Phylogenetics of Pinus Subsection Cembroides Engelm. (Pinaceae) Inferred from Low-Copy Nuclear Gene Sequences

Authors: Montes, José Rubén, Peláez, Pablo, Willyard, Ann, Moreno-Letelier, Alejandra, Piñero, Daniel, et al.

Source: Systematic Botany, 44(3) : 501-518

Published By: The American Society of Plant Taxonomists

URL: https://doi.org/10.1600/036364419X15620113920563

Systematic Botany (2019), 44(3): pp. 501–518

DOI 10.1600/036364419X15620113920563
Date of publication August 6, 2019

Phylogenetics of *Pinus* Subsection *Cembroides* Engelm. (Pinaceae) Inferred from Low-Copy Nuclear Gene Sequences

José Rubén Montes,[1] Pablo Peláez,[1] Ann Willyard,[2] Alejandra Moreno-Letelier,[3] Daniel Piñero,[4] and David S. Gernandt[1,5]

[1]Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado Postal 70-233, Ciudad de México, 04510, México; ruben.montes@st.ib.unam.mx; pablo.pelaez@cinvestav.mx
[2]Biology Department, Hendrix College, 1600 Washington Ave, Conway, Arkansas 72032, USA; willyard@hendrix.edu
[3]Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, 04510, México; amletelier@ib.unam.mx
[4]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70-275, Ciudad de México, 04510, México; pinero@unam.mx
[5]Author for correspondence (dgernandt@ib.unam.mx)

Communicating Editor: Sasa Stefanovic

Abstract—*Pinus* subsection *Cembroides* comprises approximately 15 taxa distributed from the southwestern United States to south central Mexico. Despite previous phylogenetic studies based on morphology, nuclear ribosomal DNA, and plastid DNA, we still lack a robust phylogenetic hypothesis and clear delimitation for the closely-related species within the group. We studied the evolutionary relationships within subsection *Cembroides* and explored incomplete lineage sorting and reticulation using low-copy number nuclear genes. Concatenation and multispecies coalescent phylogenies were inferred from samples representing all taxa from subsection *Cembroides* and outgroups corresponding to the closely-related subsections *Balfourianae*, *Nelsoniae*, *Gerardianae*, and *Krempfianae*. The concatenation and coalescence-based trees mainly agreed with one another in recovering *Pinus* subsection *Cembroides* as monophyletic and in recovering similar relationships among species as in previous plastid DNA-based studies. Phylogenetic position and admixture analysis suggest that *P. californiarum* should be treated as a separate species from *P. monophylla*. Furthermore, our results support recognizing *P. fallax* as a species rather than as an infraspecific taxon of *P. monophylla* or *P. edulis*. The ASTRAL-III tree was consistent with the presence of very high levels of ILS in the group of pinyon pines with small cones. Analyses that account for both incomplete lineage sorting and reticulation identify some unexpected hybridization scenarios that were not reported in the literature.

Keywords—Coalescence, pinyon pine, reticulation, target enrichment.

Pinus subsection *Cembroides* Engelm. is a clade of North American pinyon pines occurring in arid or semi-arid environments from the southwestern United States to south central Mexico (Critchfield and Little 1966). The species of this subsection are small to medium-sized trees or shrubs characterized by secondary leaves with deciduous fascicle sheaths. Except for *P. rzedowskii* Madrigal, Caball.M., all taxa have enlarged seeds with a thickened sclerotesta (Madrigal and Caballero 1969) and are functionally wingless. The morphological characters that have been used in species identification in this group include the number of needles per fascicle, needle length and width, the distribution of stomata on the adaxial and abaxial leaf surfaces, and cone morphology (Malusa 1992; Farjon and Styles 1997). *Pinus* subsection *Cembroides* is classified in section *Parrya* Mayr together with two other North American subsections, *Balfourianae* Engelm. and *Nelsoniae* Burgh (Gernandt et al. 2005). *Pinus* subsection *Nelsoniae* is monotypic, with the pinyon pine, *P. nelsonii* Shaw distinguished by persistent fascicle sheaths and connate needles (Little and Critchfield 1969; Gernandt et al. 2001). Section *Parrya* is classified together with section *Quinquefoliae* Duhamel in subgenus *Strobus* Lemmon (Gernandt et al. 2005).

Seeds of *P. cembroides* subsp. *cembroides* Zucc. and *P. edulis* Engelm. are important sources of food in Mexico and the United States (Lanner 1981; Farjon and Styles 1997). The high nutritive value in seeds encourages interactions among pinyon pines, rodents, and corvid birds. The needles are sometimes used for medical treatments (Lanner 1981). The International Union for Conservation of Nature (IUCN 2017) lists five pinyon pine taxa as vulnerable or endangered and the Mexican government lists nine as protected (SEMARNAT 2010; Table 1).

Pinus subsection *Cembroides* has been the focus of several phylogenetic studies (Malusa 1992; Farjon and Styles 1997; Gernandt et al. 2001, 2003; Parks et al. 2012; Flores-Rentería et al. 2013) and the number of recognized species and infraspecific taxa differs in recent works (Malusa 1992; Eckenwalder 2009; Farjon and Filer 2013; Table 2). Phylogenetic results in the subsection have varied due to the use of different morphological, ecological, and molecular characters and differences in sampling (Malusa 1992; Farjon and Styles 1997; Gernandt et al. 2001, 2003, 2005; Syring et al. 2005). Based on a cladistic analysis of morphological and ecological characters, Malusa (1992) divided subsection *Cembroides* into a group of eight species with small seed cones and a second group of four species with large cones. In a restriction site study of noncoding plastid DNA, Pérez de la Rosa et al. (1995) recovered *P. nelsonii* as separate from subsection *Cembroides*. A cladistic analysis of morphological characters in Neotropical species of *Pinus* subgenus *Strobus* by Farjon and Styles (1997) recovered the small-coned pinyons as monophyletic, and the four large-cones species as paraphyletic to subsection *Strobus*. Two of the large-cone species, *P. nelsonii* and *P. pinceana* Gordon, formed a clade, leading the authors to classify them together in subsection *Nelsoniae*. In contrast, Price et al. (1998) classified *P. nelsonii* and *P. pinceana* in subsection *Cembroides* and *P. rzedowskii* in the monotypic subsection *Rzedowskianae* Carvajal.

More recently, phylogenetic relationships of pinyon pines have been inferred from sequences of plastid and nuclear ribosomal DNA (Gernandt et al. 2001, 2003, 2005; Parks et al. 2012; Ortiz-Medrano et al. 2016). Gernandt et al. (2001) reported phylogenetic analyses of the ITS region of nrDNA. The authors found that divergent copies of the ITS region in individuals of the same species do not group together. Nevertheless, results from the ITS region study and two subsequent phylogenetic studies using plastid DNA sequences corroborated the separation of *P. nelsonii* from the

TABLE 1. Taxa classified in *Pinus* subsection *Cembroides* with geographic distribution and conservation risk category. * Subject to special protection, and ** Listed as endangered by the Mexican government (SEMARNAT 2010). † Vulnerable and †† Endangered by the IUCN (2017).

