MOLECULAR BIOLOGY AND EVOLUTION

Fossil Calibration of Molecular Divergence Infers a Moderate Mutation Rate and Recent Radiations for *Pinus*

Ann Willyard, John Syring,† David S. Gernandt,‡ Aaron Liston,* and Richard Cronn§*

*Department of Botany and Plant Pathology, Oregon State University; †Department of Biological and Physical Sciences, Montana State University; ‡Centro de Investigaciones Biológicas, Universidad Autónoma del Estado de Hidalgo, Pachuca, Hidalgo, Mexico; and §Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon

Silent mutation rate estimates for *Pinus* vary 50-fold, ranging from angiosperm-like to among the slowest reported for plants. These differences either reflect extraordinary genomic processes or inconsistent fossil calibration, and they have important consequences for population and biogeographical inferences. Here we estimate mutation rates from 4 *Pinus* species that represent the major lineages using 11 nuclear and 4 chloroplast loci. Calibration was tested at the divergence of *Pinus* subgenera with the oldest leaf fossil from subg. *Strobus* (Eocene; 45 MYA) or a recently published subg. *Strobus* wood fossil (Cretaceous; 85 MYA). These calibrations place the origin of *Pinus* 190–102 MYA and give absolute silent rate estimates of $0.70–1.31 \times 10^{-9}$ and $0.22–0.42 \times 10^{-9}$ substitutions·site^{-1}·year^{-1} for the nuclear and chloroplast genomes, respectively. These rates are approximately 4- to 20-fold slower than angiosperms, but unlike many previous estimates, they are more consistent with the high per-generation deleterious mutation rates observed in pines. Chronograms from nuclear and chloroplast genomes show that the divergence of subgenera accounts for about half of the time since *Pinus* diverged from *Picea*, with subsequent radiations occurring more recently. By extending the sampling to encompass the phylogenetic diversity of *Pinus*, we predict that most extant subsections diverged during the Miocene. Moreover, subsect. *Australes*, *Ponderosae*, and *Contortae*, containing over 50 extant species, radiated within a 5 Myr time span starting as recently as 18 MYA. An Eocene divergence of pine subgenera (using leaf fossils) does not conflict with fossil-based estimates of the *Pinus–Picea* split, but a Cretaceous divergence using wood fossils accommodates Oligocene fossils that may represent modern subsections. Because homoplasy and polarity of character states have not been tested for fossil pine assignments, the choice of fossil and calibration node represents a significant source of uncertainty. Based on several lines of evidence (including agreement with ages inferred using calibrations outside of *Pinus*), we conclude that the 85 MYA calibration at the divergence of pine subgenera provides a reasonable lower bound and that further refinements in age and mutation rate estimates will require a synthetic examination of pine fossil history.

Introduction

Morphological and molecular analyses of the pine genus (*Pinus*; Pinaceae) reveal conflicting estimates of age and evolutionary rates that are difficult to reconcile. *Pinus* contains 2 monophyletic subgenera, *Pinus* (diploxylon or "hard pines") and *Strobus* (haploxylon or "soft pines"), diagnosable by 2 versus 1 fibrovascular bundle per leaf (Gernandt et al. 2005). Origin of the genus is thought to date to the Early Cretaceous (Millar 1998), whereas estimates for the divergence of the subgenera range from the Late Cretaceous (Millar 1998) to the Mid-Eocene (Miller 1976). Limited morphological differentiation in the approximately 110 species has been attributed to an exceptionally slow rate of change, but morphological homoplasy (Gernandt et al. 2005) and retention of ancestral molecular polymorphism (Syring et al. 2007) are common. Several studies have reported slow molecular divergence rates for pines (Krupkin et al. 1996; Dvornyk et al. 2002; Geada López et al. 2002; Brown et al. 2004; Sokol and Williams 2005; Ma et al. 2006), but these rates appear inconsistent with estimates of per-generation deleterious mutation rates which are known to be at least 10-fold higher in pines than in self-compatible annual flowering plants (Kärkkäinen et al. 1996; Klekowski 1998). In contrast, mutation rate estimates comparable to angiosperms have been reported for pines based on antigenic distances (Prager et al. 1976) and for a retrotransposon (Kossack and Kinlaw 1999).

Key words: molecular evolution, *Pinus*, silent substitution rates, chronogram, fossils.

E-mail: rcronn@fs.fed.us.

Mol. Biol. Evol. 24(1):90–101. 2007
doi:10.1093/molbev/msl131
Advance Access publication September 22, 2006

Unlike many plant groups, *Pinus* is prominent in paleofloras (Miller 1976; Millar 1998; Price et al. 1998), has a long history of taxonomic inquiry (Price et al. 1998; Gernandt et al. 2005), and extensive genomic resources are available (Brown et al. 2001; Temesgen et al. 2001; Chagné et al. 2003; Komulainen et al. 2003; Krutovsky et al. 2004). These enhance its usefulness for studying the interrelationship between morphological and molecular evolution. However, the paleontological data cannot be utilized to full advantage until critical tests of fossil–phylogenetic associations are conducted (e.g., Magallón and Sanderson 2005). The earliest fossil attributable to *Pinus*, *Pinus belgica* (Alvin 1960), is a Cretaceous ovulate cone apparently originating from the Wealden Formation in Belgium 145–125 MYA. Attempts to incorporate this and other *Pinus* fossils into molecular phylogenetic analyses have taken strikingly different approaches. For calibration, *P. belgica* has been placed at the divergence between *Pinus* and other modern Pinaceae genera (Wang et al. 2000; García-Gil et al. 2003) or at the divergence of subgenera (Sokol and Williams 2005; Ma et al. 2006). Because Alvin (1960) further ascribed *P. belgica* to subsect. *Pinus*, its age has been commonly applied to the divergence between representatives from the sections of subg. *Pinus* (Krupkin et al. 1996; Dvornyk et al. 2002; Geada López et al. 2002; Brown et al. 2004; Eckert and Hall 2006). Alternative calibrations used 195 MYA based on a presumed Jurassic origin of the genus that lacks explicit fossil evidence (Kutil and Williams 2001) or 45 MYA as the divergence of subgenera based on the earliest fossils representing both subgenera (Kossack and Kinlaw 1999). As a consequence of inconsistent calibration, synonymous mutation rate estimates vary 50-fold, ranging from angiosperm-like (2.8×10^{-9} substitutions·site^{-1}·year^{-1} for "*gypsy*"-like retrotransposons; Kossack and Kinlaw

1999) to far slower than angiosperms (0.05×10^{-9} for *Adh*; Dvornyk et al. 2002). Similarly, divergence rates calculated for pine chloroplast DNA (cpDNA; 0.06×10^{-9}; Krupkin et al. 1996) are among the slowest reported for any plant.

Here we evaluate hypotheses concerning the age of *Pinus* and explore the impact of different calibrations on estimated absolute mutation rates across 11 nuclear and 4 chloroplast loci. A phylogenetic framework was created with an exemplar from each of the 4 monophyletic sections (a "quartet" of species). Using a critical evaluation of the fossil record, we test 2 calibrations at the crown, inferred as the divergence between subg. *Strobus* and subg. *Pinus*, and represented by competing hypotheses for the oldest fossils from subg. *Strobus* (putative lower age of 85 MYA and an upper bound of high certainty at 45 MYA). We also use our multilocus data set to recalculate ages and rates with the calibration scenarios used in recent studies (Krupkin et al. 1996; Dvornyk et al. 2002; Geada López et al. 2002; Brown et al. 2004; Eckert and Hall 2006; Ma et al. 2006) to test whether highly heterogeneous rates are an intrinsic feature of the loci and taxa sampled in those studies or if the variation can be attributed to calibration.

Materials and Methods

Plant Materials

Each subgenus comprises 2 monophyletic groups—sections *Trifoliae* and *Pinus* from subg. *Pinus* and sections *Parrya* and *Quinquefoliae* from subg. *Strobus*. For this study, 1 species was chosen from each section: *Pinus taeda* L. (Mississippi, United States) for section *Trifoliae*; *Pinus thunbergii* Parl. (Cheonnam, South Korea) or *Pinus merkusii* Jungh. & deVriese s.l. (Thailand) for section *Pinus*; *Pinus monticola* D. Don (Oregon, United States) for section *Quinquefoliae*; and *Pinus nelsonii* Shaw (Nuevo León, Mexico) for section *Parrya*. Haploid genomic DNA from seed megagametophyte tissue was isolated for amplifying nuclear DNA (nDNA) (FastDNA, Qbiogene, Irvine, CA), whereas leaf DNA was the source for cpDNA (Gernandt et al. 2005).

Loci Evaluated

Nuclear loci included in this study map to 7 of 12 linkage groups in *Pinus* (see table S1, Supplementary Material online). Represented are LG2, *aquaporin*; LG3, chlorophyll-binding protein type II precursor (*LHC-CAB*) and late embryogenesis abundant–like protein (*LEA*-like); LG5, arabinogalactan-like protein (*AGP6*) and *ferritin*; LG6, phenylalanine ammonia lyase (*PAL1*); LG7, 4-coumarate:CoA ligase (*4CL*); LG9, cinnamyl alcohol dehydrogenase (*CAD*) and open stomata (*OST1*); LG10, chloroplast-localized Cu–Zn superoxide dismutase (SOD_{chl}); and the unmapped early response to dehydration (*ERD3*). Loci from LG3, LG5, and LG9 are sufficiently distant such that they are effectively unlinked (Krutovsky et al. 2004). Chloroplast loci included in this study are *matK*, *rbcL*, the *rpL20*/*rpsS18* intergenic spacer, and the *trnV* intron. See table S1, Supplementary Material online for primer and amplicon descriptions.

