Phylogenetic Relationships of *Diploxylon* Pines and Genetic Variation of *Pinus tropicalis*

（マツ属複維管束類（*Diploxylon*）の系統関係と *Pinus tropicalis* Morelet の遺伝的変異）

Dissertation of Ph.D. degree

Submitted by

Geada López Gretel

United Graduate School of Agricultural Sciences, Ehime University, Japan
Department of Forest Resources
Laboratory of Forest Resource Biology

2003

…Genius is One percent inspiration and

Ninety percent perspiration …

Thomas Edison

CONTENTS

ABSTRACT

The molecular systematics and evolution of the *Diploxylon* pines (subgenus *Pinus*) was examined using plastids sequences. The plastid DNA sequences of *rbc*L, *mat*K, the *trn*V intron, the *rpl*20-*rps*18 spacer, and *trn*L-*trn*F spacer for 47 species of subgenus *Pinus*, representing all recognized subsections, were inspected. A total of 4358-bp were used to assess the relationships using MP, NJ and ML algorithms. The topologies showed that species in subgenus *Pinus* splits into two distinct lineages, corresponding to Eurasia and North America. The Eurasian lineage comprised the section *Pinus* and was further differentiated into two clades; the Mediterranean pines (subsections *Canarienses*, *Halepenses*, *Pinea* and *Pinaster*), and subsection *Pinus*. Two North American pines, *P. tropicalis* (Cuban pine) and *P. resinosa* (American red pine), are the typical members of subsection *Pinus*, but did not cluster together, suggesting that these species might have migrated to America independently. The Pyrenean taxon, *P. uncinata* was a close relative to the European *P. mugo* and *P. uliginosa*, which confirms its position among the members of subsection *Pinus*. The second lineage comprised the section "New World hard pines". Subsection *Contortae* occupied the basal position followed by the monophyletic subsection *Ponderosae*. A strongly supported clade of the southern U.S. members of *Australes* was separated from the other complex clade of the remaining subsections *Attenuatae, Oocarpae, Leiophyllae* and *Australes* (Florida/Caribbean species), making subsection *Australes* paraphyletic. The intersectional relationships within the latter clade were poorly resolved but subsection *Attenuatae* was defined monophyletic clade. The endemic Cuban pines, *P. cubensis* and *P. maestrensis*, and the two varieties of *P. caribaea* var. *bahamensis* and *hondurensis* were clustered as sister groups within the later clade. The divergence times for each subsection were estimated from the *rbc*L sequence data. Based on the age of the *Diploxylon-like* fossil and the sequence divergence for *rbc*L between the Eurasian and North American sister lineages (0.0095 ± 0.0021), the substitution rate was estimated to be $3.65 \pm 0.81 \times 10^{-11}$ per site per year. Application of this rate to all pairs of the clades indicated an approximate lineage-divergence date of 104 MY for subsection *Contortae* ((with 95% confidence limits of 61 -169 MY), 52 MY for *Ponderosae* (29 - 87 MY), 37 MY for *Australes* southern U.S. (22 - 69 MY), 21 MY for *Attenuatae* (7 - 43 MY), and 20 MY for the *Oocarpae-Leiophyllae-Australes* (Florida/ Caribbean species) clade (11 - 38 MY). Within the Eurasian clade, the split between subsection *Pinus* and subsections *Pinaster-Canarienses-Pineae-Halepenses* was estimated to be 64 MY (27 -119 MY).

The intraspecific genetic variation of *Pinus tropicalis* Morelet was studied using the *cp*DNA sequences, *cp*SSR (microsatellite) and RAPD data sets. A total of 106 individuals were examined from seven populations. By the *cp*DNA sequences of *trn*L intron and *trn*T-*trn*L spacer (990-bp), eight haplotypes were recognized. Average nucleotide diversity among haplotypes was very low (0.0003). AMOVA analysis revealed a genetic structure in the populations of *P.*

tropicalis; high level of genetic differentiation among populations (Φ_{ST} = 0.607), high level of genetic differentiation among groups (Φ_{CT} = 0.4997) and low level of differentiation within the groups (Φ_{SC} = 0.057) when the populations was divided into three groups, i.e., Eastern (Galalón), Central (Viñales and Pilotos), and North-western (La Jagua, San Juan, Bartolo, Mina Dora). These results suggest that gene flow (via pollen) is limited among geographically distant populations. The minimum-spanning network showed that haplotypes VI, VII and VIII from the population Pilotos were grouped and linked two groups of haplotypes: one with the haplotypes I and II, solely in the Eastern population and other with the haplotypes IV, V and III, from the North-western populations. The results of the *cp*SSR were consistent with those of the *cp*DNA sequences, while the level of the gene differentiation was lower in the *cp*SSR (θ = 0.18). No differentiation was detected among populations La Jagua, Bartolo and Mina Dora, and these populations were moderately or little differentiated from the Pilotos population. The nuclear marker (RAPD) revealed similar level of gene diversity for each populations (H_T = 0.230 and H_S = 0.166) and a great differentiation among populations (G_{ST} = 0.209). This clearly indicates that those populations have not gone into a serious bottleneck in the recent past but the current pattern of genetic variation is the mere reflect of the history of the species. All the analyses hypothesize that Pilotos could have served as Miocene refugia given that all their haplotypes were located in the interior node of the minimum-spanning network, it posses the most frequent and 'consensus *cp*SSR' repeats, and the larger value of gene diversity. In addition, analyses illustrate Galalón population was distantly related to the Central and North-western populations being a unique population; therefore, this population should be considered as an Evolutionary Significant Unit (ESU) and a Management Units (MU) as well.

Acknowledgements

This work was carried out at the Department of Forest Resource Biology, Ehime University. I warmly thank my supervisor Professor Ko Harada for introducing me the interesting field of plant genetics and evolution. I am also grateful for his guidance and support during my studies and for his patience with me. I thank the Department for providing excellent working facilities.

I am indebted to all the institutions that assisted me in my Ph. D. program: Pinar del Río University, Ehime University, Forest Product Research Institute, Japan, Kamigamo Experimental Station of Kyoto, High Tatras Park staff in Slovakia, Madrid University, Spain for providing me the sample and valuable suggestions. I would like to thanks the former Head of the Department of Biology, Pinar del Río University, Dr. Esteban García Quiñones and Dr. Mayte Alvarez Huerta had been a great source of encouragements and persuasion for this Program. To all my Cuban colleagues for the support they have given me during my Ph. D. course. My thanks go to Professor Ben LePage, Pennsylvania University for his comments and revision in the *Pinaceae* phylogeny. I thank Dr. Aaron Liston, Oregon University for his valuable recommendation of subgenus *Pinus* phylogeny.

Especially, I thank Dr. Koichi Kamiya who never believed me but taught patiently how to proceed in the phylogenetic analyses, and for all the inspiring and very often but always pleasantly confusing discussions we had. Special thanks also go to my Ms. Le Giang who lifts me up and cheers me whenever the sun did not want to shine for me; for giving a chance to share many ideas and dreams as well. I thank Ngyen Ha, Kashimura Seiichi and Yoneda Reiji for discussions concerning everything beneath the heaven. Our coffee-tea-something else breaks (CTS-breaks) were relaxing moments during these years. It was very therapeutic to notice that I was not the only “crazy-loco” in the group.

I would like to extend my thanks to everybody in the Department and collateral Departments for making my working days so pleasant and for tolerating my “**NOISE**”. **Sorry people, it is in my genes**.

Finalmente, dedico mi tesis a lo mas valioso de mi vida… Mi familia, a Iosvany Sánchez, a mi madre y mi hermana pues: ...“La ciencia puede descubrir lo que es cierto, pero no lo que es bueno, justo y humano”.

1. General Introduction

1.1. Molecular Systematics

For centuries, naturalists have tried to detect, describe and explain diversity in the biological world; this endeavor is known as systematics. The formalization of a hierarchical system of nomenclature by Linnaeus (1753) established the framework for describing and categorizing biological diversity. This hierarchical system was initially independent of the evolutionary theory, and that is why the early evolutionists opposed the Linnaean system and the Aristotelian essentialism it incorporated. However, the Linnaean system prevailed, and later on evolutionists (e.g., Lamarck 1809, Darwin 1853) simply adopted the system to produce a classification based on the Phylogenetic relationships. Primary attempts to reconstruct phylogenetic history were based on few (or any) criteria, and estimates of phylogeny were little and partially credible assertions on certain taxonomical groups. During the first 50 years of the XX century, systematics were concerned more with problems of species, speciation and geographic variation than with problems of phylogeny, and this words does not appear until 1942 in the book *Evolution: the Modern Synthesis* by Huxley. Etymologically, the word "phylogeny" comes from the Latin *Phylo- phylum*: race, tribe, kind; and *genesis*: resulting form or relating to gene, origin. On the other hand, Systematics signifies arrangement of something according to a system or rules in hierarchical manner, and the term is mainly used by the taxonomists. From this time forth, botanist and zoologist began to define objective methods for reconstructing evolutionary history based on share attributes of extant species and fossil and among groups of organisms, but is not till 1960s when these methods, with the new techniques of the Biotechnology (i.e., isoezymes and DNA sequencing) established the indestructible vinculum between Systematics and Phylogeny.

There is an important synergism between studies of molecular systematics and molecular evolution, which will has a valuable indirect effect on the population genetics studies as well. Molecular systematics applies genetic markers to make inferences about population processes and phylogeny, and at the same time creates a source of comparative database for specific

genes or proteins that will be used for evaluate rates, processes, and constrains on molecular change through time (Kimura 1983, Nei 1987, Li 1997), which can be translated into more informed use of molecular markers in population genetics and phylogenetic analyses. However, controversies can be encountered and generate more doubts than solutions. Those controversies include opinions about the relative value of molecular *versus* morphological data, the kind of data that should be collected, the algorisms to analyzing the data, the constancy in the rate of substitution and of course, the neutrality of some molecular markers. The final words have not been pronounced yet and therefore we must direct with precaution any inference until new powerful methods does not become available.

The field of Molecular Systematics covers both intraspecific variations, traditionally a subject of population genetics and interspecific diversity, traditionally the field of phylogenetics. This linkage is essential to the integration of molecular evolution and systematics and has been enhanced by the use of the allelic genealogies at both levels that predicts the effects of genetic drift, mutation, migration and selection (i.e., the evolutionary forces). For example, if the rate of nucleotide substitution is enough for the allelic genealogy to be estimated, then we can infer about historical population size, gene flow, and selective processes (Slatkin and Maddison 1989, Slatkin 1991, Hartl 2000).

There exists a broad spectrum of methods for analyzing variation at population level; the old techniques such as alloenzymes electrophoresis and cytogenetics and the advance one such as DNA sequencing or techniques of fragment analysis (e.g., RFLP, RAPD, AFLP, microsatellite, etc), which need to be assayed according to the defined problem. Alongside the advance in molecular techniques, have been improvements in the analysis of molecular variation within and among species. Within species, the ability to obtain gene tree based on the coalescence theory (Hudson 1990) and the analysis of the phytogeography (Avise 1994, Templeton et al. 1998) has encouraged the sophistication in the analysis of Population Genetics and Conservation Genetics. Among the species, the ability to obtain a more accurate phylogenetic tree and investigate the patterns of molecular evolution and rate/process of substitutions, arising as new challenging branch named Molecular Evolution (Moritz and Hillis 1997).

1.2. Population Genetics and Molecular Systematics

A central thesis of the G. L Stebbin in the book *Variation and Evolution in Plants* can resume the framework of the evolutionary biology and the stretched vinculum with population genetics.

> Individual variation, in the form of mutation and gene recombination, exists in all population; … the molding of this raw material…into variation on the level of populations by means of natural selection, fluctuation in population size, random fixation and isolation is sufficient to account for all of the differences, both adaptive and non-adaptive, which exist between related races and species……
>
> The problem of the evolutionist is … evaluating on the basis of all available evidence the role which each of these know forces has played in any particular evolutionary line…

This highlights the notion that to understand evolution we must examine its action at the level of populations within species. How the genetic variation is generated, organized and maintained in natural population? How many of this variation is apportioned within and among populations? In other words, the Population Genetics concerns. Altogether knowing the extent of genetic variation within and between population or species and understanding what kind of variation is important for forming new species; we are elucidating the evolutionary process of speciation and history of a group of organisms. Therefore, the Population Genetics occupies an essential place in the Biology being core point of many disciplines such as Molecular Biology, Systematics, Genetics and Conservation Biology. The Conservation Biology is an emergent and rapidly developing discipline that integrates aspects of biogeography, demography, ecology, economics, evolution, genetics and systematics to assist in solving critical problems in preserving biotic diversity (Smith and Wayner 1996). Conservation genetics aims to derive strategies for the long-term maintenance of the genetic variability of species, since genetic variability is associated to both the fitness of each individual and the long-term adaptability and evolutionary potential of the populations. Whereas conservation practices in the past has been

largely irresponsible, the science of conservation biology has at its core a practical, theoretical and empirical thrust, thereby promoting the strategic elements of conservation planning.

Whereas mutation is the crucial source of all genetic variation, many other factors determine the distribution and maintenance of this variation. These factors include sexual reproduction, natural selection, migration, and genetic drift in the populations. Population bottlenecks may contribute to inbreeding, with consequent implications for fitness (inbreeding depression). The counterpart to this is outbreeding depression which can occur as a consequence of moving individuals into the ranges of other populations of their species and also by mixing populations from different sources. Both have significant implications for conservation management, especially in the use of concepts such as effective population size and population viability.

1.3. The Genus *Pinus*

The first gymnosperms arose in the Middle Devonian (~365 MY). Fossilized cones have shown that ancestors of Pinaceae family evolved by the Middle-Jurassic (~160 MY). Pinaceae is divided into 11 genera, and more than half the species in the Pinaceae are included in the genus *Pinus* (111 species) (Farjon 1996). The genus *Pinus* has a largest distribution range in boreal, coastal and montane forest of the Northern Hemisphere, and where they occur, pines form dominant component of the ecosystem. In addition, the genus has tremendous ecological and economic importance throughout. Overexploitation and other human pressure are threatening the survival of many of natural population of pines, although pines are also growing in commercial plantations in both within and outside the natural ranges.

During the early part of the Cretaceous (nearly 130 MY) pines diversified into two subgenera, *Strobus* (*Haploxylon* or soft pines) and *Pinus* (*Diploxylon* or hard pines) (Mirov 1967, Little and Critchfield 1969, Price et al. 1998). Several sections and further subsections have evolved since the diversification of these two subgenera (Krupkin et al. 1996, Price et al. 1998). After the diversification of the subgenera *Pinus* and *Strobus* pines migrated throughout the middle latitudes of the Northern Hemisphere super-continent, Laurasia. Major environmental changes in the early Cretaceous led to a splitting of several subsections of *Pinus* into northern

refugial populations in western Siberia, mid-latitude populations in eastern Asia, and southern refugial populations in other parts of Asia and Europe (Kremenetski et al. 1998, Willis et al. 1998). Intensive mountain building events together with climate changes created the environmental heterogeneity that drove the radiation of pine taxa in several areas which became secondary centres of diversification of *Pinus* (e.g. Mexico and north-eastern Asia). At the end of the Eocene (55 - 37 MY) the genus *Pinus* diversified further due to climatic changes (Richardson and Rundel 1998). The impact of the Eocene had the effect of dissecting the genus and concentrating pines into widely disjunct regions. During the Pleistocene (1.7 - 0.01 MY) pine populations and species shifted first south, then north following the glacial and interglacial periods. The climatic fluctuation at the Pleistocene may have played an important role in speciation or at least in the preservation of distinctive genotypes (Richardson and Rundel 1998). The last 10 000 years after the last glacial period have shaped the current distribution of pines. Nowadays, the natural distribution of pines range from arctic and subarctic regions of North America and Eurasia south to subtropical and tropical regions of Central America and Asia, and one species extending even south of the equator (*P. merkusii*) (Mirov 1967, Price et al. 1998).

Pinus is a classical Latin name for pines, and was applied by Linnaeus (1753) in his *Species Plantarum* to groups of species (*P. cembra*, *P. sylvestris*, *P. pinea*, *P. taeda* and *P. strobus*). Incessantly, since the Linnaean primitive classification, the systematists have tried to identify and classify the genus. The first classification of Du Monceau (1755; see Little and Critchfield 1969) subdivided *Pinus* into three section based on the number of needle per fascicle: *Bifoliis*, *Trifoliis* and *Quinquefoliis* corresponding to the present subgenus *Pinus* (the first and second) and subgenus *Strobus* (for the latter), and the North American taxa were treated as *Trifoliis* section. Further, the subdivision of the two subgenera is considerably more controversial, and a large number of classifications have been proposed, dealing either with the entire genus. The complex nomenclature history of the genus reflects both changes over time in the conventions, rules of assigning names and ranks to taxa, and differences in philosophy of classifications; therefore the delineation of sections and subsections have been puzzled. Early classification of the genus are artificial and do not reflect phylogenetic relationships or evolution of characters because they are based on a limited number of characters or sampling scope.

Others, though intentionally phylogenetic seem to classify the species in groups defined by implausible homoplasies (e.g., serotinous cone, deciduous fascicle sheath scale, development of the seed cone and grass stage of the seedling, etc) (Shaw 1914, Pilger 1926, Little and Critchfield 1969, Van der Burgh 1973, Klaus 1989). Whereas the latter classifications have tried to integrate many type of data from reproductive morphology, anatomy, secondary products, crossability, etc.

At subsectional level, several morphological schema have been propounded; however, no consensus have be reached due to the limited scope in the sampling of taxa and dissimilarity in the morphological characters used for comparisons, and trouble seems to be more profound in the circumscription of the North American taxa. Although a considerable amount of information on the complex group of Mesoamerican pines have been presented by Martínez (1948), Little and Critchfield (1969), Perry (1991), Farjon (1984) and Farjon and Styles (1997), the acknowledgment and description of many taxa are inconsistent. The pine flora of this region is visibly in need of further detailed study using both morphological and molecular approach.

1.4. Goals of this work

The present study has the following aims: first, to assess the evolutionary relationships of *Diploxylon* pines (Chapter 2: section 2.1. “Molecular Phylogeny of subgenus *Pinus*”); considering the biogeography of this subgenus, divergence times of the major subsections inferred from the nucleotide sequences data and the route of migrations of *Pinus* with special emphasis on the North American continent (Chapter 2: section 2.2. “Divergence Time and Biogeography of *Pinus*”). I also present a new outline of the *Diploxylon* pines based on morphological and molecular data (Chapter 2: section 2.3. “Outline of the Classification of subgenus *Pinus*”). Second, to provide the morphological and molecular evidence for the recognition of the four Cuban taxa as well as their relationships (Chapter 3: “Relationships and Status of the Cuban pines”). Third, to characterize the geographical and genetic structure in the natural populations of the tropical pine using the *cp*DNA sequence and the intra-specific variation among and within populations through the *cp*SSR variation and RAPD for formulating the conservation strategies (Chapter 4: “Population Genetics of *Pinus tropicalis* Morelet”).

2.1. Molecular Phylogeny of subgenus *Pinus*

2.1.1. Introduction

In the present study, I focused on subgenus *Pinus*, which has 71 recognized species, 46 of which are found mainly in North America (Farjon and Styles 1997). Although various morphological and anatomical classification schemes have been proposed for the sections and subsections of subgenus *Pinus*, the relationships among the subsections and their evolutionary processes are still being debated (Mirov 1967, Little and Critchfield 1969, Farjon 1984, Rushforth 1987, Klaus 1989, Perry 1991, Farjon and Styles 1997, Price et al. 1998). For instance, there are many differences in reported delineations of sections and subsections (see review in Price et al. 1998). Despite the large number of morphological characters, the high levels of homoplasy in many morphological characters and their plesiomorphic nature have contributed to the differences among classification schemes. Thus, a comprehensive phylogenetic analysis of subgenus *Pinus* is hampered by the paucity of discrete characters, which are also scarce in the genus *Pinus* compared to those in other plant groups (Farjon and Styles 1997).

Several molecular studies have been conducted on representative species of both subgenera of conifers. These studies have revealed a large genetic distance between the two subgenera as well as a lower level of genetic variation in subgenus *Pinus* (Strauss and Doerksen 1990, Govindaraju et al. 1992, Moran et al. 1992, Wang and Szmidt 1993, Krupktin et al. 1996, Wang et al. 1999, Wang et al. 2000). However, these studies have been limited in terms of taxonomic sampling and/or geographic scope, particularly in subgenus *Pinus*. The study by Krupkin et al. (1996) based on chloroplast DNA (*cp*DNA) restriction analysis of 18 *Diploxylon* pines showed that the distinctive division within the subgenus was between the North American species (except *P. resinosa*) and the Eurasian species. The molecular analysis of ITS sequences by Liston et al. (1999a) involved a broad sampling of *Pinus* subsections and covered a wide range of geographic regions. The ITS data strongly supported the notion of the existence of a distinctive group of North American pines. However, the Eurasian species, i.e.,

section *Pinus* (composed of subsections *Pinus*, *Pinaster* and *Pinea*) was a paraphyletic clade, in which a monophyletic subsection *Pinus* was moderately supported. Even though, the topologies of the recovered phylogenetic trees gave weak support for many of the clades. The striking discrepancies between the two studies were due to the incongruence in affiliations of North American subsections, particularly in the basal position of subsection *Contortae*. Subsection *Contortae* emerged as a strongly supported sister group to all of the North American species (Krupkin et al. 1996); whereas in the analysis of Liston et al. (1999a) the "New World hard pines" were divided into two well supported subgroups, subsection *Ponderosae* and a clade of the remaining subsections. Liston et al. (1999a) also found the Himalayan *P. roxburghii* had a sister relationship to the American subsections and that it was paraphyletic to the Asian and Mediterranean hard pines. On the other hand, Wang et al. (1999) assessed the relationships of Eurasian pines using *cp*DNA sequences of *rbc*L, *mat*K, *trn*V intron and the *rpl*20-*rps*18 spacer, and their results showed that the Mediterranean *Diploxylon* pines formed one clade and that the Asian members of subsection *Pinus* formed another. The Himalayan *P. roxburghii* was found to be a strongly divergent taxon from all the remaining Eurasian hard pines, suggesting its association with North American pines. However, the sister relationship between this large North American clade and the Himalayan pine still requires confirmation by *cp*DNA sequence analysis (Liston et al. 1999b, Wang et al. 1999). Moreover, there have been difficulties in determining the phylogenetic positions of rare endemic taxa such as *P. tropicalis*, *P. caribaea*, *P. cubensis*, and *P. maestrensis*, which are seldom included in phylogenetic studies. It is expected that there will be further changes in the numbers and delimitation of species, sectional groups and subsectional groups as more molecular phylogenetic data become available (Farjon and Style 1997, Price et al. 1998). Thus, more North American species should be incorporated into the *cp*DNA data sets in new molecular phylogenetic studies, and the results of these molecular studies should be compared and combined with morphology-based phylogenetic analyses (Liston et al. 1999b).

In this study, I analyzed sequences from *rbc*L, *mat*K, the *trn*V intron, the *rpl*20-*rps*18 spacer and the *trn*L- *trn*F spacer for all of the Eurasian pines and most of the North American pines. Our main objectives were: 1) to examine the phylogenetic relationships of subgenus *Pinus* at the inter and intra-sectional levels, focusing on the controversial relationships among

the North American subsections; 2) to re-examine the classification of several uncertain species such as *P. roxburghii*, *P. uncinata*, *P. uliginosa* and Cuban pines; 3) to compare the rates of sequence divergence of coding and non-coding regions in *Pinus*.

2.1.2. Materials and Methods

2.1.2.1. *Sample Species*

Taxonomic sampling of subgenus *Pinus* included 47 species representing 10 subsections according to the classification of Price et al. (1998). I examined 31 North American pines, 10 European pines, five Asian pines and the Himalayan pine, *P. roxburghii*. In order to complete the phylogenetic relationships of Eurasian pines I added the previous sequences by Wang et al. (1999). The sources of samples are listed in Table 1. *Pinus parviflora* (subgenus: *Strobus*), were chosen as an outgroup given that this species has been resolved as the sister group of the *Diploxylon* pines (Wang et al. 1999).

2.1.2.2. *DNA Extraction, Amplification and Sequencing*

Total DNAs were extracted from 200 mg of dried needles using a modified CTAB method (Doyle and Doyle 1990), treated with RNase, and purified by phenol. Five regions in *cp*DNA (*rbc*L, *mat*K, *trn*V intron, *rpl*20-*rps*18 region and *trn*L-*trn*F intergenic spacer region) were amplified by PCR. The double-stranded DNAs of *rbc*L, *trn*V intron, and the *rpl*20-*rps*18 region were amplified using the primers designed by Wang et al. (1999), and Taberlet et al. (1991) for the *trn*L-*trn*F spacer, respectively. The amplification products were directly sequenced after purification using a GENECLEAN KIT III (BIO 101). The *mat*K sequencing was performed using the following internal primers newly designed by us except the first four primers: *mat*K-F1, 5'-GAA CTC GTC GGA TGG AGT G-3'; *mat*K-R1, 5'-GAG AAA TCT TTT TCA TTA CTA CAG TG-3'; *mat*K-F2, 5'-CGT ACT TTT ATG TTT ACA GGC TAA-3'; *mat*K-R2, 5'-TAA ACG ATC CTC TCA TTC ACG A-3' (Wang et al. 1999); *mat*K-F3, 5'-GAG TT CGC TGG AAA GAT CCA TG-3'; *mat*K-R3, 5'-CCT TCT TAG AGG ATA AAT TCC CGC AT-3'; *mat*K-F4, 5'-CTC GGA TCC AGC GAC GAA AGG TTC-3'; *mat*K-R4, 5'-ATT GCG ATC CAA ATC CAT TAG CT GA-3'; *mat*K-F5, 5'-

CGG AAC TTC CAG GAC CAA TGT AAG-3'; *mat*K-R5, 5'-CCG ACT AGA TCG CAC CAT GTA TTT-3'. PCR amplifications were accomplished at 94^{o}C for 3 min for the initial denaturation followed by 30 cycles of denaturation at 94^{o}C for 45 sec, annealing at 50^{o}C (*mat*K) or 55^{o}C (*rbc*L, *trn*V and *rpl*20-*rps*18) for 1 min, extension at 72^{o}C for 2 min, and a final extension for 5 min at 72^{o}C. The PCR products were precipitated with ethanol and used as a template for the sequence reaction. The sequencing was carried out using an ABI 310 Genetic Analyzer (Applied Biosystems, Inc) with an ABI BigDye Terminator Cycle Reaction Kit following the manufacturer's instructions.

2.1.2.3 *Sequence Alignment and Data Analysis*

Alignments were performed using Clustal X (Thompson et al. 1997) and further edited by eye. The phylogenetic analyses were completed using single and combined data sets. The phylogenetic analysis were performed with PAUP* 4.0.b8 (Swofford 1999) using maximum parsimony (MP), neighbour-joining (NJ; Saitou and Nei 1987) and maximum likelihood (ML) algorithms. Indels were treated as missing data in each analysis. Heuristic searches with random sequence addition with 100 replicates; tree-bisection-reconnection (TBR) and ACCTRAN branch length optimisation were used for MP analysis. To evaluate the relative robustness of the clades found in the parsimony analysis, 250 bootstrap (BS) replicates (Felsenstein 1985) were calculated. The NJ tree was constructed based on Kimura's two-parameter model (Kimura 1980) with 1000 bootstrap replicates. The HKY85 model with ASIS addition sequence was used for the ML heuristic searches. The transition:transversion ratio was estimated from the sequence data. In addition, the phylogenetic relationships among all the Eurasian pines were assessed utilizing the MP algorithm, combining our sequences and the sequences deposited by Wang et al. (1999) of *rbc*L, *mat*K, *trn*V intron and *rpl*20-*rps*18 region.