Taxon	Distribution
Pinus californiarum D.K.Bailey	California (CA), Baja California (BC)
Pinus cembroides subsp. *cembroides* Zucc.	Arizona (AZ), New Mexico (NM), Texas (TX), Chihuahua (CH), Coahuila (CL), Durango (DG), Hidalgo (HG), Jalisco (JC), Nuevo León (NL), Querétaro (QO), San Luis Potosí (SP), Sonora (SR),
Pinus cembroides subsp. *orizabensis* D.K.Bailey††	Puebla (PL), Tlaxcala (TL), Veracruz (VZ)
Pinus culminicola Andresen & Beaman** ††	Coahuila (CL), Nuevo León (NL)
Pinus discolor D.K.Bailey & Hawksw	Arizona (AZ), New Mexico (NM), Durango (DG), San Luis Potosí (SP), Sonora (SR)
Pinus edulis Engelm.	Arizona (AZ), Colorado (CO), Nevada (NV), New Mexico (NM), Utah (UT), Texas (TX), Wyoming (WY)
Pinus fallax (Little) Businský	Arizona (AZ), Utah (UT), New Mexico (NM)
Pinus johannis M.F.Robert*	Coahuila (CO), San Luis Potosí (SP), Zacatecas (ZS)
Pinus lagunae (Passini) D.K.Bailey* †	Baja California Sur (BS)
Pinus maximartinezii Rzed.* ††	Durango (DG), Zacatecas (ZS)
Pinus monophylla Torr. & Frém.*	Arizona (AZ), California (CA), Idaho (ID), Nevada (NV), Oregon (OR), Utah (UT)
Pinus pinceana Gordon**	Coahuila (CL), Hidalgo (HG), Querétaro (QO), San Luis Potosí (SP), Zacatecas (ZS)
Pinus quadrifolia Parl. ex Sudw. *	California (CA), Baja California (BC)
Pinus remota (Little) D.K.Bailey & Hawksw. *	Texas (TX), Chihuahua (CH), Coahuila (CL), Nuevo León (NL)
*Pinus rzedowskii*** †	Michoacán (MICH)

other pinyon pines. Plastid DNA studies also have suggested that *P. johannis* M.F.Robert and *P. discolor* D.K.Bailey & Hawksw., are not infraspecific taxa of *P. cembroides* but instead close relatives of *P. culminicola* Andresen & Beaman (Gernandt et al. 2003, 2005; Parks et al. 2012; Ortiz-Medrano et al. 2016).

Some, but not all the results from phylogenetic studies have been followed in subsequent taxonomic treatments (e.g. Eckenwalder 2009; Debreczy and Rácz 2011; Farjon and Filer 2013). Farjon and Filer (2013) recognized that *P. nelsonii* and *P. pinceana* are not closely related, but that *P. nelsonii* belongs to a more distant group, and they recognized that *P. pinceana* is closely related to *P. maximartinezii* Rzed. Farjon and Filer (2013) treated *P. johannis* as a subspecies of *P. cembroides* although this has been contradicted by phylogenetic analysis of plastid DNA. Another plastid DNA study left in question whether the single-leaf pinyon pines are monophyletic, recovering *P. californiarum* D.K.Bailey as sister to *P. edulis* rather than to *P. monophylla* Torr. & Frém. (Gernandt et al. 2007). *Pinus californiarum* has been treated as a synonym of *P. monophylla* for sharing a single needle, or separated as *P. californiarum* or *P. monophylla* var. *californiarum* (D.K.Bailey) Silba based principally on differences in the number of resin canals, number of stomatal lines, and diameter of the needle (Silba 1990; Farjon and Styles 1997; Price et al. 1998).

TABLE 2. Two classifications of the species recognized in *Pinus* subsection *Cembroides*.

Farjon and Filer (2013)	Gernandt et al. (2005)
Pinus cembroides subsp. *cembroides*	
Pinus cembroides subsp. *cembroides* var. *bicolor*	*Pinus cembroides*
Pinus cembroides subsp. *lagunae*	
Pinus cembroides var. *orizabensis*	*Pinus culminicola*
Pinus culminicola	*Pinus discolor*
Pinus edulis	*Pinus edulis*
Pinus maximartinezii	
Pinus monophylla	*Pinus johannis*
Pinus pinceana	
Pinus quadrifolia	*Pinus maximartinezii*
Pinus remota	*Pinus monophylla*
Pinus rzedowskii	*Pinus pinceana*
	Pinus quadrifolia
	Pinus remota
	Pinus rzedowskii

Introgressive hybridization and gene flow have been reported in species of *Pinus* subsection *Cembroides* (e.g. Mirov 1967; Lanner 1974a, 1974b; Lanner and Phillips 1992; Malusa 1992). Some populations of *P. edulis* (predominantly two needles per fascicle) and *P. monophylla* (predominantly single-needled) occur in sympatry (Fig. 1) in the eastern Great Basin where trees of *P. edulis* with both single needles and two needles per fascicle have been observed (Lanner 1974a). In a study of natural hybridization in pinyon pines in northwestern Arizona, Lanner and Phillips (1992) analyzed variation in morphological characters over different years. Based on differences in the frequency of needle and resin canal numbers at 22 sites, they concluded that bidirectional introgression was occurring between *P. edulis* and *P. monophylla*, and proposed that overlapping phenology, wind-dispersed pollen, and weak premating barriers were responsible.

Lanner (1974b) proposed that *P. quadrifolia* Parl. ex Sudw. originated from hybridization between *P. monophylla* (treated here as *P. californiarium*; predominantly single-needled) and a previously undescribed species with five needles per fascicle, *P. juarezensis* Lanner (treated here as a synonym of *P. quadrifolia*). According to Lanner (1974b), this would explain extreme needle number variation in *P. quadrifolia*, a characteristic that is frequently observed in pine artificial hybrids when parents differ in this character (Keng and Little 1961). The geographical distributions of the taxa overlap broadly (Fig. 1), with sympatric populations common in Baja California (e.g. in the Sierra Juárez), suggesting that putative hybrids of several types co-exist in a hybrid swarm (Lanner 1974b). Farjon and Styles (1997) reported that pollen dispersal occurs in April and May in *P. monophylla* and in March and April in *P. quadrifolia*, which would allow for interspecific gene flow. Nonetheless, the proposal to recognize *P. juarezensis* as one of the parental species has not been widely accepted. *Pinus juarezensis* was considered a synonym of *P. quadrifolia* by subsequent authors (Farjon and Styles 1997; Gernandt et al. 2003; Eckenwalder 2009). Plastid DNA from three samples of *P. quadrifolia* (two of which were from the type locality of *P. juarezensis*) grouped together and formed a sister group to *P. monophylla* from California, indicating that *P. quadrifolia* has not captured pollen of *P. californiarum* (Gernandt et al. 2003).

