American Journal of Botany 92(12): 2086–2100. 2005.

EVOLUTIONARY RELATIONSHIPS AMONG *PINUS* (PINACEAE) SUBSECTIONS INFERRED FROM MULTIPLE LOW-COPY NUCLEAR LOCI[1]

JOHN SYRING,[2] ANN WILLYARD,[2] RICHARD CRONN,[3] AND AARON LISTON[2,4]

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331 USA; and [3]Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, Oregon 97331 USA

Sequence data from nrITS and cpDNA have failed to fully resolve phylogenetic relationships among *Pinus* species. Four low-copy nuclear genes, developed from the screening of 73 mapped conifer anchor loci, were sequenced from 12 species representing all subsections. Individual loci do not uniformly support either the nrITS or cpDNA hypotheses and in some cases produce unique topologies. Combined analysis of low-copy nuclear loci produces a well-supported subsectional topology of two subgenera, each divided into two sections, congruent with prior hypotheses of deep divergence in *Pinus*. The placements of *P. nelsonii*, *P. krempfii*, and *P. contorta* have been of continued systematic interest. Results strongly support the placement of *P. nelsonii* as sister to the remaining members of sect. *Parrya*, suggest a moderately well-supported and consistent position of *P. krempfii* as sister to the remaining members of sect. *Quinquefoliae*, and are ambiguous about the placement of *P. contorta*. A successful phylogenetic strategy in *Pinus* will require many low-copy nuclear loci that include a high proportion of silent sites and derive from independent linkage groups. The locus screening and evaluation strategy presented here can be broadly applied to facilitate the development of phylogenetic markers from the increasing number of available genomic resources.

Key words: incongruence; nuclear genes; phylogeny; Pinaceae; *Pinus*.

Pinus L. is one of the 11 genera of Pinaceae, a family that is monophyletic among gymnosperms (Hart, 1987; Chaw et al., 1997). *Pinus* is by far the largest genus in the family, and its approximately 110 species comprise ca. 50% of the Pinaceae (Farjon, 1998). Unique morphological features such as needle-like leaves clustered in fascicles (shoot dimorphism) and woody ovulate cone scales with specialized apical regions (umbos), in combination with molecular evidence (Wang et al., 2000b; Liston et al., 2003), indicate that members of the genus *Pinus* are well differentiated from related genera such as *Cathaya* Chun and Kuang and *Picea* A. Dietrich.

The extensive utility (Le Maitre, 1998) and ecological significance (Richardson and Rundel, 1998) of *Pinus* has made it the focus of numerous molecular evolutionary studies (reviewed in Price et al., 1998; Liston et al., 2003). Among these studies, a variety of data types have been used to infer relationships, including allozymes (Karalamangala and Nickrent, 1989), restriction fragment length polymorphisms (Strauss and Doerksen, 1990; Krupkin et al., 1996), anonymous DNA markers (Dvorak et al., 2000; Nkongolo et al., 2002), DNA sequencing (Liston et al., 1999; Wang et al., 1999; Gernandt et al., 2005), and various combinations of these types (Wu et al., 1999). The most inclusive studies have utilized chloroplast DNA sequencing (Krupkin et al., 1996; Wang et al., 1999; Geada López et al., 2002), and a cpDNA-based (1555 bp of *matK*, 1262 bp of *rbcL*) phylogenetic hypothesis for 101 species of *Pinus* has recently been published (Gernandt et al., 2005). Phylogenetic analyses using the nuclear ribosomal DNA ITS region have also sampled nearly half of the species (Liston et al., 1999, 2003; Gernandt et al., 2001).

With the wealth of data available for *Pinus*, it is perhaps surprising that key phylogenetic relationships remain unresolved. Comparisons between the most comprehensive chloroplast (Gernandt et al., 2005) and nuclear ribosomal ITS (Liston et al., 2003) data sets highlight the limits to our current understanding. In these analyses, 12 of 17 commonly resolved clades are congruent, although many of the clades show low bootstrap support. Incongruence between these estimates occurs in several cases, some marked by strong branch support and others being poorly supported. Examples include (1) the conflicting yet robust resolutions of subsect. *Contortae*, either as the sister lineage to the remaining members of sect. *Trifoliae* (cpDNA) or as a more derived lineage within sect. *Trifoliae* (nrDNA); (2) the poor resolution among subsects. *Gerardianae*, *Krempfianae*, and *Nelsoniae* (Wang et al., 2000a; Gernandt et al., 2003; Zhang and Li, 2004); and (3) the monophyly or paraphyly of sect. *Pinus*. Even in regions of coarse topological agreement, both analyses are plagued by a nearly universal lack of resolution among terminal taxa in the species-rich subsections *Australes*, *Pinus*, *Ponderosae*, and *Strobus*. Despite ca. 30 published studies over the past two decades (reviewed in Price et al., 1998; Liston et al., 2003), a well-resolved phylogeny of *Pinus* remains a work in progress.

Resolving relationships among recently diverged taxa is a pervasive problem in molecular systematics, and it is compli-

[1] Manuscript received 28 March 2005; revision accepted 13 September 2005.

The authors thank D. Neale, G. Brown, and K. Krutovsky (USDA Forest Service, Institute of Forest Genetics, Davis, CA) for assistance selecting conifer anchor loci, D. Gernandt (Universidad Autónoma del Estado de Hidalgo, Mexico) for sharing cpDNA data prior to publication; and K. Farrell for laboratory assistance. R. Businský (Silva Tarouca Research Institute for Landscape and Ornamental Gardening, Průhonice, Czech Republic), M. Gardner, F. Inches, and P. Thomas (Royal Botanic Garden, Edinburgh, Scotland), D. Gernandt (Universidad Autónoma de Hidalgo, Mexico), D. Johnson (USDA Forest Service, Institute of Forest Genetics, Placerville, CA), R. Sneizko (USDA Forest Service, Dorena Genetic Resource Center, Cottage Grove, OR), and D. Zobel (Oregon State University) generously contributed seed and needle tissue. Funding was provided by the National Science Foundation grant DEB 0317103 to A.L. and R.C., and the USDA Forest Service Pacific Northwest Research Station.

[4] Author for correspondence (e-mail: listona@science.oregonstate.edu)

cated by an array of biological phenomena (use of inappropriate molecules for recent divergence events, rapid radiations, reticulation) and analytical shortcomings (e.g., inability of cladistic models to accommodate species non-coalescence) (Small et al., 2004). In *Pinus*, nrITS and cpDNA gene trees are largely unresolved at this scale of divergence. For example, a combined analysis of four cpDNA loci (>3500 bp) yielded a six-species polytomy in a 10-taxa sample of subsect. *Strobus* (Wang et al., 1999) and a combined analysis of four cpDNA loci (>3200 bp) yielded a seven-species polytomy for the seven sampled members of subsects. *Australes*, *Leiophyllae*, and *Oocarpae* (Geada López et al., 2002). Similarly, 2400-bp of nrDNA sequence failed to resolve relationships among subsects. *Cembroides*, *Nelsoniae*, and *Balfourianae*, and the nrDNA gene tree contains a 13-species polytomy for the 15 species in subsect. *Cembroides* (Gernandt et al., 2001). The lack of resolution in cpDNA-based studies results from insufficient sequence variation (Gernandt et al., 2005), while intragenomic polymorphism in nrITS (arising from weak concerted evolution or a failure of intraspecific coalescence; Gernandt et al., 2001) limits the utility of nrDNA. Thus, cpDNA and nrDNA are unlikely to provide sufficient character support for a robust phylogeny of *Pinus*. Even if either data source were to provide satisfactory resolution, there is conflict between these two sources of data (such as the placement of subsect. *Contortae*) that is not easily resolved.

Recent studies using low-copy nuclear loci (Harris and Disotell, 1998; Springer et al., 2001; Cronn et al., 2002; Malcomber, 2002; Cronn et al., 2003; Rokas et al., 2003; Alvarez et al., 2005) reveal the power of multiple independent markers for disentangling phylogenetic relationships at low taxonomic levels. These studies suggest that independent markers that sample a range of substitution rates and patterns (Yang, 1998) can provide greater resolution than cpDNA and nrDNA (alone or combined) for resolving rapid radiations, relationships among closely related species, and complex historical hybridization events (Harris and Disotell, 1998; Springer et al., 2001; Cronn et al., 2002; Malcomber, 2002; Cronn et al., 2003; Small et al., 2004). The large nuclear genome of pines (22–37 pg/2C; Grotkopp et al., 2004) hinders the development and application of low-copy nuclear markers for phylogenetic applications, as multigene families (Kinlaw and Neale, 1997) and repetitive and retrotransposon DNA (Friesen et al., 2001) are abundant. Despite these obstacles, genomics and genetic linkage mapping efforts directed towards *P. taeda* (loblolly pine; Brown et al., 2001) and *P. radiata* (Monterey pine, Devey et al., 2004) have revealed a large number of low-copy markers that show orthology across pine species (Brown et al., 2001). Because these conifer anchor loci show strong evidence for positional orthology, they make promising candidates for population and phylogenetic applications in pines and other members of Pinaceae (Brown et al., 2004).

In this paper, we explore the potential of four conifer anchor loci to resolve evolutionary relationships among the 11 subsections recognized by Gernandt et al. (2005). Each subsection forms a well-supported clade in the intensively sampled cpDNA-based study. Species selected for this study represent the most ancient divergence events (e.g., subgenera, sections) and more recent divergence events (e.g., among subsections). The phylogenetic resolution and support provided by these low-copy nuclear loci (individually and combined) are compared with existing hypotheses derived from chloroplast DNA (Gernandt et al., 2005) and nuclear ribosomal DNA (Liston et al., 2003). Analysis of the concatenated cpDNA, nrITS, and low-copy nuclear data sets provides a revised estimate of pine subsectional phylogeny and is used to analyze how conflicts among data sets are resolved in a combined analysis.

MATERIALS AND METHODS

Plant materials—Twelve species of *Pinus* (Appendix) were sampled as ''exemplars'' to represent all 11 subsections recognized by Gernandt et al. (2005). Two of these subsections, subsects. *Nelsoniae* and *Krempfianae*, are monotypic. Species from the remaining subsections were chosen because of their economic and ecological importance (e.g., *P. taeda*, *P. radiata*, *P. ponderosa*, *P. contorta*, and *P. monticola*) or because of the availability of fresh tissue. All analyses were rooted with *Picea* based on morphological (Hart, 1987; Farjon, 1990) and molecular (Wang et al., 2000b) phylogenetic studies of Pinaceae. DNA was isolated from fresh or dried needles using the FastPrep DNA isolation kit (Qbiogene, Carlsbad, California, USA). Haploid megagametophyte tissue was used as the DNA source for select amplifications (*P. roxburghii*, *P. taeda*, and *P. gerardiana*; Appendix), particularly in cases where heterozygosity for length polymorphisms resulted in poor quality sequences.

Locus selection, amplification, and sequencing—Over 150 low-copy conifer anchor loci are available as possible candidates for cross-species amplifications in *Pinus* (Brown et al., 2001, 2004; Temesgen et al., 2001; Krutovsky et al., 2004), all of which have been mapped in *P. taeda* (Brown et al., 2001; Temesgen et al., 2001). From this list, we screened 73 loci using three selective criteria. First, primers had to readily amplify the target locus across the diversity of the genus, as represented by *P. taeda* (sect. *Trifoliae*), *P. merkusii* s.l. or *P. thunbergii* (sect. *Pinus*), *P. monticola* (sect. *Quinquefoliae*), and *P. nelsonii* (sect. *Parrya*). Second, the amplified product needed to be ~500 bp or larger in length so that they could contain a reasonable number of variable sites. To provide a check for orthology, our third criterion was that amplification products needed to be homogeneous, with only a single prominent band upon amplification, and a single nucleotide sequence if amplified from haploid megagametophte tissue. Once sequences were obtained, our final criterion was to exclude all loci with frame shifts or exon / intron deletions (evidence of pseudogenization), or unexpectedly high rates of divergence (evidence of potentially paralogous comparisons).