The average numbers of nucleotide substitutions with their standard deviations were calculated for each region with MEGA 2.1 (Kumar et al. 2001) based on the Jukes-Cantor model (Jukes and Cantor 1969). In addition, the numbers of synonymous and nonsynonymous substitutions in *rbc*L and *mat*K genes were estimated using the method of Nei and Gojobori (1986).

2.1.3. Results

2.1.3.1 *Sequence Characterization*

The detected sequence of the *rbc*L gene included 1256 bp for all OTUs, representing 88% of the coding region. The 1256 bp sequence corresponds to positions 43152 to 44408 in the *P. thunbergii* chloroplast genome (Wakasugi et al. 1994). The average nucleotide compositions were 0.29 (T), 0.26 (A), 0.25 (G), and 0.20 (C). The transition:transversion ratio was 1.53. The primers used for *mat*K allowed us to sequence 1667 bp, including 116 bp of the *trn*K intron and 1554 bp (100%) of *mat*K coding region, corresponding to positions 1616 to 3262 in *P. thunbergii*. A total of three indels of equal length (six-bp) were found in this region. The average base compositions of the *mat*K region were 0.33 (T), 0.31 (A), 0.18 (G), and 0.18 (C). The transition:transversion ratio was 1.68. A total of 492 bp of the *trn*V intron from positions 47492 to 47984 in *P. thunbergii*, representing 91% of the total region was examined. There were no indels in the subgenus *Pinus*, and the complete aligned sequence was 494 bp when outgroups were included. The average nucleotide compositions were 0.30 (T), 0.32 (A), 0.21 (G), and 0.17 (C). The transition:transversion ratio was 1.61. The *rpl*20-*rps*18 region contains 165 bp of the 3'- sequence of the *rpl*20 gene, 256 bp of the spacer and 89 bp of the 5'- sequence of *rps*18, corresponding to positions 31403 to 31913 in *P. thunbergii*. No indels were detected in the subgenus *Pinus* in either coding region, while there were several indels in the spacer. The complete aligned sequence was 521 bp, and the average nucleotide compositions were 0.33 (T), 0.34 (A), 0.16 (G), and 0.17 (C). The transition:transversion ratio was 1.72. The length of the *trn*L-*trn*F spacer varied from 401 to 392 bp, from the position 67797 to 68177 in *P. thunbergii*. Several indels, particularly poly-T variations, which were excluded from analysis, were detected. One informative indel of five-bp in all the members of subsection *Pinus*, two indels in *P. pinaster* and *P. halepensis,* and one indel in 33 bp in *P. radiata* were found. The complete aligned sequence, which contains nine parsimonious informative sites, was 425-bp in length. The average nucleotide compositions were 0.37 (T), 0.32 (A), 0.14 (G) and 0.17 (C). The transition:transversion ratio was 1.61

2.1.3.2. *Sequence Divergence*

In *rbc*L, 28 of 40 polymorphic sites were phylogenetically informative. In the 3'-flanking region of *mat*K within the *trn*K intron, three polymorphic sites were observed, and two of them were informative. The *mat*K gene was the most variable region (5%); the numbers of variable and informative sites were 74 and 42, respectively. In the *trn*V intron, the numbers of polymorphic and informative sites were 12 and seven, respectively, but no sequence variation was found within the members of *Oocarpae, Attenuatae, Australes* and *Leiophyllae*. The coding region of *rpl*20 had four informative sites out of six polymorphic sites; all of them were nonsynonymous substitutions and three of them were in the Eurasian clade (section *Pinus*). In the *rpl*20-*rps*18 spacer, six of 13 variable sites were informative, whereas no variable sites were detected in the *rps*18 coding region. In the *trn*L-*trn*F spacer, the numbers of polymorphic and informative sites were 31 and 12, respectively. Commonly, most of the substitutions separated the two major lineages in subgenus *Pinus*, namely, the Eurasian and North American clades.

The average number of nucleotide substitutions per site was calculated for each region (Table 2). The average number of overall nucleotide substitutions (K_o) in *mat*K (0.00944 ± 0.00120) was 1.5-times higher than that in *rbc*L (0.00651 ± 0.00130), e.g., the average number of nucleotide substitutions at nonsynonymous sites (K_a) in *mat*K (0.01260 ± 0.00281) was 3-times higher than that in *rbc*L (0.00442 ± 0.00126), whereas the average number of nucleotide substitutions at synonymous sites (K_s) in *mat*K (0.00970 ± 0.00299) was close to that seen in *rbc*L (0.01287 ± 0.00376). The average number of nucleotide substitutions in both non-coding regions was 2.5-times lower than the average number of nucleotide substitutions at synonymous sites in *rbc*L and *mat*K. Interestingly, comparison of the average numbers of nucleotide substitutions in the four regions of the two lineages revealed similarly low divergence within each lineage, except for *mat*K. The sequence divergence of *mat*K in the Eurasian clade was 2.5-times higher than that in the North American clade, and the sequence divergence at nonsynonymous sites was 3-times higher in the Eurasian clade than that in the North American clade (Table 2).

2.1.3.3. *Phylogenetic Relationships of Subgenus Pinus*

The phylogenetic reconstructions based on *trn*V intron, *rpl*20-*rps*18 regions, *trn*L-*trn*F spacer alone resolved the sectional divergence, but the subsectional relationships were poorly resolved due to an insufficient number of informative sites. Those based on *rbc*L and *mat*K data revealed the relationships within subsections (data not shown), but weakly supported by bootstrap analysis. I combined the data (4357 bp) in order to increase the bootstrap support of internal branches and thus to obtain a more refined phylogeny.

MP analysis yielded four most-parsimonious trees of 367 steps (CI = 0.807; RI = 0.916). The topologies of the MP trees (Fig. 1) were essentially identical to those of the NJ (Fig. 2) and ML trees (Fig. 3). They only differed in the relative positions of *P. resinosa, P. nigra*, *P. tropicalis* and *P. massoniana* within the clade of subsection *Pinus*. All of the trees showed that species in subgenus *Pinus* were split into two distinct lineages, corresponding to Eurasia (BS = 93%) and North America (BS = 95%) with 18 and 29 species, respectively. The Eurasian lineage was differentiated further into two clades; one included all of the members of subsection *Pinus* except for *P. pinaster* and *P. heldreichii*, strongly supported by bootstrap (BS > 85%). This clade included two North American pines, *P. resinosa* and *P. tropicalis. Pinus resinosa* (American red pine) was the sister to *P. nigra* (European black pine) (BS = 100%) but was separated from *P. tropicalis* (Cuban pine). *Pinus thunbergii, P. densata* and *P. luchuensis* formed a moderately resolved monophyletic group (70 < BS < 85%). *Pinus mugo*, *P. uliginosa*, *P. uncinata* emerged as a well-resolved monophyletic group (BS > 85%). *P. densiflora*, *P. sylvestris* and its variety var. *siribica* formed a moderately supported monophyletic group (50% < BS < 85%). The second clade of the Eurasian lineage comprised the Mediterranean pines and the Himalayan pine, *P. roxburghii*. This clade was only moderately supported in the MP tree (BS = 70%) but was strongly in the NJ tree (BS = 95%). *Pinus halepensis* appeared as a sister taxon to the rest of the Mediterranean species, and *P. pinea, P. pinaster, P. roxburghii* and *P. canariensis* were monophyletic but received low bootstrap support (BS < 50%).

In the North American clade, the strongly supported (BS > 85%) monophyletic clade for the species *P. contorta*, *P. banksiana* and *P. virginiana* (i.e., subsection *Contortae*) emerged as a sister group of all the North American pines. A well-supported clade for *P. ponderosa*, *P.*

douglasiana, *P. jeffreyi*, *P. engelmannii* and *P. coulterii* (i.e., subsection *Ponderosae*) appeared to be the second (BS = 84%), and the remaining species split into two major clades. One of the clades included a strongly supported clade for *P. elliottii*, *P. pungens*, *P. serotina*, *P. rigida* and *P. taeda* (i.e., *Australes* southern U.S.) (BS > 85%), and the other contained a large-complex clade for the remaining species of Californian closed-cone pines, Mesoamerica, Florida-Caribbean (i.e., subsections *Attenuatae-Oocarpae-Australes-Leiophyllae*). Although the relationships among species within the last clade were not fully resolved, a strongly supported monophyletic group (BS > 94%) for *P. attenuata*, *P. radiata* and *P. muricata* (i.e., subsection *Attenuatae*) was clearly visualized. Moreover, moderately supported groups (50% < BS < 85%) were perceived for the eastern Cuban species (*P. cubensis* and *P. maestrensis*) among the complex of *Oocarpae-Leiophyllae-Australes* (i.e., Florida/Caribbean species). Consequently, subsection *Australes* became paraphyletic.

2.1.3.4. *Phylogenetic Relationships of Eurasian Pinus*

The phylogenetic reconstruction was based on the combined data set of *rbc*L (1256-bp), *mat*K (935-bp), *trn*V intron (494-bp) and *rpl*20-*rps*18 region (521-bp) for 25 Eurasian pines yielded 12 most-parsimonious trees of 190 steps (CI = 0.874; RI = 0.869). Overall, the Eurasian pines (section *Pinus*) includes two well-supported clades (Fig. 4), one included all of the members of subsection *Pinus* except for *P. pinaster* and *P. heldreichii*, strongly supported by bootstrap (BS > 85%) and the Mediterranean pines. *Pinus hwangshanensis*, *P. tabuliformis*, *P. thunbergii*, *P. kesiya*, *P. luchuensis*, *P. densata* and *P. yunannensis* formed a well-resolved monophyletic group (BS = 94%). *Pinus mugo*, *P. uncinata*, *P. uliginosa, P. densiflora*, *P. sylvestris* and its variety var. *siribica* formed a moderately supported monophyletic group (50% < BS < 85%). *Pinus merkusii* was placed on a separated long branch, sister to the remaining members of this subsection. The second clade comprises the Mediterranean pines and Himalayan pine *P. roxburghii*. *Pinus heldreichii* appeared as a sister taxon to the rest of the Mediterranean species. Within this clade subsection *Halepenses* (*P. brutia* and *P. halepensis*) was monophyletic (BS = 97%).

2.1.4. Discussion

2.1.4.1. *Phylogenetic Relationships of Subgenus Pinus*

The inferred phylogeny using *rbc*L, *mat*K, *trn*V and *rpl*20-*rps*18 clearly showed that *Diploxylon* split into two lineages, one comprising mainly the Eurasian species (with the exception of *P. tropicalis* and *P. resinosa*), representing subsections *Pinus*, *Canarienses*, *Halepenses* and *Pinea*, and the remaining large North American group. Thus, our molecular phylogeny confirmed the view of Price et al. (1998), who identified two sections in subgenus *Pinus*: section *Pinus* (mainly for Eurasian species) and the section informally named “New World hard pines” for the North American species.

Pinus tropicalis, an endemic Cuban pine, has been considered as a member of subsection *Pinus* (Mirov 1967, Rushforth 1987) and a close relative of the northeast North American pine *P. resinosa* due to its geographical distribution (Little and Critchfield 1969, Van der Burgh 1973, Price et al. 1998). *Pinus tropicalis* has also been associated with subsections *Oocarpae* (Klaus 1989) and *Australes* (Farjon 1984) because of its growth characteristics. Our results clearly indicate that *P. tropicalis* is a typical member of subsection *Pinus*. Thus, the “grass-stage” seedlings seen in *P. tropicalis* (Perry et al. 1998) and various New World pines have apparently evolved in parallel. Parsimony analyses using ITS recovered *P. tropicalis* and *P. resinosa* in subsection *Pinus* (Liston et al. 1999b) but could not resolve the relationship between them. Our data do not support the idea of a close relationship between *P. tropicalis* and *P. resinosa* and indicate that these species have reached America independently (Axelrod 1986). Indeed, the only reported “*sylvestris-like*” pollen denoted as a *P. tropicalis* ancestor is from north central Cuba and dates to the Oligocene (Areces 1987), while the fossil assigned as a *P. resinosa* ancestor (resembling *P. sylvestris*) is from British Columbia and dates to the Middle Eocene (Stockey 1984). The strong affiliation between *P. resinosa* and *P. nigra* displayed in our phylogeny was first acknowledged by Little and Critchfield (1969) and supported by the ability of the North American red pine to be crossed successfully only with European black pine (Nakai et al. 1995). Similar data are lacking in the case of *P. tropicalis*. Nevertheless, additional studies with multiple accessions of *P. resinosa* and *P. nigra* are needed to confirm their affinity.

The existence of the endemic Pyrenean pine, *P. uncinata* has been debated during the last 70 years. This taxon has been synonymized as *P. mugo* (Little and Crithfield 1969, Rushforth 1987) or *P. pinaster* (Mirov 1967, Klaus 1989). Conversely, this taxon was acknowledged as *P. uncinata*, restricted to Pyrenean and western Alps by Van der Burgh (1973) and Gaussen et al. (1993). Those taxonomical discrepancies have complicated its phylogenetic position, in the subsection *Pinus* or *Pinaster*. Indeed, Price et al. (1998) emphasized that a phylogenetic approach using molecular data is necessary to corroborate the placement of *P. uncinata* in the subsection *Pinus*. The result presented here, illustrates clearly its position in subsection *Pinus* and hypothesizes its possible origin from ancestral "*P. mugo*".

P. uliginosa is a restricted taxon of the Carpathian between Poland and Slovakia; however, its nomenclature is not well accepted in many of the current classifications. The European taxonomists have emphasized the presence of this taxon as intermediate form of *P. mugo* (Jalas and Suominen 1973, Gaussen et al. 1993); whereas the others do not find consistent differences between the Slovakian and Polish population of *P. uliginosa* and *P. mugo* (Mirov 1967, Little and Critchfield 1969). Although there are enormous amount of data using karyotype, serology, pollen proteins and molecular (Saylor 1972, Pru-Glowacki et al. 1985, Kormutak 1984, Gielly and Taberlet 1994), the sytematists have been so conservative to revise the current status of this taxon. Our result as well as the previous morphological-cytological-biochemical data point toward the recognition of *P. uliginosa* as distinct taxon. Finally, the phylogeny has succeeded to delineate the relationships within the "*P. mugo* complex" (Mirov 1967, Little and Critchfield 1969, Van der Burgh 1973) which comprises three species *P. mugo*, *P. uncinata* and *P. uliginosa*.

Regarding the second Eurasian clade, the close relationship between *P. pinaster* and the Mediterranean pines described here is consistent with previous classifications in which all the Mediterranean pines are included in subsection *Pinaster* (Rushforth 1987, Frankis 1993). This classification is further supported by data from artificial and natural hybridization experiments (Vidakoviç 1991), electrophoretic profiles of seed proteins (Schirone et al. 1991), immunological data (Price et al. 1987), and analysis of the ITS region (Liston et al. 1999a). Thus, the inclusion of *P. pinaster* as a member of subsection *Pinus* (Price et al. 1998) needs to be reconsidered.

Our analyses placed *P. roxburghii* in the Mediterranean clade, but its close relationship with *P. canariensis* was not well resolved. Strong morphological similitude of *P. roxburghii* to *P. canariensis* permitted its classification in subsection *Canarienses* (Little and Critchfield 1969, Farjon 1984, Klaus 1989, Price et al. 1998), while Rushforth (1987) classified all the Mediterranean pines and *P. roxburghii* into subsection *Pinaster*. Although details of the relationships within Mediterranean subsections were not well resolved, our *cp*DNA result does not support the hypothesis that the "Tethyan grade" of pines gave rise to both subsection *Pinus* and the North American subsections (Liston 1997). Moreover, our data support the proposal of Mediterranean origin of *P. roxburghii* by Klaus (1989). According to him, the Mediterranean ancestor migrated along the Tethys coast to the east and reached the Himalayan region in the Upper Cretaceous-Lower Tertiary, giving rise to *P. roxburghii*. In the analysis of Eurasian pines based on these chloroplast regions (Wang et al. 1999), the taxon labelled *P. roxburghii* seems to be a misidentified North American pine, and its placement as a different lineage is erroneous. I confirmed the sequences using multiple accessions of this taxon collected in Kathmandu (courtesy of HK Saiju, MFSC, Nepal).

The relationships among the North American subsections have remained equivocal. Based on the results of restriction analysis, Strauss and Doerksen (1990) suggested a close relationship between subsections *Contortae* and *Oocarpae*, while Liston et al. (1999a) could not determine the relationships among the subsections *Leiophyllae, Oocarpae, Australes, Attenuatae* and *Contortae* by using ITS data. Furthermore, *Contortae*, based on its morphological characters, its sympatric association in the south-eastern United States and hybridization ability among some species, has been generally considered to be related to subsection *Australes* (Saylor and Koenig 1967, Price 1989). The data presented here indicate that North American pines form four clades. The monophyletic subsection *Contortae* is strongly supported as being the basal sister group of the remaining North American subsections. Krupkin et al. (1996) acknowledged the basal position of this subsection, but Liston et al. (1999a) argued that the unexpected position of *Contortae* in the *cp*DNA phylogeny was a result of a long-branch attraction (Felsenstein 1978). However, both of those studies relied on parsimony method, which is putatively more prone to the long-branch attraction phenomena (Kim 1996,

Sanderson 2000). The total congruence among the MP, NJ and ML trees regarding the position of this subsection illustrates that the basal position must not be due to long-branch attraction.

Subsection *Ponderosae* comprises the western North American pines. The taxonomy of this subsection has been remained reasonably stable since the subsections *Ponderosae* and *Sabinianae* of Little and Critchfield's (1969) classification were unified in a single subsection named *Ponderosae* according to their crossability and similarities in oleoresin chemistry (Van der Burgh 1973, Rushforth 1987); however, the acknowledgment of some Mexican taxa belonging to this subsection is still in debate (Farjon and Styles 1997, Price et al. 1998). The monophyly of this subsection was clearly demonstrated by the results of our chloroplast analysis as well as in the ITS phylogeny (Liston et al. 1999a). Nevertheless, Krupkin et al. (1996), using *cp*DNA restriction site variation, found this subsection polyphyletic probably because of the schema utilized for the classification. In addition, our study used two groups within subsection *Ponderasae*: '*Ponderosa* group' (e.g., *P. ponderosa*, *P. jeffreyi* and *P. engelmannii*), '*Pseudostrobus* group' (e.g., *P. douglasiana*); '*Sabinianae* group' (e.g., *P. coulterii*); however the *cp*DNA phylogeny could not discriminate these two groups. Therefore, the use of rapidly evolving markers such as nuclear markers is recommended to determine the relationships between groups within *Ponderosae*.

The result of our analysis indicated that subsection *Attenuatae* is a unique taxonomic group, which is consistent with allozyme data (Millar et al. 1988), *cp*DNA restriction site analysis (Hong et al. 1993b) and RAPD data (Dvorak et al. 2000). However, the sister relationship with *Australes* proposed by Krupkin et al. (1996) is not supported by the result of our *cp*DNA analyses, or results obtained using RAPD-markers (Dvorak et al. 2000) and hybridization studies (Critchfield 1962). In addition, based on the fossil record (Axelrod 1986, Axelrod and Cota 1993) and morphological similarities, Farjon (1996) suggested a common "*oocarpae-like*" ancestor for the Mesoamerican and Californian closed-cone pines. It is likely that the Californian and Mesoamerican pines were spatially and temporally contemporaneous during the Miocene and evolved under similar climatic conditions, allowing some of the morphological characters to evolve in parallel (Dvorak et al. 2000)**.**

Few changes have been made to the taxonomy of *Australes* since Barret and Golfari

(1962) separated *P. caribaea* into three varieties. Unexpectedly, *Australes* became polyphyletic in our phylogeny. The southern U.S. species appeared as a well-supported monophyletic group in our cladogram, whereas the remaining members were placed in the latter clade (Fig. 1). Similarly, Dvorak et al. (2000) found an unambiguous division of *Australes* using RAPD markers when they assessed the relationships among *Oocarpae-Australes* subsections, consisting of two main groups: southern U.S. and Florida/Caribbean. In a study by Krupkin et al. (1996), *P. taeda* was used like an element of *Australes* in their phylogeny, establishing a position similar to that exhibited in our phylogeny, however, they were unable to detect paraphyly of this subsection. On the other hand, Liston et al. (1999a), using ITS sequences, placed *P. echinata* in a poorly resolved clade of *Oocarpae-Attenuatae-Australes*. Indeed, neither Krupkin et al. (1996) nor Liston et al. (1999a) were accurate to establish the relationship of this subsection mainly because of an insufficient number of taxa in their studies.

Despite the scarcity of studies, some have recognized the presence of two groups within *Australes*. Firstly, there were ambiguities to separate the patterns of *cp*DNA in *P. palustris* and *P. echinata* but the discernible difference with them and the patterns displayed in *P. elliottii* and *P. taeda* (Wagner et al. 1992). Secondly, a common pattern of the *BamH*I restriction fragment in *cp*DNA existed in *P. oocarpa*, *P. caribaea*, *P. palustris* and *P. echinata* in contrast to that found in *P. taeda* and *P. elliottii* (Nelson et al. 1994). In addition, Nikles (1966) found on the basis of results of hybridisation tests, that the association among *P. palustris*, *P. echinata*, *P. caribaea* var. *caribaea*, was closer than that between these species and *P. elliottii* var. *elliottii*, indicating that *P. palustris*, and *P. echinata*, *P. caribaea* var. *caribaea* share a common evolutionary history.

The subsections *Oocarpae* and *Australes* (Florida/Caribbean species) have been commonly treated as a “species complex” (Mirov 1967, Stead and Styles 1984). The lower substitution rate among the species in these subsections indicates that *Oocarpae-Australes* ancestors have radiated and speciated recently. During their radiation and migration, species through hybridisation would have developed varieties, subspecies and species that were adapted to different environmental conditions in the Mexican and Central American highlands (Mirov 1967, Perry 1991).

Our results showed that *P. caribaea* var. *caribaea* differs from the other two varieties (var.

bahamensis and *hondurensis*) and support the morphological observations by Barrett and Golfari (1962) and molecular differences noted by Nelson et al. (1994) and Zheng and Ennos (1999). Consequently, *P. caribaea* var. *caribaea* should be regarded as a distinct species rather than an ecotype as suggested by Zheng and Ennos (1999).

Subsection *Leiophyllae* appeared in the latter clade with *Oocarpae*, which echoes the notion of its recent origin, probably from the Mesoamerican pines. However, based on cone and seed morphologies, it has often been associated with section *Pinus* (Farjon and Styles 1997, Farjon 1984, etc.) or classified as an independent section. Thus, those characters in *P. leiophylla*, seem to have evolved in parallel, particularly in some Mediterranean pines.

López (1982) affirmed that *P. cubensis* is the older species of *Australes* in Cuba, while *P. maestrensis* might have evolved through hybridization between *P. cubensis* and another species, probably *P. occidentalis*. However, morphological similarities and sympatry in some areas of the Sierra Maestra between *P. cubensis* and *P. maestrensis* have complicated the taxonomy and evolutionary history of these taxa. For example, *P. maestrensis* has been misidentified and classified as *P. occidentalis*, a species found in La Española (Mirov 1967, Silva 1984), while Darrow and Zanoni (1991) and Farjon and Styles (1997) concurred that *P. occidentalis* is absent in Cuba. On the other hand, Bisse (1975) and López (1982) concluded that *P. maestrensis* is endemic to the Sierra Maestra. Our data support the idea that *P. cubensis* and *P. maestrensis* are genetically different. Further studies with multiple accessions of these species are needed to clarify the interspecific relationships. *Pinus occidentalis* should be incorporated in future phylogenetic studies.

2.1.4.2. *Phylogenetic Relationships of Eurasian Pinus*

Succinctly, the Asian members of subsection *Pinus* distributed in China formed a well-supported clade, which confirmed the point of view of Mirov (1967) and Little and Critchfield (1969) concerning the common origin of all the pines distributed in China and Ryukyu Island (*P. luchuensis* and *P. thunbergii*). In addition, our result accorded notion of the Tertiary hybrid origin of *P. densata* from *P. tabuliformis* and *P. yunnannesis* (Wang et al. 1990, Wang et al. 2001).

In the clade of the Mediterranean pines, *P. heldrechii* was a sister to the remaining

members. This taxon is endemic of the south Italy and Balkans (Mirov 1967) and usually synonymized as *P. leucodermis* (Farjon 1984, Schirone et al. 1991). The taxonomic position of this taxon remains uncertain. Generally, this taxon has been regarded as close allied of *P. nigra* and *P. sylvestris* (Klaus 1989); serological data clustered this taxon among the European members of subsection *Pinus* (Prus-Glowacki et al. 1985), while seed protein analysis illustrated a middle position between Mediterranean pines and other members of subsection *Pinus* (Schirone et al. 1991). The results here, recommends its placement within the Mediterranean pines. Within the Mediterranean pines, *P. halepensis* and *P. brutia* formed a strongly-supported clade. *P. brutia* have been described as a variety of *P. halepensis* (Farjon 1984). However, the morphological, seed protein profiles, allozymes patterns and their ability to hybridize discriminate both species but also indicate their closely relationship (Frankis 1993, Schirone et al. 1991, Korol et al. 2002).

2.1.4.3 *Sequence Divergence*

Several authors (e.g., Taberlet et al. 1991, Gielly and Taberlet 1994) have suggested that non-coding regions of the plastid DNA are more suitable to protein-coding genes such as *rbc*L for phylogenetic analysis at lower taxonomic levels because they evolve faster, presumably because of lower levels of selective constraint. However, our study showed that the synonymous sites of *mat*K and *rbc*L evolve faster than do the two non-coding regions. This unusual pattern might be a result of the constraints in the intron such as compositional constraints (Bernardi and Bernardi 1986), neighbouring base composition (Morton 1997, Morton and Clegg 1995) and constraints imposed by structure (Clegg et al. 1994). Slower substitution rates in non-coding regions have also been detected in some angiosperms (Kajita et al. 1998, Bayer et al. 2000). There are several possible reasons for the apparently slow rate in the spacer: (1) Wakasugi et al. (1994) reported the presence of ORF69 in the spacer *rpl*20-*rps*18 in *P. thunbergii*, but the presence of a protein-coding gene is unlikely because some taxa contain indels that cause frameshifts, and thus the existence of this ORF in all members of subgenus *Pinus* is doubtful; (2) regions between two ribosomal protein genes are known to be under substantial selective constraints (Zurawski and Clegg 1987), and (3) the spacer has as-yet-

unknown important regulatory functions.

The *mat*K gene has been found to be more variable than other coding genes (Steele and Vilgalys 1994, Xiang et al. 1998), which might reflect its function in the splicing of a Group II intron (Hilu and Liang 1997, Vogel et al. 1999). The numbers of nucleotide substitutions at synonymous sites in *rbc*L and *mat*K genes were similar, whereas the number of substitutions at nonsynonymous sites in *mat*K was 2.7-times higher than that in *rbc*L, which might also mirror the higher functional constraints of *rbc*L (Kellogg and Juliano 1997). Wang et al. (1999) suggested heterogeneity in the substitution rate in *mat*K between subgenera *Pinus* and *Strobus*. Similarly, our data showed that the Eurasian clade exhibited a much higher rate of nucleotide substitutions than did the North American clade, particularly at nonsynonymous sites (Table 2). Heterogeneity in the substitution rate in *mat*K and the higher nonsynonymous to synonymous ratio (>1) suggest a positive selection in the lineage of Eurasian pines.

2.2. Divergence Time and Biogeography of subgenus *Pinus*

2.2.1. Introduction

Pines have a rich fossil history dating back to the Cretaceous (see Axelrod 1986, Millar 1993). The identification and interpretation of the fossil record of *Pinus* have depended heavily on the taxonomic framework of Little and Crithfield (1969). Consequently, controversies have arisen in the assignments of many fossils to subsections that may have a recent origin; for example, the presence of *Leiophyllae* and *Ponderosae* subsections in the Cretaceous (Stockey and Nishida 1986, Millar 1993). A robust phylogeny of pines based on molecular data could facilitate the determination of which subsections or lineages of *Pinus* are ancient, which may have a more recent origin, and which characters have evolved in parallel or as adaptive mechanisms. This would assist paloebotanists in their assessment of fossils as well as provide the basis for biogeographical hypothesis in *Pinus*. The objective of this study is to estimate timing of speciation for the major subsections using the molecular clock hypothesis together with paleobotanical information to shape the history of *Diploxylon* pines.