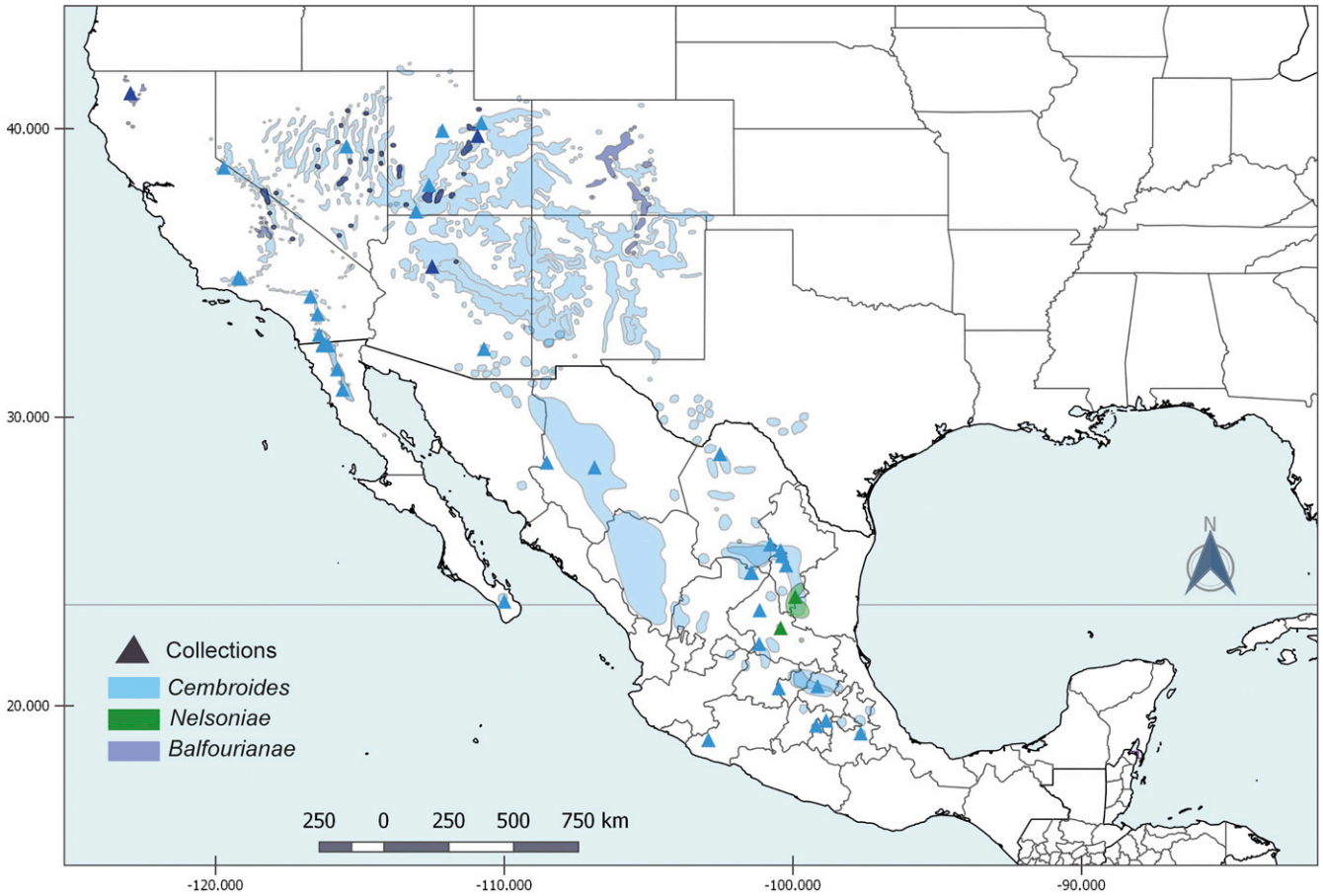

Fig. 1. Map showing the distribution of *Pinus* section *Parrya* based on Critchfield and Little (1966). Species records are based on herbarium collections and data from field studies. The individual sample sites are indicated by triangles (Appendix 1). Colors represent North American subsections that belong to section *Parrya*.

In his description of *P. remota*, Little (1968) suggested the possibility that in the past its populations had been in contact with those of *P. edulis*, permitting introgressive hybridization. The geographic distribution of *P. remota* and *P. edulis* may have been more extensive in the past (Late Quaternary), bringing the species into contact, maybe in the Chihuahuan Desert (Van Devender 1990). Additionally, there is a possible overlap in phenology between *P. remota*, which disperses its pollen between March and April, and *P. edulis*, which disperses pollen in a short period in the spring (Lanner 1970; Farjon and Styles 1997). In the cladistic analysis of morphological characters by Malusa (1992), *P. remota*, which has three needles per fascicle, formed a clade with the single-needle pinyon pines (*P. californiarum*, *P. fallax* (Businský) Little, and *P. monophylla*). The morphological character *P. remota* shares with the single-leaf pinyons is an elevated number of resin canals. An increase in the number of resin canals may have arisen in the species not by independent evolution but by hybridization, likely between *P. remota* and *P. fallax* (Malusa 1992). Plastid DNA of *P. remota* is unique, ruling out the possibility that it was acquired recently through hybridization (Gernandt et al. 2003).

Plastid DNA can provide important insights into phylogenetic relationships, but exclusive dependence on it as a phylogenetic marker can be misleading because plastid capture through introgression has been documented in diverse plant lineages (e.g. Delgado et al. 2007; Gernandt et al. 2018; Morales-Briones et al. 2018). Reticulation through hybridization and incomplete lineage sorting (ILS) also play important roles in gene tree discordance (Rieseberg and Soltis 1991; Maddison 1997) by producing very similar patterns of shared genetic diversity and obscuring phylogenetic relationships, which greatly limits our understanding of the processes of diversification in many lineages (Maddison and Knowles 2006). On one hand, weak premating barriers to gene flow can permit the introduction of alleles from other species (Mirov 1967). On the other, ILS can reduce the differentiation between species. Therefore, conifer species with long generation times (Petit and Hampe 2006) often share genetic variation (e.g. Wang et al. 2011; DeGiorgio et al. 2014). New DNA sequencing technologies and the development of coalescent-based frameworks allow us to better study hybridization and ILS. A promising approach is the use of low-copy nuclear genes to increase phylogenetic resolution and explore the problems of discordance between gene trees and species trees. For example, Syring et al. (2005) found that phylogenetic inference based on low-copy nuclear genes in *Pinus* subgenus *Strobus* increased resolution and robustness in most of the subsectional clades recovered in their study. Gernandt et al. (2018) demonstrated the utility of low-copy nuclear genes to explore ILS and reticulation at the species level in *Pinus* subsection *Australes*. For this reason, our aims were to infer the phylogenetic relationships for species of *Pinus* subsection *Cembroides* and explore the relative importance of incomplete lineage sorting and reticulation as causes of phylogenetic

discordance. We used targeted sequence capture (Gnirke et al. 2009), also known as Hyb-Seq (Weitemier et al. 2014), to characterize nuclear DNA sequences from multiple individuals per species and perform concatenated and multispecies coalescent analyses. This study represents the most complete taxonomic sampling to date for *Pinus* subsection *Cembroides* and its close relatives.