Amplification, Sequencing, and Analysis

Polymerase chain reaction (PCR) products were directly sequenced using BigDye v. 3.1 (Applied Biosystems, Foster city, CA) and visualized on an Applied Biosystems 3730 Genetic Analyzer. DNA alignments were made using ClustalW (Thompson et al. 1994) as implemented in BioEdit v. 7.0.1 (Hall 1999) or an iterative process of BlastN analysis (http://www.ncbi.nih.gov), followed by hand alignment, to align cDNA sequences from GenBank (http://www.ncbi.nlm.nih.gov/Genbank) to genomic sequences. Alignments were made at 3 levels: 1) species within a subgenus, that is, *P. taeda* to *P. thunbergii* and *P. monticola* to *P. nelsonii*; 2) the quartet of *Pinus* species, excluding unalignable regions (Syring et al. 2005); and 3) the quartet of *Pinus* species with an outgroup sequence. For 10 loci (*4CL*, *AGP6*, *aquaporin*, *CAD*, *LHC-CAB*, *PAL1*, *matK*, *rbcL*, *rpl20*/*S18*, and *trnV*; see table S1, Supplementary Material online for details), outgroup sequences from *Picea* were obtained using PCR (as described for ingroup sequences) or by searching GenBank for putative genomic or expressed sequence orthologs. Outgroup sequences were not isolated for *ERD3* and *SODchl* because of orthological concerns regarding some ingroup amplicons (see Results). Amplification of the *LEA*-like locus failed in *Picea*, so a GenBank sequence from *Pseudotsuga* was used as the outgroup. The closest GenBank matches available for *ferritin* and *OST1* (*Arabidopsis thaliana*) were used to provide a root for these 2 loci. New nucleotide sequences were submitted to GenBank (see table S2, Supplementary Material online for details); alignments are available as Supplementary Material online.

Pairwise substitution rates for silent (*d*S; synonymous plus noncoding) and nonsynonymous (*d*N) sites were calculated using DnaSP (Rozas et al. 2003) using the approximation method of Nei and Gojobori (1986) with a Jukes–Cantor correction for multiple substitutions. DnaSP (Rozas et al. 2003) was also used to calculate GC content. Models of sequence evolution (number of substitution categories, base frequencies, shape parameter, and proportion of invariant sites) were tested independently for all sites and for the silent sites (approximated by including noncoding and third-codon positions) of each locus using Modeltest 3.7 (Posada and Crandall 1998). Models selected by the hierarchical likelihood ratio tests in Modeltest 3.7 were used to obtain maximum likelihood (ML) estimates of phylogeny with PAUP* ver. 4.0b10 (Swofford 2002). To test for rate equality (i.e., clock-like evolutionary history) among lineages, we used the likelihood ratio test (Muse and Weir 1992) to compare clock-enforced and clock-relaxed likelihood scores that were obtained using the selected evolutionary models described above (1-tailed probability, $\alpha = 0.05$).

A partition homogeneity test (Cunningham 1997; 100 replicates using a heuristic search) was used to test for significant conflict between loci ($P = 0.01$) for all sites as well as silent sites from 3 different data sets (see table S3, Supplementary Material online for details). The 3 data sets are 1) 9 nuclear loci with different outgroup species (*Picea*, *Pseudotsuga*, or *Arabidopsis*; *4CL*, *AGP6*, *aquaporin*, *CAD*, *ferritin*, *LEA*-like, *LHC-CAB*, *OST1*, *PAL1*; 8,346

bp; 6,098 silent bp); 2) 5 nuclear loci with *Picea* as the outgroup (*4CL*, *AGP6*, *aquaporin*, *CAD*, *LHC-CAB*, 3,973 bp; 2,348 silent bp); and 3) 4 chloroplast loci with *Picea* as the outgroup (*matK*, *rbcL*, *rpL20/S18*, *trnV*; 3,903 bp; 1,914 silent bp). Because data sets did not exhibit significant conflict, they were concatenated, and models of sequence evolution were selected for concatenated data sets as described above. To investigate the impact of increased taxon density on rate and divergence estimates, we examined data from Syring et al. (2005) that include 4 loci evaluated in this study (*4CL*, *AGP6*, *LEA*-like, *LHC-CAB*; 5,338 bp; 3,766 silent bp) for 12 pine species serving as exemplars for subsections with *Picea* as the outgroup. Evolutionary model selection, likelihood ratio tests, and partition homogeneity tests were performed on each locus and on the concatenated silent sites as described above.

Fossil Calibration

Because all pine fossil reports that were considered as calibration sources lack radiometric dating, ages were adjusted to the midpoint of the currently assigned range for their geological Epoch or Stage (Gradstein and Ogg 2004). *Pinus* pollen, pollen cones, and ovulate cone casts lack diagnostic characters for subgeneric identification (Miller 1976; Phipps et al. 1995), so only anatomical reports from ovulate cones, leaves, and wood were considered for calibration sources. The earliest *Pinus* fossil, *P. belgica* (145–125 MYA; Alvin 1960), shows affinities to subg. *Pinus* (Miller 1976) and has been considered part of subsect. *Pinus* (Millar 1998). However, the collection locality for *P. belgica* (Wealden Formation, Belgium) was inferred from the lignitic state and adhering particles (Alvin 1960), leaving its geographic and stratigraphic origins uncertain. There is an approximate 35-Myr gap until the next *Pinus* fossils (100–75 MYA ovulate cones from North America, Japan, and Europe), and these fossils also have affinity to subg. *Pinus* (Fliche 1896; Alvin 1960; Robison 1977; Blackwell 1984; Miller and Malinky 1986; Stockey and Nishida 1986; Saiki 1996). Because these early fossils show symplesiomorphic features and lack synapomorphies to support infrageneric nodes, they cannot be used as reliable calibration sources (Magallón and Sanderson 2001).

The node corresponding to the divergence of the subgenera is well supported with both morphological and molecular synapomorphies (Gernandt et al. 2005). Because all of the oldest *Pinus* fossils have affinity to subg. *Pinus*, the first fossil representing subg. *Strobus* supports the minimum age of divergence of the subgenera. The appearance of subg. *Strobus* dates to the Late Cretaceous based on permineralized wood anatomy (Santonian, 83.5–85.8 MYA, midpoint approximately 85 MYA, *Pinuxylon* sp.; Meijer 2000), or to the Mid-Eocene based on either leaf anatomy (37.2–48.6 MYA, midpoint approximately 45 MYA, *Pinus similkameenensis*; Miller 1973), or ovulate cones (ca. 43 MYA; Axelrod 1986). It is important to note that Cretaceous fossils have been attributed to subg. *Strobus* (Jeffrey 1908; Stopes and Kershaw 1910; Penny 1947), but these fossils are not useful for calibration because they have been reassigned to subg. *Pinus* or to other genera (see Discussion). Therefore, we calibrated the divergence of the subgenera with a putative lower age of 85 MYA based on wood anatomy and an upper age of 45 MYA based on leaves and ovulate cones.

Rates and Ages

Absolute rates of silent-site changes in the concatenated data sets were estimated using 2 approaches. First, we computed silent-site divergence (dS) for comparisons between subgeneric representatives using 9 nuclear and 4 chloroplast loci. Lower and upper rate estimates (i.e., 2 calibration ages) were calculated using the formula $\mu = d\mathrm{S}/2T$, where μ is the silent divergence·site^{-1}·year^{-1}, dS is the mean of silent substitutions·site^{-1} (weighted by number of sites in each locus), and T is time in years. Second, we calculated ML branch lengths (using the selected model of sequence substitution described above; see table S3, Supplementary Material online for details) from clock-enforced phylograms using silent sites from pine quartet alignments of nDNA and cpDNA. Silent-site data sets were used for age inference to maintain consistency with silent rate estimation and to minimize the potential for selection to distort the evolutionary history or rate at a locus. Ages were also inferred using "all" nucleotide sites to provide corroboration of silent-site estimates. To account for departures from clock-like behavior in the 12-species data set, we used a non–clock-enforced tree based on silent sites to assign dates and to estimate a range of local rates using penalized likelihood (PL), with a smoothing factor estimated using cross-validation (r8s ver. 1.70; Sanderson 2002).