2.2.2. Material and Methods

2.2.2.1. *Estimation of Evolutionary Rate and Divergence Time*

The constancy in the rate of evolution was assessed through the relative-rate test (Li and Bousquet 1992) for *rbc*L using *P. parviflora* as a reference taxon. The test was performed using RRTree software (Robinson et al. 1998). The test compares the rates of synonymous and nonsynonymous substitutions between pairs of evolutionary lineages. Each major clade was assumed to be a different lineage for the pairwise comparisons. The rate of substitution per site per year (r) is a function of the time of divergence (T) and the number of nucleotide substitutions per site or sequence divergence value (D_{XY}): $r = D_{XY} / 2T$ (Nei 1987). In this study, the molecular clock was calibrated by dating the earliest known fossil possessing unequivocal *Diploxylon*

structures to an age of 130 million years (MY; Alvin 1960; see Figure 2). Considering the stochastic process of nucleotide substitutions, the mean divergence time and 95% confidence intervals between clades were estimated for *rbc*L using the method of Haubold and Wiehe (2001).

2.2.3. Results

2.2.3.1. *Substitution Rates and Divergence Times*

Nucleotide substitution rates are not perfectly clock-like and are thought to be influenced by life history factors and/or generation times (Li 1993, Xiang et al. 2000). The relative rate test performed on *rbc*L indicates that the rates both on nonsynonymous and synonymous sites were not significantly different among the major clade ($P > 0.05$ for all comparisons), supporting the molecular clock hypothesis in subgenus *Pinus* (data not shown). Thus, based on the age of the *diploxylon-like* fossil (130 MY) and the sequence divergence for *rbc*L between the Eurasian and North American sister lineages (0.0095 ± 0.0021), the substitution rate was estimated to be $3.65 \pm 0.81 \times 10^{-11}$ per site per year. Application of this rate to all pairs of the clades indicated an approximate lineage-divergence date of 104 MY for subsection *Contortae* (with 95% confidence limits of 61 MY and 169 MY), 52 MY for *Ponderosae* (with 95% confidence limits of 29 MY and 87 MY), 37 MY for *Australes* southern U.S. (with 95% confidence limits of 22 MY and 69 MY), 21 MY for the Californian closed-cone pine (*Attenuatae*) (with 95% confidence limits of 7 MY and 43 MY), and 20 MY for the *Australes-Oocarpae-Leiophyllae* clade (with 95% confidence limits of 11 MY and 38 MY). Within the Eurasian clade, the split between subsection *Pinus* and subsections *Canarienses-Pineae-Halepenses* was estimated to be 64 MY (with 95% confidence limits of 27 MY and 119 MY). On the other hand, using the average number of nucleotide substitutions per site between the North American and Eurasian clades (including *P. resinosa* and *P. tropicalis*) indicates a divergence time of 136 MY (with 95% confidence limits of 70 MY and 240 MY).

2.2.4. Discussion

2.2.4.1. *Divergence Time and Paleobotanical Interpretations*

The utility of molecular divergence data has provided an important tool for evolutionary studies, especially when the lack of fossil records hampers interpretation. Here I presented a re-evaluation of the time of divergence for the main subsections based on *rbc*L sequence data.

The division of subgenus *Pinus* into two main lineages, section *Pinus* (Eurasian) and the "New World hard pines" (North American), presented here confirms a biogeographic influence in cladogenesis. Our results support some current paleobotanical interpretations; however, there are some discrepancies, probably due to incomplete fossil records and the state of conservation.

As one example of an inconsistency between paleontological and molecular date, a fossil record dated to the late Cretaceous included subsections *Leiophyllae* and *Ponderosae* (Stockey and Nishida 1986, Millar 1993), which according to our analysis have recent origins. In addition, the reported fossils attributable to the Californian closed-cone pine from an "*oocarpae-like*" ancestor (Axelrod 1986) might be an error. Thus, a re-examination of this fossil is needed. Moreover, our data indicate *Oocarpae* and *Australes*, but not *Attenuatae* share a common ancestor. However, our data are in agreement with the fossil record regarding the divergence of the subsection *Attenuatae* in the Oligocene (38 - 26 MY) (Axelrod 1986, Hong et al. 1993b, Millar 1993). Similarly, our estimation of 52 MY for subsection *Ponderosae* concurs with the fossil with unequivocal morphology of *Ponderosae* dated to the Paleocene (Axelrod 1986).

By the late Cretaceous, pines had reached the eastern and western edges of Laurasia (Millar 1998) following an east-west migratory route. Eurasian subsections appear to have evolved during the early Cretaceous and migrants to North America arrived during the mid-Cretaceous (Millar 1998). The North American pines, *P. resinosa* and *P. tropicalis* do not appear to be part of the lineage that gave rise to the American subsections ("New World hard pines") (Fig. 1). They were instead on the clade of subsection *Pinus*, which appears to have diverged from the progenitor of the "New World hard pines" in the early Cretaceous. Thus, the North American hard pines comprise two independent lineages. The very long branch that separates

the common ancestor of the “New World hard pines” from the section *Pinus*: subsection *Pinus* suggest a long period of isolation between eastern and western North America likely because of the expansion of the Western Interior Seaway. The dramatic decrease of temperature from the Middle Miocene and the cordillera in western North America became effective barriers to biotic interchange between eastern and western North America (Tiffney 2000). The *Diploxylon* ancestor used different corridors to reach America. *Pinus resinosa* and *P. tropicalis* ancestors (resembling *P. sylvestris*) might have used the North-Arctic land bridge (DeGeer Route) (LePage and Basinger 1995), which was not impeded until late Miocene (Tiffney and Manchester 2001). Although they are the only members of subsection *Pinus* in North America, I speculate they migrated into North America independently. I calculated the time of divergence between taxa, *P. tropicalis* and *P. resinosa*, giving a value of 61~75 MY. Indeed, the only reported *P. resinosa-like* ancestor dated back to Middle Eocene (Stockey 1984), whereas the fossil of pollen “*sylvestris-like*” as *P. tropicalis*-ancestor dated back to Oligocene (Areces 1987).

Meantime, the Beringian corridor might have been used for the progenitors of the “New Word hard pines” subsections. The first subsection with North American origin was *Contortae*, however this subsection has a limited fossils records that do not begin until late Eocene and are not abundant until Pleistocene (Axelord 1986). These contradictory findings could be the result of a narrow pre-Pleistocene distribution of *Contortae* in North America (Axelord 1986, Millar 1998). This subsection might not have expanded much to the east until very recently (in the forms of *P. virginiana* and *P. clausa*), as revealed by the sparse fossil records until Pleistocene (Farjon and Styles 1997, Axelrod 1986).

During the periods of active mountain building and climatic change that characterized Early- to- Mid-Paleogene (34 - 65 MY), pines extended and drifted into many refugia (Axelrod 1986, Millar 1998). The second subsection that seems to have evolved in the western part of North America is *Ponderosae*. Although there are few Tertiary fossils resembling cones and seeds of different extant species assigned to this subsection, most of them were found in the Rocky Mountains, north of the U.S.-Mexican border, and were dated from Eocene to Miocene (Axelrod 1986). These Eocene records may represent marginal populations that were centred further north and expanded southward during the cold periods (Millar 1998). The cladogram

suggests that the immediate ancestor of the *Australes* of southern U.S. could be originated from *Ponderosae*. Indeed, the oldest fossils of the extant *Ponderosae* are similar in the external features of cones to *P. taeda*; however, the majority of fossils can be ascribed to either *Australes* or *Ponderosae* (see Stokey 1984, table 1; Axelrod 1986, Millar 1998).

Likewise, our phylogeny indicates that *Attenuatae*, *Oocarpae-Leiophyllae* and *Australes* share a common ancestral origin or one group served as a progenitor to the other that probably occurred during Eocene. Indeed, the severe climatic changes that occurred during the Eocene and Early Oligocene caused successive expansions and contractions of pine populations, resulting in the creation of several refugia (Millar 1993, Millar 1998). It can be assumed that multiple radiations occurred from the Eocene refugia: to the west to originate the Californian closed-cone pines (*Attenuatae*), to the south into Mexico to derive the Mesoamerican pines (*Oocarpae-Leiophyllae*) and to the southeast resulting in the establishment of the *Australes* members of southern U.S. (i.e., *P. elliottii*, *P. pungens*, *P. serotina*, *P. rigida* and *P. taeda*).

Based on fossils records (Axelrod 1986, Axelrod and Cota 1993) and morphological similarities, Farjon (1996) suggested a common "*oocarpae-like*" ancestor for the Mesoamerican and Californian closed-cone pines. It is likely that the Californian and Mesoamerican pines were spatially and temporally contemporaneous during the Miocene and evolved under similar climatic conditions allowing some of the morphological characters to evolve in parallel (Dvorak et al. 2000). Elements of *Oocarpae* moved south from northern Mexico into Central America, and the *Australes* shifted east during Miocene-Pleistocene and entered the eastern part of Cuba, and they might have served as Pleistocene refugia of the *Australes* (Florida/Caribbean species). Double refugia in *Australes* have been proposed in several occasions (Watts 1983, Axelrod 1986, Wells et al. 1991), one of them in Texas/Mexico (continental *Australes*) and the other further south, perhaps in the Caribbean Sea; however, our results suggest that they occurred independently *in tempo*. The first one took place during Eocene (Texas/Mexico) and the second during Pleistocene for the Florida/Caribbean species (Cuba). Because southern Florida was submerged until Pleistocene, settlement of pines in this area was not possible. The present approach supports the idea that the colonization of *Australes* in southern Florida occurred as a relatively recent event (Little and Dorman 1954, Dvorak et al. 2000) and that the establishment

of pines in south Florida was a result of multiple migrations of *Australes* from both the north and south (Mirov 1967, Dorman 1976).

2.3. Outline of the Classification of subgenus *Pinus*

New treatment of subgenus *Pinus* based on the light of the combined sequence and morphological-cytological-biochemical data is presented in the following outline. The subsection delineations largely follow the format of Price et al. (1998), which represent a harmonic compendium of previous classifications and treatments.

Subgenus *Pinus*: Cone scale with dorsal umbo and a sealing band. *Seedwings* articulate. Two fribrovascular bundles per leaf.

Other features: Resin canals in the intermediate or inner portion in the leaves; fascicles of 2-5 (rarely up to 8) leaves, leaf sheath are persistent (except in subsection *Leiophyllae*); leaf stomata on both ventral and dorsal surfaces; the bases of the leaf bracts are decurrent, leaving a rough branch after the leaves are shed (except in some species of subsection *Ponderosae*: group *Pseudostrobus*); the bark is generally thick and fissured; wood is generally harder and yellowish and with more pronounced annual growth rings; growth of spring shoots is either uninodal or multimodal; high percentage of α-pinene compared with the β-pinene.

Section Pinus:

Subsection *Pinus* (Eurasia, NE North America and Cuba):Leaves in fascicle of 2(-3); small cones maturing in 18 month with moderately flexible scale, opening early and falling completed from the branch. Wood with large (“fenestriform”) ray cell pits; 2 pairs (other pines have only 1 pair) of heterobractial chromosomes (in which the long arm is more than 2 times the length of the short arm); shoots always uninodal, high percentage of α-pinene.

Species*: P. tropicalis, P. hwangshanensis, P. tabuliformis, P. thunbergii, P. kesiya, P. densata, P. yunannensis, P. luchuensis, P. densiflora, P. sylvestris, P. uliginosa, P. mugo, P. uncinata, P. nigra, P. resinosa, P. massoniana, P. merkusii, P. thunbergii*

Mediterranean subsections (*Halepenses, Canarienses, Pinaster, Pineae*): Bright (often yellowish) green leaves in fascicle of 2-3. Glossy red-brown cones to nut-brown with thick, stiff, woody scales, often long-persistent. Wood with small ray cell pits, only 1 pair of heterobractial chromosomes (except in *Halepenses*)

Subsection *Halepenses* (S Europe, W Asia and N Africa): Leaves in 3 per fascicle,

2 pairs of heterobractial chromosomes.

Species: *P. halepensis, P. brutia*

Subsection *Canarienses* (Canary Island and Himalayas): Leaves in 3 per fascicle. Seed are long nondetachable (adnate) to the wings.

Species: *P. canariensis, P. roxburghii*

Subsection *Pinaster* (N Africa): Leaves in 3 per fascicle.

Species: *P. pinaster, P. heldreichii*

Subsection *Pinea* (S Europe): Very large seed with very rudimentary wings, cones mature in 36 months. Wingless seeds

Species: *P. pinea*

Section *Trifoliea* (North American Pines): All the North American pines belong to this section (except for *P. tropicalis* and *P. resinosa*). I propose to denominate this section as *Trifoliae* given by the fact that all the North American pines exhibit 3 leaves per fascicle. In addition, at the present time still there is not formal name for describing this section and the usage of the “New World hard pines” by Price et al. (1998) seems to be ambiguous (Michael Frankis personal communication). Leaves commonly in 3 (-4-5) per fascicle. Vigorous shoots mostly multimodal (except in subsection *Ponderosae*, uninodal). Branching not candelabra-like (except in *Ponderosae* and *Leiophyllae*). 1 pair of heterobractial chromosomes.

Subsection *Contortae* (North America): Cones small, symmetrical or oblique, and often closed when ripe (serotinous). Leave short (< 8 cm) and 2 per fascicle persistent 5-7 years.

Species: *P. banksiana, P. contorta, P. virginiana*

Subsection *Ponderosae* (W North America, S to Central America): This subsection represents 13 species, 11of them endemic. Erect uninodal shoots with spreading leaves resembling a chimney sweep’s brush. Branching candelabra-like, with upcurved branches. Cone symmetrical or slightly oblique maturing in 18 months, open when ripe and some cones basal scales remain on the branch after cone has fallen.

Species: *P. ponderosa, P. douglasiana, P. jeffreyi, P. engelmannii, P. coulterii*

The groups’ delimitation is basically done on the basis of the morphology and

geographically distribution (Mirov (1967) and Martínez (1948) scheme, described three great pine complexes of America: a *Ponderosa* complex, a *Montezumae* complex, and a *Pseudostrobus*).

'*Montezumae* Group': 5 needles per fascicles, occurs mostly along the Transvolcanic Belt, with some species occurring along the Sierra Madre Oriental and southward into Guatemala.

'*Pseudostrobus* Group': 5 needles per fascicles, also occurs primarily along the Transvolcanic Belt, Sierra Madre Oriental and southward into Guatemala, El Salvador and Honduras. No presence of “grass-stage”. The bases of the leaf bracts are non-decurrent. Needles more slender than those of the *Montezumae* group. (This group was not sampled)

'*Ponderosae* Group': 3 needles per fascicles, occurs mostly in north-western Mexico in the Sierra Madre Occidental.

'*Sabinianae* Group': Restricted to California, 3 or 5 needles per fascicle ascending or spreading, persisting 3 - 4 years. Seed wing reduced in length but detachable.

Subsection *Australes* (E U.S., Western Indies and Central America): Fairly straight branching, not candelabra-like, multinodal long shoots (more than 1 branch whorl per growing season on vigorous shoots), serotinous and persistent cone. Cones mature after 16-20 month releasing the winged seeds. It distributed in southern United States and Caribbean Islands.

'*Taedae* Group' (SE U.S.): This groups can be defined by molecular data and supported with some morphological features such as cones serotinous, 3(-2) per fascicle, ascending to spreading, persisting 3 years, Seed cones maturing in 2 years

Species: *P. taeda, P. pungens, P. serotina, P. rigida, P. elliottii*

'*Caribbean* Group' (Florida and Western Indies, and Central America): The species occur in Florida and Caribbean islands, and in Central America. The cone non-serotinous, articulated seed (except in *P. palustris*, *P. c.* var. *hondurensis* and *bahamensis*, where are adnate) and with 3 (-5) needles per fascicle (except in *P. cubensis* with 2 needles).

Species: *P. cubensis, P. caribaea, P. echinata, P. palustris, P. maestrensis*

Subsection *Attenuatae* (California, U.S.): Cones strongly serotinous and closed-cone (only opening on burning) in whorls, hard-heavy and very asymmetric, persistent 6 to 20 (-40) years, maturing in 22-24 month. Seed cone maturing after 2 years, 2 (-3) needles per fascicles, persistent 2-5 years, straight or slightly curved, twisted.

Species: *P. attenuata, P. radiata*, *P. muricata*

Subsection *Oocarpae* (Mexico, S to Central America): This subsection corresponds to the Mesoamerican pines. Strong similarity with *Australes* and *Attenuatae* makes problematic is delimitation.

'Group *Oocarpa*': Mesoamerican closed-cone pines, leave in 3-5 per fascicle. Cones are often oblique but very serotinous.

Species: *P. oocarpa, P. patula*

'Group *Teocote*': Meosamerican opened-cone pines. 3-(4-5) leaves persistent 2-3 years, in dense, the sheaths are much reduced 8-15mm, the cone more less symmetrical.

Species: *P. lawsonii*, *P. teocote, P. herrerae*

Subsection *Leiophyllae* (Mexico and adjacent SW U.S.): Deciduous fascicle sheath scale. Cones mature in 36 months. Very small seed “wingless” seeds.

Species: *P. leiophylla*.

3. Relationships and Status of the Cuban Pines

3.1. Introduction

The Neotropics pines comprised more than 60 species (Farjon and Styles 1997), from which 47 are distributed in Mexico, Central America and Caribbean Basin. The habitats varied from savanna communities at sea levels, dry pine-oak forest or hardwood montane or low elevation forest. The Caribbean Basin species comprised six taxa (Farjon 1984) basically distributed in lowland savanna habitats and four of then are Cuban endemics taxa. These savannas are dominated by the widespread *P. caribaea* (including its three varieties), which extends from Bahamas Inlands through Cuba to Belize, Honduras and Nicaragua. In Cuba, the typical variety is *P. caribaea* var. *caribaea* restricted to the western part of the Isle of Cuba (low elevation of Pinar del Río) and savanna of Isle of Pines. Likewise, *P. tropicalis* also restricted to western part of Cuba occurring at low-elevations and sandy savannas in form of a monotypic forest or in a clinal association with *P. caribaea* var. *caribaea* (Fig. 5).

In the eastern Cuba, reside the endemic pines *P. cubensis* and *P. maestrensis* (Fig. 5). However, the acknowledgement of the later taxon is not consistent by all the taxonomists and in many cases the latter is synonymized with the former (sees Farjon 1984, Farjon and Styles 1997, Price et al. 1998). *Pinus cubensis* extends from Sierra de Nipe eastward to Baracoa at sea level, forming a pure forest. Whereas, *P. maestrensis* occurs only in Sierra Maestra in forms of isolated keys separated each other by hardwood forest, and has a wide altitudinal distribution (e.g., from 200 m to 1800 m).

From the systematic point of view, there is so much debate about the position of the Cuban *Diploxylon* pines. It was finally confirmed the taxonomical position of the tropical pine (*P. tropicalis*) in section *Pinus* (Geada et al. 2002). While, the remaining three taxa lay on the section “New World hard pines”, and its relationships have not been addressed previously. Nowadays, an obscure question that persists is the recognition of *P. maestrensis* as new taxon, resulting from the hybridization between *P. cubensis* and *P. occidentalis* or as an ordinary variety of *P. cubensis*. From the evolutionary point of view, one of the important issues is that

concerns the routes of migration of the pines in the Caribbean basin region. Therefore, how do they expand through Cuba? Which is the oldest species in the Caribbean region, and which the time of divergence (speciation) for these taxa? In order to answer all of these questions I analyzed three non-coding regions on the *cp*DNA and corroborated this result with the morphological data collected by López during 1980-1998 and the group Pines Protection (University of Pinar del Río) where the author is a member.

3.2. Materials and Methods

Molecular markers: Three individuals per species were selected for the sequence analysis. The *trn*L intron, *trn*T-*trn*L and *trn*L-*trn*F spacer were amplified by PCR using the universal primers described by Taberlet et al. (1991). PCR products were directly sequenced after purification using a GENECLEAN KIT III (BIO101). DNA sequencing was performed using an ABI 310 Genetic Analyzer (Perkin Elmer) with an ABI BigDye Terminator Cycle Reaction Kit following the manufacturer's instructions. The sequences were aligned visually but no sequence variation was detected among the three sampled individuals per species. The average number of nucleotide substitutions was calculated based on the method of Jukes and Cantor (1969). A neighbor-joining tree (NJ; Saitou and Nei 1987) was constructed using MEGA ver. 2.1 (Kumar et al. 2001) (i.e., cladistic approach = phylogeny). The divergence time (*T*) between the species was estimated using substitution rate per site per year (*r*) (Geada et al. 2002) and the nucleotide divergence among sequences (D_{XY}). In order to estimate *r*, the molecular clock was calibrated by dating the unequivocal *P. cubensis* fossil to 1 MY (Pleistocene; Areces 1987).

Morphological data: I re-examined the seven populations of *P. tropicalis* (La Jagua, Galalón, Bartolo, Mina Dora, Viñales, San Juan and Pilotos), nine of *P. caribaea* (Las Cañas, Laguna Catalina, Arroyo Colorado, La Güira, Cajálbana, Rancho Mundito, Los Palacios, El Salón and Viñales), eight *P. cubensis* (Pinares de Mayarí, Cabonico, Cayo Güin, Palma Clara, La Tagua, La Corea, El Jardín and Moa), and four of *P. maestrenisis* (Sofia, La Alcarraza, Los Números, Loma del Gato), which were sampled by López (1982). The phenogram generated by López (1982) using 16 morphological characters (e.g., cone length, cone diameter, cone

length/width, umbo width, umbo length, umbo thickness, umbo length/width, number of scale per cone, fascicle diameter, fascicle length, fascicle length/width, sheaths length, number of needle per fascicle, number of cotyledons, branching and shoots), based on the coefficient of correlation of Sokal and Sneath (1963) (i.e., phenetic approach = numerical taxonomy) was compared with our phylogenetic tree.

3.3. Results

A total of 1372 bp of aligned sequences was examined (Fig. 6). Six out of eight polymorphic sites were phylogenetically informative but resulted from the divergence of *P. tropicalis* (subsection *Pinus*) from the others three species (subsection *Australes*). However, among the Cuban pines one informative indel of seven-bp was detected in eastern pines (*P. cubensis* and *P. maestrensis*) compared to the western taxa (*P. tropicalis* and *P. caribaea*). The NJ tree (Fig. 7) shows that *P. tropicalis* is a sister taxon of the remaining pines. *Pinus cubensis* is a closely related to *P. maestrensis* but not to *P. caribaea*. The nucleotide divergence between *P. caribaea* and *P. cubensis* was estimated to be 0.00222 ± 0.0002 based on the non-coding, and the time of differentiation between both taxa was estimated to be 2 MY with 95% confidence intervals of 10 MY and 0.7 MY, and between *P. cubensis* and *P. maestrensis* was estimated to be less than 20 000 years.

The morphological data unambiguously discriminates each OTUs, and the phenogram shows that the Cuban pines split into two main clades. One of them grouped all the populations of *P. tropicalis* and *P. caribaea*, and the other all the populations of the *P. maestrensis* and *P. cubensis*. The populations of *P. tropicalis* were more distant from *P. caribaea* than the populations of *P. cubensis* and *P. maestrensis*.

4.4. Discussion

Although both approaches were congruent to differentiate each OTU, the striking difference relied on the clustering pattern. The molecular characters showed subsectional relationships among the pines, whereas the morphological data displayed a geographical

pattern (Fig. 5). The morphological data basically divided the two clusters into: western populations (*P. tropicalis* and *P. caribaea*) and eastern populations (*P. cubensis* and *P. maestrensis*).

It is advisable that all approaches could provide congruent results but it is not hold in the majority of the case, thus, one simple way to overcome this difficulty is to find out the homologies between both approaches (Moritz and Hillis 1997). The results clearly demonstrate that the morphological data was strongly biased to the geographic distribution of the species, which can reflect the adaptation to distinct environments. Several authors have often alluded the close affiliation between *P. caribaea* and *P. cubensis* (Fors 1947, Hudson 1960, Little and Dorman 1954, Mirov 1967), and have hypothesized that they might have reached Cuba simultaneously using the same route of migration (Mirov 1967). However, morphological data by López (1982) and molecular estimating the divergence time to be 2 MY (Pleistocene) suggest that they unlikely diverged in Cuba. *Pinus caribaea* probably diverged in Mexico/Central America and entered into Cuba through Bahamas or Yucatan (Farjon 1996). On the other hand, López (1982) pointed out that *P. cubensis* is a well-adapted taxon with a little variation in the taxonomical characters and ecologically exigent (e.g., associated to a type of soil on certain type of rock); while, *P. caribaea* is a young plastic taxon with a large morphological variation. All of these arguments incline me to believe that *P. cubensis* is the oldest species among the Caribbean Basin members of *Australes*. Many reason make me believe that *P. cubensis* and *P. caribaea* do not share the same history; 1)- *P. caribaea* is confined to the western part of Cuba (Pinar del Río and the Isle of Pine) and there is no evidence of its distribution in the Central part of Cuba (Berry 1934, Samek 1967, Areces 1987), and 2)- this species does not hybridize with *P. cubensis* (Bisse 1975) and does with the other two varieties (*bahamensis* and var. *hondurensis*) as well as with the allied taxa of subsection *Oocarpae*. Therefore, I accord with the idea that three species might have enter to Cuba from the Miocene for *P. tropicalis* (see Chapters 2 and 4), during the late Pliocene-early Pleistocene for *P. cubensis* and middle Pleistocene for *P. caribaea* (López 1982). Additionally, the idea of the possible hybridization between *P. caribaea* and *P. tropicalis* emerged from the results of the morphological approaches and the assumption by Mirov (1967). It is, however, refused on the light of my results and the fact that the

crossability is a mere plesiomorphic character which seems to be kept below sectional level in pines (Price et al. 1998). Regarding to *P. maestrensis*, the molecular data supports the close affiliation with *P. cubensis* and possibly speciated through hybridization between *P. cubensis* × *P. occidentalis* (see section 2.1.). Finally, the morphological characters such as number of needles per fascicle, number of cone scale, diameter of the cone, length of the needle were distinctive for the identification of each one of the species but have been scarcely studied in the Cuban pines due to the difficulties to access new specimens (Farjon and Styles 1997).

Here, I present the some of the botanical features that can help to identify the four species (López 1982).

Pinus tropicalis: *Tree*, medium to large, height to 20-30 m, dbh. to ca. 80-90 cm. Trunk, monopodial, terete, erect and slender, in forest stand the clear bole for 2/3-3/4 of height. *Branches* of first and order sparse, thick, spreading to ascending. *Seedling*: 5-6 cotyledons, “grass stage” as an adaptation to the ground fire. *Needles* in fascicle only of 2. *Seed scale* ca. 100-120, *Cone* asymmetric, length of 10cm and 3 cm of diameter, *Needle* length 20-30 cm, longevity of 2-3 years, green yellowish and very rigid. *Seed wing* articulate. *Branches* of first and higher orders sparse, spreading ascending, forming an irregular and open crown.

Pinus caribaea: *Tree* large, height to 25-30 m, dbh. to ca. 70-80 cm. Trunk, monopodial, very slender and erect, in forest stand the clear bole for 2/3-3/4 of height. *Needles* in fascicle only of 3 (-4). *Cone* asymmetric, length 5-12 cm and 1.3-4 cm of diameter, *Needle* length 15-25 cm, longevity of 2-3 years, dark green and flexible. *Seedling*: 6 (-8-9) cotyledons, with an elongated stem, primary leaves green ascending soon replaced by the secondary. *Seed scale* ca. 120-170. *Seed wing* adnate. *Branches* of first order slender, spreading or ascending; of second and higher orders similar or dropping.