Materials and Methods

Taxon Sampling—We sampled 60 individuals, with most taxa in subsections *Cembroides* and *Nelsoniae* represented by multiple populations (Appendix 1). Vouchers were deposited in the Herbario Nacional de México (MEXU), Instituto de Biología, Universidad Nacional Autónoma of México and the Oregon State University Herbarium (OSC). The individuals represent all taxa of *Pinus* subsection *Cembroides* recognized by Gernandt et al. (2005). For outgroups we included three individuals representing two of three species of subsection *Balfourianae* and three of *P. nelsonii*, the only member of subsection *Nelsoniae* (Fig. 1). These two subsections were recovered as the sister group of subsection *Cembroides* in previous phylogenetic analyses and together with subsection *Cembroides* are classified in section *Parrya* (Gernandt et al. 2005). We also included representative species from section *Quinquefoliae*: subsections *Gerardianae* Loudon (2), *Krempfianae* Little and Critchfield (1), and *Strobus* as more distant outgroups (Appendix 1).

DNA Extraction and Quantification—For extraction of genomic DNA, we followed the modified CTAB method of Doyle and Doyle (1987) for diploid leaf tissue and used a Wizard genomic DNA purification kit (Promega, Madison, Wisconsin) for haploid seed megagametophyte tissue. We used a Nanodrop spectrophotometer 2000/2000c (Thermo Scientific, Waltham, Massachusetts) to measure the absorbance maxima ratio. Samples with 800 ng or more of DNA and an A260/A280 between 2.0 and 2.2 were selected for sequencing. We used a Qubit fluorometer v. 3.0 and dsDNA HS assay kit (Life Technologies, Carlsbad, California) to measure DNA concentration.

Probe Design—Details on probe design, library preparation, sequencing, and gene assembly were described by Gernandt et al. (2018). Briefly, a total of 1045 putative single copy nuclear genes were screened for probe design from *Pinus* species (*P. taeda* L., *P. pinaster* Aiton, and *P. sylvestris* L.; Neves et al. 2013; Willyard et al. 2007). The exon sequences were submitted to Arbor Biosciences (Ann Arbor, Michigan, USA), and 120 bp RNA bait sequences were used to perform a BLAST search on the *P. taeda* draft genome v. 1.0 (Neale et al. 2014; Wegrzyn et al. 2014).

Illumina Library Preparation and Target Enrichment—Genomic libraries were prepared with between 100 and 500 ng of DNA per sample. The DNA was fragmented into ca. 250 bp with a bioruptor and barcode adapters were ligated for sequencing on the Illumina using a TruSeq library prep kit (Illumina, San Diego, California). Libraries were pooled into 24 samples in equimolar ratios and enrichment was carried out with MYbaits biotinylated RNA baits following the manufacturer's protocol v. 2.3.1 (Arbor Biosciences, Ann Arbor, Michigan). The samples were combined in equal concentrations (48×) and sequenced using an Illumina Hi-Seq 2500 with the 100 bp module with paired reads.

Data Selection—We used two different data sets as input for phylogenetic analyses. The principal data set consisted of genes assembled with HybPiper v. 1.2 (Johnson et al. 2016). Data are available in the Dryad Digital Repository (Montes et al. 2019). A second data set consisted of single nucleotide polymorphisms (SNPs) identified with SAMtools (see below). For both, Illumina reads were filtered in Trimmomatic v. 0.36 (Bolger et al. 2014). For nuclear genes, a total of 60 paired R1 and R2 files in *fastq* format were filtered in Trimmomatic, removing bases at read ends with qualities $<$ Q20 using a 4 bp sliding window, and removing reads with a length $<$ 30 bp following trimming (Weitemier et al. 2014). Only reads with both pairs surviving were assembled into individual alignments with HybPiper (Johnson et al. 2016). HybPiper used BWA v. 0.7.1 (Li and Durbin 2009) to align reads to the reference nuclear gene sequences. SAMtools v. 0.1.19 (Li et al. 2009) was used to sort the reads into separate directories for each gene. The pipeline subsequently used SPAdes v. 3.10.1 (Bankevich et al. 2012) for de novo assembly of each gene individually using the retrieved reads. A total of 969 genes from 996 targets were assembled successfully for at least one sample.

The resulting gene assemblies were imported into Geneious v. R9 (Kearse et al. 2012). Individual nuclear gene alignments were performed with MAFFT v. 7.0 (Katoh et al. 2002). We used the following criteria proposed by Gernandt et al. (2018) to filter the multiple sequence alignments: 1) missing one or more samples, 2) fewer than 50% of sites, 3) pairwise similarity less than 93%, and 4) putative paralogs. The first three criteria were applied in Geneious, whereas the paralogs were detected when assembling the only two haploid references, *P. cembroides* USA and *P. bungeana* CA, with HybPiper (Johnson et al. 2016). Paralogs script identified contigs with lengths $\geq$ 85% of the reference sequence, indicating multiple long-length matches. After filters, we excluded 665 of the 969 genes assembled. The remaining 304 genes were carried forward for phylogenetic analysis.

Phylogenetic Analysis Using Low-Copy Number Nuclear Genes—We analyzed a concatenated alignment of 304 nuclear genes for 60 individuals with maximum likelihood in RAxML-HPC v. 8.2.10 (Stamatakis 2014). We performed 1000 heuristic searches for the best tree applying a general time reversible model with the gamma parameter (GTR + G) separately to each gene. Bootstrapping was performed with 1000 replicates as part of the heuristic search with RAxML. The best tree was imported into FigTree v. 1.4.0 for further editing (Rambaut 2012). Samples with unstable topological positions ("rogues") were identified using the maximum dropset size in RogueNaRok (Aberer et al. 2013), available as an online webserver (http://rnr.h-its.org/). These individuals were excluded from the concatenated alignment and multispecies coalescent analyses.

