a nuclear genome related to *Tacamahaca* (balsam poplars; Figures 2 and 5). This suggests that *P. tristis* (Himalayan balsam poplar) may have derived from an ancient hybridization event with an ancestor of North American cottonwoods as the maternal (cpDNA) donor and, probably the immediate ancestor of the lineage comprising species of section *Tacamahaca* (North American *P. angustifolia, P. balsamifera, P. trichocarpa*, Chinese *P. cathayana* and *P. szechuanica*) as a paternal donor. The introgressant status of *P. tristis* was also suggested earlier by Smith (1988). Based on the results of chloroplast RFLP studies and a preliminary analysis of nuclear rDNA RFLP, he concluded that *P. tristis* is an introgressant or hybrid of the *P. nigra* (cpDNA) lineage and the Asian portion of section *Tacamahaca*. Again, however, the cpDNA sequence data in my study links *P. tristis* with the North American cottonwood (*P. deltoides* and *P. fremontii*) lineage. The nuclear sequence data suggest an affinity to species of the section *Tacamahaca* especially to North American *P.*

balsamifera and *P. trichocarpa*. *Populus tristis* was first introduced to the North American continent in the late 1920s by the late Dr. F.L Skinner from a male clone growing at Kew Gardens in England (Dickmann & Stuart, 1983). This species has been tested in the United States and Canada on a number of soil types and survival has been high, but growth has been variable (Dickmann & Stuart, 1983). As discussed before, the relative inter-fertility between poplar species is a reflection of their genetic similarity. Thus, the high morphological similarity of *P. tristis* to North American *P. balsamifera* and recent development of the highly valued hybrid clone *P. tristis* X *P. balsamifera* (Dickmann & Stuart, 1983) strengthen the proposed model for the evolution of the Himalayan poplar.

Populus szechuanica, a species native to China, is another source of discrepancies in the *Populus* taxonomy and its placement in section *Tacamahaca* is controversial. The cpDNA RFLP study clustered this species with species of section *Populus* and *P. nigra*. Conversely, rDNA RFLP analysis showed a close affinity between balsam poplar and *P. szechuanica* (Smith, 1990). My nuclear rDNA data confirmed the rDNA-based RFLP analysis result and clustered *P. szechuanica* in the same monophyletic clade consisting of some *Tacamahaca* species including *P. tristis* (Figure 5). The cpDNA, however, showed a very different pattern. In the cpDNA based most parsimonious tree, *P. szechuanica* occupied a basal position to the lineage comprising members of section *Aigeiros* and *P. tristis* (Figure 2). However, the cladistic analysis of 76 morphological characters of buds, leaves, inflorescences, male and female flowers and fruits put *P. szechuanica* in the same clade as other *Tacamahaca* species, which is paraphyletic to the *Aigeiros* lineage

(Eckenwalder, 1996). Therefore, the evolutionary history of *P. szechuanica* may be similar to that of *P. tristis* as discussed above.

Eckenwalder (1996) proposed that natural hybridization, while commonplace in genus *Populus* over at least the last several million years, has had little effect on speciation in this group, relative to factors promoting divergence. Eckenwalder's model of evolution for *Populus* (Eckenwalder, 1996) is based on cladistic analysis of morphological characters. He proposed an evolutionary pattern involving phases of ecological radiation and geographical vicariance. According to his model, a rapid (effectively simultaneous) radiation into distinct habitats characterized the three sections: *Tacamahaca*, *Aigeiros*, and *Populus*. These sections then underwent relatively rapid allopatric speciation. The implication is that the two main phases of radiation, the emergence of the advanced sections and of species within them, were so rapid that the exact sequence of splittings (or speciation) is largely intractable due to either conflicting character distribution (Eckenwalder, 1996), or possible reticulate evolution. Nevertheless, our data and phylogenetic trees suggest ancient intersectional hybridization or reticulate evolution for at least three of the taxa, namely *P. nigra, P. tristis* and *P. szechuanica*. Hence, natural hybridization in poplars, particularly intersectional hybridization merits further investigation at the molecular level for a better understanding of the evolutionary history of genus *Populus*.

The putative hybridization and introgression events, mentioned above, must have predated the start of the Miocene about 23.5 million years ago (Parrish, 1987), when the Northern land mass Laurasia had broken apart, opening up the Atlantic Ocean and consequently separating Eurasia from North America. In the putative hybrids studied, at

least one of the parental lineages (*Aigeiros*' cottonwoods) is native to North America and the other putative parental species is native to Eurasia with no geographical sympatry. Because of the short viability periods of poplar pollen and seeds, as well as unfavourable chemical and physical conditions, cross-oceanic hybridization is unlikely (but cannot be ruled out). Thus, hybridization likely occurred while North America and Europe/Asia were contiguous within the Laurasian land mass. Following hybridization, introgression to one of the parental lineage or geographical isolation of the hybrid population through long distance seed dispersal, followed by local adaptation could have led to significant phenotypic divergence.

Finally, for the two other taxa collected from the Montreal Botanical Garden, identified as *P. songarica* and *P. roegneriana*, I couldn't find any information about either the history or the natural distribution of the trees. The botanical garden had obtained them from the Sheridan Nursery in Montreal (1973), and the Boyce Thompson Arboretum, USA (1937) respectively, and could not provide further information. However, the cpDNA data of *P. songarica*, showed a close affinity to *P. laurifolia* from section *Tacamahaca* and its rDNA data placed the species with *P. maximowiczii* and *P. laurifolia*. Thus, even though the relationships among these species are not resolved, they occupied the basal position to a clade comprising species of section *Aigeiros* and other members of *Tacamahaca* (though with less than 50% confidence value; Figure 2 and 5). *Populus songarica* therefore, could be a variety or cultivar of *P. laurifolia*. Both cpDNA and rDNA sequence data showed close relationships between *P. roegneriana* and species in section *Aigeiros*. It would be risky to recognize this taxon as a variety of *P.*

deltoides or *P. fremontii*, as to my knowledge there is no support in the literature for this classification.

5. Conclusions

The overall conclusions of this study are that although *Populus* is a readily defined genus consisting of equally well-marked, ecologically coherent sections. At least two such sections, *Tacamahaca*, and *Aigeiros* each lack strict monophyletic origin. Species of section *Aigeiros* are monophyletic based on rDNA data, but polyphyletic with respect to their chloroplast lineage. Both cpDNA and rDNA sequences data suggest a polyphyletic origin for section *Tacamahaca*. However, species of section *Populus* are monophyletic. The lineage comprising species of section *Populus* is distinct from the two other sections and the lineage of Asiatic balsam poplars (*P. simonii, P. laurifolia,* and *P. maximowiczii*) of section *Tacamahaca* is diverged from other members of the section. Incongruency between nuclear DNA- and chloroplast DNA-based phylogenetic trees suggests reticulate evolution in genus *Populus*. The fossil evidence and current biogeographic distribution patterns of poplars suggest that the probable hybridization underlying the origin of *P. nigra, P. tristis, and P. szechuanica* likely predated the break up of the Laurasian land mass in Miocene.