Results

Sequence Variation across Loci, Genomes, and Lineages

Alignments of 11 nuclear and 4 chloroplast loci from 4 *Pinus* species included 11,481 bp and 10,643 bp from subg. *Pinus* and *Strobus*, respectively (table 1). Mean lengths of nuclear sequences were shorter than chloroplast sequences (690 bp vs. 972 bp in subg. *Pinus*), but the larger sample of nuclear loci yielded a substantially larger nDNA data set (e.g., 7,594 bp nDNA vs. 3,887 bp cpDNA in subg. *Pinus*). In total, 65% of nDNA and 77% of cpDNA was exonic, whereas the noncoding portion included introns (*4CL*, *aquaporin*, *CAD*, *ERD*3, *ferritin*, *LEA*-like, *OST1*, SOD_{chl}, and *trnV*), 5′ (*OST1*) and 3′ (*ferritin*) untranslated regions, and an intergenic spacer (*rpl20/S18*). Noncoding regions provided 51% of nDNA silent sites (1,959 bp vs. 1,882 bp at third-codon positions) and 40% of cpDNA silent sites (876 bp vs. 1,289 bp at third-codon positions). Nine of the mapped nuclear loci lacked heterozygosity when amplified from haploid megagametophyte tissue, providing evidence for target specificity and orthology. Two loci—SOD_{chl} and *ERD3*—showed evidence of paralogy and were excluded from rate analyses. These loci were unexpectedly heterogeneous in subg. *Pinus*, indicative of gene duplication or nonspecific priming. The SOD_{chl} sequence from *P. monticola* also lacked 4 expected introns, raising the possibility that it is a reverse transcribed pseudogene. GC% showed considerable variation across loci. For nDNA, the average GC content was 47.3% (range = 32.9–66.7%), with coding slightly higher (51.3%; range = 40.5–67.0%) and noncoding

Table 1
Length (*L*), Noncoding Length (NC), Silent (*d*S; synonymous plus noncoding), and Nonsynonymous (*d*N) Substitutions per Site across 11 Nuclear and 4 Chloroplast Genes within Subg. *Pinus*, Subg. *Strobus*, and across Subgenera

	Section *Trifoliae–Pinus*				Section *Quinquefoliae–Parrya*				Subg. *Pinus–Strobus*			
Locus	*L*	NC	*d*S	*d*N	*L*	NC	*d*S	*d*N	*L*	NC	*d*S	*d*N
nDNA												
4CL	996	321	0.040	0.002	740	110	0.083	0.009	700	75	0.189	0.013
AGP6	444	0	0.047	0.003	465	0	0.086	0.016	444	0	0.147	0.015
Aquaporin	653	338	0.039	0.000	624	311	0.058	0.004	571	258	0.100	0.003
CAD	360	246	0.050	0.000	612	498	0.075	0.011	272	158	0.124	0.017
ERD3[a]	656	202	0.162	0.046	687	234	0.070	0.015	634	198	0.104	0.024
Ferritin	767	367	0.042	0.018	725	517	0.051	0.033	690	482	0.129	0.039
LEA-like	1,115	1,037	0.040	0.018	967	889	0.051	0.000	386	308	0.087	0.146[a]
LHC-CAB	701	0	0.072	0.006	701	0	0.091	0.017	701	0	0.205	0.029
OST1	641	476	0.036	0.000	642	477	0.064	0.000	640	475	0.063	0.000
PAL1	409	0	0.021	0.010	409	0	0.065	0.016	409	0	0.142	0.029
SOD_{chl}[a]	852	658	0.060	0.043	194	6	0.153	0.153	200	5	0.127	0.129
Mean[b] (SD)	690	331	0.042 (0.013)	0.005 (0.007)	615	277	0.063 (0.015)	0.013 (0.010)	513	178	0.118 (0.045)	0.020 (0.013)
cpDNA												
matK	1,539	107	0.008	0.013	1,545	113	0.020	0.013	1,545	113	0.039	0.036
rbcL	1,262	0	0.023	0.004	1,262	0	0.023	0.004	1,262	0	0.064	0.005
rpL20/S18	540	254	0.003	0.016	521	231	0.011	0.008	511	221	0.051	0.012
trnV	546	543	0.007	0.000[a]	549	543	0.009	0.000[a]	548	542	0.015	0.000[a]
Mean[b] (SD)	972	226	0.010 (0.008)	0.009 (0.005)	969	222	0.015 (0.006)	0.009 (0.004)	967	219	0.038 (0.019)	0.021 (0.014)
Total	11,481	4,549			10,643	3,929			9,513	2,835		

[a] Values are excluded from mean due to possible paralogy (SOD_{chl}, *ERD*3), short partition length (*trnV*), or unusually high *d*N (> 9 SD from mean; *LEA*-like).
[b] *d*S and *d*N means are weighted by number of sites.

considerably lower (35.8%; range = 23.6–44.5%). For cpDNA, the average GC content was 37.6% (range = 31.4–45.4%), and the GC content in coding (37.8%; range = 32.5–45.4%) and noncoding (36.2%; range = 34.4–39.2%) regions were very similar.

Sequence divergence in nDNA and cpDNA showed 2 important trends. First, although silent substitutions per site (*d*S) averaged approximately 3-fold higher in nDNA than cpDNA, mean nonsynonymous substitutions per site (*d*N) are almost identical among 8 nuclear (0.020) and 4 chloroplast (0.021) loci (table 1). Second, divergence at nDNA was significantly greater between representatives of subg. *Strobus* (mean *d*S = 0.063) than representatives of subg. *Pinus* (mean *d*S = 0.042; $F = 15.19$, $P = 0.001$). This trend was apparent in cpDNA, but the difference was not significant (*d*S *Strobus* = 0.015, *d*S *Pinus* = 0.010; $F = 0.982$, $P = 0.359$).

Fossil-Calibrated Mutation Rates and Divergence Dates

Single-locus ML phylogenies (results not shown) using "all" characters in the quartet alignments showed the expected species relationships in all cases except *OST1* (section *Quinquefoliae* is sister to section *Parrya* and subg. *Pinus*) and *ferritin* (ingroup nodes were unresolved). Rooting of these 2 loci with *Arabidopsis* is likely responsible for the topological differences, and we excluded these loci from our chronograms. Rate equivalence (as inferred by the likelihood ratio test) was statistically supported for all individual loci except *LHC-CAB*, *ferritin*, and *PAL1* using all characters and for all individual loci except *PAL1* using "silent" sites (see table S3, Supplementary Material online for details). A concatenated alignment of the 5 clock-like loci that share *Picea* as outgroup (*4CL*, *AGP6*, *aquaporin*, *CAD*, and *LHC-CAB*) exhibited rate equality using either silent sites or all sites (see table S3, Supplementary Material online for details). Partition homogeneity tests among these loci (data not shown) indicated that these data sets do not reflect conflicting topologies, supporting concatenation. Rate equivalence was supported for all sites as well as silent sites for each cpDNA locus and for a concatenated cpDNA alignment (see table S3, Supplementary Material online for details). Rates were inferred from silent sites for each genome; ages were calculated using all-site and silent-site data sets.

Based on a calibration with 85 or 45 MYA, absolute silent mutation rates (μ) for 9 nuclear loci average 0.70 or 1.31×10^{-9} silent substitutions·site^{-1}·year^{-1} for nDNA, and 0.22 or 0.42×10^{-9} for cpDNA (table 2). To address the impact of unequal base frequencies and among-site rate heterogeneity on rate estimates, the absolute silent mutation rate (μ) was also calculated from ML branch lengths for each calibration test (table 3). ML-based rates are marginally

Table 2
Estimated Absolute Mutation Rates (μ; substitutions ·site^{-1}·year^{-1}) Based on Silent (*d*S) and Nonsynonymous (*d*N) Substitutions in Comparisons between subg. *Pinus* and subg. *Strobus*. Rates Are Averages of 9 Nuclear Loci or 4 Chloroplast Loci, Using Divergence Dates of 85 or 45 MYA. Values in Parentheses Are 1 SD

			μ ($\times 10^{-9}$)	
Genome	Rate	*d*	*T* = 85 MYA	*T* = 45 MYA
Nuclear	*d*S	0.118 (0.045)	0.70 (0.27)	1.31 (0.50)
	*d*N	0.020 (0.013)	0.12 (0.04)	0.22 (0.08)
Chloroplast	*d*S	0.038 (0.018)	0.22 (0.11)	0.42 (0.20)
	*d*N	0.021 (0.016)	0.12 (0.06)	0.23 (0.11)

higher than *d*S-based rates for nDNA ($\mu = 0.75$ or 1.41×10^{-9}), whereas cpDNA rates are almost identical ($\mu = 0.21$ or 0.40×10^{-9}).

Chronograms based on silent sites from nDNA and cpDNA reveal identical topologies, but disparities in the branch lengths (especially at deeper nodes) yield different estimated divergence dates for *Pinus* lineages (fig. 1 and table 3). Most notably, estimates from the 5-locus nDNA data set predict older divergence events between *Pinus* and *Picea* (190–102 MYA) than does cpDNA (164–136 MYA). Both genomes show sections of subg. *Strobus* diverging before those of subg. *Pinus*, although nDNA estimates are more ancient (48–25 for subg. *Strobus* vs. 30–16 MYA for subg. *Pinus*) than those from cpDNA (37–19 MYA for subg. *Strobus* vs. 25–13 MYA for subg. *Pinus*). Divergence dates inferred from the same data sets using all nucleotides were nearly identical at sectional divergences but somewhat younger for the *Pinus–Picea* node (table 3). Divergence dates were also estimated using silent sites in the clock-enforced data set that included 4 additional nuclear loci (*ferritin*, *LEA*-like, *OST1*, and *PAL1*; 6,098 bp). The use of different outgroups for 3 of these loci precluded us from estimating the divergence of *Pinus* from *Picea* (node D), but estimated sectional divergences within subg. *Pinus* (31–16 MYA; node A) and subg. *Strobus* (48–25 MYA; node B) were not different from estimates based on 5 clock-like loci (table 3).