Among the loci meeting these criteria, we chose four that show considerable diversity in length, ratio of noncoding to coding sequence, and function. The first, *IFG1934*, is derived from a loblolly pine cDNA clone that maps to the distal end of linkage group 3 in *P. taeda* (Krutovsky et al., 2004; GenBank X67714). This locus corresponds to a chlorophyll A/B binding protein type II 1B precursor gene that was first characterized in *P. sylvestris* (Jansson and Gustafsson, 1990; EMBL X14506). The second locus, *AGP6*, has high sequence identity to an arabinogalactan-like protein that maps to linkage group 5 from *P. taeda* (Krutovsky et al., 2004; GenBank AF101785) and is associated with secondary cell wall formation in differentiating xylem (Zhang et al., 2003). As a class, AGP-like proteins are proteoglycans rich in hydroxyproline (HyP) and proline (Pro), which serve as targets for *O*-glycosylation (Showalter, 2001). The third locus is 4-coumarate: CoA ligase (*4CL*), an enzyme in the phenylpropanoid pathway that serves as a precursor pathway to lignin biosynthesis (Zhang and Chiang, 1997) and that maps to linkage group 7 of *P. taeda* (Krutovsky et al., 2004). For the purpose of aligning and describing our sequences, we used the full-length *4CL* sequence from *P. taeda* (GenBank U39405; Zhang and Chiang, 1997) as a reference. The fourth locus, *IFG8612*, is derived from a loblolly pine cDNA clone that maps to linkage group 3 in *P. taeda* (Krutovsky et al., 2004; C. S. Kinlaw, Institute of Forest Genetics, unpublished data; GenBank AA739606), approximately 64 cM distant from *IFG1934*. This locus shows high identity with a Late Embryogenesis Abundant (*LEA*)-like gene identified in *Pseudotsuga menzeisii* (Iglesias and Babiano, 1999; GenBank AJ012483).

Primers used to amplify loci were published previously (Brown et al., 2001; Temesgen et al., 2001; Zhang et al., 2003) and are shown in Fig. 1. In some

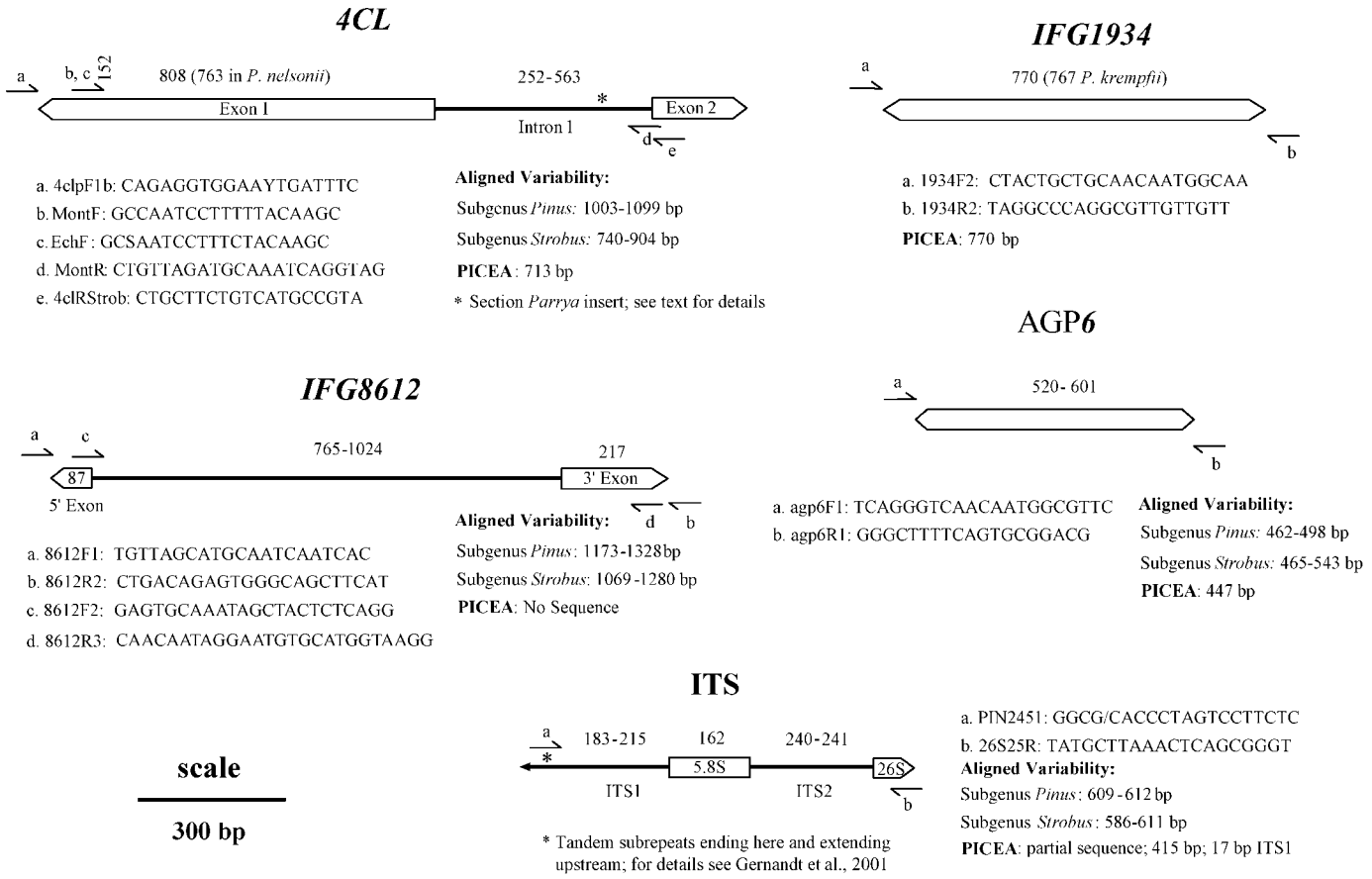

Fig. 1. Gene diagrams for the nuclear loci included in this study. Shown is the size (in bp) of each locus and the location of exons (open bars) and introns or noncoding regions (lines). Arrow heads on exons and introns indicate that the locus or domain extends beyond our sample. Primers used for amplification and sequencing are indicated by small arrows above and below gene diagrams; primer names and sequences are associated with each diagram. Amplicon lengths for *Picea* amplicon and ranges for *Pinus* subgenera are given.

cases, it was possible to amplify universally across *Pinus* with a single primer combination; for most loci, however, it was necessary to develop subgenus-specific primers. Amplification protocols generally followed those outlined in Brown et al. (2001) and Temesgen et al. (2001). This involved an initial preheating step at 80°C for 2 min, followed by 35 cycles of 94°C for denaturation, a 62–52°C ''Touch-Down'' annealing step (−1°C/cycle for 10 cycles; Don et al., 1991), and 72°C extension for 1 min per kb of amplification product. PCR reactions (25 μL) utilized ~50–200 ng of DNA template, 0.4 μM each primer, 0.2 mM dNTP, 1.5–2.0 mM $MgCl_2$, 0.13 μg/μL BSA, and 2 units of *Taq* polymerase (Fisher Scientific, Pittsburgh, Pennsylvania, USA) in 1× buffer (10 mM Tris-HCl [pH 9.0], 50 mM KCl). Amplification products were resolved on 1.5% TAE agarose gels containing 0.2% crystal violet; bands were visualized in white light, excised, and purified using QIAquick Gel Extraction kits (QIAGEN, Valencia, California, USA). Purified products (~100 ng) were sequenced using the ABI-Prism Big Dye Terminator Mix (ver. 3.1; Applied Biosystems, Foster City, California, USA). Products were resolved on an Applied Biosystems 3100 Genetic Analyzer capillary sequencer at the Oregon State University Center for Gene Research and Biotechnology Central Services Laboratory. Automated traces were evaluated using the BioEdit Sequence Alignment Editor 5.0.6 (Hall, 1999). In some instances, length polymorphism heterozygosity and/or poor amplification precluded direct sequencing of PCR products. These products were cloned into pGem-T Easy (Promega, Madison, Wisconsin, USA) following the manufacturer's recommendations.

The current study incorporates two previously published data sets. The first includes two cpDNA regions, maturase K (*matK*) and ribulose 1,5-bisphosphate carboxylase large subunit (*rbcL*), which have been sequenced for 101 pine species and *Picea* (Gernandt et al., 2005). The second data set includes a 650-bp portion of the nuclear ribosomal DNA internal transcribed spacer (nrDNA-ITS) region that was sequenced from 47 pine species and *Picea* (Liston et al., 1999; Liston et al., 2003). The nrITS data set was modified from the original study by adding a sequence from *P. monticola* (amplified and sequenced using methods described in Liston et al., 1999). Two sequences included in the original reference were replaced; these include *P. krempfii* (replaced with GenBank AF305061) and *P. echinata* (replaced with GenBank AF367378; Chen et al., 2002). Three species used in our low-copy gene analysis are replaced by closely related taxa for nrITS; subsect. *Balfourianae* is represented by *P. aristata* (*P. longaeva* in our study; GenBank AF03700), subsect. *Australes* is represented by *P. attenuata* (*P. radiata* in our study; GenBank AF03702), and *Picea sitchensis* is replaced by *Picea rubens* (from Germano and Klein, 1999; GenBank AF136610). Due to uncertain homology between subgenera within nrITS1, a stretch of ca. 114 nucleotides (representing positions 95–321 in the global alignment) was aligned by subgenus.

Sequence alignment and data analysis—Sequences were aligned by ClustalW (in BioEdit) using default parameters. Exon regions aligned with ease, as ingroup and outgroup sequences showed minimal length variation. In contrast, genus-wide alignments of noncoding sequences were non-trivial. For these, separate alignments were carried out for each subgenus using an outgroup from the alternative subgenera (*P. monticola* for alignment of subg. *Pinus*, *P. roxburghii* for subg. *Strobus*). Regions of shared homologous sequence were analyzed across sectional exemplars using the ''align two sequences'' (bl2seq) option of NCBI BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). For this analysis, the following criteria were set: word size = 15,

dropoff value for percentage of shared nucleotides = 50%, open gap penalty = 5, and gap extension penalty = 2. These subgeneric alignments were used as a guide in attempting to construct global alignments. Final adjustments were made by eye to minimize the number of required indels. Alignments are available at TreeBase (SN2289; http://www.treebase.org).

Data from individual genes were analyzed independently so that we could evaluate variation on a locus-by-locus basis. Alignment statistics (determined using MEGA 2.1; Kumar et al., 2001) included the average number of characters, number of variable and parsimony informative characters, average within-group *p*-distance, and average base composition. Indels were counted for each species and averaged across taxa.

Phylogenetic analyses were performed on individual data sets using maximum parsimony (MP; PAUP* version 4.0b10; Swofford, 2002). Most parsimonious trees were found from branch-and-bound searches, with all characters weighted equally and treated as unordered. Branch support was evaluated using the nonparametric bootstrap (Felsenstein, 1985), with 1000 replicates and tree-bisection-reconstruction (TBR) branch swapping. In cases when two or more most parsimonious trees were recovered, the MP tree that most closely (exactly, except for *IFG1934*) matched the maximum likelihood tree topology was presented. In these cases, phylogenies were inferred using maximum-likelihood estimates with PAUP*. Maximum-likelihood parameters, such as substitution model, base frequencies, shape of gamma distribution, and proportion of invariant sites, were optimized using ModelTest 3.5 (Posada and Crandall, 1998).

The nuclear genes used in this study evolve in an environment characterized by recombination and independent assortment. Additionally, three of the four genes (*AGP6, IFG1934, 4CL*) contain a high proportion of non-synonymous sites so they may respond to selection in addition to mutation, drift, and migration. For these reasons, congruent phylogenetic signal among loci is important to evaluate prior to combining data sets for simultaneous analysis (Bull et al., 1993). To assess congruence among pairs of individual data sets, we used the MP-based incongruence length difference test (ILD; Farris et al., 1994) in a pairwise fashion on all six loci. Null distributions were generated using 1000 replicates of data sets pruned of invariant and uninformative characters. This test was implemented using PAUP* (''partition homogeneity'' test) with the threshold for significance at $P \leq 0.01$, following the recommendations of Cunningham (1997). Because the ILD test may be prone to Type I errors (see Cunningham, 1997; Yoder et al., 2001; but see Hipp et al., 2004), we also used the Templeton or Wilcoxon signed-rank test (WSR; Templeton, 1983) to further explore incongruence inferred in the ILD test. For this test, all most-parsimonious trees recovered for individual genes were used as constraint topologies. *P* values were adjusted using sequential Bonferroni correction (Rice, 1989), and results were interpreted conservatively, as the test returning the highest *P* value (least significant; Miyamoto, 1996; Johnson and Soltis, 1998) is accepted as the experiment-wide *P* value. Results from Templeton tests were considered indicative of heterogeneity among data sets only if significance was observed in both directions (Johnson and Soltis, 1998). Because a clear consensus has yet to be reached on the relative merits of ''total evidence'' vs. ''conditional'' prior agreement approaches for data combination (Huelsenbeck et al., 1996; Sanderson and Shaffer, 2002), we chose to explore both of these approaches.