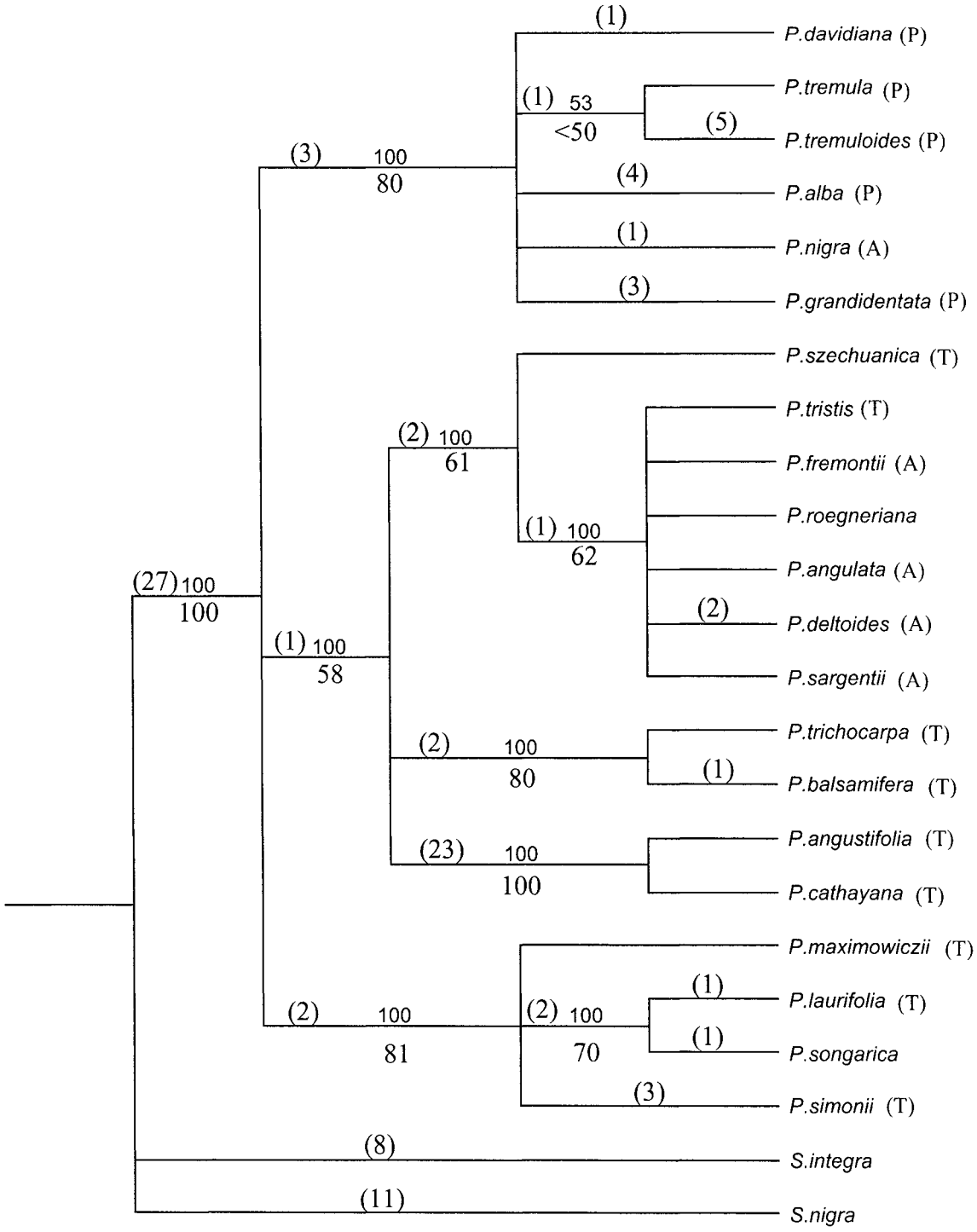


Figure 2: The majority rule consensus tree of 30939 equally parsimonious trees (tree length 118; CI=0.924) based on three non-coding regions of *trnT-trnF* of cpDNA sequences. Numbers above branches show the frequency of occurrence in 50% majority rule consensus, and numbers below branches indicate bootstrap % values. Numbers in brackets show branch lengths (number of nucleotide substitution). A = *Aigeiros*; P = *Populus*; T = *Tacamahaca.*

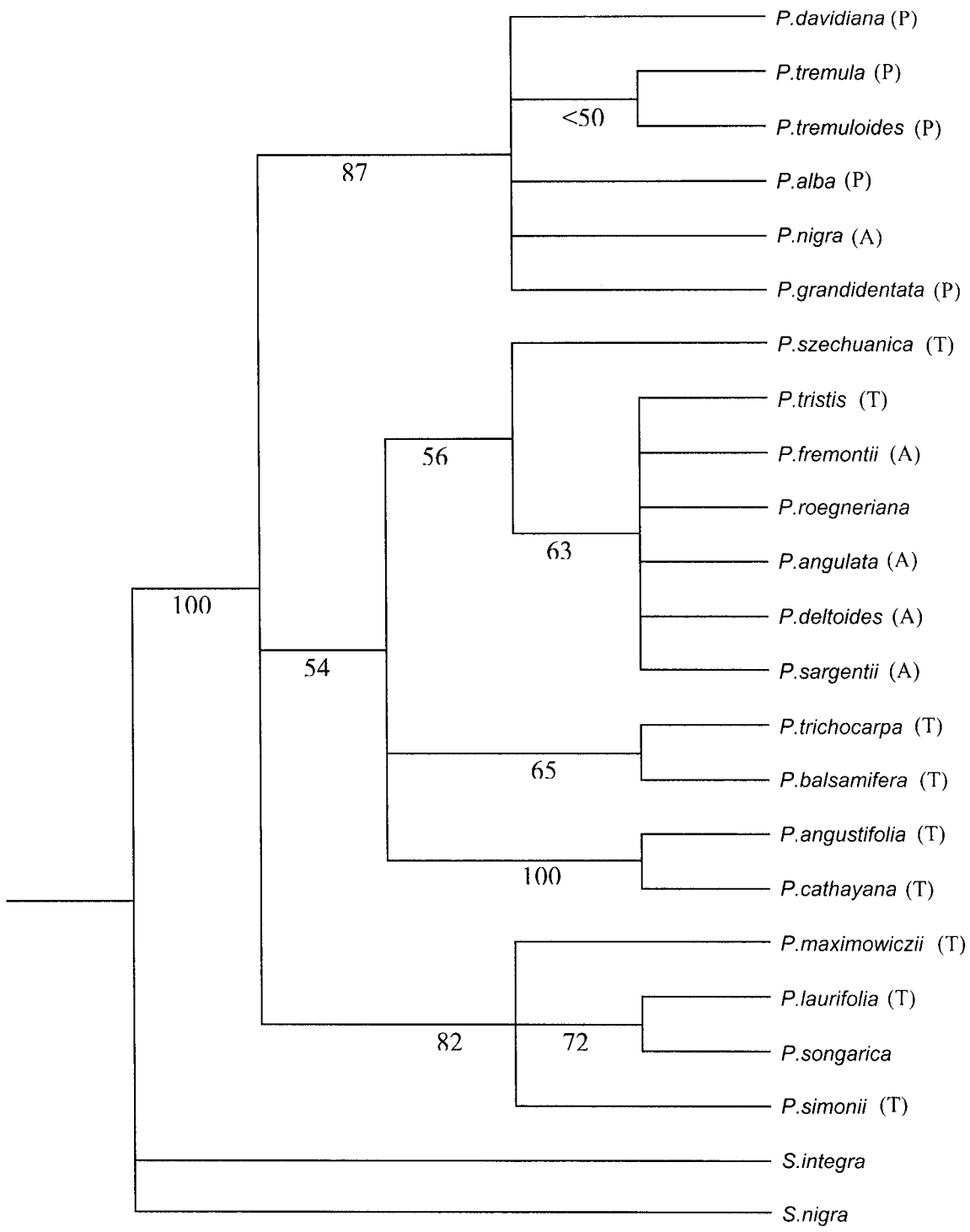


Figure 3: Maximum likelihood tree based on three non-coding regions of *trnT-trnF* of cpDNA sequences. Numbers below branches show bootstrap % values. A = *Aigeiros*; P = *Populus*; T = *Tacamahaca*.