Pinus cubensis: *Tree* medium to large, height to 25-30 m, dbh. to ca. 50-60 cm. Trunk, monopodial, erect and slender. *Needles* in fascicle only of 2, very rare 3. *Cone* asymmetric, length 5-9 cm and 1.5 – 2.5 cm of diameter, *Needle* length 6-15 cm, longevity of 3-4 years, dark green and very rigid. *Seeding*: 6-8 cotyledons. *Seed scale* ca. 90-110. *Seed wing* usually adnate (with a membrane covering one side) or articulate (with held to seed by two claws), grey-brown with a graphite-like tinge. *Branches* of first order slender, spreading irregular; branches of higher

orders similar, forming a small irregular open crown.

Pinus maestrensis: *Tree* large, height to over 30 m, dbh. to ca. 100-150 cm. Trunk, monopodial, erect. *Needles* in fascicle only of 3 (4) (5). *Cone* symmetric, length 5-9 cm and 1.5 – 2.5 cm of diameter, *Needle* length 10-18 cm, longevity of 4 years, glaucous-green and very flexible. *Seed wing* adnate or articulate. *Seedling*: 7(-8-9) cotyledons. *Branch*es are lean; of first order slender, spreading irregular; branches of higher orders similar, forming a small irregular open crown.

4. Population Genetics of *Pinus tropicalis* Morelet

4.1. Phytogeography of *Pinus tropicalis* Morelet revealed by *cp*DNA variation

4.1.1. Introduction

Genetic variation in natural populations is a major concern of evolutionary biologists because the amount and distribution of genetic variation is likely to affect the evolutionary potential of species and/or populations. In the last three decades enormous amount of data on genetic diversity in natural plant populations has been accumulated and correlations between genetic diversity and various attributes of species such as endemism, mating system and geographic distribution has been examined (Hamrick and Godt 1989). Although such genetic data has been collected for a wide variety of wild plant species, data on tropical to subtropical and non-continental species are relatively limited (Maki 2001). Insular endemic plants have garnered the attention of many evolutionary biologists (Stuessy and Ono 1998) because they often have unique characteristics that differ from their continental congeners relatives, and remarkable adaptive radiation is often found in insular endemic taxa (Fransisco-Ortega et al. 1997). In addition, island species are considered to be more prone to extinction, due to the genetic paucity in populations (Frankham 1997). Island populations tend to have a lower level of genetic diversity than do continental populations, and therefore in the viewing of the conservation biology, it is meaningful to examine genetic variation in island endemic species.

Tropical pine (*Pinus tropicalis* Morelet.) is an endemic insular species distributed naturally in the western Cuba, Pinar del Río and Isla de la Pinos. *Pinus tropicalis* and *P. resinosa* represent relic species of the Eurasian lineage of *Diploxylon* pines (section *Pinus*) in North America (Geada et al. 2002), and their occurrence in diverse climates in America implies independent entries, probably at Early Paleogene to Late Cretaceous for the ancestor of *P. tropicalis*, and Miocene for *P. resinosa* in north-eastern U.S. and Canada (Axelrod 1986). Therefore, *P. tropicalis* signifies an endemic-aged taxon among the existing

pines but very little is known about the genetic variation in its natural populations (Price et al. 1998).

Despite the large range of distribution in Pinar del Río the natural populations in Isla de Pinos have disappeared (Census of the Academia de Ciencias de Cuba, 1998). Historically, it was distributed in the northern part of Pinar del Río at lower elevation, but at present it occurs as discrete populations located in mountain areas at high altitudes in the North-eastern, Central and North-western regions, where populations seem to be more continuous. Tropical pine is the focus of increased attention because of its economical and ecological importance and declining abundance over much of its range. The declining is a consequence of the overexploitation of timber, invasion and plantation of *Pinus caribaea* var. *caribaea* and encroachment of grasses associated with the fires exclusion. These phenomena seem to have altered the size and genetic structure of the natural populations, and pose a serious threat to the development of mature woodland. As a result, most of the stands of *P. tropicalis* have been decimated; however, studies on the genetic variability of the tropical pine have been limited to very few inherited quantitative traits. Knowledge of the genetic variation in this species is important to develop appropriate strategies for the *in situ* conservation and regeneration of forests, and also for testing whether genetic diversity will be lost through sampling for *ex situ* conservation, which is possible in other species (Hamrick and Godt 1996). In addition, for a single-species conservation planning, it is important to identify the component of evolutionary lineages in order to retain the maximum genetic diversity, and to incorporate information on historical population processes (Moritz 1994, Moritz et al. 1997).

Molecular markers of organelle DNAs have proved to be useful for resolving phylogeographic patterns and inferring the routes of expansion of species (Avice 1994). In plants, however, chloroplast DNA (*cp*DNA) is thought to evolve slowly, and has generally been used for studies at higher taxonomic levels. Recently, moderate and high levels of genetic variation have been detected between populations of certain species (Lu et al. 2001, Grivet and Petit 2002, Marshall et al. 2002) especially in the non-coding regions of the *cp*DNA, which has become an appropriate marker for tracking migration routes in association with geological history (Walter and Epperson 2001, Richardson et al. 2002). In this study, I

sequenced two non-coding regions of the *cp*DNA, *trn*T-*trn*L spacer and the *trn*L intron, and used them as markers to estimate the phylogeographic pattern and the genetic variation of *P. tropicalis*. Because the *cp*DNA does not recombine and it is paternally inherited in pines (Ennos 1994) distinct lineages may be represented in the data, including information about the dispersal of pollen and seed.

4.1.2. Material and Methods

Study sites and sampling: According to a report by the Institute of Systematics and Ecology of Cuba (ISEC), there are seven natural populations of *P. tropicalis* in Pinar del Río, Cuba: one population in the northeast of Pinar del Río (Galalón), two populations in the central region (Pilotos and Viñales), and four populations in the northwestern region (La Jagua, Bartolo, San Juan and Mina Dora) (Fig. 1, Table 1). Needle samples were collected from individual trees of the populations, Galalón, Pilotos, La Jagua, Bartolo and Mina Dora, and used seedlings from germinated seeds of the remaining two populations we collected. Total DNA was isolated by the CTAB method (Doyle and Doyle, 1990).

Sequence analysis: The *trn*T-*trn*L spacer and *trn*L intron were amplified by PCR using the universal primers described by Taberlet et al. (1991). PCR products were directly sequenced after purification using a GENECLEAN KIT III (BIO101). DNA sequencing was performed using an ABI 310 Genetic Analyzer (Perkin Elmer) with an ABI BigDye Terminator Cycle Reaction Kit following the manufacturer's instructions. The sequences were aligned visually. Because the chloroplast genome is haploid and does not undergo recombination, it can be viewed as a single locus, and thus the two regions were combined in order to derive a haplotype of each individual. The neighbor-joining tree was generated with the identified haplotypes using MEGA ver. 2.1 (Kumar et al. 2001). The minimum spanning network was constructed with the option implemented in ARLEQUIN version 2.000 (Schneider et al. 2000).

Population genetic analysis of the cpDNA sequence variation: Inter- and intra-population genetic diversity were quantified in two fashions: at nucleotide levels using the nucleotide diversity within populations (π_S), overall populations (π_T) and coefficient of

nucleotide differentiation N_{ST}= (π_T - π_S/π_T), were estimated. At gene level using the gene diversity within-population (H_S), the gene diversity entire population (H_T), and the coefficient of gene differentiation (G_{ST}) were estimated, and the UPGMA tree was generated based on the Nei's genetic distance (1978) among populations, using GDA 1.1 (Lewis and Zaykin, 2001). The significance of genetic differentiation between populations and that among regions were tested by molecular analysis of variance (AMOVA) using ARLEQUIN version 2.000 (Schneider et al. 2000). The statistical molecular variance of Φ_{CT} (among regions, i.e., between northeastern, central and northwestern regions), Φ_{SC} (among populations within regions), and Φ_{ST} (among populations) were estimated. The gene flow (M=*N*m) was calculated under the island model, where estimated values of G_{ST} =1/2M+1 for haploid populations. The significance of the association between G_{ST} and geographical distance was determined with Mantel-test for assessing the pattern of isolation by distance.

The mean divergence time (*T*) and the 95% confidence intervals between the populations were estimated using nucleotide divergence between populations (D_{XY}) and the substitution rate per site per year (*r*) calculated by Geada-López et al. (2002) following the method of Haubold and Wiehe (2001).

4.1.3. Results

CpDNA sequences analysis: *Cp*DNA sequences for the two non-coding regions in all the individuals from the five populations were determined. Differences of one length differences in the *trn*T-*trn*L spacer, with sizes of 456 to 466 bp, caused by the insertion/deletion of a 10-bp of 5'- AGAAGGGGAG - 3' were detected. No length variation was detected in the *trn*L intron, and the length total of sequences was 524 bp. Four nucleotide substitutions were detected from the completed aligned sequences of the *trn*T-*trn*L spacer and *trn*L intron (980-990 bp). The sequences have been deposited in the DDBJ database under the accession numbers AB097059-AB097066 for the *trn*T-*trn*L spacer and AB097067-AB097074 for the *trn*L intron. Eight haplotypes were identified according to the

number of substitutions and indels (Table 4). Haplotypes II, III and VIII were respectively derived from the haplotypes I, IV and VII, differing solely by the repeat motif (5'-AGAAGGGGAG - 3') in the *trn*T-*trn*L spacer. The haplotype I was predominant in the population Galalón, and its derived haplotypes II was only found in this population. The haplotype IV was present in the populations La Jagua, San Juan, Mina Dora and Bartolo. Haplotype III was found in low frequency only in the population La Jagua, while haplotypes V was shared in the populations San Juan, Mina Dora, Bartolo and Viñales. The haplotype VI was found in the populations Viñales and Pilotos. Haplotypes VII and VIII were only found in the population Pilotos.

To improve the estimation of genealogical relationships among haplotypes a minimum spanning network was constructed by linking sequences in a hierarchical manner based on mutational changes between them (Fig. 9A). Within the minimum spanning network, most closely related chlorotypes were linked by single mutations. The network displayed that haplotypes VI and VII were in the interior nodes of the network, serving as a linker to the haplotype I with the haplotypes IV. According to the network each haplotypes can be detected by one step of mutation. Likewise, the neighbor-joining tree generated from the haplotypes sequences (Fig. 9B) showed two well-defined groups, one of them for haplotypes I, II, VII and VIII and other for III, IV, V and VI. Within the first group, haplotypes II and I were separated from the VII and VIII, while within the second group the haplotypes VI was the first to diverge followed by haplotypes III, IV and V.

Population genetic analysis of the cpDNA sequence variation: The genetic variation varied among populations, the population Galalón displayed a lower gene diversity (0.203), followed by population La Jagua (0.411) but with no nucleotide diversity ($\pi = 0.00000$), because the two-presented haplotypes in each population varied only for the repeat motif. The populations Viñales, San Juan, Bartolo, Mina Dora and Pilotos showed the higher value of gene diversity (0.268, 0.442, 0.530, 0.533 and 0.539, respectively) and nucleotide diversity (0.00027, 0.00045, 0.00053, 0.00054 and 0.00049, respectively). Most of the genetic variation was restricted to the North-western and Central populations. The nucleotide diversity, π_T and π_S were 0.00124 and 0.00026, respectively; whereas the gene diversity, H_T

and H_S were 0.762 and 0.417, respectively revealing that there was a strong nucleotide and gene differentiation among the populations (N_{ST} = 0.79 and G_{ST} = 0.45). The UPGMA tree generated from the pairwise Nei's distance among populations (Fig. 10) showed that Galaón population was distant to the other populations, in where the Central populations (Pilotos and Viñales) occupied the basal position followed by the North-western populations (La Jagua, San Juan, Bartolo and Mina Dora). The nucleotide divergence based on combined sequences was calculated to be 0.00247 ± 0.0013 between Galalón and the rest group of populations, and 0.00168 ± 0.0012 between the Central populations (Pilotos and Viñales) and the North-western populations. Using these values, the age of the population differentiation was estimated to be 33 MY with 95% confidence intervals of 10 MY and 45 MY between Galalón and rest group of populations, and 5 MY with confidence intervals of 1 MY and 18 MY between Central and North-western populations.

Highly significant ($P < 0.01$) genetic differentiation was detected among populations by AMOVA (Table 5). Of the total molecular variance, 52% variance was attributable to the population divergence and 48% was to the individual differences within populations. When the total variance was partitioned into three geographical regions, e.g., the North-eastern (Galalón), North-western (La Jagua, San Juan, Mina Dora and Bartolo) and Central (Viñales and Pilotos); the 58% was attributable to the differences among the regions (Φ_{CT} = 0.5841; $P < 0.01$), 39% was found to individual differences within population (Φ_{ST} = 0.607; $P < 0.01$), and 3% was caused by populational differences within regions (Φ_{SC} = 0.0570; $P > 0.05$). As a result, a highly significant genetic differentiation was found only among regions and within populations. This indicates that individuals of each population within North-western populations and Central population were more similar to their co-members than to individuals of other populations. Pairwise G_{ST} values revealed a significant differentiation between populations from different regions; but not in any pair of populations from the same region. The smallest values were found among the populations of North-western region (Table 5); however, this value was not significant. Essentially, the North-western populations can be considered to be a metapopulation. Consequently, the high genetic differentiation among

populations of *P. tropicalis* was due to the divergence between the North-eastern to the rest of the regions, and a similar result was gained from the Nei's genetic distance. Therefore, the estimation of the gene flow was restricted between the regions and populations, except between populations of the North-western region (Table 6). There was a significant positive correlation between the pairwise G_{ST} and the geographic distance matrices ($r = 0.68$, $P < 0.05$). While there are anomalous genetic similarities between certain pair of populations (i.e., La Jagua, San Juan, Bartolo and Mina Dora), some of the differentiation among populations can be explained by the isolation-by-distance mechanism.

4.1.4. Discussion

CpDNA variation and nucleotide diversity: The *cp*DNA sequence analysis was effective for detecting genetic variation in the *P. tropicalis*. *Cp*DNA genome is known to evolve mainly through point mutations or small indels (Clegg et al. 1994). They also found that indels tended to occur more frequently than point mutations. Although results of many studies (e.g., Desplaque et al. 2000) support the observation of frequent indels over point mutations in the *cp*DNA, the patterns of polymorphism in the *trn*T-*trn*L spacer revealed here as well as in *Cunninghamia* of *trn*D-*trn*L spacer (Lu et al. 2001) and in *Beta vulgaris* (Desplaque et al. 2000) showed a higher level of point mutations. A repeat of 10-bp-long motif, 5'-AGAAGGGGAG - 3', was found in haplotypes from different populations implied that this might be occurred independently as a result of the replication slippage (Clegg et al. 1994; Powell et al. 1995b) or other mutational mechanisms. Furthermore, in all the cases derived haplotypes (with the minisatellite) occurred in the same populations with the original ones but in low frequency, suggesting that the mutation has occurred fairly recently. Simple sequence repeats (so-called SSRs) such as mono- or di-nucleotide repeats have been reported in *cp*DNA (Powell et al. 1995a, Powell et al. 1995b). However, tandem repeats that consist of such a long motif (so-called minisatellite in Hartl 2000) are not a typical fashion in the variation of the *cp*DNA within populations. Such minisatellite variation has been detected

previously within one populations of the rare endemic species *Trigonobalanus verticillata* (Kamiya et al. 2002).

Allozymes studies have demonstrated that species with a restricted-discontinuous distribution or endemic-endangered species often have low genetic diversity relative to the widespread species with similar life histories (e.g., Hamrick and Godt 1989, Godt et al. 1997, Godt and Hamrick 1998). On the other hand, Hamrick and Godt (1996) denoted that pines often maintain relatively high levels of genetic variation at nuclear level (i.e*.,* allozymes) and displayed little genetic differentiation among populations. However, the values of genetic variation detected in natural populations of *P. tropicalis* at *cp*DNA (H_T = 0.762 and H_S = 0.417) are also similar to that found in the widespread pines such as *P. pinaster* (Ribeiro et al. 2001, 2002) and *P. sylvestris* (Provan et al. 1998), and the Californian closed-cone pines (Hong et al. 1993a) and *P. resinosa* (Walter and Epperson 2001) with a limited distribution. Alternatively, it is preferable to measure the polymorphism in terms of nucleotide diversity (π), which is the heterozygosity at nucleotide level (Nei 1987, Nei and Kumar 2000). This value is expressed as $2N_e\mu = \theta$ under the mutation and drift equilibrium, indicating 'genetic health' of a population (Fu and Li 1999). So far, studies of nucleotide variation of organelle DNA in out-crossing, long-lived plants at populational levels are limited (Matos and Schaal 2000), and are expected to have lower values of nucleotide diversity and nucleotide differentiation among populations (Dvornyk et al. 2002). However, I observed high nucleotide diversity in *P. tropicalis* (π_T = 0.00124 and π_S = 0.00026), which was also slightly higher than those values reported in *P. radiata* and *P. muricata* (Hong et al. 1993a) and *Pinus montezumae*-complex (Matos and Schaal 2000). Hong et al. (1993a) found that nucleotide diversities within populations of *P. radiata* as low as 0.002, and from 0.0011 to 0.0033 in *P. muricata*; while Matos and Schaal (2000) calculated the nucleotide diversities in the range of 0.0018 to 0.0025 in *P. hartwegii* and *P. montezumae*.

Genetic structure of the populations: Although the strong gene differentiation was observed in the *cp*DNA of *P. tropicalis*, this is not surprising for *Pinus*; for example, strong differentiation was found in *P. muricata* with G_{ST} = 0.87 (Hong et al. 1993a), in *P. resinosa* (θ = 0.56) (Walter and Epperson 2001). G_{ST} was 0.67 in *P. ponderosa* (Latta and Mitton 1999),

and G_{ST} = 0.20 in Douglas-fir (Hong et al. 1995). Nevertheless, lower values of differentiation have been reported among seven populations of *P. flexilis* (G_{ST} = 0.013) (Latta and Mitton 1997), and *P. attenuata* and *P. radiata* (G_{ST} = 0.000 - 0.011, respectively) (Hong et al. 1993a); *P. pinaster* (G_{ST} = 0.023 and 0.038) (Ribeiro et al. 2001, 2002); *P. resinosa* (G_{ST} = 0.12) (Echt et al. 1998), *P. albicaulis* (G_{ST} = 0.046) (Richardson et al. 2002). In contrast, the very little differentiation was detected in *P. banksiana* (G_{ST} = 0.020) and *P. contorta* (G_{ST} = 0.018) (Dong and Wagner 1994), and no genetic differentiation were found among populations of Canadian Douglas-fir (Viard et al. 2001) and *P. torreyana* (Waters and Schaal 1991). Therefore, genetic diversity and population structure vary considerably within and among the species, and the chief differences among pines involve geographic range, the degree of spatial isolation of populations, and the successional stage of their habitats.

Our analyses demonstrated that the genetic structure in the *cp*DNA variation of the tropical pine is strongly correlated with the geographical distribution of its haplotypes, which somehow characterized each region and populations. Templeton et al. (1995) considered three main factors that can cause a significant spatial/temporal association of haplotypes variations: the restricted gene flow, past fragmentation events and the range expansion events; whereas, Schaal et al. (1998) assured that geographical structure of the genetic variation in the population of plants is mainly the results of isolation and genetic drift, while gene flow between populations counteract differentiation. However, those factors are not mutually exclusive alternatives but they in minor or major degree under certain conditions create a geographical and/or genetic structure (Templeton 1998).

In the model of isolation by distance (Slatkin and Maddison 1989) restricted gene flow will prevent newly arisen mutations (i.e., haplotypes) to spread, and these mutations will resides as “private” alleles or haplotypes in the original populations. On the other hand, the ancestral haplotypes is older than its mutation offshoot, and thus, it must be in higher frequency and widerly distributed. This well explains the low frequency and limited distribution of the haplotypes II, III and VIII as well as its external position in the minimum-spanning network (Fig. 9A). Past fragmentation events will lead to the fixation of some distinctive haplotypes in each population, particularly because of the reduction in the effective

population size and the effect of the genetic drift. In our case, I believe that these factors have been involved in the current *cp*DNA distribution pattern. Past fragmentation events likely occurred during the Miocene-Pliocene (López 1982), and subsequently restricted gene flow among the distantly located populations. While, the range expansion events might have played a minor role in the present distribution of *P. tropicalis*, because no haplotypes were shared among geographically distant populations and the ancestral haplotypes reside in a restricted area.

Pollen is the dominant vector of gamete exchange for most temperate tree species, and determines the distribution of genotypes within and between populations (Ennos 1994). Moreover, in conifer there is usually a strong asymmetry between pollen and seed migration, and because wind-borne pollen typically moves farther than seeds. We would expect that pollen flow would be the major factor of population subdivision (Latta et al. 1999, Ribeiro et al. 2002, Ledig 1998) and restrictions in pollen movement can enhance the isolation-by-distance among populations (e.g., Turner et al. 1982). Additionally, extintion/recolonizations events may promote the creation of genetic structure in a pattern conforming to the isolation by distance (e.g., Slatkin and Maddison 1990, Whitlock and McCauley 1990).

Despite the small divergence within region shown by AMOVA, cluster analysis revealed a good relationship between populations and their geographical location. Strong genetic differentiation among regions in *P. tropicalis* would suggest that gene flow among different groups of individual must be very low. However, the G_{ST} obtained from the organelle genome might reflect long-term history of isolation, not ongoing processes; therefore, it should be regarded as mere descriptors of historical genetic structure along with other measures of the genetic diversity (Neigel 1997). With limited pollen flow and seed dispersal, the *cp*DNA genetic structure of *P. tropicalis* populations would follow the one-dimensional stepping-stone model (Kimura and Weiss 1964). Small marginal population would as a rule, maintain lower genetic variation owing to higher probabilities of genetic loss via stochastic processes. In *Pinus tropicalis*, however higher level of nucleotide and gene diversity was detected in the populations of the North-western in comparison with those observed for North-eastern. Such spatial distribution of genetic variation is thought to be associated with geological history

(Templeton 1998, Waters and Epperson 2001). The haplotype network provides insights into the migration history and distribution of the genetic variation in *P. tropicalis*. According to our results, haplotypes are tightly clustered in the network and concordantly with the geographical region without any missing intermediates haplotypes. This can be interpreted as isolation by distance rather that a past fragmentation events (Schaal et al. 1998). The presence of only two haplotypes in the north-western populations is also evocative of an ancient major bottleneck or that those populations could be originated from one ancestral haplotypes probably from Pilotos. Although at the present time the north-western populations are geographically distant, they might represent a full-size population at chloroplast levels until the Pleistocene. The contact among those populations can be maintained by the strong winds with preferential direction to southwest (Samek and Del Risco 1990) that can transport the pollen of *P. tropicalis* up to 10 km from the parental tree (Betacourt 1987). According to him, the pollen dispersal tends to be more mesokurtic in *P. tropicalis*, which favored the isolation-by-distance hypothesis, if the populations turn into a discontinuous distribution. Finally, the variation indicates strong genetic differentiation with a striking geographical pattern, and this pattern indicates that the spread of *P. tropicalis* is not only due to restricted gene flow (isolation-by-distance) but that the historical factors may also have played a major role on the current pattern.

*Historical factor*s: According with the palynological data (Areces 1987) and sequences data (Geada-López et al. 2002) the ancestor of the *P. tropicalis* entered into the Cuba during the Oligocene from the north of Pinar del Río. The orogeny of this area during the Oligocene-Early Miocene illustrated that Pinar del Río was divided into two emerged land: 1) the North-eastern including Bahia Honda unit (i.e., where lays Galalón) and small section of the current Central's Pinar del Río (i.e., where Pilotos is located), and 2) and the Alturas de Pizarra, but isolated from each other by the sea. Thus, the establishment of *P. tropicalis* in Alturas de Pizarra could not be possible until the Late Pliocene (Borhidi 1996, Areces 1987).

One possible scenario is that the founder population of *P. tropicalis*, could reside in the Central-North part of Pinar del Río, as suggested by Samek and Del Risco (1990), and that at present the Pilotos population may represent a relic population of the ancestral one (see

the minimum spanning network). Not only the haplotype network but also the higher haplotype diversity in Pilotos supported the idea of this population as a founder rather than the notion of being a population formed by the admixture of haplotypes of west and east. However, the only available information is reduced to the acknowledgement of morphologically diverged groups in the natural population of *P. tropicalis*, the north-eastern population (including Pilotos) and the North-western populations (López 1982). The ancestral population could have a wide range of distribution, expanded from the central to northeast (see Fig. 1; Samek and Del Risco 1990, pp13). According to that, La Jagua and Pilotos populations were continuous during Miocene till the late Pliocene. However, the further formation of Sierra del Rosario mountain range constituted a natural barrier, preventing homogenizing effect of gene flow, and allowed the genetic divergence between the two groups of populations, La Jagua in the north-eastern and Pilotos in the Central. Moreover, during Pleistocene with the invasion of *Pinus caribaea* and the complete emergence of Sierra de Rosario restricted further the exchange between populations east to west and imposed a strong bottleneck, thus leading to the fixation of different haplotypes in each population. For example, Haplotypes II and I present only in population Galalón, while Haplotypes VI, VII and VIII detected only in Pilotos.

Land connection of the Alturas de Pizarra and northwestern of Pinar del Río at the end of the Pliocene favored the migration and colorizations of new areas in the north-western. The current pattern of variation in the *cp*DNA in the north-western populations is a result of the expansion of *P. caribaea* to the low to middle elevations, which confined *P. tropicalis* to the most remotes and extremes sites, particularly during the Pleistocene. Certainly, habitat fragmentation and disturbance due to cleaning invasion of meadows grass altered fire regimen and anthropogenic activities during the last four centuries could also reduce plant population sizes and could lead to genetic isolation; however those effect were less pronouncing in long-lived species with a long generation time (Austerlitz 2000). Thus, the *cp*DNA well reflect the very ancient historical events (i.e., founder or bottleneck) rather than the actual events.

4.2. Genetic Diversity of *Pinus tropicalis* Morelet revealed by *cp*SSR and RAPDs

4.2.1. Introduction

Since genetic variation is a fundamental component of biodiversity, the assessment of genetic diversity within and among populations is central to single-species conservation strategies. Until more comprehensive information become available on the genes underlying variation of quantitative adaptive traits, neutral molecular markers remain as useful tools for the rapid decision making on conservation priorities (McKay and Latta 2002, Crandall et al. 2000). Nevertheless, diagnosing distinct populations and selecting priority candidates for conservation may require a more intensive sampling strategy of the genetic variation; and thus, particularly interesting areas subsequent studies should be performed with suitable high resolution markers such as microsatellites, RAPDs, ISSR, AFLP, etc.

So far, in the previous epigraph I examined the *cp*DNA sequence variation of *P. tropicalis* describing the historical factors involved in its geographical-genetic structured pattern. However, there is no information about the genetic diversity at nuclear level which is also an important step for conservation strategies. In addition, sequencing limited chloroplast regions seemed to bias the estimation of the overall variation in the plastid genome, and more extent survey is advisable, particularly with *cp*SSR.