Phylogenetic Analysis Using Coalescent-Based Methods—Coalescent analyses that accommodate ILS and hybridization were performed on the low-copy nuclear genes. Because estimating reticulation is computationally demanding, we used a nuclear gene tree as a guide for choosing a subsample of individuals representative of *Pinus* subsection *Cembroides*. The species tree was estimated with minimizing deep coalescences (MDC), a parsimony method (Maddison 1997; Than and Nakhleh 2009) and networks were estimated by maximum parsimony from gene trees. To study reticulation scenarios, we used PhyloNet v. 3.6.1 (Than et al. 2008) to analyze 22 sequences representing 22 taxa, of which 15 belong to subsection *Cembroides*. The remaining seven taxa were from the outgroup (one individual per species). We used the Perl script BeforePhylo.pl v. 0.9.0 (https://github.com/qiyunzhu/BeforePhylo.git) to produce individual gene alignment files with the reduced representation of samples. The maximum likelihood tree for each of the 304 genes was inferred with RAxML using the GTR + G model and bootstrapping with the autoMRE option. We sequentially permitted up to three reticulation events under the MDC criterion (InferNetwork_MP). For each reticulation setting, we performed 10 independent searches of 5 to 20 (-x 5 to -x 20). The network with the lowest number of extra lineages was selected and displayed graphically with Dendroscope v. 3.0 (Huson and Scornavacca 2012). Based on the results from the 22-terminal analysis, we evaluated the effect of constraining *P. remota* and *P. quadrifolia* as hybrids. We included the taxa proposed to have been involved in past reticulation events with these species. For *P. remota*, these include *P. cembroides*, *P. edulis*, and *P. fallax* (Malusa 1992; Little 1966). For *P. quadrifolia* we included *P. californiarum* (Lanner 1974b).

Species and lineage tree inference was also performed with SVDquartets (Chifman and Kubatko 2014) in PAUP* v. 4.0a150 (Swofford 2002). SVDquartets is a robust method for multilocus data that is designed to build unrooted quartets that accommodate ILS but not reticulation and evaluates the correspondence of the nodes (see Chifman and Kubatko 2014). For this method we used a NEXUS input file that included a data set block specifying the 304 gene partitions and a taxon set assigning each individual to its corresponding species. The file included 58 sequences representing the 15 subsection *Cembroides* taxa and ten individuals from the outgroup, the same as in the concatenated analysis. A maximum of 1,000,000 randomly chosen quartets were analyzed and branch support was estimated with 1000 bootstrap replicates (bs). The trees were displayed graphically in FigTree v. 1.4.0 (Rambaut 2012).

In addition, we concatenated these 304 low-copy number nuclear genes with SNPs (see below) for 51 samples to infer a species tree with SVDquartets. The species trees resulting from the nuclear gene alignment and the concatenated SNPs analyses were displayed graphically as a tanglegram with Dendroscope (Huson and Scornavacca 2012).

The species tree was also estimated with ASTRAL-III v. 5.6.3 (Mirarab and Warnow 2015; Zhang et al. 2017). We used 58 individuals representing 22 taxa, removing two samples, one because it had an unstable position in the RAxML analysis of the concatenated alignment and the other because of the amount of shared ancestry between two species observed with SNPs. Individual trees from the 304 nuclear genes were estimated with RAxML using the GTR+G model and bootstrapping with the autoMRE option. The best tree from each analysis was concatenated and used as the input file for ASTRAL-III. Branch lengths in coalescent units and local posterior

probabilities were estimated for the species tree (Sayyari and Mirarab 2016). We performed character state reconstruction by mapping of the number of leaves per fascicle as an ordered multistate character on the species tree inferred with ASTRAL-III to evaluate the origin of single-needle pinyons based on the likelihood ancestral state approach in Mesquite v. 3.6 (Maddison and Maddison 2018).

Sequence Read Alignment and SNP Calling—Single nucleotide polymorphisms constitute a valuable source of genetic variation to study evolutionary relationships in non-model organisms (Leaché and Oaks 2017). Although only 304 assembled genes were selected to carry out phylogenetic analyses, we observed that for other target genes some regions were assembled and could be used to identify genetic markers in spite of our gene filtering criteria. For this reason, we performed reference-guided SNP calling in an effort to increase the number of sites for phylogenetic analysis. This method provided an alternative way of interpreting the Illumina data compared to assembling with HybPiper and applying our ad hoc filters. It has the potential to identify more informative sites and reduce bias introduced by HybPiper, which only returns one allele per gene. Also, it might be more reliable for removing paralogs or improving homology across samples. Demultiplexed sequence reads from each of the 52 samples were evaluated using FastQC and MultiQC and filtered with Trimmomatic v. 0.36 to discard low quality and adapter sequences (Bolger et al. 2014; Ewels et al. 2016). The minimum length of reads to be kept was set to 50 bp. Paired cleaned sequences were mapped against the *Pinus taeda* genome v. 2.0 (Neale et al. 2014; Wegrzyn et al. 2014) using BWA-MEM with default parameters. The SAM files were converted to BAM files with SAMtools v. 0.1.19 (Li et al. 2009). Uniquely mapped and sorted reads were obtained for each alignment file using the view and sort routines from SAMtools. Potential PCR duplicates were discarded using the Picard tool MarkDuplicates (http://broadinstitute.github.io/picard). SAMtools mpileup and BCFtools (with the options for biallelic variants SNPs/indels, noBAQ, minimum mapping quality of 20, and minimum base quality of 25) were used for variant calling. SNPs were further filtered with VCFtools v. 0.1.13 (Danecek et al. 2011) for phylogenetic and admixture analyses.

SNP-Based Phylogenetic and Admixture Analyses—The SNPs were filtered if genotypes were not called across all samples (100%), minimum mean depth was below 5, minimum quality score was below 30, or minor allele frequency was lower than 0.05. The variant calling format (VCF) file was converted into a tab-delimited text file with VCFtools and subsequently SNPs were concatenated into a FASTA file including heterozygous sites. A multiple sequence alignment was generated with MAFFT v. 7.0 (Katoh et al. 2002) and used as input to infer a maximum likelihood tree in RAxML-HPC v. 8.0.26 (Stamatakis 2014). The maximum likelihood and bootstrap searches were carried out with the GTR+G model and a total of 1000 bootstrap replicates were computed.

For the admixture analysis, SNPs were filtered if genotypes called were below 80% across all samples, minimum mean depth was below 5, and minimum quality score was below 30. The VCF file was converted to an ordinary PLINK file (.ped) using PLINK v. 1.9 (Purcell et al. 2007). Maximum likelihood estimation of individual ancestries (population structure analysis) was carried out using ADMIXTURE v. 1.3.0 (Alexander et al. 2009). ADMIXTURE's cross-validation procedure (5-fold) was run for ancestral populations values (K) from 2–5. The Q-matrices generated for the different K values (1–4) were then clustered and plotted using CLUMPAK v. 1.1 (Kopelman et al. 2015).

Results

Filtering and Editing Sequences—Statistics for the 60 samples assembled in HybPiper are summarized in Appendix 2. Fifty samples corresponded to subsection *Cembroides* and 10 to the outgroups. Mean coverage of the 969 genes was 226.42×. We excluded 68.6% of genes as a result of the data filtering steps (Table 3), resulting in a final tally of 304 included gene alignments.