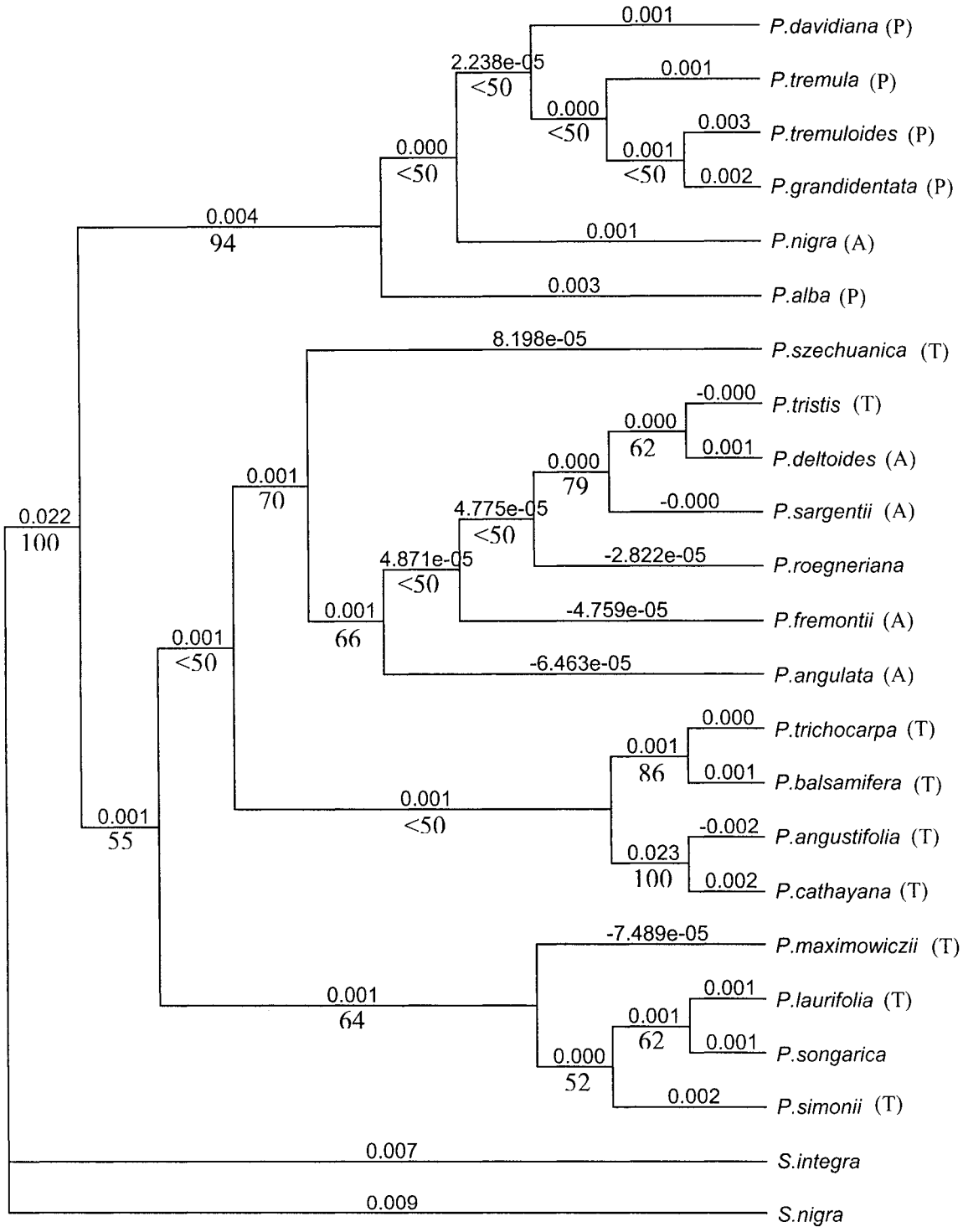


Figure 4: Neighbor-joining tree based on three non-coding regions of *trnT-trnF* of cpDNA sequences. Numbers above branches indicate branch lengths, proportional to distance measure. Numbers below branches show bootstrap % values. A = *Aigeiros*; P = *Populus*; T = *Tacamahaca.*

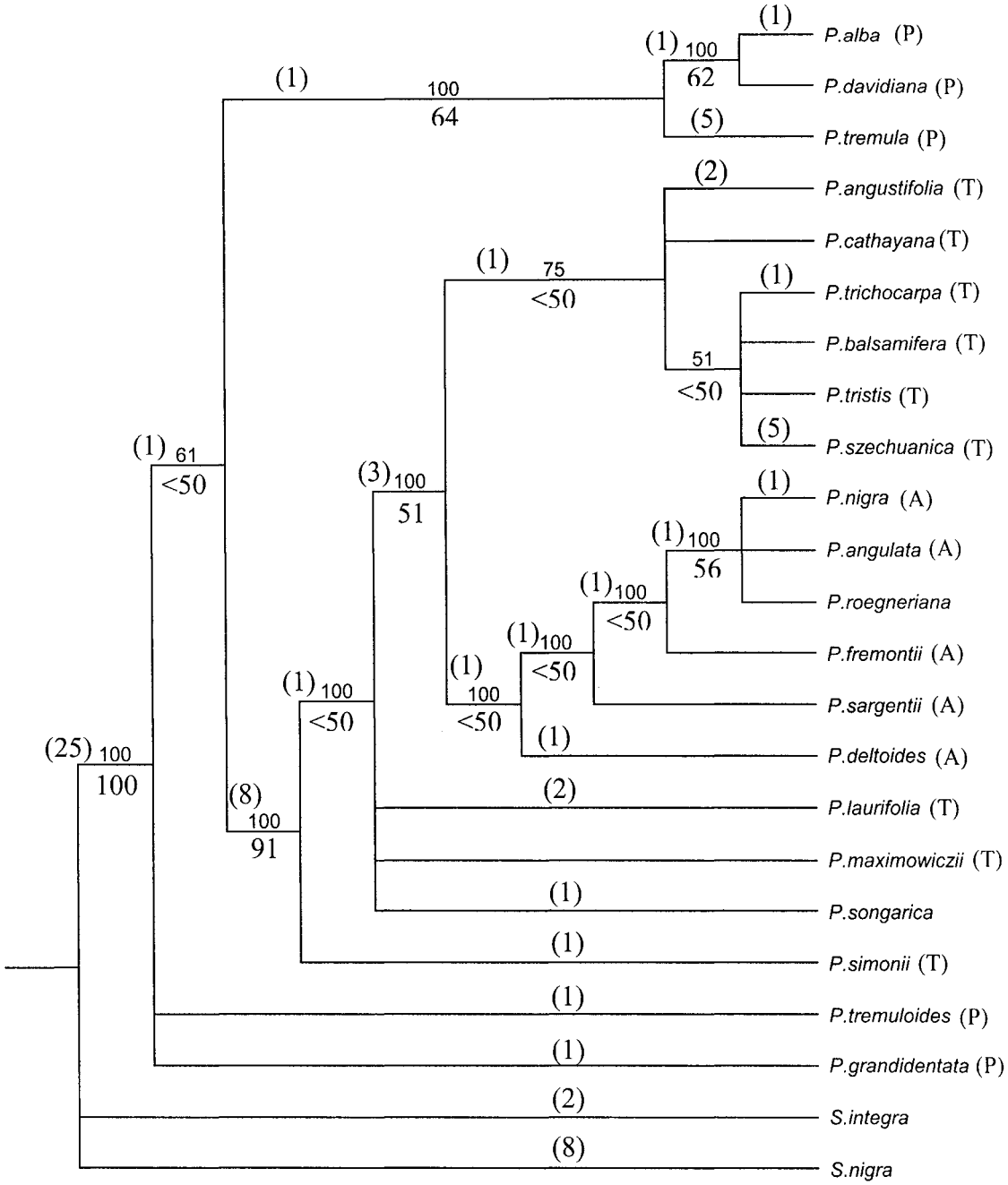


Figure 5: The majority rule consensus tree of 497 equally parsimonious trees (tree length 94; CI=0.851) based on partial *5.8S* RNA gene, *ITS1* & *ITS2*and part of *28S* subunit sequences. Numbers above branches show frequency of occurrence in 50% majority rule consensus, and numbers below branches indicate bootstrap % values. Numbers in brackets show branch lengths (number of nucleotide substitution). A = *Aigeiros*; P = *Populus*; T = *Tacamahaca.*