Chloroplast microsatellites (*cp*SSR) have demonstrated high levels of intraspecific variability and represent useful markers for population genetic analysis in small scales (Vendramin et al. 1998 and 1999, King and Ferris 1998, Provan et al. 2001). The *cp*SSR have also provided inferences about the recent evolutionary history in some tree species, especially about the postglacial migration routes and the locations of refugia (Newton et al. 1999), and more recently, to examine the phylogeographic structure (Marshall et al. 2002). Random amplified polymorphic DNA (RAPD) (Williams et al. 1990) has been widely used in population genetic studies of a large number of plant species, including pines (Szmidt et al. 1996, Bucci et al. 1997, Thomas et al. 1999, Wu et al. 1999). One of the major drawbacks of RAPDs is their sensitivity to reaction conditions, which require careful optimization (Weising

et al. 1995). RAPD have a potential limitation: their dominant nature precludes the direct estimations of allele frequency and can bias calculations of genetic diversity and population differentiation (Lynch and Milligan 1994, Isabel et al. 1995, 1999). However, the use of appropriate statistical methods has allowed improved evaluation of intraspecific diversity. The analysis of molecular variance (AMOVA) (Excoffier et al. 1992), which is not influenced by the dominance problem, has thus become an important tool for investigating the partitioning of genetic variation in dominant markers. AMOVA has been shown to give the most accurate estimate of population differentiation when conifer genotypic and phenotypic RAPD data are compared (Isabel et al. 1995, 1999), and the diversity estimates can also be obtained based on phenotype frequency using Nei's unbiased statistics (Nei 1978). In this study, RAPDs and *cp*SSR were used to assess the genetic diversity and population structure of natural populations of *P. tropicalis* in Cuba. This contributes towards the definition of genetically distinct units needed for conservation purposes.

4.2.2. Material and Methods

Study sites and sampling: I examined in this study five of the seven populations previously surveyed in the epigraph 4.1., representing the three geographical regions: one population in northeast Galalón; one populations in central, namely Pilotos; and three populations in the northwest, namely La Jagua, Bartolo and Mina Dora (Fig. 8; Table 3).

CpSSR markers: I used five chloroplast mononucleotide SSR primers (Pt36480, Pt71936, Pt10748, Pt109567 and Pt110048) designed by Vendramin et al. (1996). PCR amplifications were carried out in a total volume of 10µL containing 5ng of template DNA, 10 mM Tris-HCl, pH 8.3, with 1.5 mM $MgCl_2$, 2 mM each dNTP, 0.5 µM dUTP [R110; PRISM dUTP SET (Perkin Elmer)]; 5 pM of each primer, and 0.4 units of *Taq* polymerase (Promega). PCRs were performed a touchdown amplification protocol: two cycles with a denaturing step at 94°C for 1 min, an annealing step at 65°C for 1 min, and an extension step at 70°C for 35 s; followed by 18 cycles with denaturation at 93°C for 45 s, primer annealing at 64°C for 45 s (progressive decrease of the temperature by 0.5°C every cycle until 55.5°C was reached),

and primer extension at 70°C for 45 s. The last 20 cycles were with denaturation at 92°C for 30 s, primer annealing at 55°C for 30 s, primer extension at 70°C for 60 s, and a final extension cycle at 70°C for 5 min. The respective amplification products incubated at 95ºC for 3 min were separated using an ABI 310 Genetic Analyzer (Perkin Elmer). The data were collected using Genetic Analyzer Collection Software, and analyzed using GeneScan® 3.1 software. The fragment size was determined using GeneScan-400 [ROX] size standards loaded with the amplification products. Of the five primers, three revealed polymorphism in our samples (Pt71936, Pt109567 and Pt110048). For these, diversity measures and *F*-statistics were computed using the Lewis and Zaykin (2001) GDA program (version 1.1), calculated directly using haplotype frequencies. It is also worth noting that some special diversity statistics designed to exploit differences in the sizes (i.e., numbers of repeats) of alleles (e.g., R_{ST}, Goldstein et al. 1995) are inappropriate for our data, because loci had either only two alleles or at most three alleles with at least one of those in very low frequency. Thus, there cannot be substantial additional information to be exploited, and hence it is not helpful or necessary to assume a SMM (Stepwise Mutation Model) or TPM (Two-Phase Model), in which unknown violations may mislead. Following diversity statistics for the haplotypes in each population were calculated: number of haplotypes and the gene diversity *H* (Weir 1996). I also calculated genetic distances between pairs of populations using the measures of Nei's (1978), and Weir's (1996) θ instead of the theoretical parameter F_{ST}.

RAPD markers: RAPD reactions were carried out essentially as described by Williams et al. (1990) in 12.5µl total volumes containing, 10 mM Tris-HCl, pH 8.3 with 1.5 mM $MgCl_2$, 2.5mM dNTPs, 0.5mM of primer, and 1 U of *Taq* polymerase (Sigma), 10ng DNA. Amplification was initiated by denaturation at 94° C for 3 min, followed by 40 cycles of 94° C for 1 min, 36° C for 3 min and 72 ° C for 2 min, and a final extension step of 5 min at 72 ° C. Amplification products were separated by electrophoresis in 1.5% agarose gels. Gels were stained with 0.5 mg/ml ethidium bromide for 30 min and then photographed with a camera under UV light. A total of 42 primers from Operon Technologies, Inc. (USA) were screened using three individuals randomly chosen from the 67 surveyed but ten primers (OPC07, OPB19, OPB18, OPB03, OPB16, OPA10, OPB20, OPA12, OPC20, and OPB13) gave high

intensity, easily scorable and reproducible bands. The amplification fragments were manually scored as present (1) or absent (0). Some unclear fragments were ignored. Genetic diversity was assessed within each population and in the total population based on the proportion of polymorphic bands and average heterozygosity using program POPGENE version 1.32. Yeh et al. (1999). The relative magnitude of genetic differentiation among populations (G_{ST}) was calculated for each locus and then averaged over all loci. The number of nucleotide substitutions within populations was estimated, following the method of Clark and Lanigan (1993). The average values of pairwise distances (nucleotide diversity) within each population (π_S) and for the entire population (π_T) were calculated. The inter-populational nucleotide diversity was compared to total nucleotide diversity to give $N_{ST} = \pi_T - \pi_S / \pi_T$ (Nei and Kumar 2000). The UPGMA tree was constructed based on the pairwise distance using POPGENE (version 1.32). AMOVA (Excoffier et al. 1992) was used to estimate variance components for RAPD, partitioning the variation into Φ_{CT} (among regions; i.e., between Northeastern- Central and North-western), Φ_{SC} (among populations within regions), and Φ_{ST} (among populations). Variance components were tested statistically by nonparametric permutational procedures using 1000 permutations. All analyses were performed using ARLEQUIN version 2.000 (Schneider et al. 2000). The degree of relatedness between the genetic distance matrix generated by the two types of markers (*cp*SSR and RAPD) was determined with Mantel-test.

The contribution of genetic diversity in each population to species levels and genetic uniqueness of a specific population (X) was determined using ‘‘jackknife’’ analysis (Jaquish and El-Kassaby 1998), firstly, removing population X_1 data from the original data set (i.e., - X_1); secondly, estimating the average of G_{ST} and *I* (genetic identity *I*; Nei 1978) for the new data set, and thirdly comparing the estimates obtained from the original analysis to that of the six new data sets (i.e., - X_1, - X_2, - X_3, - X_4 and - X_5).

4.2.3. Results

CpSSR markers: Three (for primer pairs Pt71936, Pt109567, Pt110048) of the five SSR loci were polymorphic in our samples, having either two or three alleles. The relative sizes of fragments scored for the two monomorphic loci are 142 and 123-bp for Pt36480 and Pt107148, respectively. A total of five haplotypes was found, of a theoretical maximum of 18 (3×3×2) based on the numbers of alleles (Table 7). One of these, haplotype II, is particularly common, having a total frequency of 0.53 for the combined samples and predominant in Pilotos. All other haplotypes differ from haplotype II by a difference at a single locus, and all except two (III and V) were caused by a single base pair change. Hence, haplotype IV is clearly the consensus 'sequence' it is also the most common. Haplotype IV, is also the second most frequent (0.26), occurred in all populations except in Galalón (North-eastern). Haplotype I (0.16) was found in Pilotos and Galalón but noting that in the latter population was the predominant, while Haplotype III was a 'private' haplotype in Galalón with a frequency of 0.01. Although genetic diversity was high with a value of H_T = 0.6224 and H_S= 0.5156, a very low value was obtained in Galalón compared with the others populations (Table 8). The overall value θ (Weir 1996), which is an analogous to G_{ST} under the 'island model', was 0.18.

RAPD markers: Ten of the 42 primers tested produced a total of 51 distinct bands. Populations were similarly polymorphic for the whole set of bands scored. No band was fixed exclusively in a single population. All populations possess similar values of polymorphisms, and 32 of the 51 bands were polymorphic in the total data (Table 9). The values of gene diversity at gene and at nucleotide level were very similar for each population (Table 9). Most of the total genetic diversity (H_T = 0.215 and π_T = 0.007032) was partitioned within population (H_S = 0.166 and π_S = 0.004357). The proportion of genetic variation distributed among populations (G_{ST} = 0.23 and N_{ST} = 0.39), hence the 23 and 39% of the genetic variability was distributed among populations (Table 8). Nei's (1978) genetic identities (*I*) among pairs of populations for all loci were high and varied between 0.88 (Galalón) and 0.96 (Bartolo). The UPGMA tree based on the genetic distance showed that Galalón was "outliner" and the other

populations clustered together and in which Pilotos occupied the basal position (Fig. 11). However, the genetic distances between the populations are very small. Over all populations and loci, the number of migrants (*N*m) was 1.66, indicating high levels of gene flow. It should be noted, however, that this value of *N*m represents historical average levels of gene flow and may not represent present-day levels.

AMOVA estimates of genetic diversity within populations, between populations, within regions and between regions (Table 5) and showed that most of the total variation existed within populations (87%). Despite this, highly significant genetic structuring in *P. tropicalis* was detected among populations (Φ_{ST} = 0.209; *P < 0.01*), and a small but significant proportion of the total variation (2%) was attributed to differences between regions (Φ_{CT} = 0.171; *P < 0.05*), while the populational differences within regions were not significant (Φ_{SC} = 0.155; *P > 0.05*). This indicates that individuals of each population were more similar to their co-members than to individuals of other populations. Mostly, pairwise Φ_{ST} values revealed a larger separation between populations from different regions than any pair of populations from the same region.

As shown in Table 10, the lower value of genetic diversities among populations (G_{ST}) was observed for the analysis without Galalón and without Galalón and Pilotos; this value was even lower than that obtained from the analysis of the five populations' altogether (0.23), as well as for the remaining four analyses. The values of the gene diversity H_T and H_S did not varied much when were excluded successively the populations indicating that each population retain a similar genetic variation and identity. However, those values slightly dropped when La Jagua and Pilotos were excluded from the analysis in contraposition with the gene differentiation.

4.2.4. Discussion

Genetic diversity: Substantial variation for *cp*SSR loci was observed in the populations of *P. tropicalis*. However, the amount of variation is much lower than that for *cp*SSR loci in other pine species (see Table 11 and Powell et al. 1995a). *Cp*SSR data variation clearly indicates strong genetic differentiation with a striking geographical pattern and this can be

attributed to the evolutionary history of the species (see section 4.1. Phytogeography).

The genetic variation estimated from RAPD data is comparable to those obtained in others pines species (see Table 11) that have either restricted or widespread distribution. One overgeneralization on the population genetics is the fact that rare-endemic species has lower values of genetic diversities compared with the widespread congeners; however, the results highlight the view of Gitzendanner and Soltis (2000) that the genetic data for rare-endemic species are more informative when it is compared with their widespread ones. Therefore, some other attributes of the species maintain different levels and patterns of genetic diversities. The genetic diversity maintained within and among populations is a function of historical events and/or recent evolutionary processes. Because very little is usually known for a species' evolutionary and ecological history, explanations for the levels and patterns of genetic diversity found within and among populations rely primarily on inferences. High levels of genetic diversity for *P. tropicalis* might be attributable to the species' life history and ecological traits, such as its common, long-lived, woody, wide ranges, wind-pollinated, wind-dispersed and outcrossing species. Larger values of population differentiation in the present study ($G_{ST} = 0.209$) is comparable to the reported for other pines species (see Table 10).

*Cp*SSR *versus RAPD*: There are a limited number of studies comparing *cp*SSR estimates with any other markers (Latta and Mitton 1997). In this study, *cp*SSR markers showed a higher level of gene diversity than RAPD markers. Microsatellites are known to be highly variable within-population (Lefort et al. 1999, and references therein). Hedrick (1999) suggested that is preferable to use highly variable markers such as microsatellites, because the information they provide can be quite different to that provided by less variable markers. Although both types of marker have different properties and reveal different values of gene diversity, the trend in the genetic variation among populations was the same. Furthermore, the Mantel test showed that the genetic distance matrix calculated from the RAPD was significantly correlated with the *cp*SSR-based matrix ($r = 0.8$; $P = 0.46$), and I can infer that the gene flow through pollen within groups was more important than the marker-specific factors.

Concerning differentiation, the RAPD analysis exhibited similar levels of differentiation compared with the *cp*SSRs for all the populations studied. In general, population subdivision is expected to be lower for nuclear markers than for cytoplasmic markers, but several possible explanations can be invoked the results. First, one possible explanation is connected with the higher mutation rates of the *cp*SSR markers. According to Hedrick (1999) the magnitude of G_{ST} for isolated populations is strongly influenced by the amount of variation determined by the mutation rate. Second, it concerns the nature of the RAPD markers in many conifers; that is intensely scored bands are strongly associated with *cp*DNA regions (De Verno and Mosseler 1997, Wu et al. 1999, Nkongolo et al. 2002) or *mt*DNA regions (Aagard et al. 1998, Wu et al 1998; Tsumura personal communication) and therefore inflate the values of the genetic diversities and differentiation among populations. Thirdly, extensive gene flow via pollen could explain the similar among-population differentiation found using nuclear and chloroplast markers by smoothing the differences due to variations in effective population sizes and genetic drift, but this seems unlikely in our case due to the highly genetic and geographic structured in *P. tropicalis*. Fourthly, *cp*SSRs are generated by mutations at a limited number of hotspots, so they are prone to undergo identical mutations independently in different populations (see section 4.1., Doyle et al. 1998), which leads to underestimates of differentiation. Besides, some size homoplasies which have been observed in *cp*SSR can also biased the estimates (Doyle et al. 1998).

On the other hand, another factor can also inflates the estimates of the genetic diversity based on RAPD in diploid tissues is that only monomorphic loci for dominant alleles (presence band) can be observed, while the monomorphic for recessive (absence band) cannot be observed, and this will result in the overestimation of diversity. This may be the explanation for higher variability for RAPD in several studies (Aagaard et al. 1998, Wu et al. 1999), which can only be overcome if the RAPD are performed from haploid tissue (Isabel et al. 1995, Szmidt et al. 1996). Regardless of those facts, RAPD was shown to be very sensitive marker for detecting variation because of their characteristics: RAPDs result primarily from amplification of random regions of the nuclear genome. Given the fact of large size of *Pinus* genomes, ranging from 19.5 to 26.5 pg (O'Brien et al. 1996), RAPDs would

mostly reflect random amplification of non-coding repetitive DNA which are subject to weaker selective constraints (i.e., being a neutral marker).

In pines, population subdivision may be less pronounced when *cp*DNA markers are used compared with *mt*DNA markers, because wind-dispersed pollen is the main agent of gene flow (e.g., Dong and Wagner 1994, Mitton et al. 2000). Latta and Mitton (1997) observed that the population differentiation in *P. flexilis* was much higher for *mt*DNA than for *cp*DNA and RAPD, which, in turn, revealed low and similar levels of population differentiation. Moreover, in a study with *P. sylvestris* in Finland (Karhu et al. 2001), using allozymes, RAPDs and nuclear microsatellites, all markers showed very little differentiation, reflecting the high level of gene flow in the studied area. On the other hand, chloroplasts are paternally inherited in *Pinus* (Dong and Wagner 1994) and thus migrate through pollen and seed. Under island model and the isolation by distance model, a greater differentiation for uniparentally markers is expected compared with the biparental (nuclear markers) (Ennos 1994, Hu and Ennos 1997). For example, the inferred value of genetic differentiation for maternal inheritance marker (*mt*DNA) in *P. tropicalis* was estimated to be $G_{ST} = 0.45$ (see Ennos 1994). Because there is usually a strong asymmetry between pollen and seed flow in conifers, one would expect that pollen flow would be the major factor of population subdivision. Therefore, congruent results with the two types of markers were observed. Moreover, the spatial isolation of the two groups (northeastern and the central-northwestern), with the Sierra Cajálbana system constituting a natural barrier, probably prevented the possibility of the genetic homogenizing effect of gene flow, and allowed the genetic divergence between the two groups of populations due to isolation-by-distance.

Conservation strategies: The comparative analysis of the genetic diversity with different markers has provided interesting insights not only into the organization and subdivision of diversity but also into the genetic mechanism that contributed to the extant structure of the diversity. Moreover, for a single-species conservation planning, it is important to identify the component of evolutionary lineages in order to retain maximum genetic diversity and to incorporate information of historical processes (Moritz 1994, 1995). According to Moritz (1994) populations that are reciprocally monophyletic for *mt*DNA or *cp*DNA alleles but show

significant divergence in allele frequencies at nuclear loci should be declared as ESUs (Evolutionary Significant Units) to conserve. Whereas, the populations with significant divergence of allele frequencies ad with a high levels of diversity at nuclear or *cp*DNA, regardless of the phylogenetic distinctiveness of alleles should be named as Management Units (MUs), and being the nucleus in the *ex situ* conservation programs (Moritz 1994, Newton et al. 1999). Therefore, one of the major sampling strategies for conservation of the endemic such as *P. tropicalis* is to identify the appropriate ESUs as well as its geographical placement. Sampling guidelines for conservation in the last 15 years have two main directions: 1) to concentrate efforts for preserving small number of population that would represent the total diversity of the species (Hamrick 1993, Petit et al. 1998); 2) to promote the conservation of population which contains uniqueness in characteristics (Davidson and El-Kassaby 1997, Moritz 1994). The results presented here, clearly indicate that strategies that integrate both tendencies are necessary in *P. tropicalis*.

Because of the characteristics the tropical pine, if few populations were strategically placed only on the basis of patterns of genetic variation, much of the variability among population could be lost. The low G_{ST} and high *I* obtained without Galalón analysis may indicate that this population harbors some unique genetic characters and they should receive additional attention for conservation purposes. Additionally, I propose that Pilotos should be incorporate to the conservation plans as Management Units (MUs) because this population seems to be the ancestral of *P. tropicalis* retaining a large genetic variation at *cp*DNA and at nuclear genomes. To conclude with the conservation issue, I would also recommend complementing this study with parallel investigations on adaptive traits and the economic traits before definite strategy will be adopted.

5. References

Aagaard JE, K V Krutovskii, SH Strauss 1998 RAPD markers of mitochondrial origin exhibit lower population diversity and higher differentiation than RAPDs of nuclear origin in Douglas-fir. Mol Ecol **7:** 801-812.

Alvin KL 1960 Further conifers of the Pinaceae from the Wealden Formation of Belgium. Mem. Inst Rey Sci Nat Belgique **146:** 1-39.

Areces A 1987 Consideraciones sobre la supuesta presencia de *Pinus sylvestris* L. en el Oligoceno de Cuba. Publicación Centro Investigaciones del Petróleo. Serie Geológica **2:** 27-40.

Austerlitz F, S Mariette, N Machon, P-H Gouyon, B Godelle 2000 Effects of colonization processes on genetic diversity: differences between annual plants and tree species.Genetics **154:**1309-1321.

Avise JC 1994 Molecular markers, natural history and evolution. New York Chapman & Hall.

Axelrod DI 1986 Cenozoic history of some western American pines. Ann Missouri Bot Gard **73:** 565-641.

Axelrod DI, J Cota 1993 A further contribution to closed-cone pine (*Oocarpae*) history. Amer J Bot **80:** 743-751.

Barret WH, GL Golfari 1962 Descripción de 2 nuevas variedades de pino del caribe (*Pinus Caribaea* Morelet). The Caribbean forester **23:** 59-71.

Bayer R J, CF Puttock, SA Kelchner 2000 Phylogeny of South African Gnaphalieae (Asteraceae) based on two noncoding chloroplast sequences. Amer J Bot **87:** 259-272.

Bernardi G, G Bernardi 1986 Compositional constrains and genome evolution. J Mol Evol **24:** 1-11.

Berry EW 1934 Pleistocene plants of Cuba. Torrey Bot. Club Bull **61:** 237-240.

Betancourt A 1987 *Silvicultura especial de árboles maderables tropicales*. Científico-Técnico. Havana. Cuba. pp 456.

Bisse J 1975 Nuevos árboles de la flora de Cuba. Ciencias Serie 10, Bot **2:** 1-23.

Borhidi A 1996 *Phytogeography and Vegetation Ecology of Cuba*. eds Akademiai Kiado.

Budapest. pp. 752.

Bucci G, GG Vendramin, L Lelli, F Vicario 1997 Assessing the genetic divergence of *Pinus leucodermis* Ant. Endangered populations: use of molecular markers for conservation purposes. Theor Appl Genet **95:** 1138–1146.

Bucci G, M Anzidei, A Magaghiele, GG Vendramin 1998 Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. Mol Ecol **7:** 1633-1643.

Clark AG, MS Flannigan 1993 Prospects for estimating nucleotide divergence with RAPDs. Mol Biol Evol **10:** 1096-1111

Clegg MT, BS Gaut, LG Jr Gerald, BR Morton 1994 Rates and pattern of chloroplast DNA evolution. Proc Natl Acad Sci USA **91:** 6795-6801.

Crandall KA, ORP Bininda-Emonds, GM Mace, RK Wayne 2000 Considering evolutionary processes in conservation biology. Trends Ecol Evol **15:** 290-295.

Critchfield WB 1962 Hybridization of the southern pines in California. Proc For Genet Workshop, pp. 25-27, Macon, G.A.

Darrow WK, T Zanoni 1991 Hispaniolan pine (*Pinus occidentalis* Swartz) a little known subtropical pine of economic potential. Commow Forest Rev **69:** 133-146.

Davidson R, YA El-Kassaby 1997 Genetic diversity and gene conservation of pacific silver fir (*Abies amabilis*) on Vancouver Island, British Columbia. Forest Genetics **4:** 85-98.

Desplanque B, F Viard, D Forcioli, J Bernard, P Saumitou-Laprade, J Cuguen, H Van Dijk 2000 The linkage disequilibrium between *cpDNA* and mtDNA haplotypes in *Beta vulgaris* subsp *maritima* (L.): the usefulness of both genomes for population genetic studies. Mol Ecol **9:** 141-154.

De Verno LL, A Mosseler 1997 Genetic variation in red pine (*Pinus resinosa*) revealed by RAPD and RAPD-RFLP analysis. Can J For Res **27:** 1316–1320.

Diaz V, M Muñiz, E Ferrer 2001 Random amplified polymorphic DNA and amplified fragment length polymorphism assessment of genetic variation in Nicaragua populations of *Pinus oocarpa*. Mol Ecol **10:** 2593-2603.

Dong J, DB Wagner 1994 Paternally inherited chloroplast polymorphism in *Pinus*: estimation

of diversity and population subdivision, and test of disequilibrium with a maternally inherited mitochondrial polymorphism. Genetics **136:** 1187-1194.

Dorman KW 1976 The genetics and breeding of Southern Pines. USDA For. Serv Agric Handb No. 471. 407p.

Doyle J, L Doyle 1990 Isolation of plant DNA from fresh tissues. Focus **12:** 13-15.

Doyle JJ, M Morgante, SV Tingey, W Powell 1998 Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). Mol Biol Evol **15:** 215–218.

Dvorak WS, AP Jordon, GP Hodge, JL Romero 2000 Assessing evolutionary relationships of pine in the *Oocarpae* and *Australes* subsections using RAPD markers. New Forests **20:** 163-192.

Dvornyk V, A Sirvio, M Mikkonen, O Savolainen 2002 Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol **19:** 179-188.

Echt CS, LL DeVerno, M Andezi, GG Vendramin 1998 Chloroplast microsatellite reveal population genetics diversity in red pine, *Pinus resinosa* Ait. Mol Ecol **7:** 307-316.

Ennos RA 1994 Estimating the relative rate of pollen and seed migration among plant populations. Heredity **80:** 584-593.

Excoffier L, PE Smouse, JM Quattro 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:** 479-491.

Farjon A 1984 Pines: Drawings and Descriptions of the Genus. E. J. Brill and W. Backhuys. Leiden. Netherlands.

Farjon A 1996 Biodiversity of *Pinus* (Pinaceae) in Mexico: Speciation and paleo-endemism. Bot J Linn Soc **121:** 365-385.

Farjon A, BT Styles 1997 Pinus. Flora Neotropica Monograph 70, New York Botanical Garden, NY.

Felsenstein J 1978 Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool **27:** 401-410.

Felsenstein J 1985 Confidence limits on phylogenies: an approach using the bootstrap.

Evolution **39:** 783-791.

Fors AJ 1947 El pino macho en las lomas de Trinidad, Cuba. Caribbean Forest. **8:** 125-129.

Frankham R 1997. Do island populations have less genetic variation than the mainland populations? Heredity **78:** 311-327.

Frankis MP 1993 Morphology and affinities of *Pinus brutia*. Pages 11-18 *in*: "International Symposium of *Pinus brutia* Ten." (Secretariat of the Organizing Committee, Ed), Ministry of Forestry, Ankara.

Fransisco-Ortega JA, DJ Crawford, A Santos-Guerra, RK Jansen 1997 Origin and evolution of *Argyranthemum* (Asteraceae: Anthemideae) in Micronesia. Page 407-432 *in* TJ Givnish and KL Systma eds. Molecular evolution and adaptive. Cambridge University Press, Cambridge UK.

Fu YX, WH Li 1999 Coalescing into the 21st century: An overview and prospects of coalescent theory. Theor Popul Biol **56:** 1-10.

Gaussen H, VH Heywood, AO Chater 1993 *Pinus*. Page 40-44. *in:* Flora Europaea, Vol. 1.ed. TG. Tutin, N.A. Burges, AO Chater et al. Cambridge Univ. Press, Cambridge, UK.

Geada LG, K Kamiya, K Harada 2002 Phylogenetic relationships of *Diploxylon* pines (subgenus *Pinus*) based on plastid sequence data. Int J Plant Sci **163:** 737-747.

Gielly L, P Taberlet 1994 The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus *rbc*L sequences. Mol Biol Evol **11:** 769-777.

Gitzendanner M, PS Soltis 2000 Patterns of genetic variation in rare and widespread plant congeners. Amer J Bot **87:** 783–792.

Godt MJW, JL Hamrick 1998 Low allozyme diversity in *Schwalbea americana* (Scrophulariaceae), an endangered plant species. J Heredity **89:** 89-93.

Godt MJW, T Race, JL Hamrick 1997 A population genetic analysis of *Ziziphus celata*, an endangered Florida shrub. J Heredity **88:** 531-533.

Goldstein DB, AR Linares, LL Cavalli-Sforza, MW Feldman 1995 An evaluation of genetic distances for use with microsatellite loci. Genetics **139:** 463–471.

Gómez A, E Aguiriano, MA Bueno, R Alía 2000 Microsatélites del cloroplasto en *Pinus pinea* L. Page 39-46 *in* I Simposio del Pino Piñonero, Actas Tomo II. Junta de Castilla y León,

Valladolid, Spain.

Govindaraju D, P Lewis, C Cullis 1992 Phylogenetic analysis of pines using ribosomal DNA restriction fragment length polymorphisms. Plant Syst Evol **179:** 141-153.

Grivet D, RJ Petit 2002 Phytogeography of the common ivy (*Hereda sp.)* in Europe: genetic differentiation through space and time. Mol Ecol **11:** 1351-1362.

Hamrick JL 1993 Genetic diversity and conservation in tropical forest. Page 1-9 *in* RM Drysdale, SET John, and ACYapa eds. Proceedings of the ASEAN-Canada symposium on genetic conservation and production of tropical tree seed, ASEAN-Canada Forest Tree Seed Center, Muaklek, Saraburi, Thailand.

Hamrick JL, MJ Godt 1989 Allozyme diversity in plant species. Page 43-63 *in*, AHD Brown, MT Clegg, AL Kahler, BS Weir. eds Plant Population Genetics, Breeding and Germplasm Resources. Sinauer, Sunderland, Mass.