Cross-validation yielded an optimum smoothing factor of 63 for PL in the 12-taxon data set. In general, results from PL indicate more recent divergence events than predicted from the analysis of only 4 exemplar species (fig. 2 and table 3). For example, using an 85 MYA calibration, subg. *Strobus* sections are predicted to have diverged 37 MYA or 10 Myr more recently than predicted from 4-taxon comparisons (fig. 1). Similarly, subg. *Pinus* sections are predicted to have diverged 28 MYA, which is 2 Myr more recent than indicated by 4-taxon estimates (fig. 1). An important finding highlighted by this additional taxon sampling is that modern pine subsections radiated in a narrow time span in the relatively recent past. For example, using the 85 MYA calibration, the 3 subsections of *Trifoliae* (fig. 2*b*, nodes E–G), with approximately 51 extant species, radiated within a 5 Myr time span starting 18 MYA.

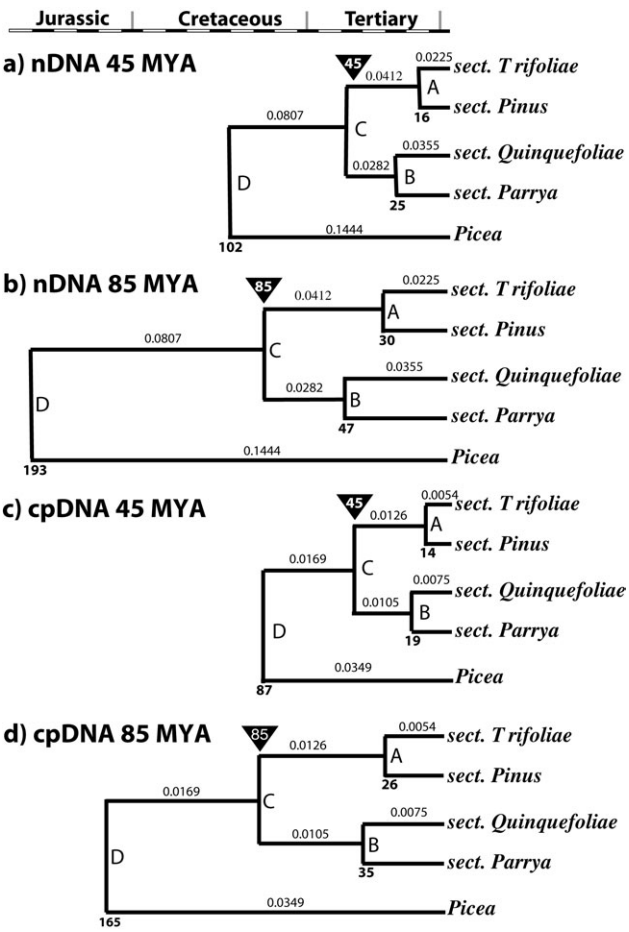

FIG. 1.—ML chronograms of major lineages of *Pinus* using silent sites from 5 nuclear loci (2a, 2b; 2,348 bp) or 4 chloroplast loci (2c, 2d; 1,914 bp). Calibration is at node C with either 45 MYA, based on leaves (2a, 2c), or 85 MYA, based on wood (2b, 2d). Branch lengths are shown above each branch, and estimated ages are shown below each node. See Supplementary Material online for tree statistics.

Discussion

Mutation rate estimates are used widely for hypothesis testing in molecular, population, and evolutionary genetic studies (Muse 2000), and they are increasingly used to estimate genetic parameters of conifers (Dvornyk et al. 2002; García-Gil et al. 2003; Brown et al. 2004; Ma et al. 2006). Numerous estimates of mutation rates have been made for angiosperms (e.g., Wolfe et al. 1987, 1989; Gaut et al. 1996; Koch et al. 2000; Clark et al. 2005), but comparable conifer estimates are limited to studies of single-locus variation across multiple species (Geada López et al. 2002), multilocus variation in a single species (García-Gil et al. 2003; Brown et al. 2004), or multilocus variation in a few closely related species (Ma et al. 2006). The perspective added by our synthesis of multiple nuclear and chloroplast loci based on exemplar taxa and 2 fossil calibration points indicates that divergence times separating pine lineages have been frequently overestimated, with a concomitant underestimation of absolute mutation rates (Krupkin et al. 1996; Dvornyk et al. 2002; Geada López et al. 2002; Brown et al. 2004; Sokol and Williams 2005; Ma et al. 2006). This new perspective supports a relatively recent radiation of extant pine sections in the Early Miocene, a hypothesis that has been proposed (e.g., Miller 1973; Strauss and Doerksen 1990) (but often ignored) to explain the modest morphological and genetic divergence between pine species.

Sources of Error

Confidence intervals (CI) for molecular clock estimations must consider the simultaneous uncertainties of rate and time that are confounded within divergences, as well as the age and placement of calibration points. Because fossils document minimum divergence times, the elapsed time from taxon origin to the earliest fossil cannot be known. Additional uncertainty is introduced by variation in deposition, discovery, and the quality of our understanding of the age of

Table 3
Estimated Divergence Dates for *Pinus* Nodes Shown in Figure 1 Based on Silent Sites (or all sites) from nDNA or Chloroplast DNA. Calibration Is at Node C at 85 or 45 MYA. Ages and Silent Mutation Rates (μ) Are Estimated Using ML Branch Lengths for Data Showing Rate Equality or Using PL for the Nonclocklike 12-Taxon Data set

				Ages for Nodes (MYA)				
Data set[a]	Outgroup	Method	Sites	A	B	C	D	μ ($\times 10^{-9}$)
Nuclear								
5 Loci—4 taxa	*Picea*	ML	Silent	30	48	85	190	0.75
			All	29	47		172	—
			Silent	16	25	45	102	1.41
			All	15	25		91	—
9 Loci—4 taxa	Multiple[b]	ML	Silent	31	48	85	NA	0.66
				16	25	45	NA	1.24
4 Loci—12 taxa	*Picea*	PL	Silent	28	37	85	NA	0.74
				15	20	45	NA	1.39
Chloroplast								
4 loci—4 taxa	*Picea*	ML	Silent	25	37	85	164	0.21
			All	27	33		136	—
			Silent	13	19	45	87	0.40
			All	14	18		72	—

NOTE.—NA, not available.
[a] Data set composition described in text.
[b] *OST1*, *ferritin* = *Arabidopsis thaliana*; *LEA*-like = *Pseudotsuga menziesii*.

the source formation as well as the taxonomic affinity of the fossil. There are promising approaches for estimating CI for divergence times (e.g., Kumar et al. 2005; Yang and Rannala 2006), but these will require multiple robust fossil calibrations for *Pinus*. In addition, molecular divergence rates may vary across the genome and across lineages. We addressed this complex problem using several approaches to provide a perspective on the relative impact of various factors on the error ranges of our rate estimates.

First, both of the fossils selected for calibration lack radiometric dating. Because midpoints of the ranges for geological Epochs or Stages were used to create the chronograms (figs. 1 and 2 and table 3), we recalculated divergence dates using the upper and lower bounds of these periods. The stratigraphic age ranges for wood from the Santonian *Pinuxylon* sp. (Meijer 2000) and leaves from the Middle Eocene *P. similkameenensis* (Phipps et al. 1995) are currently considered to be 83.5–85.8 MYA, and 37.2–48.6 MYA, respectively (Gradstein and Ogg 2004). The inferred age ranges for node D (cf. table 3), using these stratigraphic ranges to calibrate at node C, are 187–192 MYA and 83–109 MYA for calibrations based on wood and leaves, respectively. Inferred age ranges for node A are 29.6–30.4 and 13–17 MYA for calibrations based on wood and leaves, respectively, whereas node B ages are 47–48 and 21–27 MYA, respectively. Clearly, the use of midpoint ages introduces uncertainty in our estimates, particularly for Eocene fossils because that Epoch is much longer than the Santonian (11.4 vs. 2.3 Myr). Nonetheless, this uncertainty is small relative to the differences in alternative calibration points suggested by fossil wood (85 MYA) versus fossil leaves (45 MYA) for the origin of *Pinus* subgenera.

Second, the use of the mean substitution rates from multiple loci reduces stochastic deviation resulting from rate variation among genomic regions, and limiting the sample to silent sites reduces the potential for selection to influence rates. To assess the amplitude of this variation among our 9 nuclear loci, we also calculated absolute substitution rates using the mean *d*S plus and minus 1 standard deviation (SD). This results in $\mu = 0.43–0.96 \times 10^{-9}$ for a calibration based on wood and $0.81–1.81 \times 10^{-9}$ for leaves (cf. table 2). We also projected the "combined" effects of locus-specific rate variation (mean *d*S plus 1 SD and the upper bound for the stratigraphy; *d*S minus 1 SD and the lower bound) for each calibration. This results in "inclusive" ranges of $\mu = 0.42–0.98 \times 10^{-9}$ for a calibration based on wood and $0.75–2.2 \times 10^{-9}$ for leaves (cf. table 2). Among the 9 loci that were used for rate calculation, distortion may have been introduced by our use of different outgroups for *ferritin*, *LEA*-like, and *OST1* (see table S2, Supplementary Material online for details). Further, *PAL1* silent sites showed a significant departure from rate equivalence even with *Picea* as the outgroup. However, ages and rates inferred from the 5- and 9-locus data sets (table 3) are largely consistent. Further, data accumulated for 50 *Pinus* loci (Cronn R, unpublished data) show a median silent rate that is very similar to the silent rate described in these results. This implies that despite locus-specific rate variation, our 5-locus data set is useful as a first approximation of μ in *Pinus*.