RESULTS

Locus selection—Of the 73 loci initially screened, 61 (84% of the total surveyed) were excluded because they failed to meet the criteria outlined in the Materials and Methods. Failure to amplify the target locus in species throughout the genus was the most common problem. Loblolly pine (*P. taeda*) was the source for nearly all conifer anchor loci, so amplification success was highest for subg. *Pinus* (*P. taeda* and *P. merkusii* or *P. thunbergii*), and failure was most frequent for members of subg. *Strobus* (*P. monticola* and *P. nelsonii*). In total, 12 loci met our selection criteria (detailed in A. Willyard et al., unpublished manuscript; and R. Cronn, unpublished data). Among these, four loci were selected for phylogenetic analyses. Two loci were composed entirely of exon sequence (*IFG1934, AGP6*), and the remaining loci either contained a high proportion (*4CL*) or a small proportion (*IFG8612*) of exon sequence.

Sequence characteristics of low-copy nuclear loci in Pinus—*IFG1934*—The sequenced region corresponds to amino acids 4–260 (256 of 274 total) of the *P. sylvestris* chlorophyll A/B binding protein gene. This exon encodes an N-terminal signal transit peptide (nucleotide positions 1–108) and three α-helix motifs (aligned positions 285–375; 465–510; 630–726) and exhibits a G + C content of 60.2% (33.6% G + 26.6% C). Stop codons and insertions were absent in this alignment, although *P. krempfii* showed a 3-bp deletion (positions 145–147). The apparent structural conservation and low average divergence (mean *p*-distance = 0.0435 within *Pinus*) led to an unambiguous alignment of *IFG1934.* The alignment was 770 bp long, and it included 99 variable and 55 parsimony informative positions within *Pinus* (Table 1). Among variable sites, 23 were restricted to first codon positions, 11 to second positions, and 65 to third positions. Amino acid replacements were observed in 26 of the 256 amino acid sites (10.2%).

AGP6—We obtained an aligned sequence of 609 bp for *AGP6* corresponding to positions 31–640 of *P. taeda AGP6.* This exonic region shows a high G + C content (24.2% G + 42.3% C), and it includes 202 codons with an abundance of proline (31.5% across pines), alanine (22.6%), valine (11.9%), and threonine (11.4%). The random-coil conformation adopted by this glycosylated protein appears prone to indel events, presumably through the expansion/contraction of HyP-P-(A/V/T) motifs. The first of these repetitive motifs, ''TAPAAPTTAKP'' (residues 31–41 of the *P. taeda AGP6* translation AAF75821) is present in three (*P. nelsonii*) or two (*P. monophylla, P. radiata,* and *P. roxburghii*) near-perfect repeats, with one unit for the remaining species. A second repeat, ''PVAPAAAPTKP[A/T]P'' (residues 61–73 of AAF75821), follows the first motif and is present as two tandem units (*P. merkusii* and *P. longaeva*) or one unit (all remaining species). A third repeat, ''PPVA[T/V/A]'' (residues 109–138 of AAF75821), is present as seven near-perfect repeats in *Picea sitchensis*, but is fixed at six tandem repeats in pines. Due to the complexity of these repeats, we first translated *AGP6* sequences, aligned the amino acids using ClustalW, then converted amino acids back to the nucleotide sequence.

Of the 609 bp included in our analysis, 66 nucleotide positions were variable (10.8% overall) and 29 were parsimony informative in *Pinus* (Table 1). Among variable positions, nine were restricted to first codon positions, nine to second positions, and 48 to third codon positions. Amino acid replacements were observed at 16 of the 202 amino acid sites (7.9%). While this gene had a relatively high non-synonymous rate of substitution, eight of the 16 amino acid replacements involved four common amino acids (P ↔ A = 2; V ↔ L = 2; V ↔ A = 4), and no missense frame shifts were detected.

4CL—We obtained an aligned sequence of 1315 bp for *4CL,* corresponding to nucleotide positions 461–1513 of *4CL* from *P. taeda*. This sequence included 675 bp of exon 1 that contained 225 codons (residues 107–331 of 537 total). Indels in the *4CL* exon were restricted to a single species, *P. nelsonii*. This deletion removed 44 bp of exon 1 (aligned positions 631–

TABLE 1. Variability of low-copy nuclear genes in *Pinus* with comparisons to nrITS and cpDNA.

Gene	Region	Taxon	Aligned length[b]	Avg. length[c] (range)	%G + C	Variable sites (1/2/3)[d]	PI sites[e]	Average D[f] (SE)	Indels[g]
IFG1934	exon	subg. *Pinus*	770	770 (na)	60.4	35 (7/3/25)	10	0.0172 (.003)	0 (0)
		subg. *Strobus*	770	769.5 (767, 770)	59.9	55 (11/8/36)	17	0.0269 (.004)	0.2 (3)
		genus *Pinus*	770	769.8 (767, 770)	60.2	99 (23/11/65)	55	0.0435 (.005)	0.1 (3)
AGP6	exon	subg. *Pinus*	540	477.5 (462–498)	66.8	15 (3/2/10)	4	0.0094 (.003)	1.7 (35.7)
		subg. *Strobus*	582	503.5 (465–543)	66.1	44 (4/8/32)	11	0.0352 (.006)	1.7 (47.1)
		genus *Pinus*	609	490.5 (462–543)	66.5	66 (9/9/48)	29	0.0375 (.006)	2.4 (37.8)
4CL	exon 1	subg. *Pinus*	675	675 (na)	55.1	22 (3/5/14)	4	0.0122 (.003)	0 (0)
		subg. *Strobus*	675	667.5 (630, 675)	54.3	32 (5/4/23)	6	0.0196 (.004)	0.2 (45)
		genus *Pinus*	675	671.3 (630, 675)	54.7	75 (13/9/53)	36	0.036 (.005)	0.1 (45)
	intron 1	subg. *Pinus*	472	358.7 (328–417)	28.1	32	8	0.0363 (.007)	10.5 (7.0)
		subg. *Strobus*	365	201.7 (110–229)[h]	27.5	38	15	0.0759 (.017)	2.5 (11.3)
IFG8612	exon A + B	subg. *Pinus*	304	304 (na)	44.5	15 (5/4/6)	2	0.0186 (.004)	0 (0)
		subg. *Strobus*	304	304 (na)	44.4	19 (4/7/8)	6	0.0272 (.009)	0 (0)
		genus *Pinus*	304	304 (na)	44.5	37 (11/10/16)	20	0.0311 (.009)	0 (0)
	intron A	subg. *Pinus*	1177	963.3 (869–1024)	42.8	80	11	0.0298 (.004)	9.7 (11.0)
		subg. *Strobus*	1263	908 (765–976)	40.0	124	18	0.0418 (.007)	13.0 (27.5)
nrITS	ITS1, 5.8S, ITS2	subg. *Pinus*	626	610.7 (609–612)	55.9	51	9	0.0307 (.004)	3.3 (1.3)
		subg. *Strobus*	635	602 (586–611)	54.8	73	17	0.0418 (.005)	11.8 (2.5)
		genus *Pinus*[a]	515	515 (497–508)	55.6	99	55	0.0644 (.007)	7.8 (1.7)
cpDNA	*rbcL* + *matK*	subg. *Pinus*	2806	2806 (na)	40.2	72 (21/23/28)	21	0.0105 (.001)	0 (0)
		subg. *Strobus*	2806	2806 (na)	40.5	58 (15/8/35)	20	0.0086 (.001)	0 (0)
		genus *Pinus*	2806	2806 (na)	40.4	162 (43/42/77)	88	0.0177 (.002)	0 (0)

[a] Excludes positions 95–321 of ITS1 in the global alignment, which were aligned by subgenus as noted in Materials and Methods.
[b] Aligned length in base pairs determined using outgroups from alternative subgenera (*P. monticola* or *P. merkusii*) or using *Picea* in the case of genus totals. Outgroups not used in determining all other statistics presented in this table.
[c] Number of characters per species averaged over all species in the alignment; na refers to no variability across species.
[d] Total variable positions in alignment (parsed by codon position 1, 2, or 3).
[e] Number of parsimony informative sites.
[f] D = *p*-distance, the proportion of nucleotide sites at which paired sequences are different.
[g] Average number and length of indels. Number of inferred indels were counted for each species in an alignment and averaged over total number of species. Numbers in parentheses are the average length of the indels in base pairs.
[h] Not incorporating the section of excluded alignment (see Materials and Methods for details).

675, inclusive), including the exon 1–intron 1 splice site present in all other species. Despite this disruption, the gene may be functional if a GC dinucleotide pair at positions 627–628 is used as the splice site; this would result in a 19 amino acid deletion that retains the normal reading frame. Of the 675 exon nucleotides included in our analysis, 75 positions were variable and 36 were parsimony informative within *Pinus* (Table 1). Among variable positions, 13 were found in first codon positions, nine in second positions, and 53 in third codon positions. Amino acid replacements were observed at 22 of the 225 amino acid sites (9.8%).

Also included in *4CL* was intron 1, a region characterized by eight-fold length variation across the included taxa. The shortest *4CL* intron originated from *Picea sitchensis* and was 38 bp in length. Within *Pinus*, the shortest introns were found in members of Sect. *Quinquefoliae*, and were 227–229 bp in length. The largest introns were present in members of sect. *Parrya* (particularly *P. monophylla*, 539 bp), with much of the expansion due to the gain of an A/T-microsatellite and regions of low complexity (runs of A or T). Length differences in the *4CL* intron were accompanied by substantial variation in nucleotide composition, because the percentage A + T ranged from 70.2% in *P. taeda* to 76.3% in *P. nelsonii*. An insertion in the *4CL* intron was observed at 895 bp in the global alignment, and it was restricted to members of sect. *Parrya* (*P. nelsonii*, *P. longaeva*, and *P. monophylla*). The region varies in length from 252 (*P. nelsonii*) to 323 bp (*P. monophylla*), and it appears to be composed of superimposed indels of uncertain homology. Alignment of this stretch was sufficiently problematic to warrant its exclusion from all analyses.

As might be expected from such a dynamic region, intron alignments across divergent taxa were difficult. The effect of indels and low sequence complexity is apparent in pairwise BLAST comparisons between *P. taeda* (subg. *Pinus*), *P. monticola* (subg. *Strobus*), and four species representing the sections of the genus (Fig. 2). Local alignments between species from a common subgenus (e.g., *P. taeda* vs. *P. ponderosa* or *P. merkusii*) show relatively long stretches of easily aligned high-quality sequence (cut-off criteria defined by a word size of 15 and sequence identity ≥50%), with only short regions of nonhomologous indels. In contrast, alignments across subgenera returned short stretches of alignable sequence (<40 bp), or else failed entirely to identify blocks meeting cut-off criteria (e.g., *P. taeda* vs. *P. nelsonii*; *P. monticola* vs. *P. ponderosa*). Due to the uncertain homology between subgeneric introns, we aligned *4CL* exon sequences genus-wide, but aligned intron sequences only within subgenus (except adjacent to the flanking sequence where homology was apparent). Of the 472 aligned intron bases included in our subg. *Pinus* alignment, 32 positions were variable and eight parsimony informative (Table 1). Of the 365 aligned intron bases included in our subg. *Strobus* alignment, 38 were variable and 15 were parsimony informative.