Phylogenetic Analyses Using Nuclear Genes—The alignment of 304 concatenated nuclear genes was 222,129 bp in length and included 16,503 parsimony informative sites and 13,639 variable but parsimony uninformative sites. The RAxML analysis recovered *Pinus* subsection *Cembroides* as monophyletic (100% bs). Rooting with section *Quinquefoliae* resulted in recovering subsections *Nelsoniae* and *Balfourianae* sister to one another and in turn sister to subsection *Cembroides* (Fig. 2). This section *Parrya* clade of exclusively North American taxa was recovered with high bootstrap support (100% bs). In the outgroup, the subsections *Krempfianae* and *Gerardianae* were united (100% bs). In the subsection *Cembroides* clade, eight of 13 taxa represented by multiple individuals were recovered as exclusive lineages (*P. culminicola*, *P. johannis*, *P. fallax*, *P. maximartinezii*, *P. monophylla*, *P. pinceana*, *P. remota*, and *P. rzedowskii*; Fig. 2). The bootstrap values for most branches in the phylogeny were high. The three large-cone pinyon pines with a restricted geographical distribution in Mexico formed a well-supported clade, in which *P. maximartinezii* and *P. pinceana* formed a monophyletic group (100% bs) sister to *P. rzedowskii* (100% bs). The large-cone clade was sister to a clade of the remaining (small-cone) species of subsection *Cembroides*. In the small-cone clade, *P. johannis* and *P. discolor* were sister to *P. culminicola* (94% bs). Five individuals of *P. remota* were sister to the clade of *P. johannis*, *P. discolor*, and *P. culminicola* (92% bs). *Pinus cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. lagunae* were paraphyletic to the clade of *P. discolor*, *P. johannis*, *P. culminicola*, and *P. remota*, with individuals of *P. cembroides* subsp. *cembroides* from Chihuahua and San Luis Potosí forming a clade with two individuals of *P. lagunae* (Baja California Sur), and *P. cembroides* subsp. *orizabensis* (Puebla) sister to an individual of *P. cembroides* subsp. *cembroides* from Hidalgo. The two individuals of *P. fallax* formed a group with one sample of *P. edulis* (93% bs), all from Utah (USA). *Pinus monophylla*, *P. californiarum*, and *P. quadrifolia* were recovered as a clade with high bootstrap support (98% bs). *Pinus monophylla* was recovered as monophyletic and sister to *P. quadrifolia*, and both as sister to a single sample of *P. californiarum* from Baja California.

Only one sample, *Pinus monophylla* UT1 (DSG478), was identified as having an unstable position with RogueNaRok (dropset size 3.0). We did not include it in the subsequent coalescence analyses with MDC, SVDquartets, and ASTRAL-III, and it was removed from the concatenated analysis in RAxML. Also, one sample of *P. californiarum* (CA2; DSG403) was removed from the coalescence and concatenated nuclear genes analyses to avoid the probability of bias in the evolutiionary relationships due to the proportions of shared ancestry observed for this sample with SNPs (see Fig. 3A).

SNP-Based Phylogenetic and Admixture Analyses—Mapping the Illumina reads to the genome of *P. taeda* identified 26,499

Table 3. Results of the data after filtering. [a]One or more samples were not assembled to the reference sequence (996 genes). [b]Fewer than 50% of sites identical. [c]Pairwise identity less than 93%. [d]The HybPiper script identified contigs with lengths ≥ 85% of the reference sequence, indicating multiple long-length matches (see Johnson et al. 2016).

	Missing data[a]	Percent of alignment columns identical[b]	Pairwise identities[c]	Paralogs[d]	Eliminated by visual inspection	Total
Genes excluded	160	299	30	103	73	665
Percent	16.5%	30.9%	3.1%	10.6%	7.5%	68.6%