Finally, we derived CIs for age estimates using the method of Haubold and Wiehe (2001), which uses nonoverlapping pairs of phylogenetic distances to infer the unknown mutation rate for the other pair. Using this method at node A (fig. 1*b*), with subg. *Strobus* as the reference and ages calculated from the 85 MYA crown calibration, yields 95% CIs of 14–67 MYA for the divergence of subg. *Pinus*. Applying this technique to node B (with subg. *Pinus* as the reference) yields intervals of 22–106 MYA for subg. *Strobus*. The magnitude of these intervals far exceeds error ranges introduced by stratigraphic ranges, intralocus rate variation, or even fossil choice between permineralized wood (85 MYA) or leaves (45 MYA). These large uncertainties clearly highlight the challenge inherent

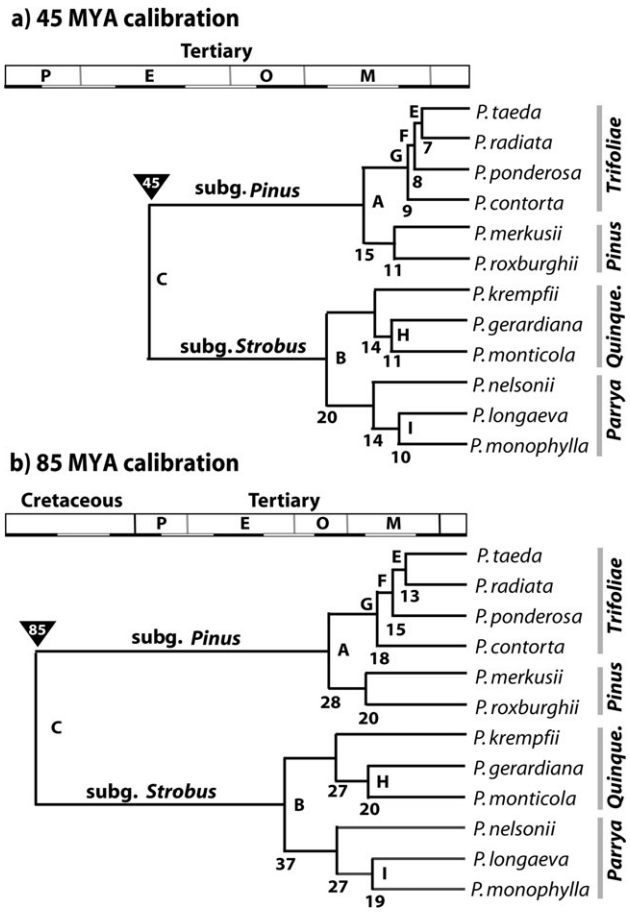

FIG. 2.—PL chronograms of 12 taxa using silent sites from 4 nuclear loci (3,766 bp), based on calibration at node C with *a*) 45 MYA or *b*) 85 MYA. Estimated ages are shown below each node. See Supplementary Material online for tree statistics. Nodes E–I represent subsections referred to in the text: E, *Australes*; F, *Ponderosae*; G, *Contortae*; H, *Strobus*; and I, *Balfourianae*/*Cembroides*. All nodes have greater than 70% bootstrap support except for nodes E and F, which collapse in the strict consensus tree.

in the simultaneous uncertainties of rate and time over evolutionary time scales.

Although the magnitude of these sources of error highlight the importance of considering molecular rate variation and stratigraphic uncertainty, past calibration scenarios applied to *Pinus* show that incorrect fossil assignment can be a far more dramatic source of error. For example, the common practice of using *P. belgica* to calibrate the divergence of sections within subg. *Pinus* (node A, fig. 3*e*) pushes the divergence of pine subgenera (node C) to 339 MYA and the divergence of *Pinus–Picea* (node D) to 758 MYA in the Precambrian. The attendant silent mutation rate in the nuclear genome is exceptionally slow ($\mu = 0.19 \times 10^{-9}$). This unrealistic (but commonly cited) calibration is clearly responsible for many of the exceptionally low μ values reported for pine nDNA and cpDNA (Krupkin et al. 1996; Dvornyk et al. 2002; Geada López et al. 2002; Brown et al. 2004; Ma et al. 2006). This calibration can also cause substantial distortion in biogeographical interpretations. For example, Eckert and Hall (2006) recently hypothesized dispersal and vicariance events using *P. belgica* to calibrate the divergence of sections within subg. *Pinus* (node A, fig. 3*f*) in a cpDNA data set. In doing so, they fail to consider the unrealistic estimate that this calibration indicates for the divergence of *Pinus* and *Picea* (i.e., 720 MYA) based on our clock-like cpDNA branch lengths (fig. 3*f*). In the same manner, associating *P. belgica* with the divergence of pine subgenera also conflicts with the fossil record. Such a calibration (e.g., 136 MYA, Sokol and Williams 2005; 130 MYA, Ma et al. 2006) indicates a *Pinus–Picea* split of approximately 300 MYA (fig. 3*d*) and produces absolute silent mutation rates of $\mu = 0.47–0.49 \times 10^{-9}$. These projected ages for *Pinus* (Carboniferous or Early Permian, respectively) predate *P. belgica* by more than 150 Myr. In summary, the errors resulting from incorrect fossil/node association far exceed the cumulative error estimated with CIs that account for the confounding effect of simultaneous rate and time variation. In contrast, using *P. belgica* to represent the divergence of *Pinus* and *Picea* in our clock-like data set (node D, fig. 3*c*; Wang et al. 2000; García-Gil et al. 2003) places the divergence of *Pinus* subgenera at 63 MYA. This estimate lies between the earliest putative subg. *Strobus* wood and leaf fossils.

Implications of a Moderate Tempo for Pine Mutation Rates

Several authors have argued that a high mutation rate is necessary to account for the extremely high level of inbreeding depression found in conifer species capable of partial self-fertilization (Lande et al. 1994; summarized in Scofield and Schultz 2006). Because plants do not segregate a germ line, somatic mutations may accumulate in meristems and be incorporated into gametes. Hence, longevity and large stature contribute to a large number of mitoses, both of which may serve to elevate the per-generation mutation rate (Scofield and Schultz 2006). Evidence for a high per-generation mutation rate in *Pinus* compared with other plant groups is provided by observations of the frequency of chlorophyll-deficient mutants. Assuming an equal number of loci capable of mutating to chlorophyll deficiency and equivalent lengths across loci, the per-generation rate of deleterious mutations for *Pinus sylvestris* ($U = 1–3 \times 10^{-2}$; Kärkkäinen et al. 1996) is comparable to the mutation rate for *Rhizophora mangle*, another long-lived woody perennial ($U = 1.5 \times 10^{-2}$; Klekowski and Godfrey 1989). These mutation rates are approximately 100-fold higher than the average for 10 annual flowering plant species ($U = 1–3 \times 10^{-4}$; Klekowski 1992).

Mutation rates in *Pinus* can be evaluated by comparing absolute rates derived from fossil calibrations in other plant groups. Our 85 MYA calibration provides an estimate for nDNA ($\mu = 0.70 \times 10^{-9}$ synonymous substitutions·site^{-1}·year^{-1}; table 2) that is 14-fold higher than previous *Pinus* estimates (Dvornyk et al. 2002), but approximately 4-fold slower than the rate for palms ($\mu = 2.61 \times 10^{-9}$; Gaut et al. 1996) and 7- to 40-fold slower than the range reported for herbaceous angiosperms ($\mu = 5–33 \times 10^{-9}$; Gaut et al. 1996; Koch et al. 2000; Clark et al. 2005). Similarly, our estimated rates for chloroplast silent sites ($\mu = 0.22 \times 10^{-9}$ synonymous substitutions·site^{-1}·year^{-1}; table 2) are approximately 6-fold slower than those reported for angiosperm cpDNA ($\mu = 1.1–1.6 \times 10^{-9}$; Wolfe et al. 1987). Expressed on a per-year basis, these comparisons