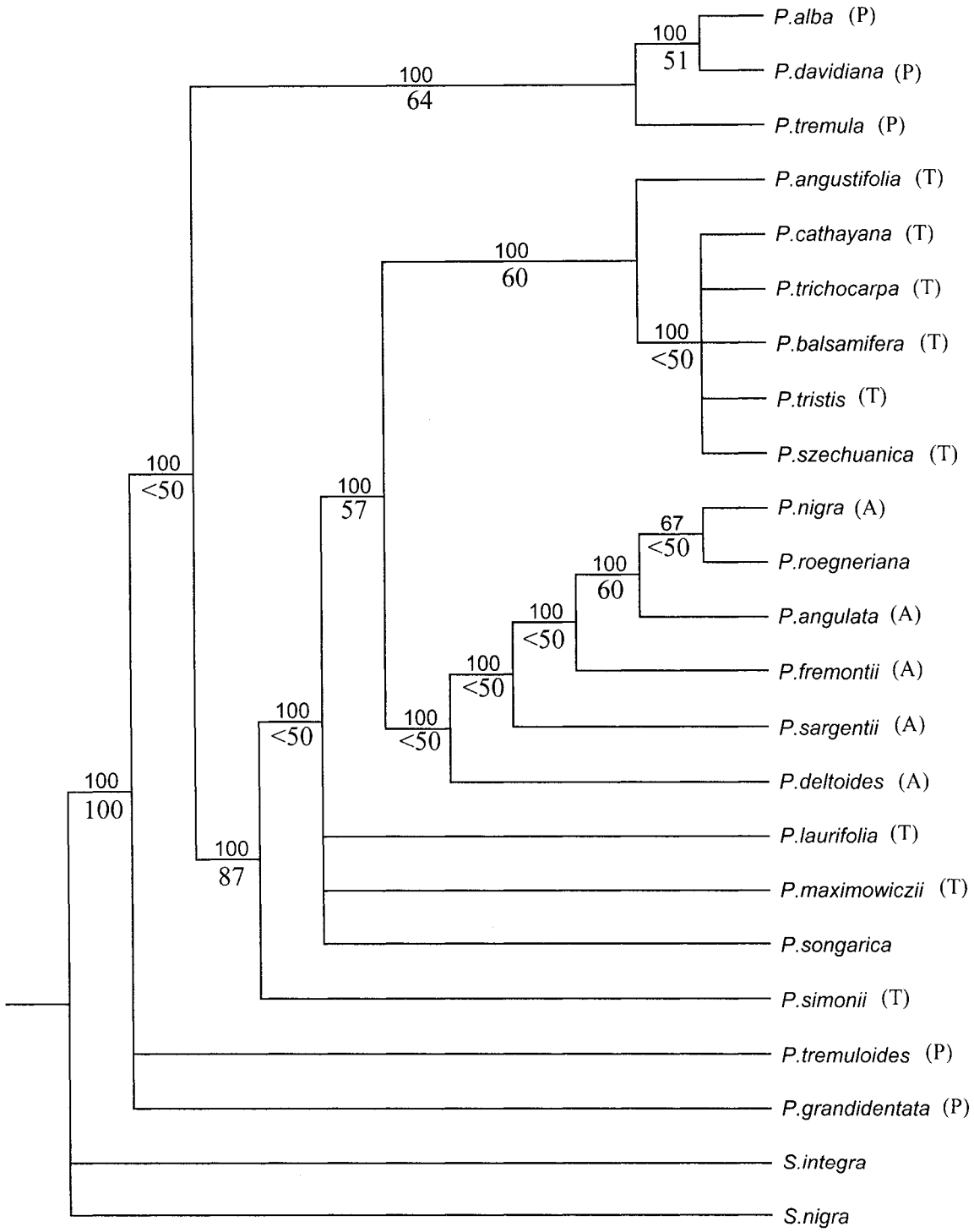


Figure 6: The majority rule consensus tree of three maximum likelihood trees based on partial *5.8S* RNA gene, *ITS1* & *ITS2* and part of *28S* subunit sequences. Numbers above branches show frequency of occurrence in 50% majority rule consensus, and numbers below branches show bootstrap % values. A = *Aigeiros*; P = *Populus*; T = *Tacamahaca*.

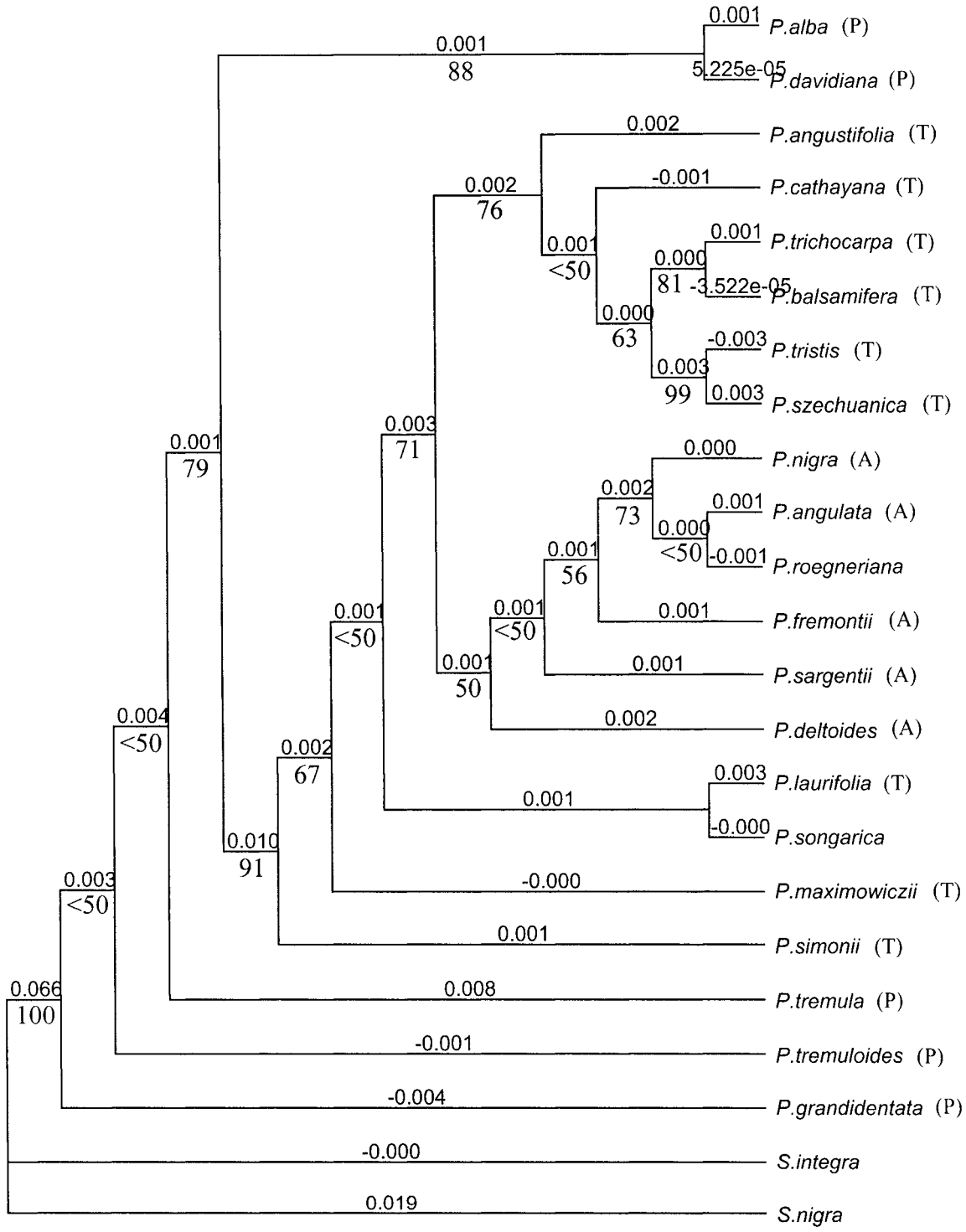


Figure 7: Neighbor-joining tree based on partial *5.8S* RNA gene *ITS1* & *ITS2* and part of *28S* subunit sequences. Numbers above branches indicate branch lengths, proportional to distance measure. Numbers below branches show bootstrap% values. A = *Aigeiros*; P = *Populus*; T = *Tacamahaca.*

Aigeiros			Tacamahaca			Populus		
deltoides								
1	*fremontii*							
1	1	*nigra*						
1	1	1	*angustifolia*					
1	0	3	1	*Balsamifera*				
1	1	1	1	1	*Trichocarpa*			
2	2	2	2	2	2	*tremuloides*		
2	0	2	0	2	0	1	*grandidentata*	
2	2	2	2	2	2	1	1	*alba*

Figure 8: Natural hybridization among North American species of *Populus*.
(Eckenwalder, 1996)

allopatric: 0

sympatric: 1 = hybrids present,

2 = no hybrids formed,

3 = unknown

Literature Cited

Anderson E. (1953) Introgressive hybridization. Logical Review and Biological Proceedings of the Cambrige Philosophical Society. 28: 280-307.