Hamrick JL, MJ Godt 1996 Effects of life history traits on genetic diversity in plant species. Phil Trans Roy Soc London Biol Sci **351:** 1291-1298.

Hamrick JL, MJ Godt, SL Sherman-Broylers 1992 Factors influencing levels of genetic diversity in woody plant species. New Forests **6:** 95-124.

Hartl DL 2000 *A Primer of Population Genetic* 3rd. Sinauer Associates, New York, USA.

Haubold B, T Wiehe 2001 Statistics of divergence time. Mol Biol Evol **18:** 1157-1160.

Hedrick PW 1999 Highly variable loci and their interpretation in evolution and conservation. Evolution **53:** 313-318.

Hilu KW, H Liang 1997 The *mat*K gene: sequence variation and application in plant systematics. Amer J Bot **84:** 830-839.

Hong YP, VD Hipkins, SH Strauss 1993a Chloroplast DNA diversity among trees, populations and species in the Californian closed-cone pines (*Pinus radiata*, *P. muricata* and *P. attenuata*). Genetics **135:** 1187-1196.

Hong Y-P, AB Krupkin, SH Strauss 1993b Chloroplast DNA transgresses species boundaries and evolves at variable rates in the Californian closed-cone pines (*Pinus radiata, P. muricata,* and *P. attenuata*). Mol Phylogenet Evol **2:** 322-329.

Hong YP, B Ponoy, JE Carlson 1995 Genetic diversity and phylogeny in Douglas-fir based on

RFLP and RAPD (DAF) analysis of nuclear, chloroplast and mitochondrial genomes. Page 276-266 *in* Population Genetics and Genetic Conservation of Forest Trees, ed PH Baradat, WT Adams,G Muller-Starck. SPB Academic Publishing, Amsterdam.

Hu XS, RA Ennos 1997 On estimation of the ratio of pollen to seed flow among plant populations. Heredity **79:** 541-552.

Hudson RH 1960 The anatomy of the genus *Pinus* in relation to its classification. J Inst Woody Sci **6:** 26-46.

Hudson RR 1990 Gene genealogies and the coalescent process. Oxford Surv Evol Biol **7:** 1-44.

Isabel N, J Beaulieu, J Bousquet 1995 Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. Proc Natl Acad Sci USA **92:** 6369–6373.

Isabel N, J Beaulieu, P Thériault, J Bousquet 1999 Direct evidence for biased gene diversity estimates from dominant Random Amplified Polymorphic DNA (RAPD) fingerprints. Mol Ecol **8:** 477-483.

Jalas J, J Suominen 1973 Atlas *Florae Europaea, 2 Gynospermae*. Cambridge Univ. Press, Cambridge, UK.

Jaquish B, YA El-Kassaby 1998 Genetic variation of western larch in British Columbia and its conservation. J Heredity **89:** 248–253.

Jukes TH, CR Cantor 1969 Evolution of proteins molecules. Pages 21-132 *in* HN Munro, ed. Mammalian Proteins Metabolism. Academic Press, New York.

Kajita T, K Kamiya, K Nakamura, H Tachida, R Wickneswari, Y Tsumura, H Yoshimaru, T Yamazaki 1998 Molecular phylogeny of Dipterocarpaceae in Southeast Asia based on nucleotide sequences of *mat*K, *trn*L intron, and *trn*L-*trn*F intergenic spacer region in chloroplast DNA. Mol Phylogenet Evol **10:** 202-209.

Kamiya K, K Harada, MC Mahani, AM Latiff 2002 Genetic variation of *Trigonobalanus verticillata*, a primitive species of Fagaceae, in Malaysia revealed by chloroplast sequences and AFLP markers. Gen Genet Syst **77:** 177-186.

Karhu A 2001 Evolution and application of pine microsatellite. Ph.D dissertation. University of

Oulu. Finland. pp52.

Kellogg EA, ND Juliano 1997 The structure and function of Rubisco and their implications for systematics studies. Amer J Bot **84:** 413-428.

Kerr AC, MA Iturralde-Vinent, AD Saunders, TL Babbs, J Tarney 1999 A new plate tectonic model of the Caribbean: Implication from a geochemical reconnaissance of Cuba Mesozoic volcanic rock. Geol Soc Amer Bull **111:** 1581-1599.

Kim J 1996 General inconsistency conditions for Maximum Parsimony: Effects of branch lengths and increasing numbers of taxa. Syst Biol **45:** 363-374.

Kimura M 1980 A simple method for estimating evolutionary rates of base substitutions. J Mol Evol **16:** 111-120.

Kimura M 1983 *The Neutral Theory of Molecular Evolution*. Cambridge Univ. Press, Cambridge, UK.

Kimura M, GH Weiss 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49:** 313-326.

King RA, C Ferris 1998 Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. Mol Ecol **7:** 1151–1161

Klaus W 1989 The Mediterranean pines and their history. Plant Syst Evol **162:** 133-163.

Kormutak A 1984 Some cytological and biochemical aspects of interspecific incompatibility in pines (*Pinus* sp). Page 1-92 *in:* Acta Dendrologica.

Korol L, G Shalar, G Schiller 2002 Genetic variation within *Pinus halepensis* Mill. Provenances growing in different environments in Israel. *in:* “Proceedings of the Symposium of Population and Evolutionary Genetics of Forest Tree”. Stará Lesná, Slovakia. August 25-29.

Kremenetski CV, K Liu, GM MacDonald 1998 The late Quaternary dynamics of pines in northern Asia. Page 95-106. *in*: Richardson DM ed. Ecology and Biogeography of *Pinus*. Cambridge University Press, UK.

Krupkin AB, A Liston, SH Strauss 1996 Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus,* Pinaceae) from chloroplast DNA restriction site analysis. Amer J Bot **83:** 489-498.

Kumar S, K Tamura, IB Jakobsen, M Nei 2001 Mega2.1: Molecular Evolutionary Genetics Analysis software. Bioinformatics **17:** 1244-1245.

Latta RG, JB Mitton 1997 A comparison of population differentiation across four classes of gene markers in limber pine (*Pinus flexilis* James). Genetics **146:**1153-1163.

Latta RG, JB Mitton 1999 Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. Evolution **53:** 769-776.

Latta RG, YB Linhart, D Fleck, M Elliot 1999 Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. Evolution **52:** 61-67.

Lee SW, FT Ledig, DR Johnson 2002 Genetic at allozymes and RAPD markers in *Pinus longaeva* (Pinaceae) of the White Mountains, California. Amer J Bot **89:** 566-577.

Lee SW, YY Kim, JO Hyun, ZS Zim 1997 Comparison of genetic variation in *Pinus densiflora* population by allozyme and RAPD analysis. Korean Journal of Breeding 29:72-83 [in Korean with English summary].

Lefort F, C Echt, R Streiff, GG Vendramin 1999 Microsatellite sequences: a new generation of molecular markers for forest genetics. Forest Genetics. **6:** 15–20.

Legid TF 1998 Genetic variation in *Pinus*. Pages 251-280 *in* DM Richardson, ed. Ecology and Biogeography of *Pinus*. Cambridge Univ. Press, Cambridge, UK.

LePage BA, JF Basinger 1995 Evolutionary history of the genus *Pseudolarix* Gordon (Pinaceae). Int J Plant Sci **159:** 367-390.

Lewis JF, G Draper 1990 Geology and tectonic evolution of northern Carbbean margin. Page 77-133 *in*: Dengo G, JE Case eds The Caribbean region: Boulder, Colorado, Geological Society of America. The Geology of North America. v.H.

Lewis PO, D Zaykin 2001 Genetic Data Analysis: Computer Program for the Analysis of Allelic Data, Version 1.0 [d16c]. Available at http://lewis.eeb.uconn.edu/lewishome/software.html

Li P, J Bosquest 1992 Relative-rate test for nucleotide substitutions between two lineages. Mol Biol Evol **9:** 1185-1189.

Li W-H 1993 So, what about the molecular clock hypothesis? Curr Opin Genet Dev **3:** 896-901.

Li WH 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, Massachusetts, USA.

Linnaeus C 1753 Species Plantarum. edn 1. Stockholm: Impensis Laurentii Salvii.

Liston A 1997 Biogeographic relationships between the Mediterranean and North American flora: Insights from molecular data. *Lagascalia* **19:** 323-330.

Liston A, DS Gernandt, TF Vining, CS Campbell, D Piñero 1999b Molecular Phylogeny of Pinaceae and *Pinus*. *in:* "Proceedings of the International Conifer Conference". Ashford, England, August 23-26.

Liston A, WA Robinson, D Piñero, E R Alvarez-Buylla 1999a Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phylogenet Evol **11:** 95-109.

Little EL, KW Dorman 1954 Slash pine (*Pinus elliottii*) including south Florida slash pine: Nomenclature and description. US For Serv SE For Exp Sta Paper No.36.

Little ELJr, WB Critchfield 1969 Subdivisions of the genus *Pinus* (pines). USDA Forest Service, Miscellaneous Publication Number 1144, Washington DC.

López A 1982 Variabilidad del género *Pinus* (Coniferales: Pinaceae) en Cuba. Acta Botánica Cubana **12:** 1-32.

Lu S-Y, C-I Peng, Y-P Cheng, K-H Hong, T-Y Chiang 2001 Chloroplast DNA phylogeography of *Cunninghamia konishii* (Cupressaceae), an endemic conifer of Taiwan. Genome **44:** 797-807.

Lynch M, BG Milligan 1994 Analysis of population genetic structure with RAPD markers. Mol Ecol **3:** 91–99.

Maki M 2001 Genetic differentiation within and among island populations of the endangered plant *Aster miyagii* (Asteraceae), an endemic to the Ryushu Island. Amer J Bot **88:** 2189-2194.

Marshall HD, C Newton, K Ritland 2002 Chloroplast phylogeography and evolution of highly polymorphic microsatellite in lodgepole pine (*Pinus contorta*). Theor Appl Genet **104:** 346-378.

Martínez M 1948 Los pinos mexicanos. 2nd edn. Ediciones Botas. Mexico City.

Matos JA, BA Schaal 2000 Chloroplast evolution in the *Pinus montezumae* complex: a

coalescent approach to hybridization. Evolution **54:** 1218-1233.

McKay JK, G Latta 2002 Adaptive population divergence: markers, QTL and traits. Trends Ecol Evol **17:** 285-291.

Millar CI 1993 Impact of the Eocene on the evolution of *Pinus* L. Ann Missouri Bot Gard **80:** 471-498.

Millar CI 1998 Early evolution of pines. Pages 69-91 *in* DM Richardson, ed. Ecology and Biogeography of *Pinus*, Cambridge Univ. Press, Cambridge, UK.

Millar CI, SH Strauss, TH Conkle, RD Westfall 1988 Alloenzyme differentiation and biosystematics of the California closed-cone pines (*Pinus,* subsection *Oocarpae*). Syst Bot **13:** 351-370.

Mirov NT 1967 The genus *Pinus*. Ronald Press Company, New York, NY.

Mitton JB, BR Kreiser, GE Rehfeldt 2000 Primers designed to amplify a mitochondrial *nad*1 intron in ponderosa pine, *Pinus ponderosa*, limber pine, *P. flexilis*, and Scots pine, *P. sylvestris*.

Moran GF, D Smith, JC Bell, R Appels 1992 The 5S RNA genes in *Pinus radiata* and the spacer region as a probe for relationships between *Pinus* species. Plant Syst Evol **183:** 209-221.

Moritz C 1994 Defining evolutionary significant unit for conservation. Trends Ecol Evol **9:** 373-375.

Moritz C, DM Hillis DM 1997 Molecular Systematics: Context and Controversies. Page 1-13. *in*: DM Hillis, Moritz C, Mable BK, ed. Molecular Systematics. Sinauer Assoc. Inc. Publishers. USA.

Moritz C, L Josep, C Cunningham, CJ Schneider 1997 Molecular perspective on historical; and subtropical fragmentation of Australian tropical and subtropical rainforest: implications for conservation. *in*. Laurent W, R Bierregard. Page 442-454 eds Tropical Forest Remnants: Ecology, management and Conservation of fragmented Communities. Chicago University Press Chicago.

Morton BR 1997 The influence of neighboring base composition on substitutions in plant chloroplast coding sequence. Mol Bio Evol **14:** 189-194.

Morton BR, MT Clegg 1995 Neighboring base composition is strongly correlated with base substitution bias in a region of the chloroplast genome. J Mol Evol **41:** 597-603.

Mosseler A, KN Egger, GA Hughes 1992 Low levels of genetic diversity in red pine confirmed by random amplified DNA markers. Can J For Res **22:** 1332-1337.

Nakai I, K Mitsueda, S Oochata 1995 Affinity relationships based on the inter-specific hybridization in the genus *Pinus* and natures on the growth of the hybrids. Bull Tokyo Univ For **67:** 1-18.

Nei M 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89:** 583–590.

Nei M 1987 *Molecular Evolutionary Genetics*. Columbia University Press. New York. pp 512.

Nei M, T Gojobori 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Bio Evol **3:** 418-426.

Neigel JE 1997 A comparison of alternative strategies for estimating gene flow from genetic markers Ann Rev Ecol Syst **28:** 105-128.

Nelson CD, WL Nance, DB Wagner 1994 Chloroplast DNA variation and within taxonomic varieties of *Pinus caribaea* and *Pinus elliottii*. Can J For Res **24:** 424-426.

Newton AC, TR Allnut, ACM Gillies, AJ Lowe, RA Ennos 1999 Molecular phylogeography, intraspecific variation and the conservation of tree species. Tree **14:** 140-145.

Nickles DG 1966 Comparative variability and relationship of Caribbean Pine (*Pinus caribaea* Mor.) and Slash Pine (*Pinus elliottii* Engelm.). Ph.D. Thesis. NC State University. 201p.

Nkongolo KK, P Michael, WS Gratton 2002 Identification and characterization of RAPD markers inferring relationships among pines species. Genome. **45:** 51-58

O'Brien IE, DR Smith, RC Gardner, BG Murray 1996 Flow cytometric determination of genome size in *Pinus*. Plant Sci **115:** 91–99.

Perry JP Jr, A Graham, MD Richardson 1998 The history of pines in Mexico and Central America. Pages 137-1491 *in* DM Richardson, ed. Ecology and Biogeography of *Pinus*, Cambridge Univ. Press, Cambridge, UK.

Perry JPJr 1991 The pines of Mexico and Central America. Timber Press, Inc. Portland, Oregon.

Petit RJ, A El-Mousadik, O Pons 1998 Identifying populations for conservation on the basis of genetic markers. Conservation Biol **12:** 844-855.

Pilger R 1926 *Pinus*. Page 331-342. *in*: A Engler & K Prantl. ed. *Die Natürlichen Pfalnzenfamilien*. Vol XIII. Leipzing. Engelman.

Powell W, M Morgante, R McDevitt 1995a Hypervariable microsatellite provide a general source of polymorphic DNA markers for the chloroplast genome. Curr Biol **5:**1023-1029.

Powell W, M Morgante, R McDevitt, GG Vendramin, J Rafalski 1995b Polymorphic simple sequence regions in chloroplast genome: applications to the population genetics of pines. Proc Natl Acad Sci USA **99:** 7759-7763.

Price R A, A Liston, SH Strauss 1998 Phylogeny and systematics of *Pinus*. Pages 49-68 *in* DM Richardson, ed. Ecology and Biogeography of *Pinus*. Cambridge Univ. Press, Cambridge, UK.

Price RA 1989 The genera of Pinaceae in the southeastern United States. J Arnold Arbor **70:** 247-305.

Price RA, J Olsen-Stojkovich, JM Lowensteins 1987 Relationships among the genera of Pinaceae: an immunological comparison. Syst Bot **12:** 91-97.

Provan J, N Sorano, NJ Wilson 1998 Gene-pool variation in caledonian and European Scots pine (*Pinus syvestris* L.) revealed by chloroplast simple-sequence repeats. Proc Roy Soc London, Serie B-Biological Science **265:** 1697-1705.

Provan J, W Powell, P Hollingsworth 2001 Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol **16:** 142-148.

Prus-Glowacki W, J Szweykowski, R Nowak 1985 Serotaxonomical investigation of the European pine species. Silvae Genet **34:** 4-5.

Ribeiro MM, C Plomion, R Petit, GG Vendramin, AE Szmidt 2001 Variation in chloroplast single-sequence repeats in Portuguese maritime Pine (*Pinus pinaster* Ait.). Theor Appl Genet **102:** 97-103.

Ribeiro MM, S Mariette, GG Vendramin, AE Szmidt, C Plomion, A Kremer 2002 Comparison of genetic diversity estimates within and among populations of maritime pine using

choroplast simple-sequence repeat and amplified fragment length polymorphism data. Mol Ecol **11:** 869-877.

Richardson BA, SJ Brunsfeld, NB Klopfenstein 2002 DNA from bird-dispersed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). Mol Ecol **11:** 215-227.

Richardson DM, PW Rundel 1998 Ecology and biogeography of *Pinus*: and introduction. Page 3-46. *in*: Richardson DM. ed. Ecology and Biogeography of *Pinus*. Cambridge University Press, Cambridge.UK.

Robinson M, M Gouy, C Gautier, D Mouchiroud 1998 Sensitivity of the relative-rate test to taxonomic sampling. Mol Biol Evol **15:** 1091-1098.

Rushforth K 1987 Conifers. London, UK.

Saitou N, M Nei 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:** 406-425.

Samek V 1967 Elementos de la silvicultura de los pinares. Acad Cien Cuba. La Habana. pp 104.

Samek V 1973 Pinares de Cajálbana. Estudio Sinecológico. Ser. Forestal Academica de Ciencias **13:** 3-56.

Samek V, E Del Risco 1990 Los pinares de la provincia Pinar del Río, Cuba. ed Academia, La Habana. Cuba.

Sanderson ML, JM Wojciechowski, J-M Hu, T Sher Khan, SG Brady 2000 Error, bias and long-branch attraction in data for two chloroplast photosystem gene in seed plants. Mol Biol Evol **17:** 782-797.

Saylor LC 1972 Karyotype analysis of the genus *Pinus* –subgenus *Pinus*. Silvae Genetica **21:** 155-163.

Saylor LC, RL Koenig 1967 The slash × sand pine hybrid. Silvae Genetica **16:** 134-138.

Schaal BA, DA Hayworth, KM Olsen, JT Rauscher, WA Smith 1998 Phylogeographic studies in plants: problems and prospects. Mol Ecol **7:** 465-474.

Schnieder S, D Rosseli, L Excoffier 2000 Arlequin 2.000: A software for population genetic analysis. Genetic and Biometry Laboratory. University of Geneva, Switzerland.

Schirone B, G Piovesan, R Bellarosa, C Pelosi 1991 A taxonomic analysis of seed proteins in *Pinus ssp* (Pinaceae). Plant Syst Evol **178:** 43-53.

Shaw GR 1914 The genus *Pinus.* J Arnold Arbor no 5. Reverside Press. Cambridge.

Silva J 1984 An International census of the Coniferae, I. Phytologia Mem **7:** 1-79.

Slatkin M 1991 Inbreeding coefficient and coalescence times. Genet Res Camb **58:** 167-175.

Slatkin M, WP Maddison 1989 A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123:** 603-613.

Slatkin M, WP Maddison 1990 Detecting isolation by distance using phylogenies of gene. Genetics **126:** 249-260.

Smith TB, RK Wayner 1996 *Molecular Genetics Approaches in Conservation*. Oxford Univ. Press. USA.

Sokal RR, PHA Sneath 1963 Principles of numerical taxonomy. ed. WH Freeman. San Fransisco. USA. Pp 359.

Stead JW, BT Styles 1984 Studies of the Central American pines: a revision of the '*pseudostrobus*' group (Pinaceae). Bot J Linn Soc **89:** 249-275.

Steele KP, R Vilgalys 1994 Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. Syst Bot **19:** 126-142.

Stockey RA 1984 Middle Eocene *Pinus* remains from British Columbia. Bot Gaz **144:** 262-274.

Stockey RA, M Nishida 1986 *Pinus harborensis* sp. nov. and the affinity of permineralized leaves from the Upper Cretaceous of Hokkaido. Amer J Bot **64:** 1856-1866.

Strauss SH, AH Doerksen 1990 Restriction fragment analysis of pine phylogeny. Evolution **44:** 1081-1096.

Stuessy TF, M Ono 1998 *Evolution and Speciation of Island Plants*. Cambridge Univ. Press, Cambridge UK

Swofford DL 1999 "PAUP: Phylogenetic Analysis Using Parsimony", version 4.0b8. Sinauer Associates, Sunderland, MA.

Szmidt AE, X Wang, M Lu 1996 Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* (L.) using haploid tissue analysis. Heredity **76:** 412–420.

Taberlet P, L Gielly, G Pautou, J Bouvet 1991 Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol **17:** 1105-1109.

Templeton AR 1998 Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol Ecol **7:** 381-397.

Templeton AR, E Routman, CA Phillips 1995 Separating population structure from population history: a cladistic analysis of geographical distribution of mitochondria DNA sequence data III. Cladogram estimation. Genetics **140:** 767-782.

Thomas BR, SE McDonald, M Hicks, DL Adams, RB Hodgetts 1999 Effects of reforestation methods on genetic diversity of lodgepole pine: an assessment using microsatellite and randomly amplified polymorphic DNA markers. Theor Appl Genet **98:** 793-801.

Thompson JD, TJ Gibson, F Plewniak, F Jeanmougin, DG Higgins 1997 Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **24:** 4876-4886.

Tiffney B H 2000 Geographic and climatic influences on the Cretaceous and Tertiary history of Euramerican floristic similarity. Acta Univ Carol Geol **44:** 5-16.

Tiffney BH, SR Manchester 2001 The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the northern hemisphere Tertiary. Int J Plant Sci **162:** S53-S17.

Turner MC, JC Stephens, WW Anderson 1982 Homozygosity and patch structure in plants populations as a result of nearest-neibor pollination. Proc Natl Acad Sci USA **79:** 203-207.

Van der Burgh J 1973 Hölzer der niederrheinischen Braunkohlenformation, 2. Hölzer der Braunkohlengruben “Maria Theresia” zu Herzogenrath, “Zukunft West” zu Eschweiler und “Victor” (Zülpich Mitte) zu Zülpich. Nebs einer systematisch-anatomischen Bearbeitung der Gattung *Pinus* L. Rev Paleob Palynol **15:** 73-275.

Vendramin G, L Lelli, P Rossi, M Morgante 1996 A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. Mol Ecol **5:** 595-598.

Vendramin GG, B Degen, RJ Petit, M Anzidei, A Madaghiele, B Ziegenhagen 1999 High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. Mol Ecol **8:** 1117–

1126

Vendramin GG, M Anzidei, A Madaghiele, G Bucci 1998 Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. Theor Appl Genet **97:** 456-463.

Vidakoviç M 1991 Conifers: Morphology and Variation, revised English ed., Graficki Zavod Hrvatske, Zagreb, Croatia.

Virad F, YA El-Kassaby, A Ritland 2001 Diversity and genetic structure in population of *Pseudotsuga menziesii* (Pinaceae) at chloroplast microsatellite loci. Genome **44:** 336-344.

Vogel J, T Börner, WR Hess 1999 Comparative analysis of the splicing of the complete set of chloroplast group II introns in three higher plant mutants. Nucleic Acids Res **27:** 3866-3874.

Wagner DB, WL Nance, CD Nelson, T Li, RN Patel, DR Govindaraju 1992 Taxonomic patterns and inheritance of chloroplast DNA variation in a survey of *Pinus echinata*, *Pinus elliottii*, *Pinus palustris*, and *Pinus taeda*. Can J For Res **22:** 683-689.

Wakasugi T, SI Tsudzuki, M Shibata, M Sugiura 1994 A physical map and clone bank of the black pine (*Pinus thunbergii*) chloroplast genome. Plant Mol Biol Rep **12:** 227-241.

Walter R, BK Epperson 2001 Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. Mol Ecol **10:** 103-111.

Wang X-Q, DC Tank, T Sang 2000 Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol **17:** 773-781.

Wang X-R, AE Szmidt 1993 Chloroplast DNA-based phylogeny of the Asian *Pinus* species (Pinaceae). Plant Syst Evol **188:** 197-211.

Wang X-R, Y Tsumura, H Nagasaka, AE Szmidt 1999 Phylogenetic relationships of Eurasian pines (*Pinus,* Pinaceae) based on chloroplast *rbc*L, *mat*K, *rpl*20-*rps*18 spacer and *trn*V intron sequences. Amer J Bot **86:** 1742-1753.

Waters ER; Schaal BA 1991. No variation is detected in the chloroplast genome of *Pinus torreyana*. Can J For Res **21:** 1832-1835.

Watts WA 1983 A vegetation history of the eastern United States 25 000 to 10 000 years ago.

Page 294-310 *in:* SC Porter ed The Late Pleistocene, Late-Quaternary Environmental of the United States. University of Minnesota Press, Minneapolis.

Weir BS 1996 *Genetic Data Analysis II*. Sinauer Associates. Sunderland. MA.USA.

Weising K, H Nybom, K Wolff, W Meyer 1995 *DNA Fingerprinting in Plant and Fungi*. CRC Press, Boca Raton.

Wells OO, GL Switzer, RC Schimidtling 1991 Geographic variation in Mississippi loblolly pine and sweetgum. Silvae Genetica **40:** 105-118.

Whitlock MC, DE McCauley 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution **44:** 1717-1724.

William CG, O Savolainen 1996 Inbreeding depression in conifer: implications for breeding strategy. For Sci **42:** 102-117.

Williams CG, RD Barnes, I Nyoka 1999 Embryonic Genetic load for a Noetropical conifer, *Pinus patula* Schiede et Deppe. J Heredity **90:** 394-398.

Williams JGK, AR Kubelik, KJ Livak, JA Rafalski, SV Tingey 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research **18:** 6521-6535.

Willis KJ, KD Bennett, HJB Birks 1998 The late Quaternary dynamics of pines in Europe. Page 107-121 *in:* Richardson DM. ed. Ecology and Biogeography of *Pinus*. Cambridge University Press, Cambridge.UK.

Wu J, KV Krutovskii, SH Strauss 1998 Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. Genetics **150:** 1605-1614.

Wu JK, KV Krutovskii, SH Strauss 1999 Nuclear DNA diversity, population differentiation and phylogenetic relationships in the Californian closed-cone pines based on RADP and alloenzyme markers. Genome **42:** 893-908.

Xiang Q-Y, DE Soltis, PS Soltis 1998 Phylogenetic relationship of Cornaceae and close relatives inferred from *mat*K and *rbc*L sequences. Amer J Bot **85:** 285-297.

Xiang Q-Y, DE Soltis, PS Soltis, SR Manchester, DJ Crawford 2000 Timing the Eastern Asia-Eastern North American floristic disjunction: Molecular Clock corroborates

paleontological estimates. Mol Phylogenet Evol **15:** 462-472.

Yeh FC, R-C Yang 1999 *Population Genetic Analysis: PopGene, Version 1.31* Free program distributed by the authors over the internet home page http://www.ualberta.ca/~fyeh/

Zheng Y-Q, RA Ennos 1999 Genetic variability and structure of natural and domesticated populations of Caribbean pine (*Pinus caribaea* Morelet). Theor Appl Genet **98:** 765-771.

Zurawski G, MT Clegg 1987 Evolution of higher-plant chloroplast DNA encoded genes: implications for structure-function and phylogenetic studies. Ann Rev Plant Physiol **38:** 391-418.