IFG8612—We obtained an aligned sequence of 2645 bp for *IFG8612*, corresponding to nucleotide positions 222–525 of the *LEA* mRNA described from *Pseudotsuga menzeisii*. This sequence includes two partial exons that span 304 bp and code for 101 amino acids (28.3 from exon A, 71.7 from exon B). These exons show 37 variable positions within *Pinus*, of which

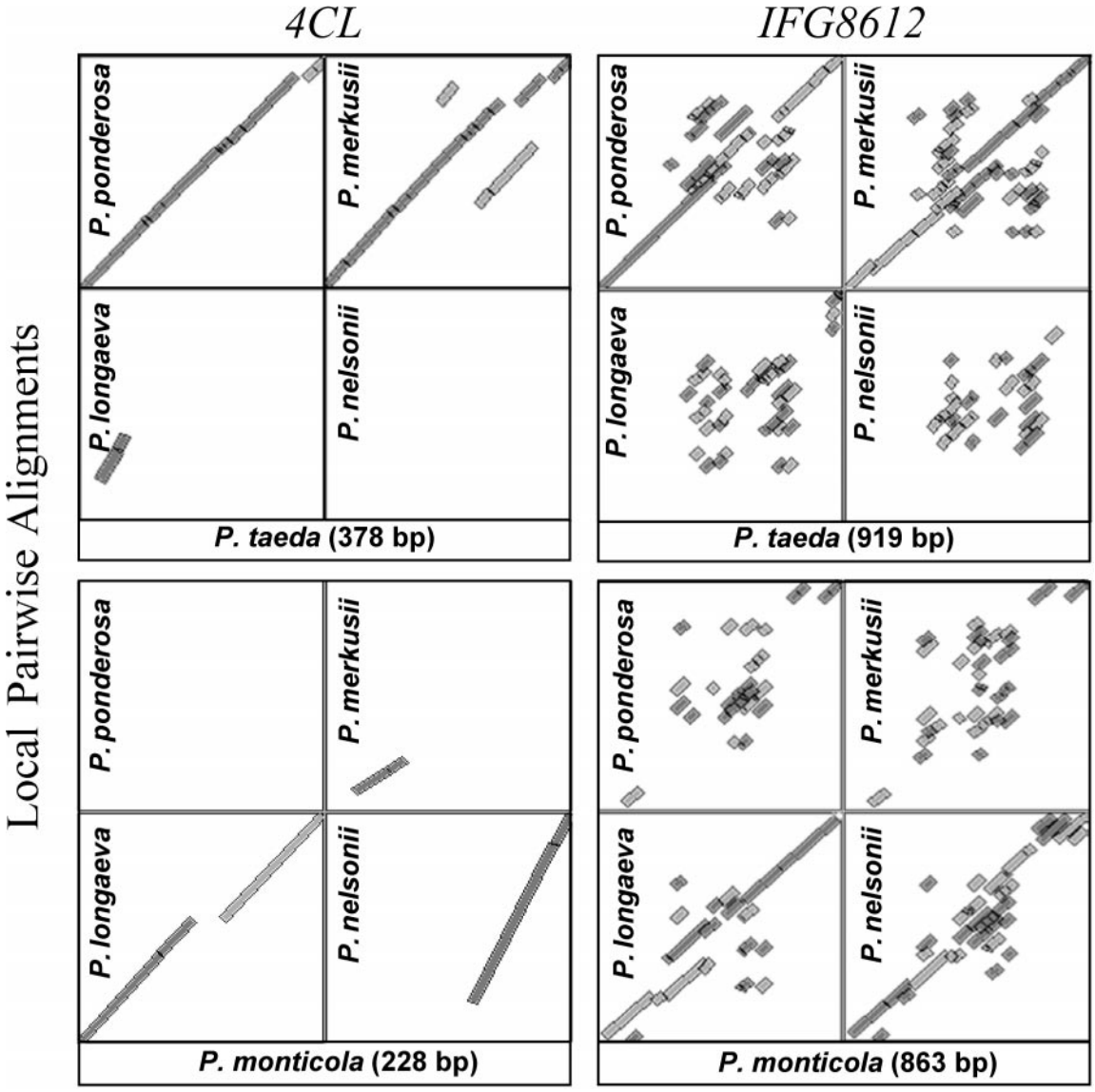

Fig. 2. Pairwise intron alignments within and between subgenera of *Pinus*. Introns from *4CL* and *IFG8612* are aligned between *P. taeda* (subg. *Pinus*; upper panels) and *P. monticola* (subg. *Strobus*; lower panels) on the *x*-axis, and exemplars from different subgenera on the *y*-axis. Diagonal lines show regions of sequence meeting BLAST criteria for minimum word size (15), similarity threshold (50%), match/mismatch weights (+1/−2) and gap open/extension penalties (5/2).

20 are parsimony informative. Among variable positions, 11 were restricted to first codon positions, 10 to second positions, and 16 to third positions. Amino acid replacements were observed at 17 of 101 amino acid sites (16.8%), of which seven were from exon A, and 10 from exon B. Nucleotide frequencies were relatively balanced in the *IFG8612* exon (24.1% G, 20.4% C, 29.0% A, 26.5% T), and indels and stop codons were absent.

The intron from *IFG8612* ranged in absolute length from 765 bp in *P. gerardiana* to 1024 bp in *P. radiata*. As observed with the *4CL* intron, alignment of the *IFG8612* intron could only be accomplished within members of the same subgenus. Local alignments between members of the same subgenus (e.g., *P. taeda* vs. *P. ponderosa* or *P. merkusii*) showed long blocks of easily aligned high-quality sequence interspersed with short repeats that showed similarity to several regions (e.g., *P. taeda* vs. *P. merkusii*; *P. monticola* vs. *P. longaeva*) (Fig. 2). Alignments across subgenera returned short blocks (<70 bp) of similarity that were composed of repeats and stretches of low sequence complexity. We aligned *IFG8612* exon sequences across the genus but aligned intron sequences by subgenus. Intron alignments for members of subg. *Pinus* were 1177 bp long and included 80 variable positions (11 parsimony informative), and aligned introns from subg. *Strobus* were 1263 bp long and included 124 variable sites (18 parsimony informative, Table 1).

Phylogenetic analyses*—Individual loci*—Most-parsimonious trees derived from individual loci show general patterns of conformity across the four low-copy nuclear loci, as well as differences reminiscent of conflicting resolutions obtained from nrITS and cpDNA (Fig. 3). For example, monophyly of

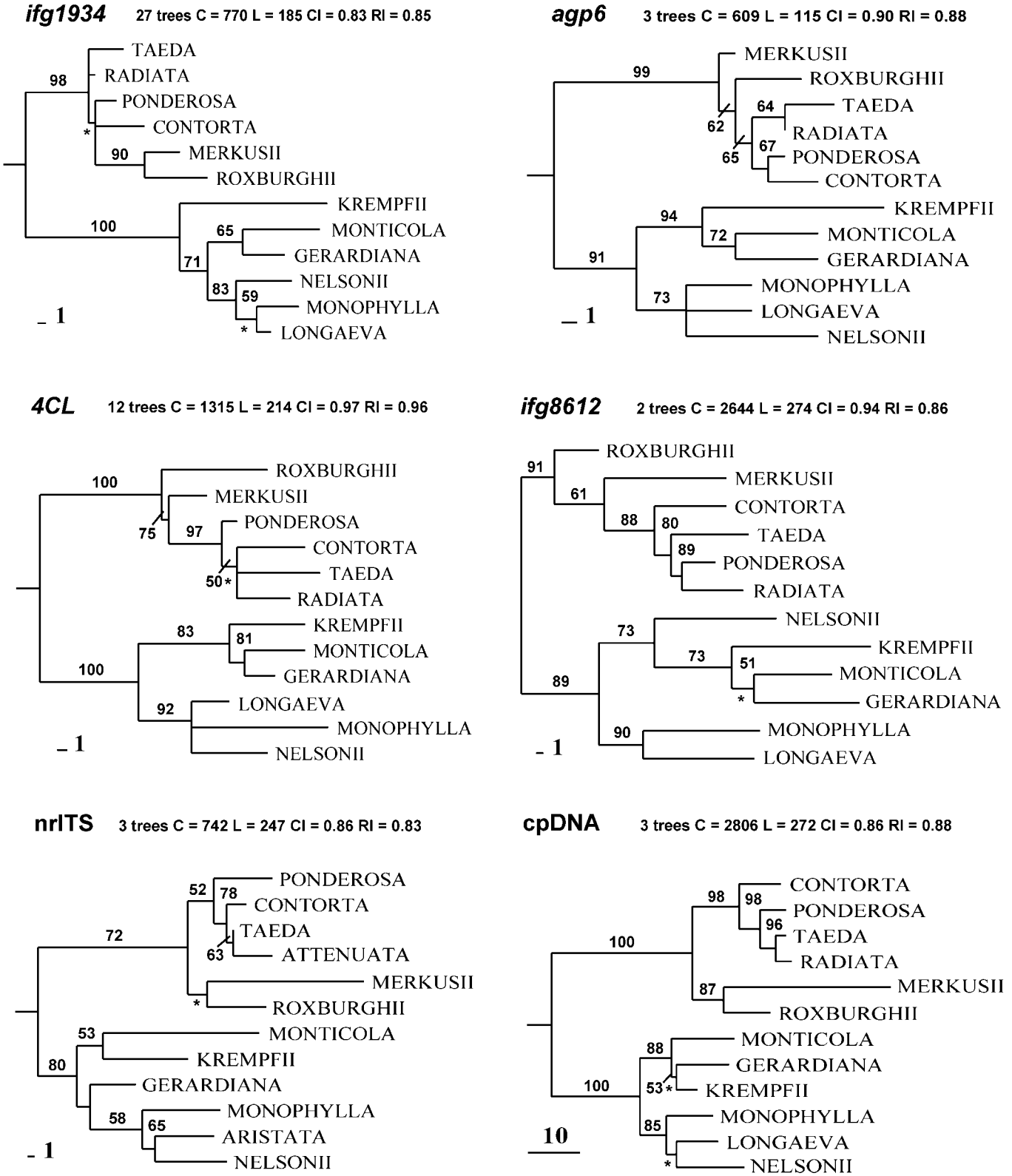

Fig. 3. Most-parsimonious trees derived from individual loci. Trees selected for presentation are those that most closely match the topology of the maximum likelihood tree. Bootstrap values from 1000 replicates and TBR branch swapping are shown near nodes. All trees are rooted with *Picea* except *IFG8612*, which is midpoint rooted; for clarity the outgroup branch has been omitted. *Pinus aristata* and *P. attenuata* are used in the nrITS phylogeny as placeholders for *P. longaeva* and *P. radiata*, respectively. C = number of characters, L = length of trees, CI = consistency index, RI = retention index. Asterisks (*) indicate nodes that collapse in the strict consensus tree.

TABLE 2. Pairwise incongruence length difference test (ILD) results. *Picea* included in the analysis.

Data set	*IFG1934*	*AGP6*	*4CL*	*IFG8612*	nrITS
AGP6	0.812				
4CL	0.610	0.702			
IFG8612	0.390	0.143	0.079		
nrITS	0.358	0.438	0.313	0.003**	
cpDNA	0.231	0.323	0.045*	0.003**	0.060

Note: * significant at $P \leq 0.05$; ** significant at $P \leq 0.01$.

subgenera *Strobus* and *Pinus* is supported by all low-copy nuclear genes (bootstrap support [BS] values from 89 to 100%) and is in agreement with both nrITS and cpDNA. Within subg. *Strobus*, sect. *Parrya* (*P. nelsonii*, *P. longaeva*, and *P. monophylla*) is supported as monophyletic in all low-copy gene trees except *IFG8612*. Data from *IFG8612* resolve *P. nelsonii* in a novel position sister to sect. *Quinquefoliae* (*P. monticola*, *P. gerardiana*, and *P. krempfii*), but this resolution shows low support (73% BS). The monophyly of sect. *Quinquefoliae* is also supported in three of the four low-copy nuclear genes. In *IFG1934*, *P. krempfii* is sister to the remaining sections of subg. *Strobus*, although this resolution has low support (71% BS). In general, the consensus resolution emerging from three of four low-copy nuclear loci is that these taxonomic sections are monophyletic. This resolution has strong support from the cpDNA data set, and the apparently conflicting resolution from nrITS involves the unique placement of *P. gerardiana* with its low character support.

Equivocal resolutions also occur in trees derived from low-copy markers, and the conflict between these markers does little to resolve discrepancies between nrITS and cpDNA. For example, the conflict between nrITS and cpDNA regarding the position of *P. contorta* remains unresolved. Likewise, within subg. *Pinus*, *P. merkusii* and *P. roxburghii*, are resolved as sister taxa by *IFG1934* with high support (90% BS). This resolution is supported by cpDNA data (87% BS), although the nrITS does not resolve the relationships of these two species. The remaining low-copy nuclear genes resolve *P. merkusii* and *P. roxburghii* as paraphyletic, with *4CL* showing moderate bootstrap support (75%).