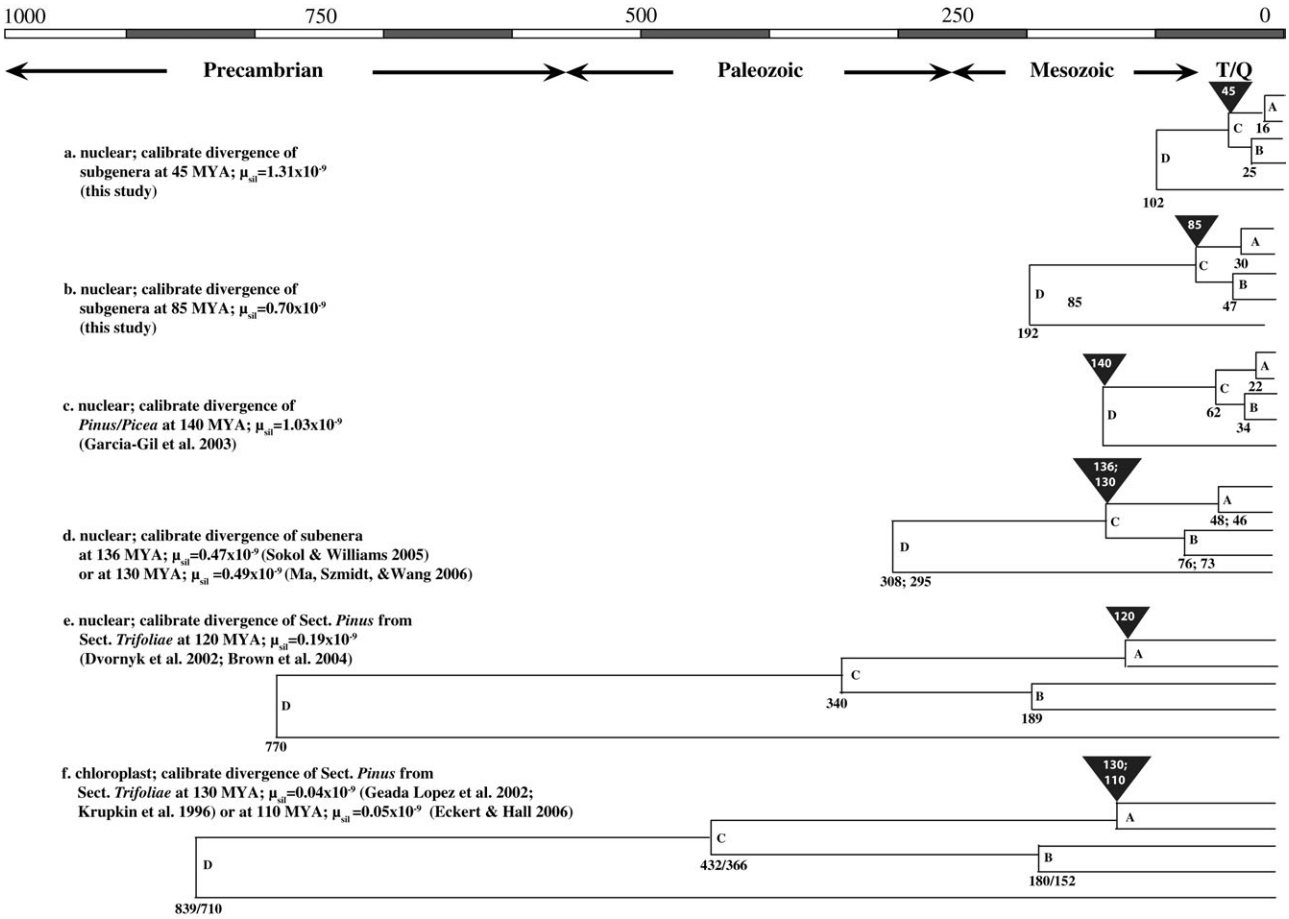

FIG. 3.—Application of some recently published fossil calibration scenarios (see Discussion), using clock-enforced branch lengths (shown in fig. 1) from our nDNA or cpDNA data sets. ML-projected ages are shown below each node; ML-calculated silent rates are given for each scenario. Node letters correspond to figure 1.

suggest that substitution rates in *Pinus* are still far slower than most angiosperms. In this context, it is important to note that molecular substitution rates can also show a "generation-time effect" (e.g., Gaut et al. 1992, 1996; Kay et al. 2006). Based on our estimate of the average rate ($\mu = 0.70 \times 10^{-9}$ synonymous substitutions·site^{-1}·year^{-1}) and correcting for longevity (assuming a 25-year generation time; Brown et al. 2004), the per-generation nuclear substitution rate for *Pinus* averages 1.75×10^{-8} substitutions·site^{-1}·generation^{-1}. This rate is nearly equivalent to the 1.5×10^{-8} substitutions·site^{-1}·year^{-1} rate inferred for short-lived Brassicaceae (Koch et al. 2000). The near-equivalence between these values closes the gap between gymnosperm and angiosperm rates relative to prior rate estimates (e.g., Dvornyk et al. 2002; Brown et al. 2004), but they still fall short of the 100-fold difference that might be expected based on chlorophyll-deficiency mutations.

Clearly, these calculations hinge on an appropriate association between a calibration time and a phylogenetic node; the choice of a node with which to associate a fossil is the largest determinant of the rate. The availability of *Pinus* fossils with synapomorphies supporting the divergence of *Pinus* subgenera (a well-supported node in a molecular phylogeny) is more robust than options available for many plant groups, and we suggest that this will be an important factor for future, more refined, rate comparisons. This study illustrates the variability of divergence rates among loci because *d*S for 9 nuclear loci ranges from 0.063 to 0.205 (table 1). Even though fossil dates were used as fixed calibration points in our calculations, each fossil represents a "minimum" age for a lineage. It follows then that the rates calculated from mean dS ($\mu = 0.70$–1.31×10^{-9} substitutions·site^{-1}·year^{-1}; table 2) represent "maximum" absolute rates, although the same caveat applies to angiosperm rates used for comparison.

Our revised estimates of mutation rates in pines have important implications for studies of population and molecular genetic parameters. For example, an effective population size (N_e) of 5.6×10^5 was calculated for *P. taeda* using 19 loci and a substitution rate ($\mu = 1.17 \times 10^{-10}$; estimated by calibrating divergence between members of section *Trifoliae* with the age of *P. belgica*; Brown et al. 2004). If we apply our estimate of $\mu = 0.70 \times 10^{-9}$ (calibrated at 85 MYA between the subgenera; table 2), we find that the predicted N_e for *P. taeda* is far lower at 9.4×10^4. This value of N_e contrasts sharply with census population estimates for *P. taeda* (which may exceed 10^{10}; Brown et al. 2004), and it suggests that this species has experienced dramatic population growth following a relatively recent genetic bottleneck in the history of this species.

Pinus Fossil Status

Despite the diversity and abundance of *Pinus* fossils (see Millar 1998), several obstacles need to be overcome before integrating additional fossils into a multipoint calibration. To date, pine fossil reports have been based on unattached organs (ovulate or pollen cones, leaves, or wood fragments). Two of these organs have yet to be found on a contiguous fossil, although species have been named based on the hypothesized common origin of separate organs. Retention of ancestral character states and homoplasy among extant *Pinus* species are sufficiently frequent that characters from a single organ are inadequate for discriminating among extant pine subsections or sections (Gernandt et al. 2005). Because *Pinus* subgenera are diagnosable by leaf vasculature (Gernandt et al. 2005) and by a combination of wood features (Van der Burgh 1973), this key divergence event provides the most reliable calibration point for *Pinus*.

Historically, fossil descriptions used a typological approach for assigning affiliations with extant taxa. This process is complicated by historical revisions of *Pinus* sectional affiliations, and it lacks the necessary phylogenetic framework for making assignments. As previously noted, the number of fibrovascular bundles per leaf is the only nonhomoplasious character known to date that diagnoses *Pinus* subgenera (Gernandt et al. 2005). However, the presence of a single fibrovascular bundle at the base of a leaf fossil may not be diagnostic for the subgenera because bundles branch distally into the 2 bundles characteristic of subg. *Pinus* (Stockey and Nishida 1986). Based on this information, *Pinus* sp. leaf fossils from Staten Island, New York (Jeffrey 1908) are now thought to represent subg. *Pinus* rather than subg. *Strobus*. Similarly, Stockey and Nishida (1986) suggest that Cretaceous *P. yezoensis* leaf fossils from Japan (Stopes and Kershaw 1910), also once believed to represent subg. *Strobus*, are more representative of *Cedrus*.

Although needles are informative within *Pinus*, ovulate cones are considered the most dependable evidence for assigning fossils to the genus. Miller (1976) outlined 4 characteristics that together define *Pinus*: inflated scale apex, bract and scale traces united at origin, all resin canals abaxial to vascular tissue in scale base, and scale strands curved on adaxial side. Critically, members of the extinct cone genus *Pityostrobus* can share at least 2 of these features, complicating the identification of Cretaceous pine ovulate cones. Indeed, using Miller's criteria, some early "pine" fossils have been reassigned to *Pityostrobus*, *Picea*, or *Cedrus*. For example, Miller and Malinky (1986) referred cone scales from the Magothy Formation of Delaware, originally described as representing subg. *Strobus* (Penny 1947) to *Pityostrobus*. Because many fossil cones are found in marine deposits, abrasion during transport can remove important cone scale features (Smith and Stockey 2002), misleading attempts at classification (Wolfe and Schorn 1989).

Evidence for an Eocene Divergence of Subgenera

Some authors have suggested a relative recent origin for subg. *Strobus* in the Eocene (Miller 1973; Phipps et al. 1995; Kossack and Kinlaw 1999), as reflected by our 45 MYA crown calibration. This hypothesis is based on numerous ovulate cone fossils assigned to subg. *Strobus* that appear in the fossil record starting approximately 40 MYA (Millar 1998), closely following the oldest known leaves attributable to subg. *Strobus* (45 MYA). Unattached leaves (with subg. *Strobus* features), deposited near ovulate cones (with subg. *Pinus* features), in the Princeton Chert in British Columbia were originally described as 2 species: *P. similkameenensis* and *P. arnoldii* (Miller 1973). More recently, Phipps et al. (1995) reinterpreted these fossils as a single species (*P. similkameenensis*) and proposed that the mosaic of characters in this hypothetical species may represent a lineage that predates the divergence of subg. *Strobus*. These authors did note that alternative explanations (e.g., organs derived from a lineage with no extant descendent or from sympatric pine species) cannot be ruled out.