APG: The Angiosperm phylogeny group (1998) An ordinal classification for the families of the flowering plants. Annals of the Missouri Botanical Garden. 85: 531-553.

Arens P., Coops H., Jansen J., and Vosman B. (1998) Molecular genetic analysis of black poplar (*Populus nigra*) along Dutch rivers. Molecular Ecology. 7: 11-18.

Argus G.W. (1986) The genus *Salix (*Salicaceae*)* in the southwestern United States. Systematic botany monographs. vol. 9. American association of plant taxonomists.

Arnold J. (1993) Cytonuclear disequilibria in hybrid zones. Annual Review of Ecology and Systematics. 24: 521-554.

Azuma T. (2000) Phylogenetic relationships of Salix based on *rbcL* sequence data. American Journal of Botany. 87(1): 67-75.

Barnes B.V. (1961) Hybrid aspens in the lower Peninsula of Michigan. Rhodora. 63: 311-324.

Barnes B.V. (1967) Indications of possible mid-Cenozoic hybridization in the aspens of the Columbia Plateau. Rhodora. 69: 70-81.

Barnes B.V. and Pregitzer K.S. (1985) Occurrence of hybrids between bigtooth and trembling aspen in Michigan. Canadian Journal of Botany. 63: 1888-1890.

Barrett J., Rajora O.P. (1993) Mitochondrial DNA variation and genetic relationships of *Populus* species. Genome. 36: 87-93.

Becerra J., Venable D.L. (1999) Nuclear Ribosomal DNA phylogeney and ITS implications for evolutionary trends in Mexican *Bursera.* American Journal of Botany. 86 (7): 1047-1057.

Birky C.W. Jr. (1995) Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. Proceedings of the National Academy of Science. USA. Vol. 92. pp.11331-11338.

Braatne J.H., Hinckly T.M. (1992) Influence of soil water supply on the physiological and morphological components of plant water balance in *Populus trichocarpa, Populus deltoides* and their F1 hybrids. In: Biology of *Populus* and its implications for management and conservation. Part I, Chapter 3. Edited by Stettler R.F., Bradshaw H.D., Jr., Heilman P.E., and Hinckley T.M. NRC Research Press, National Research Council of Canada, Ottawa, ON, Canada. pp. 57-85.

Brayshaw T.C. (1965) Native poplars of southern Alberta and their hybrids. Canada Forestry Services. Publ. 1109, Ottawa, ON.

Culling K.W. (1992) Design and testing of plant-specific PCR primer for ecological and evolutionary studies. Molecular Ecology. 1, 223-240.

Curtis S.E and Clegg M.T. (1984) Molecular evolution of chloroplast DNA sequences. Molecular biology and evolution. 1: 291-301.

Dayanandan S., Bawa K.S., and Kesseli R.V. (1997) Conservation of microsatellite among tropical trees (*leguminosae*). American Journal of Botany. 84: 1658-1663.

Dickmann D.I. (2001) An overview of the genus *Populus.* In: Poplar culture in North

America. Part A, Chapter 1. Edited by Dickman D.I., Isebrands J.G., Eckenwalder J.E, and Richardson J. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp.1-42.

Dickmann D.I., Isebrands J.G., Eckenwalder J.E., Richardson J. (2001). Poplar culture in North America. NRC Research Press, Ottawa, Ontario, Canada.

Dickmann D.I, Stuart K. (1983) The culture of poplars in Eastern North America. Department of Forestry, Michigan State University, East Lansing, Michigan.

Doyle J.J., Doyle J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin. 19: 11-15.

Doyle J.J. (1992) Gene trees and species trees: molecular systematics as one-character Taxonomy. Systematic Botany. 17, 144-163.

Eckenwalder J.E. (1997a) North American cottonwoods (*Populus,* Salicaceae) of sections *Abaso*, and *Aigeiros*. Journal of the Arnold Arboretum, Harvard University. 58: 193-208.

Eckenwalder J.E. (1977b) Systematics of *Populus L.* in southwestern North America with special reference to sect. *Aigeiros* Duby. Ph.D thesis, University of California, Berkeley, CA.

Ecknewalder J.E. (1982) *Populus xinopia hybr. nov.* (Salicaceae), a natural hybrid between the native North American *P. fremontii* S. Watts and the introduced Eurasian *P. nigra* L. Madrono. 29(2): 67-78.

Eckenwalder J.E. (1984a) Natural intersectional hybridization between North American species of *Populus* in sections A*igeiros* and *Tacamahaca*. I. Population studies of *P. parryi*. Canadian Journal of Botany. 62: 317-324.

Eckenwalder J.E. (1984b) Natural intersectional hybridization between North American species of *Populus* in sections A*igeiros* and *Tacamahaca*. II. Taxonomy. Canadian Journal of Botany. 62: 325-335.

Eckenwalder J.E. (1984c) Natural intersectional hybridization between North American species of *Populus* in sections A*igeiros* and *Tacamahaca*. III. Paleobotany. Canadian Journal of Botany. 62: 336-342.

Eckenwalder J.E. (1996) Systematics and evolution of *Populus*. In: Biology of *Populus* and its implications for management and conservation. Part1, Chapter1. Edited by Stettler R.F., Bradshaw H.D., Jr., Heilman P.E., and Hinckley T.M. NRC Research Press, National Research Council of Canada, Ottawa, ON, Canada. pp. 7-32.

Farris J.S. (1989) The retention index and the rescaled consistency index. In: MacClade 4. Maddison D.R., Maddison W.P (2001) University of Arizona. Sinauer Associate. Sunderland, Massachusett. U.S.A. pp. 384-386.

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

Fowells H.A. (1965) Silvics of forest trees of the United States. In: Eckenwalder J.E. (1984b) Natural intersectional hybridization between North American species of *Populus* in sections A*igeiros* and *Tacamahaca*. II. Taxonomy. Canadian Journal of Botany. 62: 325-335.

Frison E., Lefevre F., and Turok J. (1995) *Populus nigra* network. Report of the first meeting, 3-5 Oct. 1994, Izmit, Turkey. International plant genetic resources institute, Rome, Italy.

Guries R.P., Stettler R.F. (1976) Pre-fertilization barriers to hybridization in the poplars. Silvae Genetics. 25: 37-44.

Heinze B. (1997) A PCR marker for a *Populus deltoides* allele and its use in studying introgression with native European *Populus nigra*. Belgium Journal of Botany. 129: 123-130.

Hogenboom N.G. (1984) Incongruity: Non-functioning of intercellular and intracellular partner relationships through non-matching information. In: Cellular interactions. Edited by Linskens H.F. and Heslope-Harrison J. Springer-Verlag, Berlin. pp. 640-654.

Houtzagers G. (1937) Het geslacht *Populus* in verband met zijn betekenis voor de houttelt. In: Stettler R.F., Zsuffa L., and Wu R. (1996) The role of hybridization in the genetic manipulation of Populus. In: Biology of *Populus* and its implications for management and conservation. Part 1, Chapter 4. Edited by Stettler R.F., Bradshaw H.D., Jr., Heilman P.E., and Hinckley T.M. NRC, Research Press, National Research Council of Canada, Ottawa, ON, Canada. pp. 87-112.

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

Judd W. S., Campbell C. S., Kellogg E. A., Stevens P. F., Donoghue M. J. (2002) Plant Systematics: A phylogenetic approach. Second edition. Chapter 5. Sinauer Association, Inc. Sunderland, MA. pp. 105-119.

Keim P., Page K.N., Whitham T.G., and Lark K.R. (1989) Genetic analysis of an

interspecific hybrid swarm of *Populus*: Occurrence of unidirectional introgression. Genetics. 123: 557-565.

Kishino H., and Hasegawa M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Homonoidea. Journal of Molecular Evolution. 29: 170-179.