Conflict among low-copy nuclear genes, nrITS and cpDNA—Results from the ILD test (Table 2) show that significant incongruence among the low-copy nuclear loci is not indicated. If these comparisons are extended to include nrITS, only the nrITS vs. *IFG8612* comparison ranks as significant ($P = 0.003$), indicating a general lack of incongruence between most nuclear loci and nrITS. Comparisons between nuclear genes and cpDNA reveals that *IFG8612* shows significant incongruence with plastid data ($P = 0.003$) and that *4CL* is near the threshold of significance ($P = 0.045$).

For an alternative assessment of conflict, we evaluated the sensitivity of each data set to topological change using the parsimony-based WSR test. Results from this test (Table 3) show a number of one-way comparisons that are significant, including comparisons involving *IFG8612* (5 pairwise comparisons), nrITS (4 comparisons), cpDNA (4 comparisons), and *AGP6* and *4CL* (1 comparison each). Of these comparisons, only *IFG8612* vs. cpDNA was significant in two directions. For example, constraining cpDNA data to alternative topologies led to inferences of topological conflict with three data sets (*IFG8612*, $P = 0.004$; *4CL*, $P = 0.023$; nrITS, $P = 0.010$); conflict was only reflected in the alternative test when *IFG8612* data are constrained to the cpDNA topology ($P = 0.027$). While a significant test in one direction may be indicative of potential conflict, there have been recommendations to recognize conflict as ''significant'' only when it is observed in two directions (Johnson and Soltis, 1998). Using these criteria, the sole case of clear partition conflict involves the nuclear locus *IFG8612* and cpDNA.

In total, results from both tests (ILD, WSR) show that the low-copy nuclear gene data sets do not exhibit strong conflict and that they represent a reasonable partition for combination. The merit of combining the low-copy genes with other partitions, such as nrITS or cpDNA, is debatable because both data sets show conflict with one of the four low-copy nuclear loci (*IFG8612*). To explore the impact of potentially conflicting data in our phylogenetic analyses, we combined the low-copy data with nrITS and cpDNA.

Topological comparisons between low-copy nuclear data, nrITS, and cpDNA—Combining the four low-copy nuclear genes into a single data set (5338 aligned characters) yields three MP trees that fully resolve *Pinus* subsectional representatives (Fig. 4A). In contrast to trees derived from individual genes, many clades exhibit strong character support (>90% BS), and eight of 10 potential nodes show BS values over 70%. Subgenera *Pinus* and *Strobus* are strongly supported (100% BS), as are three of the four sections. The sole exception among sections is support for sect. *Pinus*, which is low

TABLE 3. Templeton test results of six loci and two combined data sets (one including all low-copy nuclear genes and the other a concatenation of all six loci). Outgroup rooting was used for all topologies.

		Constraint topology							
Data set	No. trees	*IFG1934*	*AGP6*	*4CL*	*IFG8612*	nrITS	cpDNA	Low copy	All loci
IFG1934	50		0.469	0.542	0.503	0.491	0.635	0.629	0.629
AGP6	6	0.152		0.180	0.170	0.057	0.182	0.393	0.393
4CL	90	0.205	0.292		0.164	0.294	0.164	0.564	0.292
IFG8612	1	0.102	0.024*	0.101		0.012*	0.027*	0.180	0.118
nrITS	4	0.791	0.774	0.791	0.415		0.944	0.992	0.828
cpDNA	2	0.078	0.137	0.023*	0.004**	0.010**		0.846	0.846
Low copy	1	0.826	0.787	1.000	0.719	0.039*	0.206		1.000
All loci	3	0.058	0.063	0.145	0.011*	0.019*	0.234	1.000	

Note: * significant at $P \leq 0.05$; ** significant at $P \leq 0.01$; boxed values are significant in both directions following sequential Bonferroni correction.

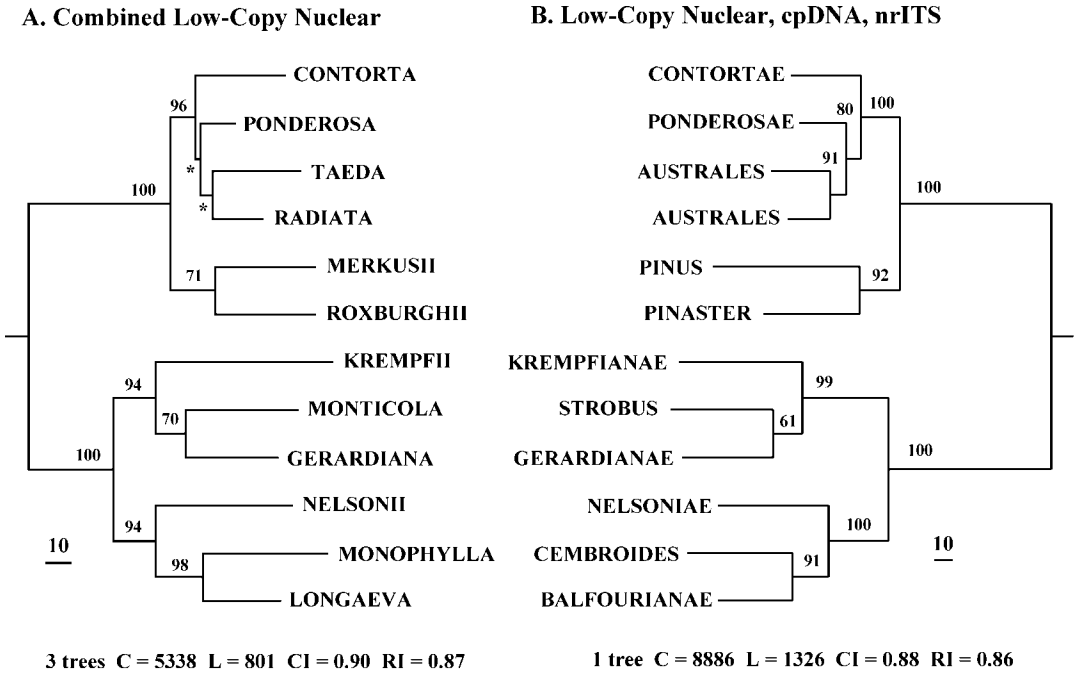

Fig. 4. Most-parsimonious trees derived from four low-copy nuclear loci combined (A), and cpDNA [*matK* + *rbcL*], nrITS, and four low-copy nuclear loci combined (B). Species names are used in (A) and their respective subsections are used in (B); the topologies are identical. Bootstrap values from 1000 replicates and TBR branch swapping are shown near nodes. Both trees are rooted with *Picea*; for clarity the outgroup branch has been omitted. C = number of characters, L = length of trees, CI = consistency index, RI = retention index. Asterisks (*) indicate nodes that collapse in the strict consensus tree.

(71% BS) in this comparison. At the level of subsectional representatives, sect. *Trifoliae* remains unresolved (Fig. 4A) with *P. contorta*, *P. ponderosa*, and *P. taeda*/*P. radiata* forming a trichotomy in the strict consensus of three MP trees (not shown).

This strongly supported resolution provides an opportunity to reassess the outstanding questions raised in previous analyses of *Pinus* nrITS and cpDNA. For example, resolution of subg. *Pinus* based on nrITS and cpDNA differs in two key aspects. First, the basal divergence inferred from previous analyses of nrITS is a grade, whereby sect. *Pinus* is paraphyletic due to the absence of monophyly between *P. merkusii* and *P. roxburghii*. This is not the case in our pruned nrITS analysis where *P. merkusii* and *P. roxburghii* are monophyletic, however, both alternatives have low support values. In contrast, cpDNA shows sect. *Pinus* to be the monophyletic sister group to sect. *Trifoliae* (87% BS). The combined low-copy nuclear data (Fig. 4A) indicate that sect. *Pinus* has moderately low support (71% BS) as a monophyletic lineage, a finding that supports the cpDNA topology (but does not explicitly contradict the pruned nrITS tree; Fig. 3). The second point of disagreement in subg. *Pinus* involves the resolution of subsect. *Contortae* (*P. contorta*), either as sister to subsect. *Australes* (*P. taeda*, *P. attenuata*) with nrITS, or as the sister lineage to the remainder of sect. *Trifoliae* with cpDNA. The combined low-copy nuclear data (Fig. 4A) clearly resolves *P. contorta* as a member of sect. *Trifoliae*, but it is one lineage in a trichotomy that includes the remaining samples from subsects. *Ponderosae*, *Australes*, and *Attenuatae*. Among the low-copy nuclear loci, only *IFG8612* provides a clear resolution of *P. contorta*, placing it sister to the remaining members of sect. *Trifoliae* as is observed with nrITS (Fig. 3). The remaining relationships within sect. *Trifoliae* of the *IFG8612* tree contradict a substantial body of prior molecular evidence (Strauss and Doerksen, 1990; Liston et al., 1999; Geada López et al., 2002; Gernandt et al., 2005), principally the paraphyly of subsect. *Australes* relative to *P. ponderosa*. Due to this lack of consistency, we consider the relationship between subsect. *Contortae* and the remaining lineages within sect. *Trifoliae* unresolved.

Resolution within subg. *Strobus* based on nrITS and cpDNA differs substantially with respect to relationships within sect. *Quinquefoliae*, specifically the placement of subsects. *Gerardianae* (*P. gerardiana*) and *Krempfianae* (*P. krempfii*). Whereas nrITS resolves subsect. *Krempfianae* sister to subsect. *Strobus* (*P. monticola*) and subsect. *Gerardianae* in a weakly supported clade with sect. *Parrya*, cpDNA resolves subsect. *Strobus* as sister to a weakly supported clade of subsects. *Gerardianae* and *Krempfianae*. Neither of these resolutions have strong support. Indeed, the subg. *Strobus* clade inferred from nrITS collapses to a polytomy if one only considers nodes with >65% bootstrap support, and the same is true for sect. *Quinquefoliae* relationships in the cpDNA tree. In contrast, the combined low-copy phylogeny reflects the consensus of individual loci by placing subsects. *Strobus*, *Gerardianae*, and *Krempfianae* in a well-supported clade (94% BS) with subsect. *Krempfianae* sister to the other two lineages (70% BS). This result adds support to recent phylogenetic and taxonomic investigations based on cpDNA that identify subsects. *Strobus*, *Gerardianae*, and *Krempfianae* as a distinct monophyletic lineage. It differs markedly, however, from the cpDNA resolution in that the morphologically distinct subsect. *Krempfianae* is resolved as sister to subsects. *Strobus* and *Gerardianae*. While support for this node is moderately low (70% BS), the fact that individual analysis of all low-copy loci show this same

TABLE 4. Comparison of three data sets used in reconstructing *Pinus* phylogeny. *Picea* not included in statistics.

Characteristic	cpDNA	nrITS	Low-copy
Number of genes	2	Tandem arrays	4
Average base pairs	2806	606	3451
Phylogenetically informative characters	88 (3.1%)	61 (10.1%)	206 (6.0%)
Linkage groups	1	1	4
Consistency index	0.86	0.86	0.90
Resolved nodes in strict consensus	8	9	8
Nodes with >70% bootstrap support (10 nodes max)	8	3	8
Nodes with >90% bootstrap support (10 nodes max)	5	0	6
%G + C	40.4	55.4	51.0
Average *p*-distance	0.017	0.064	0.038

resolution (individual BS = 51–81%; Fig. 3) adds measurably to our confidence in this finding.

The low-copy nuclear resolution of subsect. *Nelsoniae* (*P. nelsonii*) in Fig. 4A underscores the power (and the unexpected outcome) of combining data. The placement of this taxon in the low-copy tree is unambiguous, because it is included in a clade with (94% BS), and sister to (98% BS), the combined lineages of subsects. *Cembroides* (*P. monophylla*) and *Balforianae* (*P. longaeva*). This resolution conflicts with cpDNA and nrITS trees, which show uncharacteristic agreement (but low support) for subsect. *Nelsoniae* as sister to subsect. *Balfourianae*. Critically, the topology recovered in the low-copy nuclear data set is supported by only one of the individual nuclear loci, *IFG1934* (59% BS). In contrast, *AGP6* and *4CL* show these lineages as an unresolved trichotomy, and *IFG8612* places subsect. *Nelsoniae* in an unusual location sister to sect. *Quinquefoliae*. The final placement and enhanced support of the resolution of subsect. *Nelsoniae* through combined analysis is somewhat unexpected, arising through the accumulation of modest phylogenetic signal across all individual loci.