Table 1. Sources of *Pinus* samples and outgroup taxa used in this study and their GenBank Accession Numbers

Species	Accesion numbers	Geographic region	Source
Section *Pinus*	*rbc*L; *mat*K; *trn*V intron; *rpl*20-*rps*18 region; *trn*L-*trn*F spacer		
Subsection *Pinus*			
P. thunbergii Parlatore	D17510	E Asia	7
P. resinosa Aiton	AB063384; AB080945; AB063600; AB064255; AB081122	NE N America	3
P. tropicalis Morelet	AB063378; AB080920; AB063594; AB064249; AB081123	Cuba	1
P. nigra Arnold	AB063378; AB084498; AB019891; AB019928; AB081150	Europe, Medit.	3
P. mugo Turra	AB063372; AB081087; AB063588; AB064243; AB099811	Europe	3
P. sylvestris Linnaeus	AB019809; AB084492; AB019883; AB019920; AB099813	N Eurasia	3
P. sylvestris ssp. sibirica Ledebour	AB097775; AB097781; AB097792; AB097800; AB097804	Eurasia	5
P. densiflora Siebold & Zuccarini	AB019814; AB084497; AB019888; AB019925; AB099810	E Eurasia	3
P. massoniana Lambert	AB019815; AB081088; AB019889; AB019926; AB099814	C-E China, Taiwan	3
P. pinaster Aiton	AB019818; AB084493; AB019892; AB019929; AB099807	Mediterranean	6
P. densata Masters	AB097770; AB097778; AB097787; AB097795; AB097805	China	3
P. uncinata Miller ex Mirbel	AB097774; AB097779; AB097786; AB097794; AB097802	Pyrenees	6
P. luchuensis Mayr	AB097772; AB097780; AB097788; AB097796; AB097806	Ryukyu Island	3
P. uliginosa A. Neumann	AB097776; AB097782; AB097793; AB097801; AB097803	Central Europe	5
P. kessiya Gordon	AB019813; AB019850; AB019887; AB019924	S-E Asia	4
P. merkussi Junghuhn & de Vriese	AB019811; AB019848; AB019889; AB019922	S-E Asia	4
P. heldreichii Christ	AB019821; AB019858; AB019895; AB019932	Balkans, C- Europe	4

P. yunannensii Francher	AB019816; AB019853; AB019890; AB019927	Yunan	4
P. tabuliformis Carrière	AB019810; AB019847; AB019884; AB019921	Tibet, Mongolia	4
P. hwangshanensis Hsia	AB019812; AB019849; AB019886; AB019923	Central-E China	4
Subsection *Canarienses*			
P. canariensis C.Smith	AB019823; AB084494; AB019897; AB019934; AB099812	Canary Islands	1
P. roxburghii Sargent	AB064339; AB084495; AB064341; AB064342; AB099808	Himalayas	3
Subsection *Pinea*			
P. pinea Linnaeus	AB019822; AB084496; AB019896; AB019933; AB099806	Mediterranean	1
Subsection *Halepenses*			
P. halepensis Miller	AB019819; AB081089; AB019893; AB019930; AB099809	Mediterranean	1
P. brutia Tenore	AB019820; AB019857; AB019894; AB019931	Aegean, Turkey	4
Section “New World hard pines”			
Subsection *Contortae*			
P. banksiana Lambert	AB063367; AB080922; AB063583; AB064238; AB081125	N America	3
P. contorta Loundon	AB063369; AB080821; AB063585; AB064240; AB081124	W N America	2
P. virginiana Miller	AB063379; AB080923; AB063595; AB064250; AB081126	S-E N America	3
Subsection *Ponderosae*			
P. ponderosa P.&C. Lawson	AB063371; AB080924; AB063587; AB064242; AB081127	W-N America	3
P. douglasiana Martínez	AB063388; AB080925; AB063604; AB064259; AB081128	W Mexico	3
P. jeffreyi Balfour	AB080914; AB080926; AB080916; AB080918; AB081129	W Mexico	3
P. coulteri D. Don	AB097777; AB097785; AB097791; AB097799; AB097809		3
P. engelmannii Carrière	AB080915; AB080927; AB080917; AB080919; AB081130	W-C Mexico	3

Subsection *Attenuatae*			
P. attenuata Lemmon	AB063365; AB080933; AB063581; AB064236; AB081131	W-N America	2
P. muricata D. Don	AB063387; AB080935; AB063603; AB064258; AB081133	W-N America	3
P. radiata D.Don	AB063383; AB080934; AB063599; AB064254; AB081132	W-N America	3
Subsection *Australes*			
P. caribaea var.*caribaea* Barret & Golfari	AB063368; AB080940; AB063584; AB064239; AB081137	Cuba	1
P. caribaea var.*hondurensis* Barret & Golfari	AB063385; AB080942; AB063601; AB064256; AB081139	C America	3
P. caribaea var.*bahamensis* Barret & Golfari	AB063366; AB080941; AB063582; AB064237; AB081138	Bahamas	1
P. cubensis Grisebach	AB063370; AB080938; AB063586; AB064241; AB081134	Cuba	1
P. palustris Miller	AB063373; AB080937; AB063589; AB064244; AB081136	S-E America	2
P. maestrensis Bisse	AB063371; AB080939; AB063587; AB064242; AB081135	Cuba	1
P. taeda Linnaeus	AB063377; AB080928; AB064248; AB063593; AB081142	S-E America	2
P. rigida Miller	AB063376; AB080929; AB064247; AB063592; AB081146	S-E America	3
P. pungens Lamber	AB063375; AB080932; AB064246; AB063591; AB081145	S-E America	3
P. serotina Michaux	AB081076; AB080930; AB081079; AB081082; AB081143	S-E America	3
P. elliottii Engelmann	AB081075; AB080931; AB081078; AB081081; AB081144	S-E America	3
P. echinata Millar	AB081077; AB080936; AB081080; AB081083; AB081147	S-E America	3
Subsection *Oocarpae*			
P. herrerae Martínez	AB063386; AB080943; AB063602; AB064257; AB081148	C Mexico	3
P. oocarpa Schiede & Schlechtendal	AB063382; AB081084; AB063598; AB064253; AB081140	C America	3
P. patula Schlechtendal & Chamisso	AB063381; AB080944; AB063597; AB064252; AB081141	E Mexico	3
P. teocote Schlechtendal & Chamisso	AB097773; AB097783; AB097789; AB097797; AB097807	Mexico	3

P. lawsonii Roezl ex Gordon	AB097771; AB097784; AB097790; AB097798; AB097808	C Mexico	3
Subsection *Leiophyllae*			
P. leiophylla Sciede & Deppe	AB019825; AB081085; AB019899; AB019936; AB081149	Mexico	3
Outgroups			
P. parviflora Siebold & Zuccarini	AB019800; AB81086; AB019874; AB019911; AB099815	E Asia	2

Note.- 1, Pinar del Río University, Cuba; 2, Forest Product Research Institute, Japan; 3, Kamigamo Experimental Station of Kyoto, Japan; 4, Wang et al. (1999), 5, High Tatras Park (Slovakia), 6, Madrid University, Spain, 7, Wakasugi et al. (1994).

Table 2. Average number of nucleotide substitutions per site in *cp*DNA coding and non-coding regions

Locus	No. analysed sites	No. polymorphic sites	Nucleotide Diversity subgenus *Pinus*		
			Subgenus *Pinus*	Eurasian lineage	North American lineage
*rbc*L					
K_o	1256 [1254]	40 (28)	0.00651 ± 0.00126	0.00358 ± 0.00101	0.00429 ± 0.00098
K_s	309	22 (14)	0.01287 ± 0.00376	0.00633 ± 0.00252	0.00784 ± 0.00265
K_a	945	18 (14)	0.00442 ± 0.00110	0.00270 ± 0.00091	0.00309 ± 0.00088
*mat*K					
K_o	831 [816]	60 (32)	0.00944 ± 0.00120	0.00669 ± 0.00112	0.00267 ± 0.00061
K_s	187	16 (8)	0.00970 ± 0.00299	0.00702 ± 0.00244	0.00422 ± 0.00191
K_a	629	44 (24)	0.01260 ± 0.00281	0.00681 ± 0.00138	0.00235 ± 0.00057
*rpl*20-*rps*18 region	516 [510]	21 (10)	0.00571 ± 0.0070	0.00520 ± 0.00142	0.00238 ± 0.00083
*rpl*20	162 [162]	6 (4)	0.00703 ± 0.00099	0.00237 ± 0.00143	0.00303± 0.00100
*rpl*20-*rps*18 spacer	262 [256]	15 (6)	0.00652 ± 0.00098	0.00739 ± 0.00145	0.00346 ± 0.00093
*trn*V intron	492 [492]	11 (7)	0.00643 ± 0.00198	0.00568 ± 0.00196	0.00271 ± 0.00106
*trn*L-*trn*F spacer	417 [349]	31 (11)	0.00119 ± 0.00307	0.01001 ± 0.00251	0.00448 ± 0.00162

Note.- K_o = overall sites with the standard deviations; K_s= synonymous sites; K_a= nonsynonymous sites. Values within parenthesis are number of informative sites. Values within brackets are number of site excluding gaps.

Table 3. Sampled populations of *Pinus tropicalis*, number of individuals, region and exact location

Populations	No. of individuals	Region	Location	Altitude (m)
Galalón (Ga)	10	North-eastern	22^{o} 46'N; 82^{o} 24'W	250
La Jagua (LJ)	11	North-western	22^{o} 37'N; 83^{o} 37'W	150
Bartolo (Ba)	12	North-western	22^{o} 33'N; 83^{o} 51'W	100
San Juan (SJ)	20	North-western	22^{o} 23'N; 84^{o} 00'W	150
Mina Dora (MD)	10	North-western	22^{o} 30'N; 83^{o} 59'W	100
Pilotos (Pi)	24	Central	22^{o} 26'N; 83^{o} 37'W	30
Viñales (Vi)	20	Central	22^{o} 35'N; 83^{o} 13'W	200

Table 4. Distribution of the *cp*DNA haplotypes among populations

Haplotype	Position					Populations							Total
	*trn*T-*trn*L spacer			*trn*L intron		LJ	Ga	Ba	MD	Pi	Vi	SJ	
	49	115	349	451	990								
I	A	G		A	C	9							9
II	A	G	*	A	C	1							1
III	A	A	*	T	T		2						2
IV	A	A		T	T		9	7	6			16	22
V	C	A		T	T			5	5		17	4	10
VI	A	A		T	C					15	3		15
VII	A	G		T	C					7			7
VIII	A	G	*	T	C					2			2

Note-. The asterisk represented the absence of minisatelitte (-AGAAGGGGAG-) in the haplotype sequence

Table 5. Analysis of Molecular Variance (AMOVA)

Sources of variation	d.f.	Variance Component	Φ-statistics (P)
***cp*DNA sequence**			
Among populations	6	0.2148 (V_a)	Φ_{ST} = 0.5202 (*P < 0.01*)
Within populations	100	0.1981 (V_b)	
Total	106	0.4129	
Among regions	2	0.2951 (V_a)	Φ_{CT} = 0.4997 (*P < 0.01*)
Among population within regions	4	0.0120 (V_b)	Φ_{SC} = 0.0570 (*P > 0.05*)
Within populations	100	0.1981 (V_c)	Φ_{ST} = 0.6070 (*P < 0.01*)
Total	106	0.5052	
RAPD			
Among populations	4	0.8742	Φ_{ST} = 0.209 (*P < 0.01*)
Within populations	63	3.2909	
Total	67	4.1614	
Among regions	1	0.8032	Φ_{CT} = 0.171 (*P < 0.05*)
Among population within regions	3	0.6048	Φ_{SC} = 0.155 (*P > 0.05*)
Within populations	62	3.2909	Φ_{ST} = 0.290 (*P < 0.01*)
Total	67	4.6991	

Table 6. Populational G_{ST} pairwise comparison of *Pinus tropicalis* (above diagonal) and *N*m (below diagonal)

Populations	**Ga**	**LJ**	**MD**	**Ba**	**SJ**	**Pi**	**Vi**
Galalón (ga)		0.7337*	0.6235*	0.6333*	0.7131*	0.5825*	0.7561*
La Jagua (LJ)	0.18		0.1768	0.1581	0.0380	0.5413*	0.7090*
Mina Dora (MD)	0.30	2.30		0.1003	0.0462	0.4640*	0.6241*
Bartola (Ba)	0.28	2.60	15.8		0.0244	0.4623*	0.6315*
San Juan (SJ)	0.20	12.63	10.3	20.0		0.5566	0.6974
Pilotos (Pi)	0.36	0.42	0.57	0.58	0.39		0.0451
Viñales (Vi)	0.16	0.20	0.30	0.29	0.22	10.6	

Table 7. Haplotype definitions, with fragment sizes for polymorphic *cp*SSRs

Locus	**Pt71936**	**Pt109567**	**Pt11048**
Hap I	149	168	116
Hap II	150	168	115
Hap III	149	169	115
Hap IV	150	168	116
Hap V	151	168	116

Table 8. Frequencies and diversities of the *cp*SSR haplotypes within studied populations

Populations[a]	Haplotypes					*H*
	I	II	III	IV	V	
Ga	8	1	1	0	0	0.3700
Ba	0	7	0	4	0	0.5091
LJ	0	8	0	3	1	0.5302
MD	0	7	0	3	1	0.5636
Pi	3	13	0	8	0	0.6051

Note.- [a] Populations abbreviations: LJ, La Jagua; Ba, Bartolo; Ga, Galalón; MD, Mina Dora; Pi, Pilotos

Table 9. Estimates of polymorphism (*P*), diversities (*H* and π**)** and G_{ST} pairwise comparison within each of 5 populations of *Pinus tropicalis* based on 10 RAPD primers

Populations	***P* (%)**	***H* ± SD**	**π**	**Pairwise G_{ST}**			
				LJ	**Ba**	**MD**	**Pi**
Galalón	42	0.158 ± 0.020	0.004671	0.15	0.21	0.17	0.23
Bartolo	46	0.177 ± 0.021	0.004973		0.08	0.12	0.17
La Jagua	44	0.159 ± 0.019	0.003918			0.12	0.15
Mina Dora	36	0.149 ± 0.020	0.003545				0.14
Pilotos	54	0.186 ± 0.019	0.004672				
Average	44.4	0.166	0.004357				
Overall	70	0.215 ± 0.021	0.007032				
Fixation index		G_{ST} = 0.23	N_{ST} = 0.38				

Table 10. Nei's (1987) genetic diversity statistic (H_S, H_T, and G_{ST}), and genetic identity (I) after the removal of each population one by one

Populations	H_T	H_S	G_{ST}	I	Nm
- Galalón	0.2064	0.1675	0.18	0.98	2.15
- La Jagua	0.2144	0.1624	0.24	0.93	1.58
- Bartolo	0.2171	0.1673	0.22	0.92	1.67
- Mina Dora	0.2216	0.1698	0.23	0.93	1.63
- Pilotos	0.2013	0.1605	0.20	0.95	1.97
- La Jagua - Pilotos	0.1882	0.1605	0.14	0.98	3.00

Table 11. Comparison of estimates of genetic diversity and genetic differentiation in *cp*DNA, *cp*SSR and RAPD with related pines

Species populations	*cp*DNA		*cp*SSR		RAPDs		
	H	G_{ST}	H	G_{ST}	H_S	H_T	G_{ST}
Pinus longaeva					0.130[A]	0.136	0.039
Pinus attenuata	0.00[B1]	0			0.15[B2]	0.21	0.36
Pinus radiata	0.03[B1]	-0.011			0.17[B2]	0.22	0.26
Pinus muricata	0.09[B1]	0.87			0.13[B2]	0.22	0.45
Pinus densiflora					0.39[C]		
Pinus sylvestris			0.977 [D2]	0.019	0.13 [D1]	0.16	0.019
Pinus flexilis			0.36 [E]	0.013		0.14	0.025
Pinus oocarpa					0.358 [F]	0.403	0.112
Pinus resinosa			0.568 [G2]	0.121	0.00 [G1]	0.00	---
			0.300 [G3]	0.560			
Pinus hartwwgii							
Nevado de Colina		0.014[I]					
Cerro Potosí		0.419					
Ixtachihuatl/Popocatépetl		0.087					
All population		1					
Pinus montezumae							
Cerro Potosí		0.475[I]					
Ixtachihuatl/Popocatépetl		0.118					
All population		1					
Pinus pinaster							
French			0.944[J1]	0.038			
Portuguese			0.866 [J2]	0.023			

All Europe	0.812 [J3]	0.235			
Pinus albiculis	0.907[K]	0.12[K]			
Pinus halepensis	0.222[L1]	0.308	0.124 [L2]	0.138	0.102
Pinus halepenses-complex	0.257 [L1]	0.556			
Pinus pinea	0.46[M]	0.201			
Pseudostuga menzeisii	0.388 [H]	0.019	0.258	0.241	0.072

Note.- [A]Lee et al. 2002; [B1]Hong et al. 1993; [B2]Wu et al. 1999; [C]Lee et al. 1997; [D1]Szmidt et al. 1996; [D2]Provan et al. 1996; [E]Latta and Mitton 1997; [F]Diaz et al. 2001;[G1]Mosseler et al. 1992; [G2]Echt et al.1998; [G3]Water et al. 2001; [H]Viard et al. 2001; [I]Matos and Schaal 2000; [J1]Ribeiro et al. 2001; [J2]Ribeiro et al. 2002; [J3]Vendramin et al. 1998; [K]Richardson et al. 2002; [L1]Bucci et al. 2002, [L2]Korol et al. 2002; [M]Gómez et al. 2000.

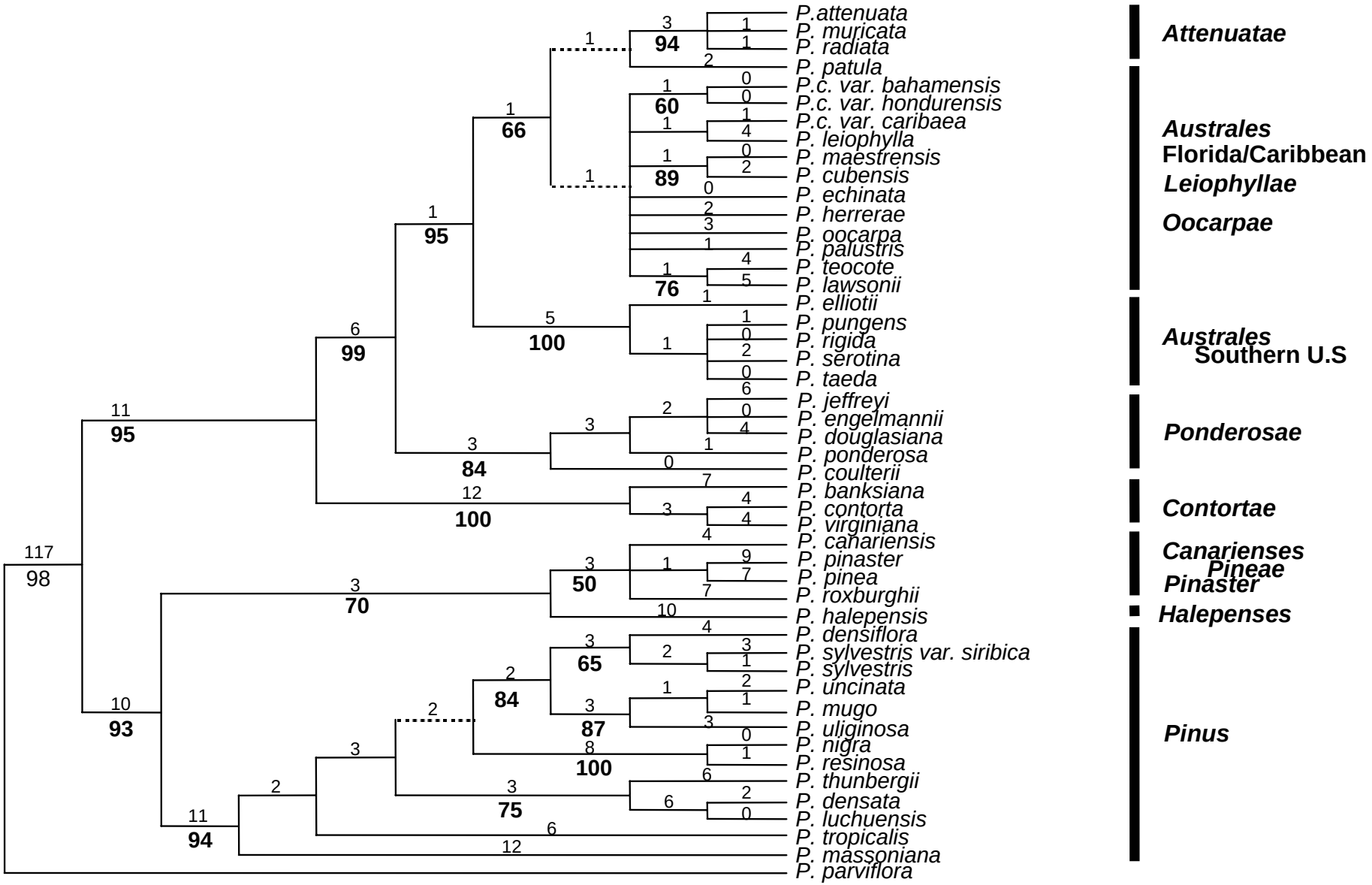

Figure 1. One of the 4 most-parsimonious trees based on the combined sequences. Bold numbers below each branch show the bootstrap values based on 250 replicates and above indicate the number of steps. Subsections are indicated on the right.

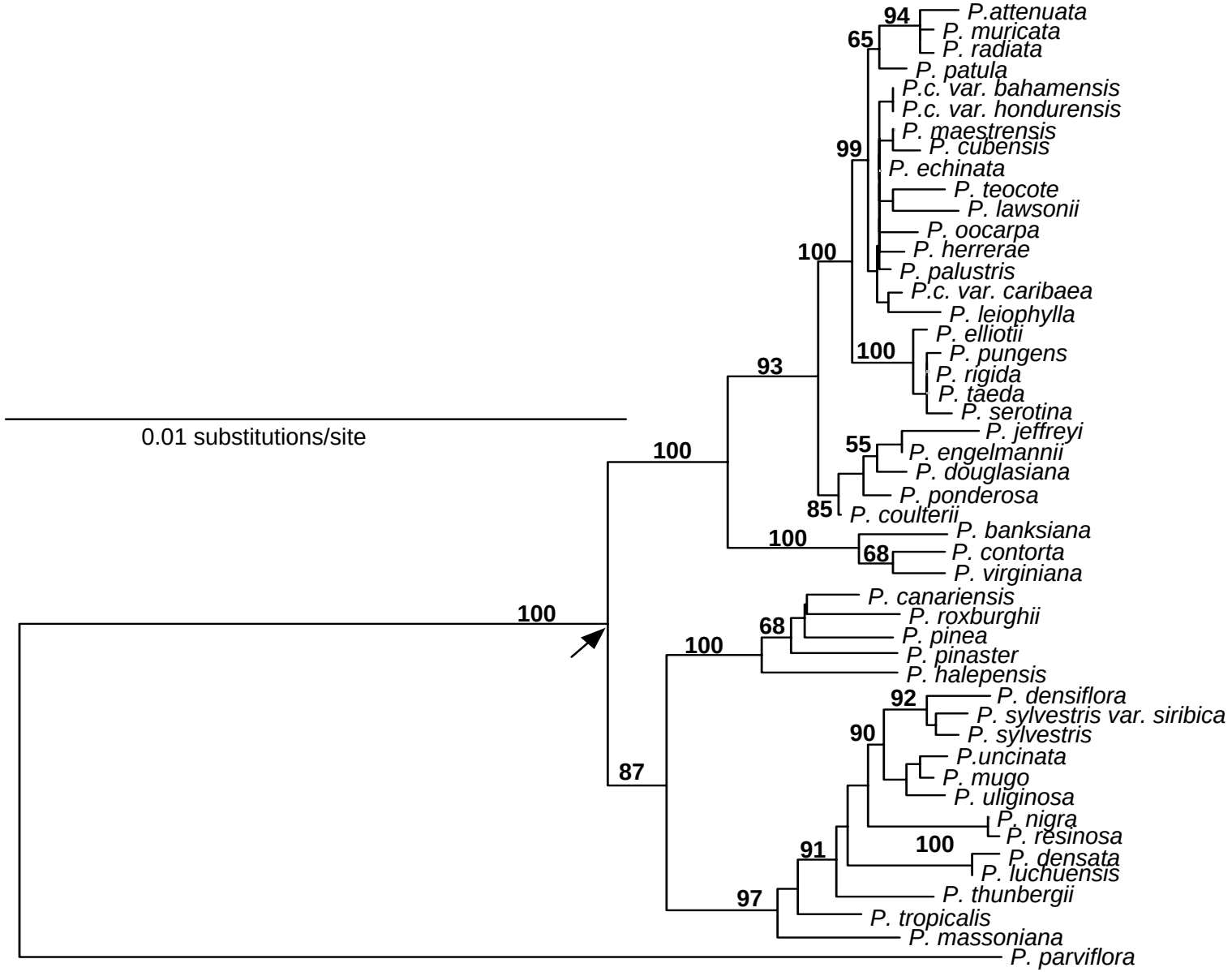

Figure 2. Neighbor-joining tree based on the combined sequences. Bold numbers above the branches show the bootstrap values based on 1000 replicates. Arrow shows the landmark event used for the calibration (130 MY).

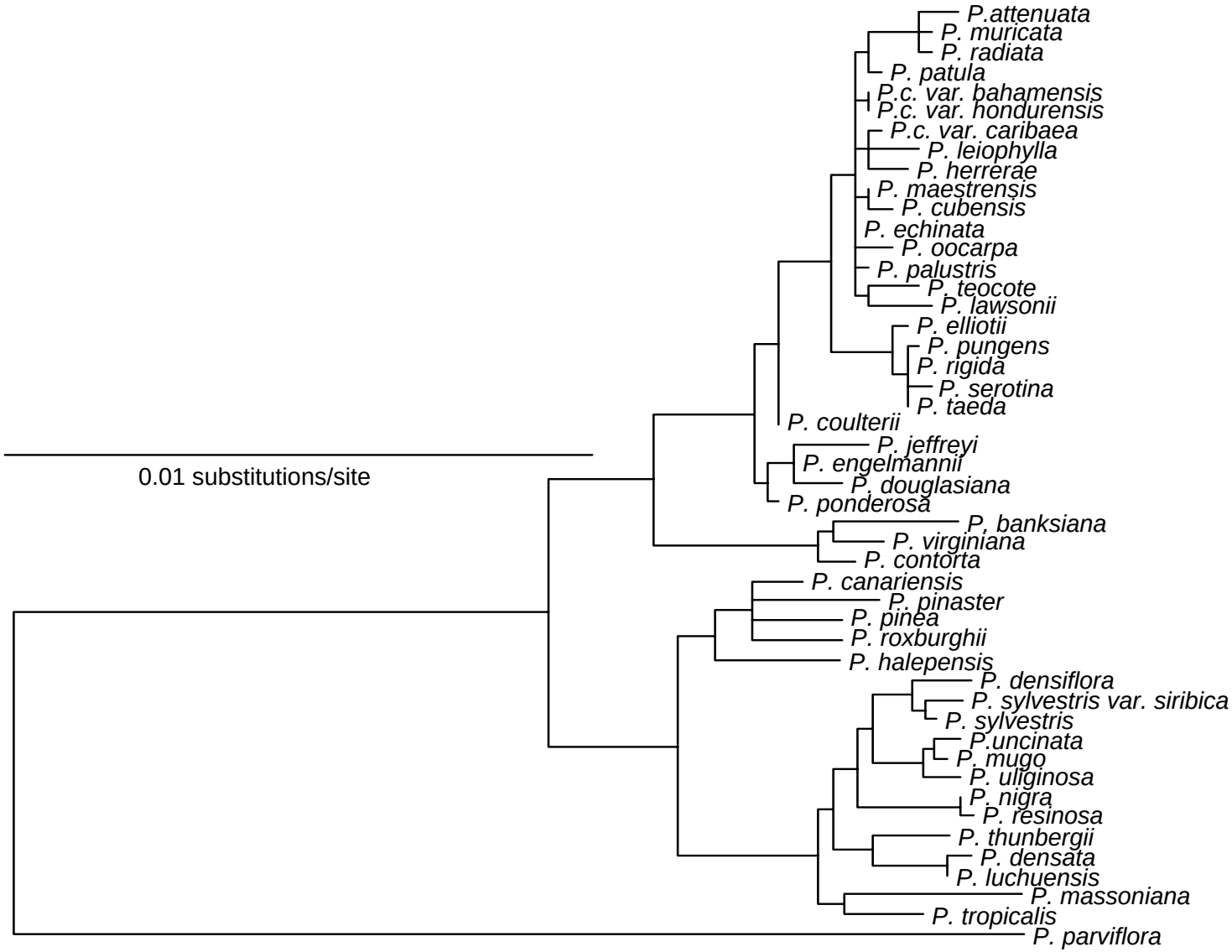

Figure 3. Maximum likelihood tree based on the combined sequences, assuming the HKY85 model of substitutions, -lnL = 8886.65224.