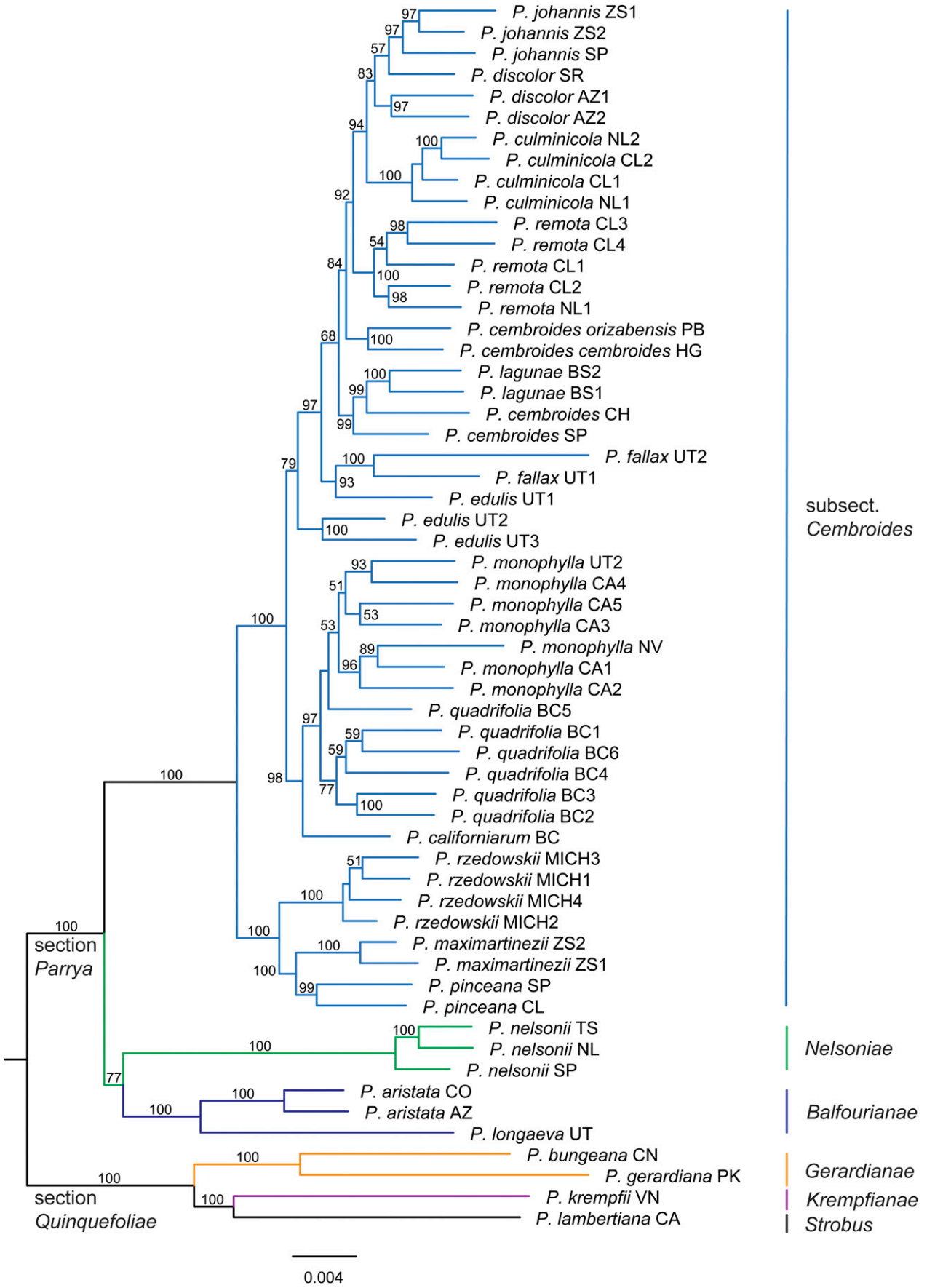

FIG. 2. Phylogeny of *Pinus* subsection *Cembroides*. Maximum likelihood tree inferred from the concatenated alignment (222,129 bp; 58 terminals and 304 genes). Bootstrap values > 50% are shown above the branches. The taxonomic subsections are represented by colors in some trees in this study (blue = *Cembroides*; green = *Nelsoniae*; navy blue = *Balfourianae*; orange = *Gerardianae*; violet = *Krempfianae*; black = *Strobus*). The sample names indicate the taxon and the state of collection. Locality codes are provided in Table 1.

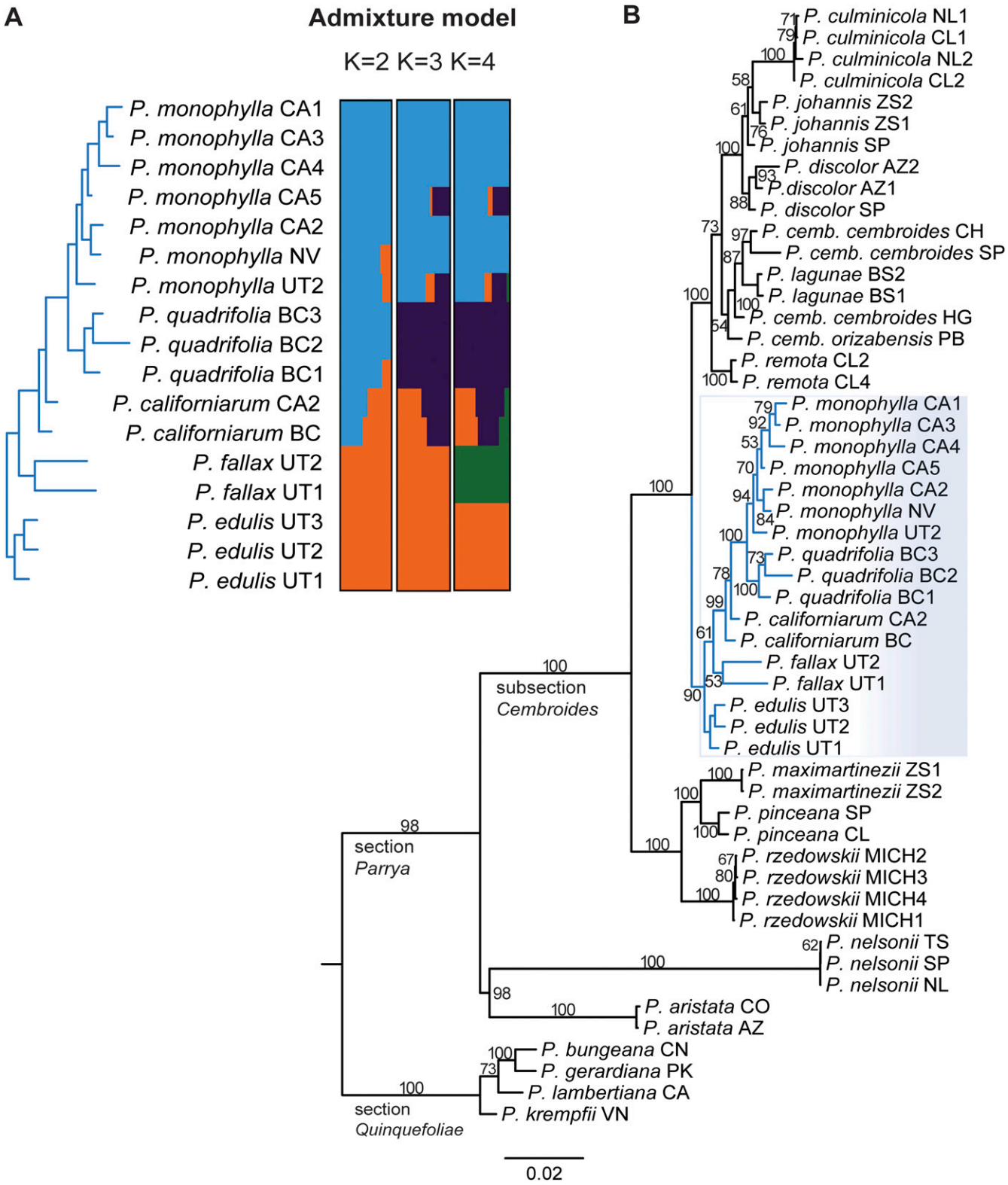

FIG. 3. Admixture analysis of a pinyon pine clade. A. Analyses were performed for *K* values ranging from 2 to 4 with a matrix containing 26,499 single nucleotide polymorphisms. Different colors represent different clusters. The combination of different colors in a bar indicates the degree of admixture. Samples in the admixture model are in the same topological order as in the maximum likelihood phylogeny which is shown at right for comparison. B. Maximum likelihood tree inferred from single nucleotide polymorphisms. Bootstrap values > 50% are shown above the branches.

SNPs genotyped in all 52 pines, including 7262 parsimony informative sites and 735 variable but parsimony uninformative sites. The resulting maximum likelihood tree inferred with RAxML (Stamatakis 2014; Fig. 3B) had a similar backbone topology as that inferred from the concatenated nuclear genes (Fig. 2), with *Pinus* subsection *Cembroides* monophyletic (Fig. 3B) and divided into two main clades comprising the large-cone pinyon pines (100% bs) and the small-cone pinyon pines (100%

bs). In the small-cone clade, *P. californiarum*, *P. monophylla*, *P. quadrifolia*, *P. fallax*, and *P. edulis* formed a well-supported monophyletic group (90% bs). *Pinus culminicola*, *P. johannis*, *P. discolor*, *P. lagunae*, *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. remota* formed another well-supported clade (100% bs). In contrast to the analysis of the HybPiper exon assembly (Fig. 2), the maximum likelihood tree inferred from SNPs grouped the two subspecies of *P. cembroides* and *P. lagunae* together in an exclusive lineage (Fig. 3B). *Pinus californiarum* individuals were sister to *P. monophylla* + *P. quadrifolia*, but the two samples of *P. californiarum* were paraphyletic in the SNP-based phylogeny, suggesting significant genetic variation within this species (Fig. 3B)