Analysis of combined nrITS, cpDNA, and low-copy nuclear data sets—As a final analysis, we combined data sets consisting of low-copy nuclear genes, nrITS, and cpDNA to explore how a global analysis would be resolved, with particular attention focused on the possible sources of conflict revealed by ILD and WSR tests (Tables 2, 3). Analysis of these five nuclear loci and one cpDNA partition as a single data set (8886 aligned characters) produces one MP tree that fully resolves *Pinus* subsectional representatives (Fig. 4B). The most striking feature of this analysis is that adding nrITS and cpDNA data did not change the topology of the combined low-copy tree. However, BS support values show that the addition of nrITS and cpDNA data improve the nuclear resolution (increased BS values) at some nodes, while adding to character conflict (decreased BS values) at other nodes. The addition of nrITS and cpDNA data uniformly increased support for sectional nodes, with nodal BS values increasing by 4 to 21%. Relationships within sect. *Trifoliae*—unresolved in the combined low-copy data set—are enhanced by the addition of cpDNA, resulting in a cpDNA-like resolution of these taxa, including the position of subsect. *Contortae*. The support values within sect. *Trifoliae* are reduced from the cpDNA tree (a result of the strong disagreement with nrITS), particularly at the node separating subsect. *Contortae* from the remainder of sect. *Trifoliae*. Topological relationships within sects. *Quinquefoliae* and *Parrya* do not change with the addition of nrITS and cpDNA data, but support values for the clades of *P. monticola*/*P. gerardiana* and *P. monophylla*/*P. longaeva* are reduced from the combined low-copy analysis due to the inclusion of conflicting characters from both nrITS and cpDNA.

DISCUSSION

Pine phylogeny and taxonomic sampling—This is the first study to confidently resolve the relationships among the subsections of subg. *Strobus*, including the positions of *P. krempfii* and *P. nelsonii*. In subg. *Pinus* the monophyly of sect. *Pinus* is weakly supported and the position of sect. *Contortae* remains equivocal. Additional taxon sampling across the genus may stabilize uncertain relationships in deeper divergence events by breaking up long branches and adding additional synapomorphic sites (Pollock et al., 2002; but see Rosenburg and Kumar, 2001). A coarse estimate of the influence of taxon density can be investigated by studying the topological differences in the cpDNA and nrITS trees presented here (Fig. 3) and their original, taxonomically dense analyses (Liston et al., 1999; Gernandt et al., 2005). In the case of cpDNA (Fig. 3), the topology is in good agreement with the topology of Gernandt et al. (2005) following pruning of 89 species. The same is true of the nrITS subsectional topology following the removal of 35 species. Minor differences in topology between trees in Fig. 3 and their original citations are restricted to nodes with low support (<61% BS). These results suggest that our limited taxonomic sampling does not bias the resolution of subsectional phylogeny.

Which loci have the greatest phylogenetic utility in Pinus?—Molecular studies of *Pinus* using cpDNA and nrDNA have returned conflicting topologies at deeper nodes and largely failed to resolve the numerous and possibly rapid radiations that comprise the terminal clades. These results highlight the need for alternative data sources that can be used to reduce phylogenetic uncertainty by increasing resolution and support. The data used in this study represent six independent molecular sources, each of which have different attributes—both positive and negative—with regard to molecular phylogenetic analysis (Tables 1, 4; Fig. 5).

Among the loci screened in this study, nrITS shows the highest average *p*-distance among pine species (0.064), and a higher proportion of phylogenetically informative sites (10.1%) than either the low-copy (6.0%) or cpDNA sequences (3.1%, Tables 1, 4). While this high degree of variation appears favorable by divergence measures, nrITS resolves pines with generally low bootstrap support values (Table 4, Fig. 3); for these exemplars, only three nodes have greater than 65% bootstrap support, and no nodes have greater than 80% sup-

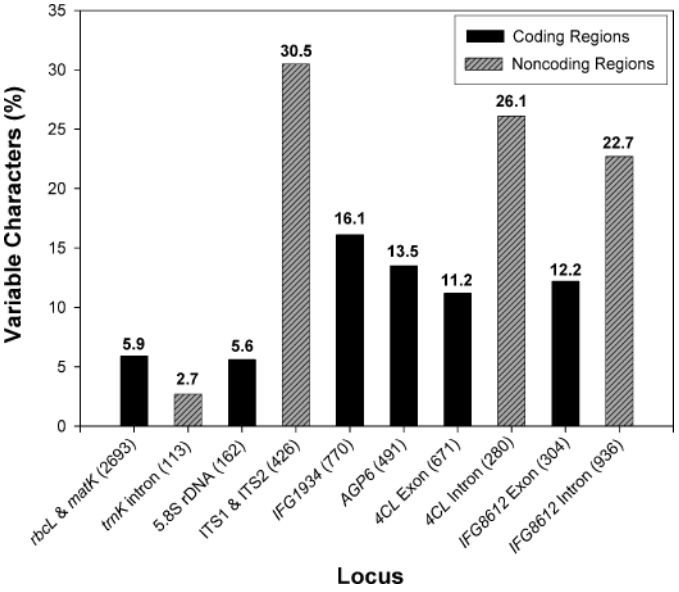

Fig. 5. Percentage variable characters by coding and noncoding regions for the six data sets examined in this study. Percentage variable characters are shown at the top of each bar; average number of characters is in parentheses after the locus name.

port. It could be argued that additional sampling of nrITS could improve resolution and support, especially since our sample of ~630 bp represents one-fifth of the 3000+ bp ITS1-ITS2 region in *Pinus* (Liston et al., 1999; Gernandt et al., 2001). Despite the promise of additional characters, Gernandt et al. (2001) showed that concerted evolution is weak in the 5′ region of ITS1. The absence of complete concerted evolution is increasingly reported in phylogenetic studies (summarized in Álvarez and Wendel, 2003; Campbell et al., 2005). In such cases, allelic heterozygosity and non-allelic (paralogous) diversity can accumulate, placing practical limitations on the use and interpretation of ITS variation in a phylogenetic context. For these reasons, we believe that nrITS is unlikely to provide new insights into *Pinus* phylogeny.

Chloroplast coding DNA generally evolves at a conservative rate, making it useful for the resolution of more ancient divergence events (Clegg et al., 1994; Small et al., 2004). Average *p*-distances among the species in this study were 0.017 for two cpDNA loci (*matK, rbcL*), a value that is about one-fourth the rate of nrITS and one-half the rate of the four nuclear genes. Where this genome fails to provide information is at cladogenic events involving the species-rich groups (Krupkin et al., 1996; Geada López et al., 2002; Gernandt et al., 2005). Increased sampling of noncoding regions has the potential to increase resolution (Shaw et al., 2005; D. Gernandt, Universidad Autonoma Estado Hidalgo, personal communication). However, even if a resolved topology is obtained with cpDNA, this genome represents a single linkage group. As has been shown, over-reliance on cpDNA (or any single data set) can provide strongly supported yet erroneous phylogenetic inferences because it is susceptible to lineage sorting of ancestral polymorphisms, as well as to chloroplast capture through introgression (reviewed in Wendel and Doyle, 1998). Both phenomena can lead to inaccurate organismal phylogenies that are difficult to verify in the absence of alternative hypotheses.

The four nuclear genes used in this study diverge at rates intermediate to nrITS and cpDNA, with locus structure being highly influential in determining overall variability. Among low-copy nuclear loci, there is an eight-fold difference between the slowest and fastest evolving regions (*AGP6* in subg. *Pinus* and *4CL* intron in subg. *Strobus*, respectively). Average pairwise distances (Table 1) across appropriate clades (genus-wide for coding regions, within subgenera for noncoding regions) indicate that exons diverge 2.1 times faster than cpDNA, and introns diverge 1.3 times faster than nrITS. Introns from *4CL* and *IFG8612* are 1.9–2.3 times more variable than their respective coding regions (Table 1).

In general, introns are attractive targets for evaluating relationships among closely related taxa because they diverge at relatively rapid rates (Small et al., 2004). This will be desirable (and necessary) as we focus on reconstructing relationships among closely related species, such as those in subsections *Australes* and *Ponderosae*. It should be noted that inferred rates of divergence alone can be misleading with respect to phylogenetic utility or informativeness. For example, the *4CL* intron shows the highest rate of divergence and greatest number of phylogenetically informative sites among all sequences examined (Table 1) yet it shows ~72% A + T. This distorted base composition may limit the resolution of taxa due to the reduced genetic alphabet and increased incidence of homoplasy. In addition, the high lability of this region leads to alignment difficulties due to A/T microsatellites and single sequence repeats (Fig. 2). We predict that introns with roughly equal base frequencies (such as *IFG8612*) will be preferred even if they show a lower rate of change.

For gaining insight into the overall phylogenetic pattern of *Pinus* subsections (a group containing both ancient and recent divergence events), a combination of coding and noncoding regions is required to obtain a (relatively) well-supported phylogeny across the genus (Fig. 4A). For example, the ancient divergence event leading to the formation of subgenera cannot be revealed using intron sequences alone (such as *IFG8612*), due to the absence of recognizable homology in these regions (Fig. 2). Exon sequences also have limitations, although in this case selective constraints limit the degree of character change in terminal taxa. Considering that the greatest phylogenetic uncertainty in *Pinus* resides in terminal lineages (e.g., sections and subsections, down to species), we suggest that future emphasis be placed on rapidly evolving noncoding regions. While length variation among more divergent taxa is likely to limit the breadth of phylogenetic comparisons (e.g., across subgenera or sections), this limitation can be circumvented by the judicious use of slower evolving exon regions.

The greatest value of adding additional nuclear genes for studying relationships among pines is that they present multiple evolutionarily independent perspectives that can be compared to hypotheses derived from nrITS and cpDNA. These loci reside on three different nuclear chromosomes, and the two loci located on the same chromosome—*IFG8612* and *IFG1934*—are sufficiently distant (>60 centiMorgans; Krutovsky et al., 2004) that they are unlikely to show linkage disequilibrium at the species level (Brown et al., 2004). The general agreement between these independent nuclear and cpDNA loci strengthens earlier findings that *Pinus* includes two main lineages (subg. *Pinus* and *Strobus*), both of which can be divided into two sublineages (sects. *Pinus* and *Trifoliae* within subg. *Pinus*; sects. *Quinquefoliae* and *Parrya* within subg. *Strobus*). Agreement between low-copy nuclear loci and

cpDNA that differs with nrITS (e.g., monophyly of subsects. *Strobus*, *Gerardianae*, and *Krempfianae*; Figs. 3, 4A) provides support for earlier inferences that have been identified by cpDNA alone (Gernandt et al., 2005). Critically, disagreement between low-copy genes, cpDNA, and nrITS (e.g., relationships in sects. *Parrya* and *Trifoliae*) highlights the most problematic nodes and underscores the need for focused sampling with respect to taxa and characters.

In light of the overall lack of resolution within subsections in prior molecular studies of pines, we suggest that useful molecular information is likely limited to noncoding regions of cpDNA and low-copy nuclear loci with a high proportion of silent sites, such as *IFG8612*. These genomic resources are readily available (Brown et al., 2004; Krutovsky et al., 2004; Small et al., 2004), and they are certain to be applied with increasing frequency in *Pinus* and related conifers. Traditionally, one of the most challenging aspects of working with low-copy nuclear genes is assessing orthology in the presence of high heterozygosity (Small et al., 2004). The availability of haploid megagametophyte tissue in *Pinus* seeds (and most gymnosperms) simplifies this task, making it possible to amplify single haplotypes (e.g., Brown et al., 2004). If amplification products from megagametophyte tissue show ''heterozygosity,'' it is a clear indication that paralogous loci are being amplified.