A 45-MYA calibration also shows surprising agreement with a recent cpDNA-based estimate of seed plant divergences using entirely different calibration sources (Magallón and Sanderson 2005). In that analysis, the divergence of Gnetophytes and Pinaceae was constrained with a fossil date of 216 MYA; estimated divergence dates for *Pinus* subgenera (ca. 50 MYA) and synonymous cpDNA substitution rates ($\mu = 0.26 \times 10^{-9}$) nearly match our results. Because fossils represent minimum ages, calibration at the divergence of *Pinus* subgenera with 45-MYA leaf fossils appears to provide a reasonable upper bound for this event.

Evidence for a Cretaceous Divergence of Subgenera

The description of a *Pinuxylon* sp. from the Aachen (Meijer 2000) may push the age of subg. *Strobus* more than 40 Myr before the oldest leaf and cone fossils, but this is not a simple determination. The cone genus *Pityostrobus* was abundant and diverse during the Cretaceous, but associated wood remains undescribed and its exact relationship to *Pinus* is unresolved (Smith and Stockey 2002). Further, the 2 characters used to affiliate the Aachen *Pinuxylon* wood with subg. *Strobus* (cross-field pitting and ray tracheid dentations) are not unequivocal synapomorphies (data not shown), even in extant pine species, indicating the possibility of retention of ancestral polymorphism or parallel evolution. Both of these tracheid features are quantitative traits for which intraspecies variation has been noted (Shaw 1914; Hudson 1960; Van der Burgh 1973) and for which the ancestral state is unclear (Hart 1987). Perhaps most challenging is that diagnosis of subg. *Strobus* fossil wood is made by determining that a specimen has "weak-to-absent" dentations. In this context, the description of the Aachen *Pinuxylon* sp. as possessing "ray tracheids faintly dentate but owing to rather poor state of preservation only locally visible" (Meijer 2000) may best be considered tantalizing evidence for the existence of subg. *Strobus* in the Cretaceous, but evidence requiring corroboration. The evidence for subg. *Strobus*–like wood in the Cretaceous is bolstered by fossil wood of about the same age that shows similarities to modern-day members of subg. *Pinus* (Blackwell 1984), hence providing wood fossils flanking the crown node. Although the possibility that the subg. *Pinus*–like fossils may simply be the ancestral state in

Pinaceae cannot be excluded, they do lend credence to the hypothesis that the Aachen *Pinuxylon* represents *Pinus* rather than another member of the Pinaceae.

The 45-MYA calibration (even when viewed as a minimum age) suggests a more recent *Pinus–Picea* split (102–87 MYA; fig. 1) than is indicated by the existence of *P. belgica* at approximately 135 MYA (Miller 1976). However, there are unresolved issues concerning the age of the genus. First, the approximately 35-Myr gap between *P. belgica* and the next *Pinus* fossil, combined with uncertainties about the source of *P. belgica*, deserves note. Second, the only other evidence for an "Early" Cretaceous origin for *Pinus* is pollen from Alaska (Langenheim et al. 1960). The age of the Kuk River fossil formation that is the source for these reported 145–100 MYA *Pinus* microfossils is now estimated to be the Mid- to Late Albian (104–97 MYA; Koteja and Poinar 2001). In addition, Erwin and Schorn (2006) recommend caution when relying on pollen without associated megafossils due to the unknown status of pollen associated with seed cone genera such as *Pityostrobus*. If one were to discount the inference of an "Early" Cretaceous age for *P. belgica* and the previous estimate for the age of the Alaskan pollen, the oldest evidence for *Pinus* would be ovulate cones from the Albian–Cenomanian (ca. 103 MYA; Fliche 1896). This does not appear to fit with the 190–164 MYA *Pinus–Picea* split implied by a Cretaceous divergence of the subgenera (fig. 1). On the other hand, the earlier age estimate for our node D (fig. 1) that results from the 85 MYA calibration appears consistent with an older divergence of *Pinus* and *Picea* because there is support for an early independent divergence of these 2 genera from the *Pityostrobus* grade (Smith and Stockey 2002). The 85 MYA calibration also agrees with a study of Pinaceae genera, which calibrated the *Pinus–Picea* split at 140 MYA, yielding an estimated age of approximately 84 MYA for the divergence of *Pinus* subgenera (Wang et al. 2000).

Perhaps the strongest evidence that subg. *Strobus* diverged in the Cretaceous are the numerous Oligocene (ca. 26 MYA) fossils that potentially represent recent lineages. Oligocene ovulate cone fossils have been affiliated with 6 different *Pinus* subsect.: *Pinaster* and *Pinus* in section *Pinus* (Mai 1986; Erwin and Schorn 2006); *Ponderosae* in section *Trifoliae* (Wolfe and Schorn 1989); *Balfourianae* and *Cembroides* in section *Parrya* (Wolfe and Schorn 1989); and *Strobus* in section *Quinquefoliae* (Mai 1986). Existence of these subsections approximately 26 MYA is in better agreement with our predicted age of these groups using an 85 MYA calibration (fig. 2*b*, nodes E–I). This assessment carries 2 important caveats. First, most subsectional assignments have been made by typological matching that considered extant taxa from only 1 continent (but see Erwin and Schorn 2006). Second, a cladistic analysis of the characters used to affiliate fossils to subsections has yet to be conducted, so their value at this taxonomic level remains speculative. Based on the framework provided by our clock-like data, many of these Oligocene fossils must reflect symplesiomorphic character states for the simple reason that they infer too ancient a divergence for *Pinus* lineages. Even our 85 MYA chronogram does not project subsections diverging early enough to support the affiliation of Late Cretaceous leaf fossils (ca. 85–78 MYA) with 3 different subsections: *Pinaster* (Stockey and Nishida 1986), *Pinus* (Robison 1977), and *Ponderosae* (Stockey and Nishida 1986). Integrative fossil validation, such as the method used to discard "outlier" fossils in a time-calibrated phylogeny of turtles (Near et al. 2005), can be used to take advantage of the numerous pine fossils, but only as putative synapomorphies supporting fossil associations are identified.

Concluding Remarks

Our interpretations of mutation rates and divergence ages are relatively insensitive to the choice of genome (nDNA vs. cpDNA), the selection of silent sites versus all sites, the number of taxa, or clock constraints (4 taxa constrained vs. 12 taxa unconstrained). Instead, the choice of fossil and node for calibration have a pronounced impact. This suggests that refinements of estimated mutation rates will benefit more from a reevaluation of the *Pinus* fossil record based on a cladistic analysis of morphological characters in extant taxa rather than increased genomic sampling. Because *Pinus* has a rich fossil record, a critical evaluation of the morphological synapomorphies that support additional phylogenetic nodes should be made a priority. The dramatic distortions of ages and rates (fig. 3) resulting from incorrect placement of fossils should serve as a cautionary note to studies of fossil-poor families. As in angiosperms (Sanderson and Doyle 2001), a relatively recent divergence of the crown group in *Pinus* may have been obscured by the antiquity of the stem lineage. Others have suggested that a recent origin for most extant pine taxa, especially in subg. *Pinus*, is the simplest explanation for low divergence rates (e.g., Strauss and Doerksen 1990; Govindaraju et al. 1992). Regardless of which absolute calibration is chosen, the relative divergence times inferred from multiple nuclear and chloroplast loci provide an important perspective for studying pine species relationships. For example, "relative" divergences between sections in subg. *Pinus* are only one-sixth of the total divergence within the genus.

We conclude that a 45-MYA subgeneric divergence may be too young but yields an upper limit for *Pinus* evolutionary rates. Because a subgeneric calibration at 85 MYA based on permineralized wood yields realistic age projections for both older and younger nodes, it provides a useful lower rate limit. Together, these rate estimates (table 2) reveal a moderate tempo for pine divergence and provide a framework that can be used to compare future conifer gene– and taxon–specific rates.

Supplementary Material

Supplementary tables S1, S2, and S3 and the alignment file are available at *Molecular Biology Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank J. Berdeen, D. Johnson, W. Kwan-Soo, and M. McGregor for providing plant collections; G. Brown, S. Gonzalez-Martinez, K. Krutovsky, D. Neale, and C. Plomion for sharing linkage maps and primer sequences;

D. Erwin, S. Manchester, G. Poinar, and J. van der Burgh for advice on fossils; S. Muse, D. Remington, R. Small, J. Wendel, and 2 anonymous reviewers for helpful comments; and K. Farrell, C. Streng, and O. Zerón Flores for laboratory assistance. Funding for this study was provided by National Science Foundation grant DEB 0317103 to A.L. and R.C., Secretaria de Educación Pública grant to D.S.G., and the USDA Forest Service Pacific Northwest Research Station. The research was conducted at Oregon State University and Universidad Autónoma del Estado de Hidalgo.

Literature Cited

Alvin K. 1960. Further conifers of the Pinaceae from the Wealden Formation of Belgium. Inst R Sci Nat Belg Mém. 146:1–39.

Axelrod D. 1986. Cenozoic history of some western American pines. Ann Mo Bot Gard. 73:565–641.

Blackwell W. 1984. Fossil ponderosa-like pine wood from the Upper Cretaceous of northeast Mississippi. Ann Bot (Lond). 53:133–136.