Leskinen E. (1999) Molecular phylogeny of Salicaceae and closely related Flacourtiaceae: evidence from *5.8S*, *ITS1* and *ITS2* of the rDNA. Plant Systemetics and Evolution. 215: 209-227.

Lefevre F., Kajba D., Heinze B., and Turok J. (2001) Black poplar: A model for gene resource conservation in forest ecosystems. The forestry Chronicle. Vol. 77, Nov. 2: 239- 244.

Little E.L. (1971) Atlas of United States trees. 1. Conifers and important hard woods. In: Eckenwalder J.E. (1984b) Natural intersectional hybridization between North American species of *Populus* in sections A*igeiros* and *Tacamahaca*. II. Taxonomy. Canadian Journal of Botany. 62: 325-335.

Maddison D.R.; Maddison W.P. (2001) MacClade 4. University of Arizona. Sinauer Associate. Sunderland, Massachusett. U.S.A.

Mitton J.B., and Grant M.C. (1996) The natural history of quaking aspens. Bioscience. 46: 25-31.

Neale D.B., Wheeler N.C. (1986) Paternal inheritance of chloroplast DNA in Douglas fir. Canadian Journal of Forest Research. 16: 1152-1154.

Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic tree. Molecular Biology and Evolution. 4: 406-25.

Page R.D., Holmes E.C. (1998) Molecular Evolution, A phylogenetic Approach. Blackwell Publishing. UK.

Page R.D. (2000) Extracting species trees from complex gene trees: reconciled trees and vertebrate phylogeny. Molecular Phylogenetics and Evolution. 14(1): 89-160.

Palmer J.D. (1985a) Chloroplast DNA and molecular phylogeny. Biological Essays 2. 263-267.

Palmer J.D. (1985b) Evolution of chloroplast and mitochondrial DNA in plants and Algae. In: Monographs in Evolutionary biology, Evolutionary genetics. Edited by MacIntyre R.J. Plenum press. N.Y. pp. 131-240.

Palmer J.D. (1986) Chloroplast DNA and phylogenetic relationships. In: DNA Systematics. Edited by Dutta S.K. Vol.2. CRC Press, Inc. Boca Raton, Florida pp. 64-76.

Palmer J.D., Jansen RK (1988) Chloroplast DNA variation and plant phylogeny. Annals of the Missouri Botanical Garden. 75: 1180-1206.

Parrish J.T. (1987) In: Origin and evolution of tropical rain forests. Chapter 11, Northern hemisphere megathermal rain forests. Edited by Morley R.J. John wiley & Sons, LTD. New York. pp. 240.

Posada D., Crandall K.A. (1998) Modeltest: testing the model of DNA substitution. Bioinformatics. 14(9): 817-818.

Rajora O.P., Danick B.P. (1995) Chloroplast DNA variation in *Populus*. Interspecific restriction fragment polymorphisms and genetic relationships among *Populus deltoides*, *P. nigra*, *P. maximowiczii*, *P. x Canadensis*. Theoretical and Applied Genetics. 90: 324- 330.

Raven P.H. (1980) Hybridization and the nature of species in higher plants. Canadian Botanical Association Bulletin. 13: 3-10

Ronald W.G. (1982) Intersectional hybridization of *Populus* sections, *Leuce-Aigeiros* and *Leuce-Tacamahaca*. Silvae Genetics. 31: 94-99.

Sears B.B. (1980) The elimination of plastids during spermatogenesis and fertilization in the plant kingdom. Plasmid. 4: 233-255.

Shimodaira H., and Hasegawa M. (1999) Multiple comparison of log-likelihoods with applications to the phylogenetic inference. Molecular Biology and Evolution. 16: 1114- 1116.

Smith R.L. (1988) Phylogenetics of *Populus* L. (Salicaceae) based on restriction site fragment analysis of cpDNA. Master thesis. University of Wisconsin, Madison.

Smith R.L., Sytsma K.J. (1990) Evolution of *Populus nigra* (sect. *Aigeiros*): Introgressive hybridization and the chloroplast contribution of *Populus alba* (sect. *Populus*). American Journal of Botany. 77(9): 1176-1187.

Stettler R.F., Bradshaw H.D., Jr., Heilman P.E. (1996) Biology of *Populus* and its implications for management and conservation. NRC, Research Press, National Research Council of Canada, Ottawa, ON, Canada.

Stettler R.F., Zsuffa L., and Wu R. (1996) The role of hybridization in the genetic manipulation of *Populus*. In: Biology of *Populus* and its implications for management and conservation. Part 1, Chapter 4. Edited by Stettler R.F., Bradshaw H.D., Jr., Heilman P.E., and Hinckley T.M. NRC, Research Press, National Research Council of Canada, Ottawa, ON, Canada. pp. 87-112.

Strauss S. (2002) Report of the poplar working group. Oregon State University.

Swofford D.L., Olsen G.J., Waddel PG., & Hillis D.M. (1996). Phylogenetic inference. In: Molecular Systematics (2nd edn). Edited by D.M. Hillis, C. Moritz & B.K. Mable. Sinauer Associate, Sunderland, Massachusetts, USA. pp. 407-514.

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

Sytsma K.J. (1990) DNA and morphology: inference of plant phylogeny. Trends in Ecology and Evolution. 5: 104-110.

Taberlet P. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology. 17: 1105-1109.

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.

Templeton A.R.(1993) Convergent evolution and non-parametric inferences from Restriction fragment and DNA sequence data. In: Statistical analysis of DNA sequence data. Edited by B. Weir. Marcel Dekker; New York. pp.151-179.

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

Villar M., Gaget M., and Dumas C. (1986) Sexual reproduction biology in *Populus*, compatibility and incompatibility. In: Smith R.L. (1988) Phylogenetics of

Populus L.(Salicaceae) based on restriction site fragment analysis of cpDNA. Master thesis. University of Wisconsin, Madison.

Villar M.,Gaget M.(1996) In: Stettler R.F., Zsuffa L., and Wu R. (1996) The role of hybridization in the genetic manipulation of *Populus*. In: Biology of *Populus* and its implications for management and conservation. Part 1, Chapter 4. Edited by Stettler R.F., Bradshaw H.D., Jr., Heilman P.E., and Hinckley T.M. NRC, Research Press, National Research Council of Canada, Ottawa, ON, Canada. pp. 87-112.

Winfield M.O., Arnold G.M., and Edwards K.J. (1998) A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the upper Severn area of the UK using AFLP markers. Molecular Ecology. 7: 3-10.

Whitham G.D., Martinsen E.M. (1996) Ecological and evolutionary implications of hybridization: *Populus*-herbivor interactions. In: Stettler R.F., Zsuffa L., and Wu R. (1996) Physiology of growth. In: Biology of *Populus* and its implications for management and conservation. Part 1, Chapter 11. Edited by Stettler R.F., Bradshaw H.D., Jr., Heilman P.E., and Hinckley T.M. NRC, Research Press, National Research Council of Canada, Ottawa, ON, Canada. pp. 247-275.

Zsuffa L. (1975) A summary review of interspecific breeding in the genus *Populus*. In: Proceedings of the 14th Meeting of the Canadian Tree Improvement Association, part 2. Canadian Forest Service, Ottawa, ON. Canada. pp. 107-123.

Zsuffa L., Giordano E., Pryor L.D., and Stettler R.F. (1996) Trends in Poplar culture: some global and regional perspectives. In: Biology of *Populus* and its implications for management and Conservation. Part 2, chapter 19. Edited by

Stettler R.F., Bradshaw H.D., Jr., Heilman P.E., and Hinckley T.M. NRC, Research Press, National Research Council of Canada, Ottawa, Canada. pp. 515-539.