Molecular data and prospects for resolving phylogenetic relationships within pines—The use of multiple independent loci to explore phylogenetic relationships in coniferous plants is a recent advancement (Wang et al., 2000b; Kusumi et al., 2002), although the use of mapped nuclear loci has only been accomplished in angiosperm groups (Chee et al., 1995; Cronn et al., 2002; Álvarez et al., 2005). Studies of this nature that compare and contrast the properties and utility of a set of loci are important in establishing cost effective approaches for advancing the resolution of evolutionary relationships within any group of taxa (see Small et al., 2004). With the burgeoning list of nuclear loci available for molecular genetic applications in economically important groups (Fulton et al., 2002; Krutovsky et al., 2004), two hurdles need to be overcome in order to gain a finer evolutionary perspective of pines: (1) reconciling the inevitable conflict that will arise as more genes are added and (2) resolving widespread, possibly rapid radiations that characterize many subsections.

The first issue—how to interpret a phylogeny in the presence of conflict—is a topic of continuing debate in molecular phylogenetics (Bull et al., 1993; Miyamoto, 1996; Suchard et al., 2003). Increasingly, phylogenetic studies are being performed on very large samples ranging from tens of genes (Cronn et al., 2002; Takezaki et al., 2004) to entire genomes (Rokas et al., 2003; reviewed in Eisen and Fraser, 2003). Results show that incongruence among gene trees is to be expected when estimates are based on a small number of genetically independent data sources; indeed, strongly supported *incorrect* phylogenies can be derived from a few genes if they show non-independence (e.g., linkage or similar functional constraints). Studies in yeast (Rokas et al., 2003) show that concatenation of a relatively small number of genes (perhaps 20) can produce stable phylogenies that minimize the influence of evolutionary noise and selective effects. In light of these findings, it seems clear that taxonomic conclusions based on a single source of molecular phylogenetic data (e.g., cpDNA; Gernandt et al., 2005) should be considered working hypotheses awaiting confirmation, rather than a final product. The number of loci that will be required to produce a stable pine phylogeny remains unknown, but given the abundance of pine ESTs, the number of comparative linkage maps in *Pinus*, and the extensive history of seed collection/gene conservation activities in the forest genetics community, a genome-wide phylogenetic estimate of the pine phylogeny based on multiple nuclear loci is a realistic goal. The growing number of cross-genera comparisons in the Pinaceae (Wang et al., 2000b; Krutovsky et al., 2004) shows that similar studies are equally tractable in other coniferous genera.

The second obstacle to a resolved pine phylogeny is the numerous, apparently rapid radiations in Sect. *Trifoliae* and subsects. *Pinus*, *Cembroides*, and *Strobus* (Krupkin et al., 1996; Wang et al., 1999; Gernandt et al., 2001; Geada López et al., 2002; Gernandt et al., 2005; A. Willyard et al., unpublished manuscript). The problem is illustrated by the 48 species of sect. *Trifoliae*. Among the four representatives of this section (Appendix), average pairwise divergence at four low-copy nuclear genes is 1.5% over 5338 aligned bases, and there are 102 variable and 6 parsimony informative sites. Including additional species in future analyses will certainly shift variable sites to the ''phylogenetically informative'' category. Nevertheless, there are fewer than 16 synapomorphies uniting these taxa, and the terminal branches leading to *P. taeda* and *P. ponderosa* (representing the species-rich subsections *Australes* and *Ponderosae*) are only 34 and 14 characters long, respectively (Fig. 4A). This amount of variation is unlikely to be sufficient to provide robust character support for resolving the 24 remaining species of subsect. *Australes* and 17 species of subsect. *Ponderosae*. As noted, adding nrITS (Liston et al., 1999, 2003) and cpDNA (Gernandt et al., 2005) could add resolution, but it seems likely that many of these groups will remain multi-species polytomies even after the acquisition of more data. From our current temporal vantage point, such divergences appear nearly instantaneous (e.g., Fig. 4A). We have calculated estimated divergence times for the four lineages of sect. *Trifoliae* using nine nuclear gene sequences (A. Willyard et al., unpublished manuscript); irrespective of the calibration method, the four lineages of sect. *Trifoliae* appear to have diverged within a brief time, perhaps as little as 2–5 million years during the Miocene. This perspective is valuable, because it suggests that many species groups diversified over a relatively narrow time frame, perhaps in response to climate change or adaptive ecotypic divergence. The near-absence of interspecific crossing barriers within these subsections (Little and Righter, 1965; Garrett, 1979; Critchfield, 1986) attests to their limited genetic divergence and close phylogenetic affinity.

A related, yet possibly more vexing, obstacle to resolving recently diverged pine groups is the apparent longevity of allele lineages in conifers. Standard phylogenetic methods rely on the assumption that intraspecific divergence is low relative to interspecific divergence; restated, the branches derived from gene geneologies need to be long compared to within-species coalescence times. This ''long branch assumption'' has recently been shown to be violated in species of *Picea* across three low-copy genes (Bouillé and Bousquet, 2005), and at nrITS in *Pinus* and *Picea* (Gernandt et al., 2001; Campbell et al., 2005). Comparisons of three nuclear loci from *Picea abies*, *P. glauca*, and *P. mariana* show that non-coalescence among species was commonplace (Bouillé and Bousquet, 2005). The estimated coalescence time between randomly selected alleles

from any two species ranged from 10 to 18 million years ago, and these estimates overlap with estimated species divergence times (13–20 mya). Similarly, preliminary genetic data gathered from the comparably aged North American five-needle pines (A. Willyard et al., unpublished manuscript; J. Syring, unpublished data) shows a similar pattern for two of the loci used in this study (*IFG8612*, *AGP6*). As might be expected, narrowly restricted species (e.g., *P. albicaulis*, a timberline endemic) exhibit species-level allelic coalescence, but widespread species (e.g., *P. lambertiana*, sugar pine) share allele lineages with other North American (even East Asian) pine species.

If trans-species-shared polymorphisms are commonplace in *Pinus* and *Picea*, it seems likely that similar trends will be revealed in other plant groups with long fossil representation, similar life histories, and large population sizes. Bouillé and Bousquet (2005) cautioned that widespread lack of species-level coalescence may impede phylogenetic estimation in conifers and other trees with similar ecological and reproductive traits. These authors ''call for caution in estimating congeneric species phylogenies from nuclear gene sequences'' in conifers. While we acknowledge that their results are likely to apply to closely related pine species (within subsections), the congruence of results from four low-copy independent data sources presented herein shows that these markers confidentally resolve phylogenetic relationships among more distantly related pines. This finding underscores the importance of historical events on the utility of nuclear markers and in the interpretation of phylogenies derived from their use. Considering the dramatic impact intraspecific polymorphism can have on phylogenetic results, we predict that a final resolution of the pine phylogeny will incorporate many sources of data (such as those described in this paper) and a mix of traditional phylogenetic approaches (parsimony, likelihood) and coalescent analyses to evaluate the relative age of species and their complex genealogical relationships.

LITERATURE CITED

ÁLVAREZ, I., R. C. CRONN, AND J. F. WENDEL. 2005. Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. *Plant Systematics and Evolution* 252: 199–214.

ÁLVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–34.

BOUILLÉ, M., AND J. BOUSQUET. 2005. Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. *American Journal of Botany* 92: 63–73.

BROWN, G. R., E. E. KADEL III, D. L. BASSONI, K. L. KIEHNE, B. TEMESGEN, J. P. VAN BUIJTENEN, M. M. SEWELL, K. A. MARSHALL, AND D. B. NEALE. 2001. Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics* 159: 799–809.

BROWN, G. R., G. P. GILL, R. J. KUNTZ, C. H. LANGLEY, AND D. B. NEALE. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences, USA* 101: 15255–15260.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Systematic Biology* 42: 384–397.

CAMPBELL, C. S., W. A. WRIGHT, M. COX, T. F. VINING, C. S. MAJOR, AND M. P. ARSENAULT. 2005. Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. *Molecular Phylogenetics and Evolution* 35: 165–185.

CHAW, S. M., A. ZHARKIKH, H. M. SUNG, T. C. LAU, AND W. H. LI. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution* 14: 56–68.

CHEE, P. W., M. LAVIN, AND L. E. TALBERT. 1995. Molecular analysis of evolutionary patterns in U genome wild wheats. *Genome* 38: 290–297.

CHEN, J., C. G. TAUER, AND Y. HUANG. 2002. Nucleotide sequences of the internal transcribed spacers and 5.8S region of nuclear ribosomal DNA in *Pinus taeda* L. and *Pinus echinata* Mill. *DNA Sequence* 13: 129–131.

CLEGG, M. T., B. S. GAUT, G. H. LEARN JR., AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

CRITCHFIELD, W. B. 1986. Hybridization and classification of the white pines (*Pinus* section *Strobus*). *Taxon* 35: 647–656.

CRONN, R. C., R. L. SMALL, T. HASELKORN, AND J. F. WENDEL. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89: 707–725.

CRONN, R. C., R. L. SMALL, T. HASELKORN, AND J. F. WENDEL. 2003. Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. *Evolution* 57: 2475–2489.

CUNNINGHAM, C. W. 1997. Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* 14: 733–740.

DEVEY, M. E., S. D. CARSON, M. F. NOLAN, A. C. MATHESON, C. TERIINI, AND J. HOHEPA. 2004. QTL associations for density and diameter in *Pinus radiata* and the potential for marker-aided selection. *Theoretical Applied Genetics* 108: 516–524.

DON, R. H., P. T. COX, B. J. WAINWRIGHT, K. BAKER, AND J. S. MATTICK. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 19: 4008.

DVORAK, W. S., A. P. JORDAN, G. P. HODGE, AND J. L. ROMERO. 2000. Assessing evolutionary relationships of pines in the *Oocarpae* and *Australes* subsections using RAPD markers. *New Forests* 20: 163–192.

EISEN, J. A., AND C. M. FRASER. 2003. Phylogenomics: intersection of evolution and genomics. *Science* 300: 1706–1707.

FARJON, A. 1990. A bibliography of conifers. Regnum Vegetabile 122. Koeltz Scientific Books, Koenigstein, Germany.

FARJON, A. 1998. World checklist and bibliography of conifers. Royal Botanical Gardens, Kew, Richmond, UK.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994. Testing significance of incongruence. *Cladistics* 10: 315–319.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

FRIESEN, N., A. BRANDES, AND J. S. HESLOP-HARRISON. 2001. Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. *Molecular Biology and Evolution* 18: 1176–1188.

FULTON, T. M., R. VAN DER HOEVEN, N. T. EANNETTA, AND S. D. TANKSLEY. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 14: 1457–1467.

GARRETT, P. W. 1979. Species hybridization in the genus *Pinus*. USDA Forest Service Research Paper NE-436. USDA. Forest Service Northeastern Forest Experiment Station, Broomall, Pennsylvania, USA.

GEADA LÓPEZ, G., K. KAMIYA, AND K. HARADA. 2002. Phylogenetic relationships of diploxylon pines (subgenus *Pinus*) based on plastid sequence data. *International Journal of Plant Sciences* 163: 737–747.

GERMANO, J., AND A. S. KLEIN. 1999. Using species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish between *Picea glauca*, *P. mariana*, and *P. rubens* from single needle samples. *Theoretical and Applied Genetics* 99: 37–49.

GERNANDT, D. S., A. LISTON, AND D. PIÑERO. 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications for molecular systematic studies of pine species complexes. *Molecular Phylogenetics and Evolution* 21: 449–467.

GERNANDT, D. S., A. LISTON, AND D. PIÑERO. 2003. Phylogenetics of *Pinus* subsections *Cembroides* and *Nelsoniae* inferred from cpDNA sequences. *Systematic Botany* 4: 657–673.

GERNANDT, D. S., G. GEADA LÓPEZ, S. O. GARCÍA, AND A. LISTON. 2005. Phylogeny and classification of *Pinus*. *Taxon* 54: 29–42.

GROTKOPP, E., M. REJMANEK, M. J. SANDERSON, AND T. L. ROST. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. *Evolution* 58: 1705–1729.

HALL, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.

HARRIS, E. E., AND T. R. DISOTELL. 1998. Nuclear gene trees and the phylogenetic relationships of the mangabeys (Primates: Papionini). *Molecular Biology and Evolution* 15: 892–900.

HART, J. A. 1987. A cladistic analysis of conifers: preliminary results. *Journal of the Arnold Arboretum* 68: 269–307.

HIPP, A. L., J. C. HALL, AND K. SYTSMA. 2004. Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. *Systematic Biology* 53: 81–89.