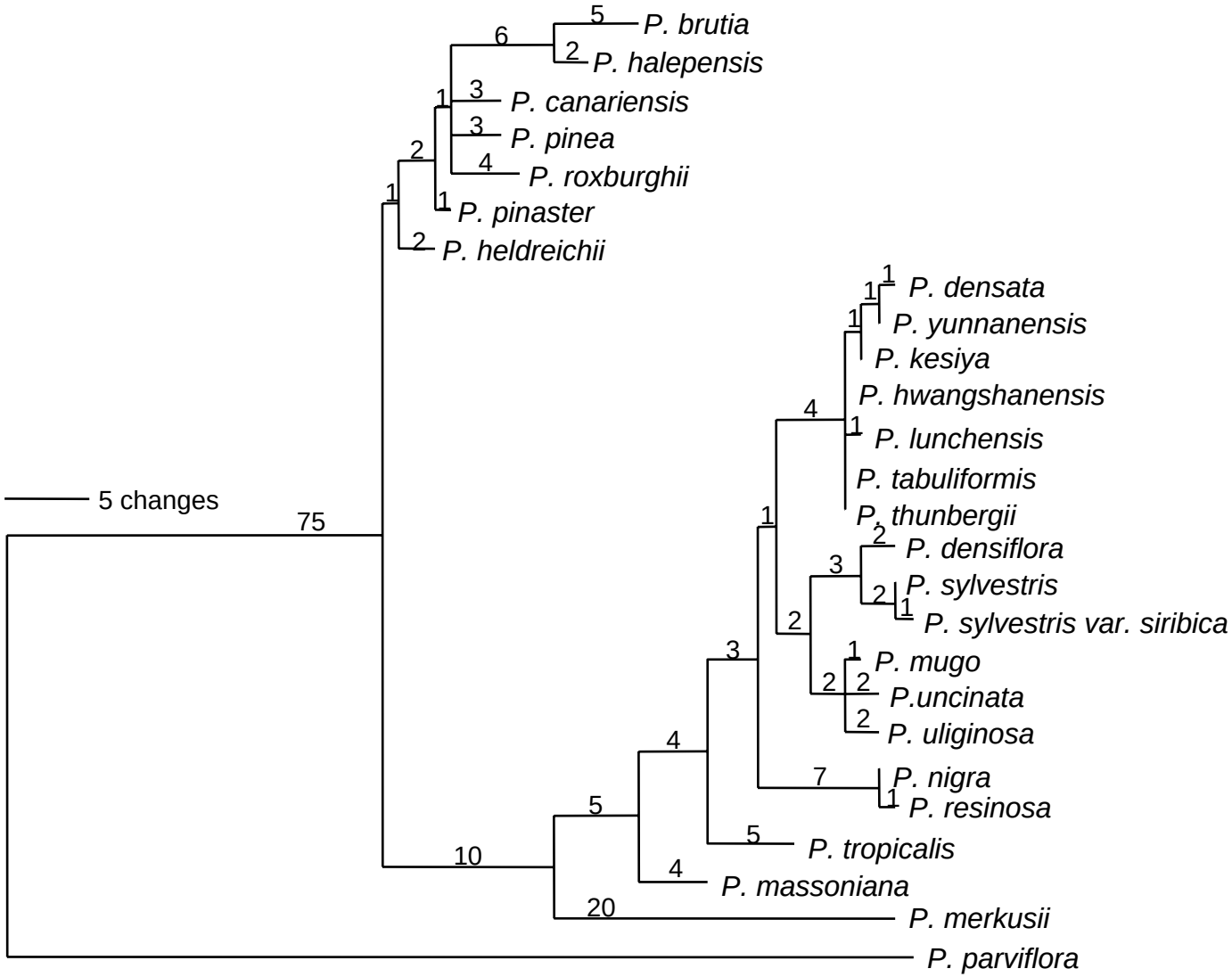

Figure 4. One of the 12 Most-parsimonious trees based on the combined sequences of *rbc*L, *mat*K, *trn*V intron and *rpl*20-*rps*18 region the Eurasian pines.

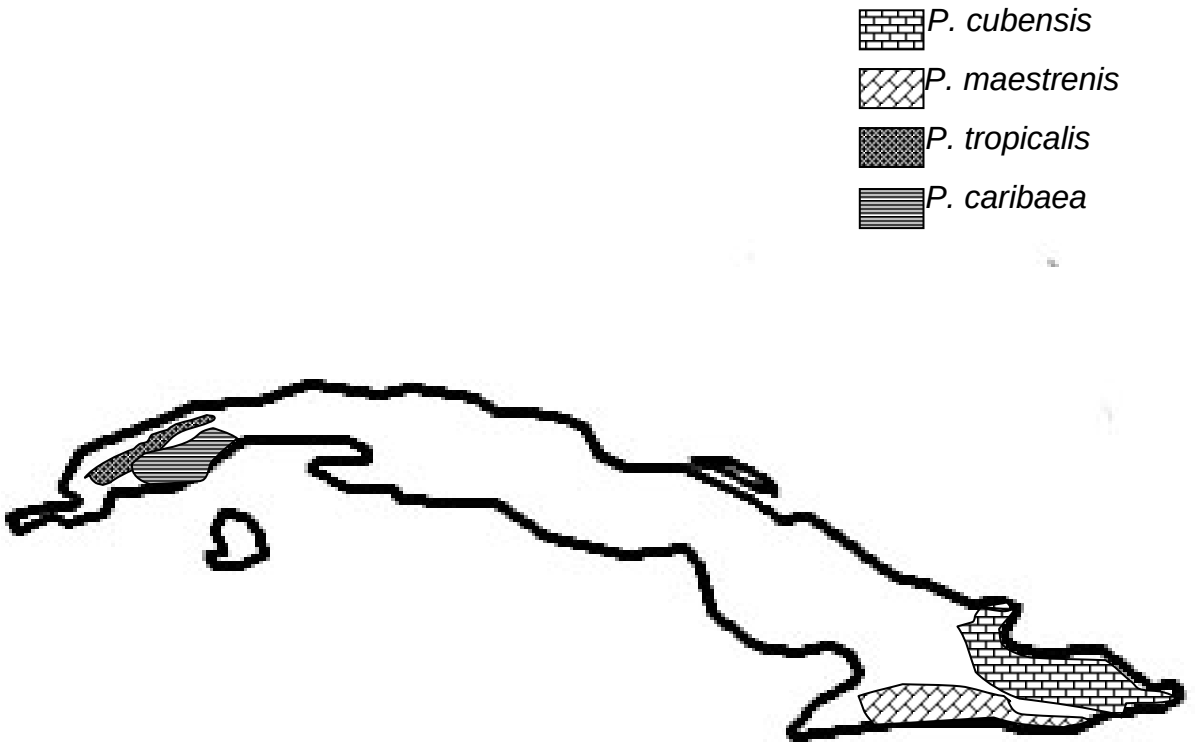

Figure 5. Natural Distribution of the Cuban pines.

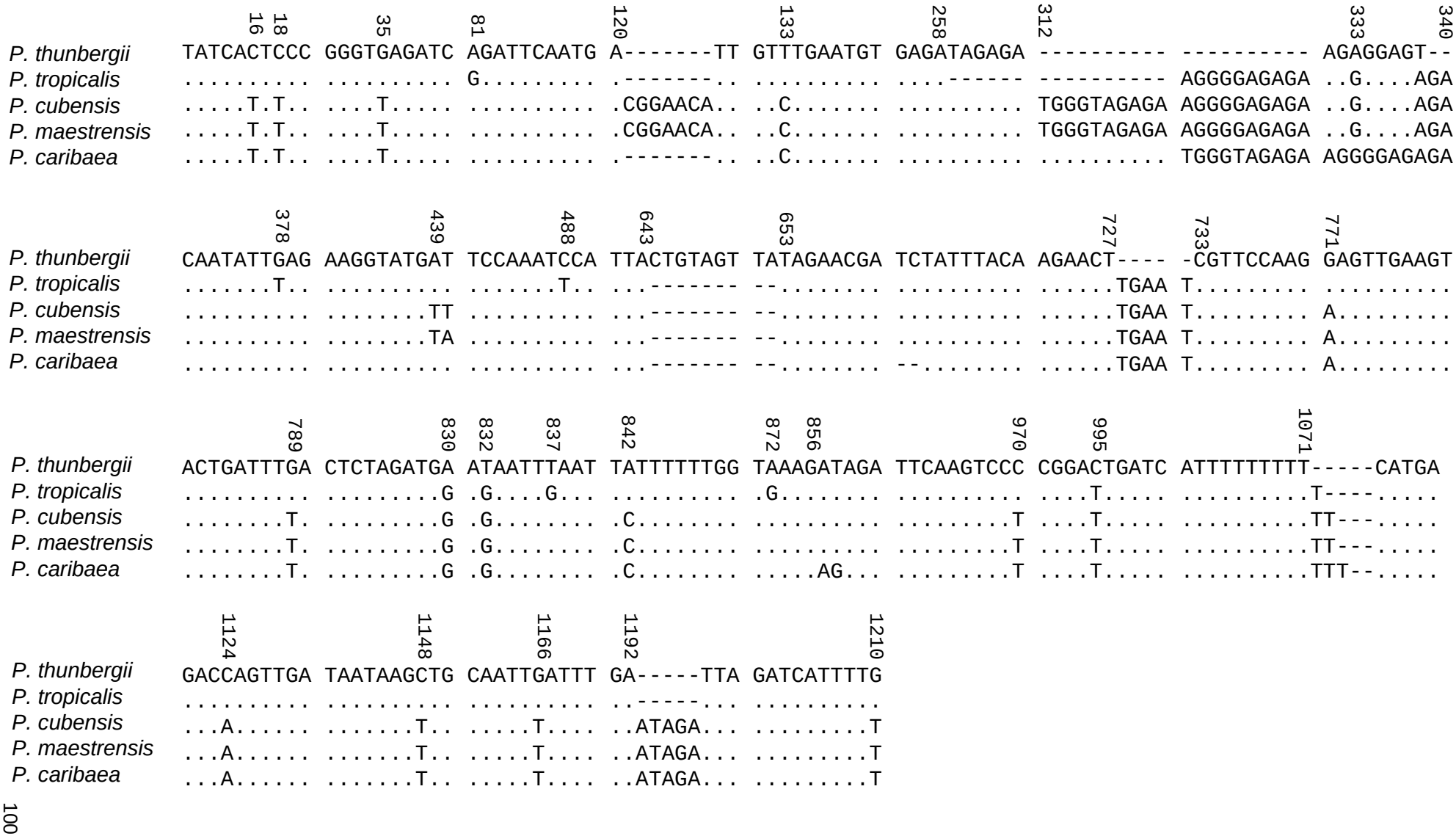

Figure 6. Positions of variables site and indels in the aligned sequences. Dashes represent the gaps and dots indicate the same nucleotide referring to *P. thunbergii* from 67797 – 69174 of the complete *cp*DNA sequences.

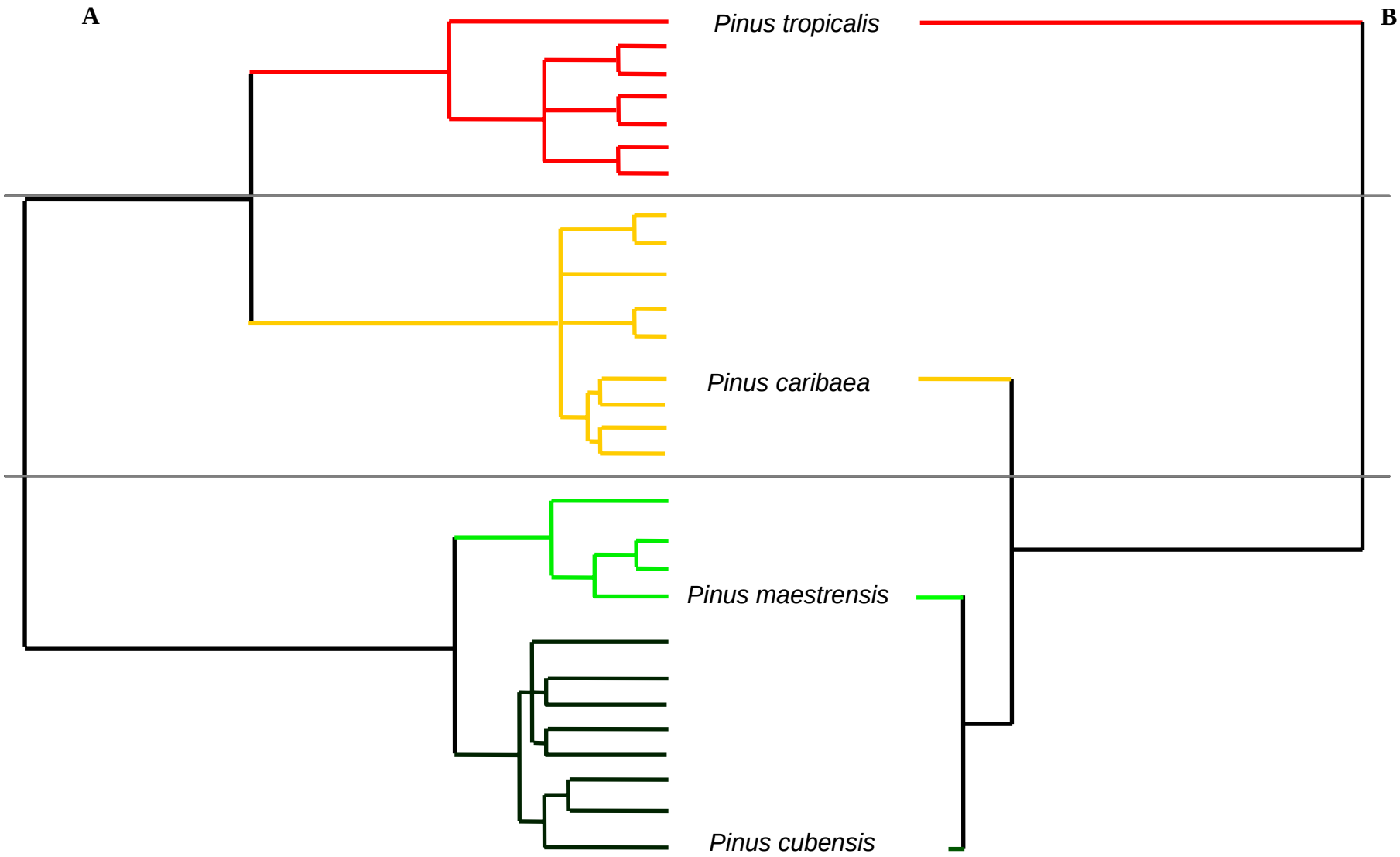

Figure 7. Cuban pines. (A) Dendrogram based on the 16 morphological characters. (B) Neighbor-joining tree based on the combined sequences of *trn*T-*trn*L spacer, *trn*L intron and *trn*L-*trn*F spacer.

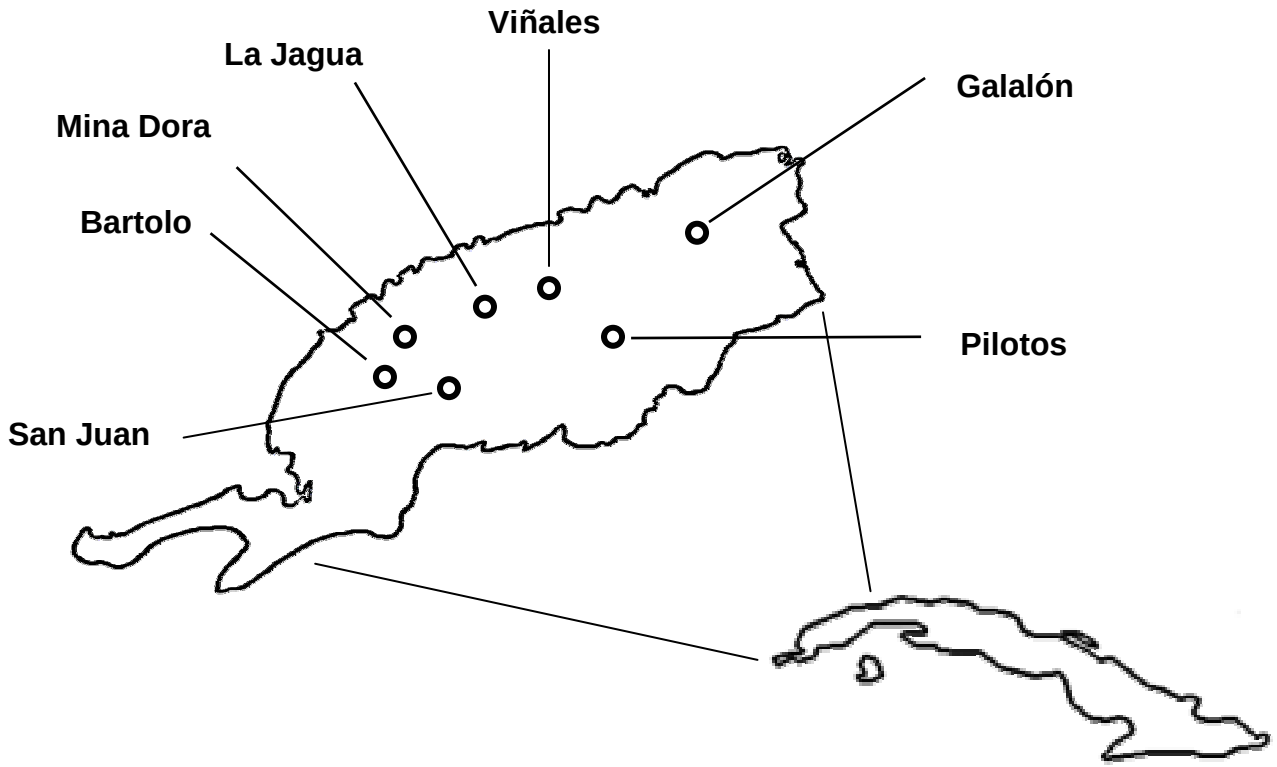

Figure 8. Geographical distribution of *Pinus tropicalis* and location of the seven natural populations sampled from North-eastern, Central and North-western part of Cuba.

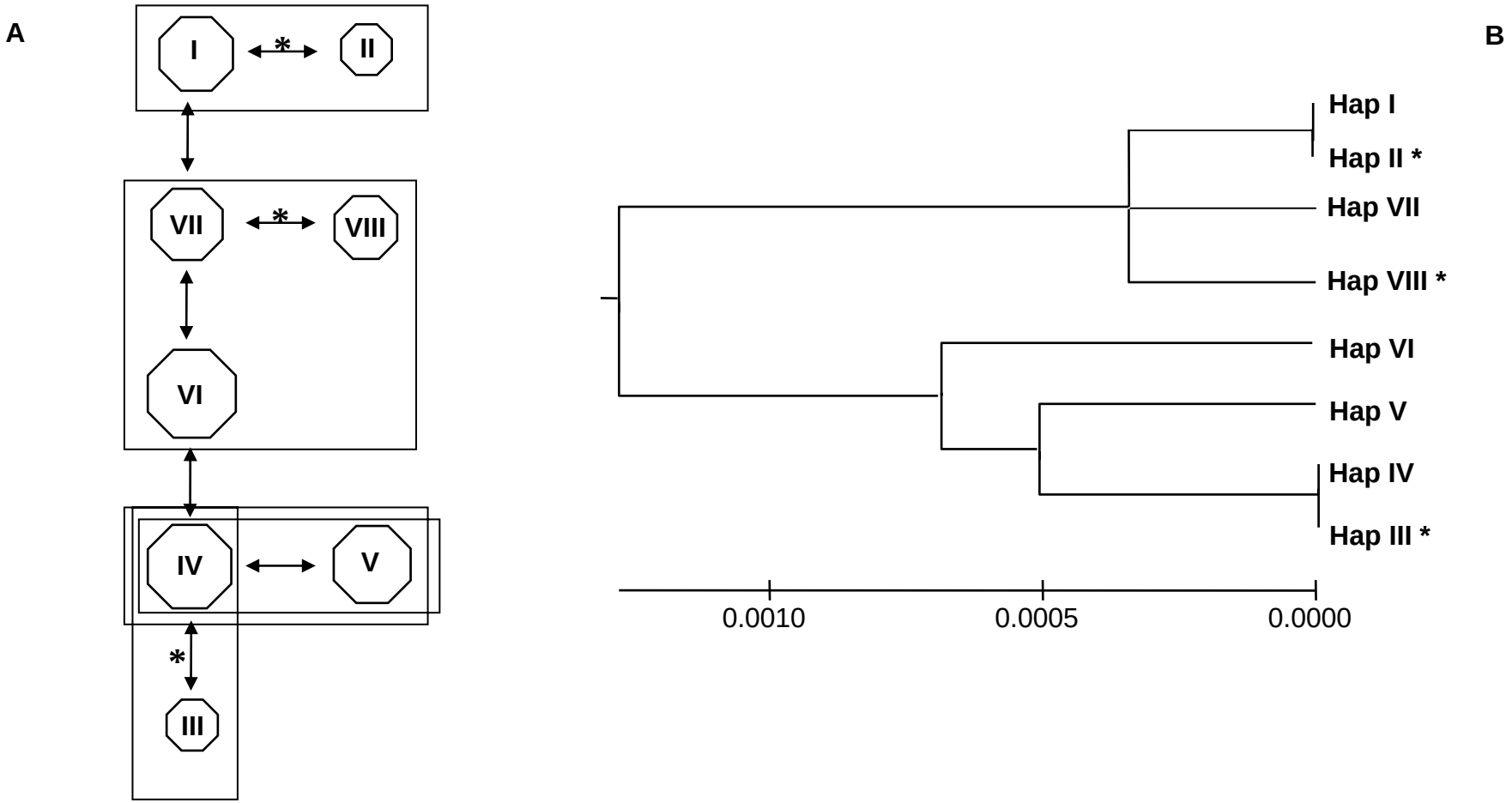

Figure 9. (A) Minimum-spanning network of the haplotypes detected in *P. tropicalis* using the *trn*T-*trn*L spacer and *trn*L intron. The asterisks represent the absence of minisatelitte (-AGAAGGGGAG-) in the haplotype sequence. (B) Neiborgh-Joining tree generated from the combined sequence data of *trn*T-*trn*L spacer and *trn*L intron of 106 individuals.

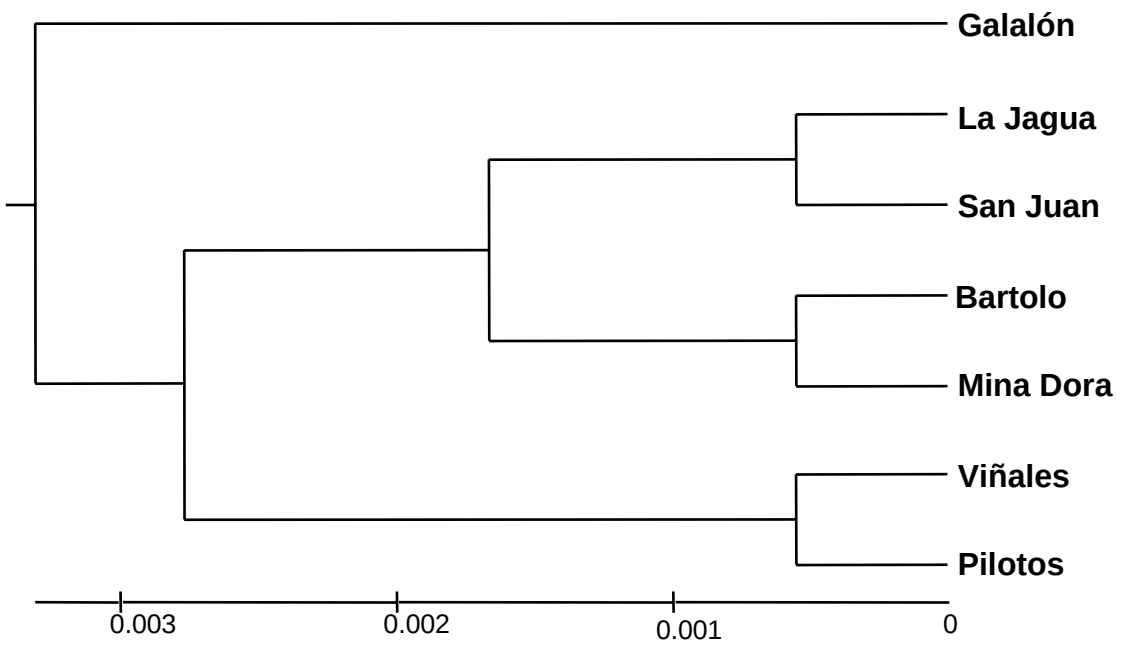

Figure 10. UPGMA tree generated on the pairwise distance among populations based on Nei's distance (1978) based on the *trn*T-*trn*L spacer and *trn*L intron sequences.

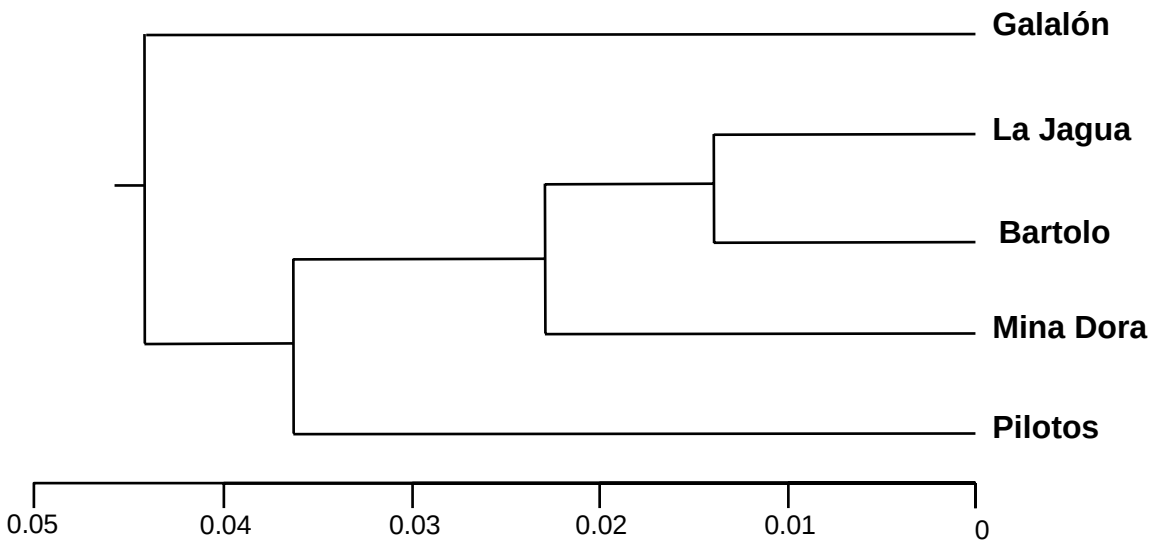

Figure 11. UPGMA tree generated on the pairwise distance among populations based on Nei's distance (1978) based on the RAPD primers.