Appendix 1 – Tables

Table 1: Sections, subsections, and species in the genus *Populus* (Eckenwalder, 1977 a, b)

Section	**Species**	**common name**
Populus (Leuce Duby)		
Subsection ALBIDAE	*P.alba*	White poplar
	P.monticola	Mexican white poplar
Subsection TREPIDAE	*P.davidiana*	Korean aspen
	P.grandidentata	Bigtooth aspen
	P.tremuloides	Trembling aspen
	P.tremula	European aspen
	P.adenopoda	Chinese aspen
	P.sieboldii	Japanese aspen
	P.tremuloides	Trembling aspen
Tacamahaca Spach.	*P.angustifolia*	Narrow-leaved cottonwood
	P.balsamifera	Balsam poplar
	P.koreana	Korean poplar
	P.cathayana	
	P.laurifolia	Laurel poplar
	P.maximowiczii	Japanese poplar
	P.simonii	Simon poplar
	P.szechuanica	
	P.trichocarpa	Black cottonwood
	P.yunnanensis	
	P.suaveolens	
	P.tristis	Himalayan balsam poplar
Aigeiros Duby	*P.deltoides*	Eastern cottonwood
	P.fremontii	Fremont cottonwood
	P.nigra	Black poplar
Turanga Bge.	*P.euphratica*	Euphrates poplar
Leucoides Spach	*P.ciliata*	Himalayan poplar
	P.heterophylla	Swamp cottonwood
	P.lasiocarpa	
	P.wilsonii	Wilson poplar
Abaso Ecken.	*P.mexicana*	Mexican poplar

Table 2: List of the species studied.

Section	Species	Voucher #	Location
Populus(Leuce Duby)	*P.alba*	MH-002	AP
	P.davidiana	MH-007	AP
	P.grandidentata	MH-013	MRNQ
	P.tremuloides	MH-024	AP
	P.tremula	MH-019	AP
Tacamahaca Spach.	*P.angustifolia*	MH-059	MAM
	P.balsamifera	MH-054	AP
	P.cathayana	MH-060	MA
	P.laurifolia	MH-061, MH-062	MA
	P.maximowiczii	MH-030	MRNQ
	P.simonii	MH-065	MBG
	P.szechuanica	MH-066	MBG
	P.trichocarpa	MH-040	MRNQ
	P.tristis	MH-067	MBG
Aigeiros Duby	*P.deltoides*	MH-049	AP
	P.deltoides,var. angulata	MH-071	MAM
	P.deltoides,var. sargentii	MH-070	MAM
	P.fremontii	MH-063	MAM
	P.nigra	MH-043, MH-044, MH-045, MH-046, MH-047	MRNQ
Other samples	*P.songarica*	MH-069	MBG
	P.roegneriana	MH-068	MBG
	Salix nigra	MH-S-001	MAQ
	Salix integra	MH-S-002	MAQ

AP = Alberta Pacific Forest Products Collection.

MRNQ = Le Ministere des Ressources Naturelles du Quebec.

MBG = Montreal Botanical Garden.

MAM = Morden Research Station and the Arboretum, Manitoba.

MAQ = Morgan Arboretum, Quebec.

Table 3: Sequences of primers used.

To amplify and sequence *trnT- trnF* three non-coding regions;(Figure 1-a)

a = 5'-cattacaaatgcgatgctct-3'

c = 5'-cgaaatcggtagacgctacg-3'

d = 5'-ggggatagagggactttgaac-3'

f = 5'-atttgaactggtgacacgag-3'

trnCR = 5'-tgttagaacaacttccattgagtctc-3'

trnDF = 5'-agtcccattctacatgtcaatatcg-3'

To amplify and sequence *ITS1* and *ITS2*; (Figure 1-b)

a = 5'-tcgtaacaaggtttccgtagg-3'

b = 5'-gctacgttcttcatcgatg-3'

ITS3 = 5'-gcatcgatgaagaacgcagc-3'

ITS 28kj = 5'-cttggacggaatttaccg-3'

Appendix 2 - Protocols

Protocol 1: DNA Extraction Procedure.

1) Obtain clean and preferably fresh leaf samples (1 to 2 leaves or 0.75 to 1.5 g). Place in a labelled plastic bag with 2 to 3 ml Extraction Buffer. Crush with rollers. Transfer extract through miracloth into a microcentrifuge tube, label, and place on ice.
2) Centrifuge the sample(s) at 6000 rpm (3000 G) for 5 minutes at room temperature. Decant the supernatant. Pellet(s) may be frozen at –80°C for future extraction, if necessary.
3) Re-suspend the pellet with 600 µl of Extraction /Nucleic Lysis (1:1) Buffer. Add RNase, if required. Then add 120µl of 5% Sarkosyl solution. Mix with a vortex mixer.
4) Heat to 65 °C for a minimum of 30 minutes. It is preferable to mix the tube(s) about halfway through the heating procedure with the vortex mixer.
5) In the fume-hood, add 600 µl of the Chloroform/Isoamyle Alcohol (24:1) mixture. Invert tube(s) to form an emulsion (gently) for about 10 minutes. Centrifuge at 7000 rpm (5000 G) for 5 minutes. Transfer upper phase to a clean microfuge tube containing 600 µl of Isopropanol. Gently invert tubes and allow the DNA to precipitate out of solution at room temperature for at least 30 minutes.
6) Centrifuge at 14000 rpm (20000 G) for 10 minutes. Decant off supernatant.

7) Wash pellet(s) with 70% ethanol. Centrifuge at14000 rpm (20000 G) for 10 minutes. Decant off supernatant and then dry in the vacuum.

8) Dissolve dried pellet(s) in 150 µl of TE Buffer.

Extraction Buffer: (500 µl)

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

*Add distilled water to volume.

Nucleic Lysis Buffer: (200 µl)

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

*Add distilled water to volume. It may be necessary to heat this solution up and then cool to aid in dissolving.

Protocol 2: PCR amplification reaction.

For 25 µl reaction:

Primer forward/reverse (5 pmol/µl)	1 µl for each
dNTP mixture (2.5 mM)	2.3 µl
$MgCl_2$ (25mM)	2.5 µl
Buffer	2.5 µl
dH_2O	14.7 µl
Taq DNA polymerase	1 µl

Buffer: (1ml)

1 M Tris (pH 9.5)	200 µl
4 M KCl	62.5 µl
Tween	5 µl
BSA	1 mg

*Add distilled water to a final volume of 1 ml.

Protocol 3: PCR thermal cycling program.

Lid preheated to 105 °C

1) Initial denaturation	94 °C for 2 minutes
2) Denaturation	94 °C for 1 minute
3) Annealing	55 °C for 30 seconds
(R = 3 °C /S)	
4) Elongation	72 °C for 1 minute
5) Go to 2	Repeat 34 times
6) Final elongation	72 °C for 4 minutes
7) Cooling and storage	4 °C

Protocol 4: 1% agarose gel recipe.

For 100 ml gel:

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

*Heat in microwave to dissolve. Then cool down to 55 °C, and add 3.3µl of ethidium bromide (10g/L).

Tris-borate EDTA (TBE) 5X(1 L)

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

*Add distilled water to a final volume of 1 L.

Protocol 5: DNA sequencing reaction recipe.

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

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

* Add distilled water to volume, if necessary.

Protocol 6: Sequencing reaction thermal cycling program.

Lid preheated to 105 °C

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

Protocol 7: Purification of products of sequencing reactions.

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

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

2) Mix the product and the cocktail with vortex mixer and leave it at room temperature for 15 minutes.
3) Centrifuge at 14000 rpm (20000 G) at room temperature for 20 minutes.
4) Aspirate the supernatant with pipette carefully.
5) Wash the pellet with 250 μl of 70% ethanol, and centrifuge at 14000 rpm (20000 G) for 10 minutes.
6) Aspirate the supernatant and repeat the wash.
7) Dry the pellet in vacuum. Pellet may be kept at $-20\,^{\circ}C$ to be sequenced later, if necessary.

Appendix 3 – Statistical Tests

Test 1: Test for homogeneity of base distribution across sequences.

Hypotheses:

H0 = Base distribution is homogenized across sequences.

HA =Base distribution is not homogenized across sequences.

Test statistic:

Chi-square test (X^2 test) of homogeneity of base frequencies across taxa to check if the observed frequencies differ significantly from the expected frequencies.

Test 2: Likelihood Ratio Test (LRT).

User defines hierarchical criteria in selecting a better substitution model for a given data set. Each and every possible combination of nucleotide substitution, among site rate variation, base frequencies, and transitions and transversions is tested, and log-likelihood score is used to evaluate the same tree under progressively more parameter rich (NESTED) models of the GTR family.