HUELSENBECK, J. P., J. J. BULL, AND C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. *Trends in Ecology and Evolution* 11: 152–158.

IGLESIAS, R. G., AND M. J. BABIANO. 1999. Isolation and characterization of a cDNA encoding a late-embryogenesis-abundant protein (Accession No. AJ012483) from *Pseudotsuga menziesii* seeds. *Plant Physiology* 119: 806.

JANSSON, S., AND P. GUSTAFSSON. 1990. Type I and Type II genes for the chlorophyll *a/b*-binding protein in the gymnosperm *Pinus sylvestris* (Scots pine): cDNA cloning and sequence analysis. *Plant Molecular Biology* 14: 287–296.

JOHNSON, L. A., AND D. E. SOLTIS. 1998. Assessing congruence: empirical examples from molecular data. *In* D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II: DNA sequencing, 297–348. Kluwer, Norwell, Massachusetts, USA.

KARALAMANGALA, R. R., AND D. L. NICKRENT. 1989. An electrophoretic study of representatives of subgenus *Diploxylon* of *Pinus*. *Canadian Journal of Botany* 67: 1750–1759.

KINLAW, C. S., AND D. B. NEALE. 1997. Complex gene families in pine genomes. *Trends in Plant Science* 2: 356–359.

KRUPKIN, A. B., A. LISTON, AND S. H. STRAUSS. 1996. Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. *American Journal of Botany* 83: 489–498.

KRUTOVSKY, K. V., M. TROGGIO, G. R. BROWN, K. D. JERMSTAD, AND D. B. NEALE. 2004. Comparative mapping in the Pinaceae. *Genetics* 168: 447–461.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, AND M. NEI. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244–1245.

KUSUMI, J., Y. TSUMURA, H. YOSHIMARU, AND H. TACHIDA. 2002. Molecular evolution of nuclear genes in Cupressaceae, a group of conifer trees. *Molecular Biology and Evolution* 19: 736–747.

LE MAITRE, D. C. 1998. Pines in cultivation: a global view. *In* D. M. Richardson [ed.], Ecology and biogeography of *Pinus*, 407–431. Cambridge University Press, New York, New York, USA.

LISTON, A., W. A. ROBINSON, D. PIÑERO, AND E. R. ALVAREZ-BUYLLA. 1999. Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Molecular Phylogenetics and Evolution* 11: 95–109.

LISTON, A., D. S. GERNANDT, T. F. VINING, C. S. CAMPBELL, AND D. PIÑERO. 2003. Molecular phylogeny of Pinaceae and *Pinus*. *Acta Horticulturae* 615: 107–114.

LITTLE, E. L. JR., AND F. I. RIGHTER. 1965. Botanical descriptions of forty artificial pine hybrids. USDA Forest Service Technical Bulletin 1345. USDA Forest Service, Washington, D.C., USA.

MALCOMBER, S. T. 2002. Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of rapid radiation in a widespread, morphologically diverse genus. *Evolution* 56: 42–57.

MIYAMOTO, M. M. 1996. A congruence study of molecular and morphological data for Eutherian mammals. *Molecular Phylogenetics and Evolution* 6: 373–390.

NKONGOLO, K. K., P. MICHAEL, AND W. S. GRATTON. 2002. Identification and characterization of RAPD markers inferring genetic relationships among pine species. *Genome* 45: 51–58.

POLLOCK, D. D., D. J. ZWICKL, J. S. MCGUIRE, AND D. M. HILLIS. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51: 664–671.

POSADA, D., AND K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

PRICE, R. A., A. LISTON, AND S. H. STRAUSS. 1998. Phylogeny and systematics of *Pinus*. *In* D. M. Richardson [ed.], Ecology and biogeography of *Pinus*, 49–68. Cambridge University Press, New York, New York, USA.

RICE, W. R. 1989. Analyzing tables of statistical tests. *Evolution* 43: 223–225.

RICHARDSON, D. M., AND P. W. RUNDEL. 1998. Ecology and biogeography of *Pinus*: an introduction. *In* D. M. Richardson [ed.], Ecology and biogeography of *Pinus*, 3–46. Cambridge University Press, New York, New York, USA.

ROKAS, A., B. L. WILLIAMS, N. KING, AND S. B. CARROLL. 2003. Genome scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.

ROSENBERG, M. S., AND S. KUMAR. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proceedings of the National Academy of Sciences, USA* 98: 10751–10756.

SANDERSON, M. J., AND H. B. SHAFFER. 2002. Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics* 33: 49–72.

SHAW, J., E. B. LICKEY, J. T. BECK, S. B. FARMER, W. LIU, J. MILLER, K. C. SIRIPUN, C. T. WINDER, E. E. SCHILLING, AND R. L. SMALL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Americn Journal of Botany* 92: 142–166.

SHOWALTER, A. M. 2001. Arabinogalactan-proteins: structure, expression, and function. *Cellular and Molecular Life Sciences* 58: 1399–1417.

SMALL, R. L., R. C. CRONN, AND J. F. WENDEL. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145–170.

SPRINGER, M. S., R. W. DEBRY, C. DOUADY, H. M. AMRINE, O. MADSEN, W. W. DE JONG, AND M. J. STANHOPE. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution* 18: 132–143.

STRAUSS, S. H., AND A. H. DOERKSEN. 1990. Restriction fragment analysis of pine phylogeny. *Evolution* 44: 1081–1096.

SUCHARD, M. A., C. M. R. KITCHEN, J. S. SINSHEIMER, AND R. E. WEISS. 2003. Hierarchical phylogenetic models for analyzing multipartite sequence data. *Systematic Biology* 52: 649–664.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods, version 4.0b10). Sinauer, Sunderland, Massachusetts, USA.

TAKEZAKI, N., F. FIGUEROA, Z. ZALESKA-RUTCZYNSKA, N. TAKAHATA, AND J. KLEIN. 2004. The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of forty-four nuclear genes. *Molecular Biology and Evolution* 21: 1512–1524.

TEMESGEN, B., G. R. BROWN, D. E. HARRY, C. S. KINLAW, M. M. SEWELL, AND D. B. NEALE. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda*). *Theoretical Applied Genetics* 102: 664–675.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37: 221–244.

WANG, X. R., Y. TSUMURA, H. YOSHIMARU, K. NAGASAKA, AND A. E. SZMIDT. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. *American Journal of Botany* 86: 1742–1753.

WANG, X. R., A. E. SZMIDT, AND H. N. NGUYEN. 2000a. The phylogenetic position of the endemic flat-needle pine *Pinus krempfii* (Pinaceae) from Vietnam, based on PCR-RFLP analysis of chloroplast DNA. *Plant Systematics and Evolution* 220: 21–36.

WANG, X. Q., D. C. TANK, AND T. SANG. 2000b. Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution* 17: 773–781.

WENDEL, J. F., AND J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. *In* D. Soltis, P. Soltis, and J. Doyle [eds.], Molecular systematics of plants II, 265–296. Chapman and Hall, New York, New York, USA.

WU, J., K. V. KRUTOVSKII, AND S. H. STRAUSS. 1999. Nuclear DNA diversity, population differentiation, and phylogenetic relationships in the California closed-cone pines based on RAPD and allozyme markers. *Genome* 42: 893–908.

YANG, Z. 1998. On the best evolutionary rate for phylogenetic analysis. *Systematic Biology* 47: 125–133.

YODER, A. D., J. A. IRWIN, AND B. A. PAYSEUR. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. *Systematic Biology* 50: 408–424.

ZHANG, X. H., AND V. L. CHIANG. 1997. Molecular cloning of 4-coumarate: coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. *Plant Physiology* 113: 65–74.

ZHANG, Y., G. BROWN, R. WHETTEN, C. A. LOOPSTRA, D. B. NEALE, M. J. KIELISZEWSKI, AND R. R. SEDEROFF. 2003. An arabinogalactan protein associated with secondary cell wall formation in differentiating xylem of loblolly pine. *Plant Molecular Biology* 52: 91–102.

ZHANG, Z. Y., AND D. Z. LI. 2004. Molecular phylogeny of section *Parrya* of *Pinus* (Pinaceae) based on chloroplast *mat*K gene sequence data. *Acta Botanica Sinica* 46: 171–179.

APPENDIX. Sampled *Pinus* species and outgroup *Picea*. Taxa are listed by subgenus. A dash indicates the region was not sampled (*IFG8612 Picea*), that sequences were taken from GenBank (*P. taeda*; see footnote), or that an alternative voucher source was used (*P. roxburghii*). Information for nrITS and cpDNA are reported in the original citations with modifications noted in the methods.

Taxon[a]; Section; Subsection; Voucher information[b]; *IFG1934, AGP6, 4CL, IFG8612.*

Subgenus *Pinus*

P. merkusii Junghuhn & de Vriese; *Pinus*; *Pinus*; Thailand, *X.R. Wang 956* (no voucher); AY617085, AY634320, AY634352, AY634337. ***P. roxburghii*** Sargent; *Pinus*; *Pinaster*; Nepal, *Kew 1979.06113* (K); —, —, AY634353, —. ***P. roxburghii*** Sargent; *Pinus*; *Pinaster*; India, *S.C. Garkofij s.n.*, RMP[c] 0416[e] (no voucher); AY617086, AY634321,—, AY634338. ***P. radiata*** D. Don; *Trifoliae*; *Australes*; California, USA, *RMP[c] 0418* (OSC); AY617083, AY634318, AY634350, AY634335. ***P. taeda*** Linnaeus; *Trifoliae*; *Australes*; Georgia, USA, *R. Price s.n.* (no voucher)[e,f]; —, —, —, AY634332. ***P. contorta*** Douglas ex Loudon; *Trifoliae*; *Contortae*; Oregon, USA, *A. Liston 1219* (OSC); AY617084, AY634319, AY634351, AY634336. ***P. ponderosa*** Douglas ex P. & C. Lawson; *Trifoliae*; *Ponderosae*; Oregon, USA, *F.C. Sorenson s.n.*, RMP[c] 0415 (OSC); AY617082, AY634317, AY634349, AY634334.

Subgenus *Strobus*

P. longaeva D.K. Bailey; *Parrya*; *Balfourianae*; California, USA, *D.S. Gernandt 03099* (OSC); AY617094, AY634329, AY634361, AY634346. ***P. monophylla*** Torrey & Fremont; *Parrya*; *Cembroides*; California, USA, *D.S. Gernandt 404* (OSC); AY617092, AY634327, AY634359, AY634344. ***P. nelsonii*** Shaw; *Parrya*; *Nelsoniae*; Mexico, *D.S. Gernandt & S. Ortiz 11298* (OSC & MEXU); AY617095, AY634330, AY634362, AY634347. ***P. gerardiana*** Wallich ex D. Don; *Quinquefoliae*; *Gerardianae*; Pakistan, *R. Businský 41105* (RILOG[d])[e]; DQ018375, DQ018373, DQ018377, DQ018379. ***P. krempfii*** Lecomte; *Quinquefoliae*; *Krempfianae*; Vietnam, *P. Thomas, First Darwin Expedition 242* (E); DQ018376, DQ018374, DQ018378, DQ018380. ***P. monticola*** Douglas ex D. Don; *Quinquefoliae*; *Strobus*; Oregon, USA, *J. Syring 1001* (OSC); AY617090, AY634325, AY634357, AY634342.

Outgroup

Picea sitchensis (Bong.) Carr.; Oregon, USA, *J. Syring 1002* (OSC); AY617096, AY634331, AY634363, —.

[a] Taxonomy follows Gernandt et al. (2005). Some taxonomists restrict *P. merkusii* to populations from the Phillipines and Sumatra, in which case the plant used here would be considered as *P. latteri* Mason.

[b] Herbarium acronyms follow *Index Herbariorum*: http://sciweb.nybg.org/science2/IndexHerbariorum.asp.

[c] RMP = Resource Management and Production Division of the Pacific Northwest Forest Science Center in Corvallis, Oregon.

[d] RILOG = Silva Tarouca Research Institute for Landscape and Ornamental Gardening, 252 43 Průhonice, Czech Republic.

[e] DNA extracted from megagametophyte rather than leaf tissue.

[f] Sequences for *P. taeda* from *IFG1934, AGP6*, and *4CL* were taken from GenBank (H75103, AF101785, U39405, respectively).