Brown G, Gill G, Kuntz R, Langley C, Neale D. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci USA. 101:15255–15260.

Brown G, Kadel E, Bassoni D, Kiehne K, Temesgen B, van Buijtenen J, Sewell M, Marshall K, Neale D. 2001. Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics. 159:799–809.

Chagné D, Brown G, Lalanne C, Madur D, Pot D, Neale D, Plomion C. 2003. Comparative genome and QTL mapping between maritime and loblolly pines. Mol Breed. 12:185–195.

Clark R, Tavaré S, Doebley J. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. Mol Biol Evol. 22:2304–2312.

Cunningham C. 1997. Can three incongruence tests predict when data should be combined? Mol Biol Evol. 14:733–740.

Dvornyk V, Sirviö A, Mikkonen M, Savolainen O. 2002. Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol. 19:178–188.

Eckert A, Hall B. 2006. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. Mol Phylogenet Evol. 40:166–182.

Erwin D, Schorn H. 2006. *Pinus baileyi* (Section *Pinus*, Pinaceae) from the Paleogene of Idaho, USA. Am J Bot. 93:197–205.

Fliche P. 1896. Etude sur la flore fossile de l'Argonne (Albien-Cénomanien). France: Nancy.

García-Gil M, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol. 12:1195–1206.

Gaut B, Morton B, McCaig B, Clegg M. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA. 93:10274–10279.

Gaut B, Muse S, Clark W, Clegg M. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. J Mol Evol. 35:292–303.

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

Gernandt D, Geada López G, Ortiz García S, Liston A. 2005. Phylogeny and classification of *Pinus*. Taxon. 54:29–42.

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

Gradstein F, Ogg J. 2004. Geologic time scale 2004: why, how and where next! Lethaia. 37:175–181.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 41:95–98.

Hart J. 1987. A cladistic analysis of conifers: preliminary results. J Arnold Arbor. 68:269–307.

Haubold B, Wiehe T. 2001. Statistics of divergence times. Mol Biol Evol. 18:1157–1160.

Hudson R. 1960. The anatomy of the genus *Pinus* in relation to its classification. J Inst Wood Sci. 6:26–46.

Jeffrey E. 1908. On the structure of the leaf in Cretaceous pines. Ann Bot (Lond). 22:207–220.

Kärkkäinen K, Koski K, Savolainen O. 1996. Geographical variation in the inbreeding depression of Scots pine. Evolution. 50:111–119.

Kay K, Whittall J, Hodges S. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. Evol Biol. 6:36–44.

Klekowski E. 1992. Review: mutation rates in diploid annuals—are they immutable? Int J Plant Sci. 153:462–465.

Klekowski E. 1998. Mutation rates in mangroves and other plants. Genetica. 102:325–331.

Klekowski E, Godfrey P. 1989. Ageing and mutation in plants. Nature. 340:389–391.

Koch M, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol. 17:1483–1498.

Komulainen P, Brown G, Mikkonen M, Karhu A, García-Gil M, O'Malley D, Lee B, Neale D, Savolainen O. 2003. Comparing EST based genetic maps between *Pinus sylvestris* and *P. taeda*. Theor Appl Genet. 107:667–678.

Kossack D, Kinlaw C. 1999. *IFG*, a gypsy-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. Plant Mol Biol. 39:417–426.

Koteja J, Poinar G. 2001. A new family, genus, and species of scale insect (Hemiptera: Coccinea: Kukaspididae, new family) from Cretaceous Alaskan amber. Proc Entomol Soc Wash. 103:356–363.

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

Krutovsky K, Troggio M, Brown G, Jermstad K, Neale D. 2004. Comparative mapping in the Pinaceae. Genetics. 168:447–461.

Kumar S, Filipski A, Swarna V, Walker A, Hedges S. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. Proc Natl Acad Sci USA. 102: 18842–18847.

Kutil B, Williams C. 2001. Triplet-repeat microsatellites shared among hard and soft pines. J Hered. 92:327–332.

Lande R, Schemske D, Schultz S. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. Evolution. 48:965–978.

Langenheim R, Smiley C, Gray J. 1960. Cretaceous amber from the Arctic coastal plain of Alaska. Geol Soc Am Bull. 71:1345–1356.

Ma X-F, Szmidt AE, Wang X-R. 2006. Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. Mol Biol Evol. 23:807–816.

Magallón S, Sanderson M. 2001. Absolute diversification rates in angiosperm clades. Evolution. 55:1762–1780.

Magallón S, Sanderson M. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. Evolution. 59:1653–1670.

Mai D. 1986. Über typen und originale tertiärer arten von *Pinus* L. (Pinaceae) in mitteleuropÜäischen sammlungen—ein beitrag zur geschichte der gattung in Europa. Feddes Repert. 97:571–605.
Meijer J. 2000. Fossil woods from the late Cretaceous Aachen Formation. Rev Palaeobot Palynol. 112:297–336.
Millar C. 1998. Early evolution of pines. In: Richardson D, editor. Ecology and biogeography of *Pinus*. Cambridge (UK): Cambridge University Press. p. 69–91.
Miller C. 1973. Silicified cones and vegetative remains of *Pinus* from the Eocene of British Columbia. Contrib Mus Paleontol Univ Mich. 24:101–118.
Miller C. 1976. Early evolution in the Pinaceae. Rev Palaeobot Palynol. 21:101–117.
Miller C, Malinky J. 1986. Seed cones of *Pinus* from the late Cretaceous of New Jersey, USA. Rev Palaeobot Palynol. 46:257–272.
Muse S. 2000. Examining rates and patterns of nucleotide substitution in plants. Plant Mol Biol. 42:25–43.
Muse S, Weir B. 1992. Testing for equality of evolutionary rates. Genetics. 132:269–276.
Near T, Meylan P, Shaffer H. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am Nat. 165:137–146.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3:418–426.
Penny J. 1947. Studies on the conifers of the Magothy flora. Am J Bot. 34:281–296.
Phipps C, Osborn J, Stockey R. 1995. *Pinus* pollen cones from the middle Eocene Princeton chert (Allenby formation) of British Columbia. Int J Plant Sci. 156:117–124.
Posada D, Crandall K. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.
Prager E, Fowler D, Wilson A. 1976. Rates of evolution in conifers (Pinaceae). Evolution. 30:637–649.
Price R, Liston A, Strauss S. 1998. Phylogeny and systematics of *Pinus*. In: Richardson D, editor. Ecology and biogeography of *Pinus*. Cambridge (UK): Cambridge University Press. p. 49–68.
Robison C. 1977. *Pinus triphylla* and *Pinus quinquefolia* from the upper Cretaceous of Massachusetts. Am J Bot. 64:726–732.
Rozas J, Sánchez-DelBarrio J, Messeguer X, Rozas R. 2003. DnaSP, DNA sequence polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
Saiki K. 1996. *Pinus mutoi* (Pinaceae), a new species of permineralized seed cone from the Upper Cretaceous of Hokkaido, Japan. Am J Bot. 83:1630–1636.
Sanderson M. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 19:101–109.
Sanderson M, Doyle J. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and *18s* rDNA data. Am J Bot. 88:1499–1516.
Scofield D, Schultz S. 2006. Mitosis, stature and evolution of plant mating systems: low-Φ and high-Φ plants. Proc R Soc Lond B Biol. 273:275–282.
Shaw G. 1914. The genus *Pinus*. Cambridge (MA): Riverside Press.
Smith S, Stockey R. 2002. Permineralized pine cones from the Cretaceous of Vancouver Island, British Columbia. Int J Plant Sci. 163:185–196.
Sokol K, Williams C. 2005. Evolution of a triplet repeat in a conifer. Genome. 48:417–426.
Stockey R, Nishida M. 1986. *Pinus harborensis* sp. nov. and affinities of permineralized leaves from the Upper Cretaceous. Can J Bot. 64:1856–1866.
Stopes M, Kershaw E. 1910. The anatomy of Cretaceous pine leaves. Ann Bot (Lond). 24:395–402.
Strauss S, Doerksen A. 1990. Restriction fragment analysis of pine phylogeny. Evolution. 44:1081–1096.
Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates.
Syring J, Farrell K, Businský R, Cronn R, Liston A. Forthcoming 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Syst Biol.
Syring J, Willyard A, Cronn R, Liston A. 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am J Bot. 92:2086–2100.
Temesgen B, Brown G, Harry D, Kinlaw C, Sewell M, Neale D. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda*). Theor Appl Genet. 102:664–675.
Thompson J, Higgins D, Gibson T. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
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. Nebst einer systematisch-anatomischen bearbeitung der gattung *Pinus* L. Rev Palaeobot Palynol. 15:73–275.
Wang X-Q, Tank D, Sang T. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol. 17:773–781.
Wolfe K, Li W-H, Sharp P. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA. 84:9054–9058.
Wolfe J, Schorn H. 1989. Taxonomic revision of the Spermatopsida of the Oligocene Creede flora, southern Colorado. US Geol Surv Bull. 1923:1–40.
Wolfe K, Sharp P, Li W-H. 1989. Rates of synonymous substitution in plant nuclear genes. J Mol Evol. 29:208–211.
Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol. 23:212–226.

Spencer Muse, Associate Editor

Accepted September 19, 2006