When the nested models are used, likelihoods for different models can be compared.

Hypotheses:

H0 (Null) = current model is true (observed fits with the expected model)

H1 (Alternative) = Current model does not fit but next simplest model may fit (observed variation does not fit with expected model)

LRT statistic (Δ):

$\Delta = \log L1 / L0$ (Where L1 = likelihood of the alternate hypothesis *H1*, and L0 = likelihood of null hypothesis *H0*)

In the LRT, the null model is a special case of the alternate model, and therefore 2Δ is approximately distributed as X^2 with degrees of freedom (d.f.) being the difference in the number of parameters in the two hypotheses.

$LRT = 2([-\ln L_{alternative}] - [-\ln L_{null}])$

Test 3: Winning Sites Test & Templeton Test (Parsimony "equivalents" to the K-H test).

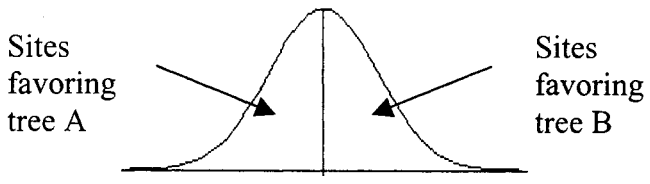
These methods (WS,TT) test the null hypothesis that the differences between two trees (A,B) are no more than expected by chance.

Winning Sites Test

- Sums the number of sites supporting tree A over tree B and vice-versa.
- Under the null hypothesis (*H0*), characters are equally likely to support tree A or tree B and a binomial distribution gives the probability of the observed difference occurring by chance.

Templeton Test

- This test is a non-parametric Wilcoxon signed ranks test of the differences in fit of characters to two trees.
- It is similar to the winning sites test but also takes into account the "magnitudes" of differences in the support of characters for the two trees.



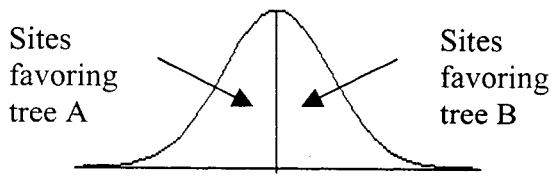
Test 4: Kishino & Hasegawa and Shimodaira-Hasegawa Tests.

- Given two trees, the difference in log-likelihoods is computed by taking each site in turn, and finding the difference in log-likelihoods for that site on the two trees.
- Similar to WS and TT tests, KH and SH tests also examine the null hypothesis that the differences between two trees are no more than expected by chance.
- If the sum of the log-likelihood differences is not significantly different from zero, the two trees are not significantly different.
- Under *H0,* the mean of the differences in steps/likelihood for each site is expected to be zero and the distribution normal. If the observed difference is significantly greater than zero, we can reject the null hypothesis and conclude that the two trees are significantly different.

Test statistic:

$$\Delta = \Sigma^{k}_{i=1} (\log L_{(k,tree1)} - (\log L_{(k,tree2)})$$

$$= \log L_{tree1} - \log L_{tree2}$$



Distribution steps/likelihood differences at each site

Appendix 4 - Methods of Phylogenetic Tree Reconstruction

Maximum parsimony:

In maximum parsimony analysis, the goal is to reconstruct phylogenetic trees invoking the fewest possible evolutionary changes at a given position of a molecular sequence. The total number of evolutionary changes in a tree (the tree length) is the sum of the number of changes at each site (Page& Holmes, 1998). The task of computer programs such as PAUP that implement the parsimony method is to calculate the length of the various candidate trees and to find the tree with the minimal length, which is the most parsimonious. Thus parsimony analysis is based on the implicit assumption that evolutionary change is rare (Maddison & Maddison, 2001). Therefore, the tree that minimizes change is likely to be the best estimate of the actual phylogeny. However, the main disadvantage of parsimony-based methods is that, under certain circumstances, the resulting tree is not consistent. In other words, even with the addition of further data, it is possible to obtain the wrong tree. The classic scenario where this might happen has been termed 'long branches attraction'. The problem of long branches attracting is most likely to occur when rates of evolution show considerable variation among sequences, or where the sequences being analyzed are quite divergent (Page & Holmes, 1998) with multiple reversals. Therefore maximum parsimony is inconsistent. Nevertheless, we can calculate certain tree statistics namely the consistency index (CI), retention index (RI), and the rescaled consistency index (RC) to evaluate the reliability of the maximum parsimony tree (Farris, 1989).

The CI for all characters on a tree is the minimum possible treelength divided by the observed treelength. If the characters in the data matrix are perfectly congruent with each other and with the tree (no homoplasy or reversals), the observed number of steps will be equal to the minimum, and the CI will be 1.00. The values of the CI range from 1 to 0, and decrease with increasing homoplasy, as reversals will contribute additional steps in the tree, increasing tree length. The RI for all characters in a tree is calculated as the (maximum possible treelength –actual treelength) / (maximum possible treelength – minimum possible treelength). Invariant characters are not included in the RI. If the characters in the data matrix are parsimony-informative and completely congruent with each other and with the tree, the RI will have a value of 1. If the data are maximally homoplastic on the tree, the RI will have a value of 0 (Maddison & Maddison, 2001). The RC for all characters on a tree is the CI multiplied by the RI. Invariant characters are not included in the RC. As with the RI, the RC ranges from 0-1, with higher RC values indicating that characters in the data set are more congruent with each other and with the tree (Maddison & Maddison, 2001).

Maximum likelihood:

Likelihood is the probability of observing the data, given a particular model. The likelihood of a tree is the probability that the data matrix observed would have been generated by a specific evolutionary process on that tree (Swofford et al., 1996). The tree that fits the observed data, as the most probable evolutionary outcome is the maximum likelihood estimate of the phylogeny. The likelihood is not the probability that the tree is the true tree, rather it is the probability that the given tree agrees with the empirical data.

The likelihood method relies on a model of the evolutionary process. Different models may make the observed data more or less probable. The structure of the model may be simple or complex and may specify the relative rates of change between the states of the characters as well as the variation in evolutionary rate from one character to the other (Maddison & Maddison, 2001). For instance, this model may include parameters for the transition/transversion ratio (TS/TV), base composition, and variation in rate among sites. The corresponding parameter values from an empirical data set could be estimated using the "Modeltest" software (Posada, 1998), which compares different nested models of DNA substitution in a hierarchical hypothesis-testing framework. Modeltest calculates the likelihood ratio test statistic in order to reject or accept different null hypotheses about the rates of DNA substitution.

Maximum likelihood is an appealing method for inferring phylogeny, as it can incorporate explicit models of sequence evolution and also permits statistical tests of evolutionary hypotheses. Furthermore, the maximum likelihood method used in phylogeny reconstruction is usually consistent, because the parameters being estimated are tree branch lengths and these do not change as more nucleotide sites are considered (Felsenstein, 1988). However, obtaining the maximum likelihood estimate for a given tree computation is an intensive process and limits the application of this method to relatively small data sets.

Distance method:

Distance-based phylogenetic trees are reconstructed based on a matrix of pair-wise distances between taxa. For nucleotide sequence data, distances are calculated from

the fraction of sites that differ between the two sequences. The phylogeny makes a prediction of the distance for each pair as the sum of branch lengths in the path from one species to another through the tree. A measure of goodness of fit of the observed distances to the expected distance is used to obtain the, preferred tree topology that minimizes the discrepancy between these two measures (Felsenstein, 1988).

Nei's (1987) "neighbor-joining" method, a commonly used distance based phylogeny inference method follows an algorithm which joins species and calculates branch lengths to find the minimum evolution tree. The latter corresponds to the tree with the shortest length, such that the length of a tree is the sum of all branch lengths), Nei's method, however, does not optimize a criterion of fit between tree and data (Felsenstein, 1988). The result obtained from this clustering algorithm often depends on the order in which the sequences are added to the growing tree. Therefore the neighbor joining strategy does not guarantee that the minimum evolution tree will be found. Nonetheless, in practice the neighbor-joining tree is often the same or very similar to the minimum evolution tree (Page & Holmes, 1998).