The interesting phylogenetic relationships for the clade comprising *P. californiarum*, *P. monophylla*, *P. quadrifolia*, *P. fallax*, and *P. edulis*, together with previous evidence of hybridization and introgression between some of these species, led us to explore the genetic structure of the individuals from this clade using ADMIXTURE. In total, 67,938 SNPs corresponding to only these five species were used. The cross-validation errors for ancestral structured populations values increased as the number of populations increased. Therefore, although the lowest cross-validation error was for K = 2, we used a cluster range of 2–4 to explore the dynamics of classifications in the different number of populations. Interestingly, for the lowest K all the samples were mainly defined by one subpopulation with the exception of the sample corresponding to *P. californiarum* CA2, which showed the strongest signs of admixture (Fig. 3A). For K = 3, samples from *P. quadrifolia* defined a third substructure and admixture patterns for sample CA2 was maintained. *Pinus edulis* and *P. fallax* were grouped differentially only at the highest K value. *Pinus californiarum* CA2 and *P. californiarum* BC1 consistently showed evidence of admixture across all three values of K. *Pinus monophylla* individuals were mainly assigned to the same population regardless of which K value was used (Fig. 3A). Also, *P. edulis* and *P. fallax* showed consistency regarding their classification in the same population for the first two K values.

Phylogenetic Results for Coalescent-Based Methods—HybPiper exon data were used to infer a species tree under the MDC criterion with Phylonet on a subset of subsection *Cembroides* individuals (22). *Pinus* subsection *Cembroides* was recovered as monophyletic (8084 lineages; Fig. 4). The analyses sequentially permitting up to three reticulation events (Fig. 4B–D) always identified gene flow within subsection *Cembroides*; none was identified among the outgroups. Allowing for a single reticulation resulted in a reduction from 8084 to 7619 lineages and involved introgression from *P. monophylla* into *P. edulis* (Fig. 4B). Allowing for two reticulations resulted in a reduction to 7453 lineages (Fig. 4C). The first reticulation involved introgression from *P. fallax* into *P. cembroides* subsp. *cembroides* (Fig. 4C: H1) and the second involved introgression from *P. edulis* into *P. lagunae* (Fig. 4C: H2). Allowing for three reticulations resulted in a reduction to 7211 lineages (Fig. 4D). The first reticulation (Fig. 4D: H1) involved introgression from *P. culminicola* into *P. quadrifolia*, the second involved introgression from a possible extinct taxon into *P. lagunae* (Fig. 4D: H2), and the third involved introgression from a possible extinct taxon into *P. cembroides* subsp. *cembroides* (Fig. 4D: H3).

Specifying *P. remota* as a hybrid under MDC resulted in a reduction from 8084 lineages inferred in the species tree (Fig. 4A) to 7683 in the network (Fig. 5A). *Pinus remota* was sister to *P. johannis* and gene flow was inferred from *P. fallax* (28% inheritance probability). Specifying *P. quadrifolia* as a hybrid resulted in a reduction to 7627 lineages (Fig. 5B). In this network *P. quadrifolia* was sister to *P. monophylla* (54% inheritance probability) and gene flow was inferred from *P. lagunae* (46% inheritance probability) (Fig. 5B).

Pinus subsection *Cembroides* was monophyletic and relationships among the outgroups were well supported in the SVDquartets and ASTRAL-III trees (Figs. 6B–8). The Mexican pinyon pine *P. nelsonii* (subsection *Nelsoniae*) and *P. aristata* and *P. longaeva* were sister to subsection *Cembroides* in the SVDquartets lineage and species trees (Figs. 6–7) but was sister to subsections *Cembroides* + *Balfourianae* in the ASTRAL-III tree (Fig. 8). *Pinus maximartinezii* and *P. pinceana* were sister and in turn sister to *P. rzedowskii* in all trees (Figs. 2–8). In the lineage tree individuals of *P. monophylla* formed a group with *P. quadrifolia* and *P. californiarum* but individuals of the same species were not recovered as exclusive lineages. Individuals of *P. edulis* also were not recovered as exclusive lineages (Fig. 6B). *Pinus lagunae* and the taxonomic varieties of *P. cembroides* were not recovered as exclusive lineages (Fig. 6), whereas *P. remota*, *P. culminicola*, and *P. johannis* were each recovered as exclusive lineages (> 95% bs).

In both the SVDquartets and ASTRAL-III analyses, *P. remota* was sister to the *P. culminicola*, *P. johannis*, and *P. discolor* clade (Figs. 7–8). The relationships among *P. discolor*, *P. culminicola*, and *P. johannis* differed between ASTRAL-III and SVDquartets (Figs. 6–8). Whereas *P. culminicola* was sister to *P. johannis* and *P. discolor* in the SVDquartets species tree, in the ASTRAL-III tree *P. discolor* was sister to *P. culminicola* and *P. johannis* (Fig. 8). The position of *P. fallax* and *P. edulis* also differed in the two coalescence analyses. Moreover, the position of *P. edulis* as sister to both varieties of *P. cembroides*, *P. culminicola*, *P. discolor*, *P. johannis*, *P. lagunae*, and *P. remota* was not well supported in the SVDquartets species tree (65% bs) indicating a weak node (Fig. 7). Branch lengths in the tree inferred with ASTRAL-III were short within subsection *Cembroides* except for the branch subtending the clade with *P. maximartinezii*, *P. pinceana*, and *P. rzedowskii*.

The likelihood results for character state reconstruction on the ASTRAL-III species tree supported a common origin of reduction to single needles in *P. californiarum, P. fallax*, and *P. monophylla* followed by a single increase in the number of needles per fascicle in *P. quadrifolia* (-log 33.50).

Comparison Between SVDquartets Analyses from SNPs and Nuclear Genes—The concatenated nuclear gene alignment was 222,129 bp in length (16,503 parsimony-informative sites) compared to 26,499 characters for the SNP data set (7262 parsimony-informative sites). The analyses performed in SVDquartets recovered species trees that were topologically similar. Trees based on both SNPs and concatenated nuclear genes (Fig. 9) recovered subsection *Cembroides* as monophyletic with high branch support and the large- and small-cone clades were sister groups in both analyses. The relationships within subsection *Cembroides* differed for several small-cone species. The bootstrap values for the analyses based on concatenated genes and SNPs were similar, but in some positions the support was higher using concatenated genes (Fig. 9B). In the concatenated gene analysis, *P. remota* was sister to the *P. culminicola* + *P. discolor* + *P. johannis* clade (96% bs), whereas with SNPs, *P. cembroides* subsp. *orizabensis* was sister to the *P. culminicola* + *P. discolor* + *P. johannis* clade

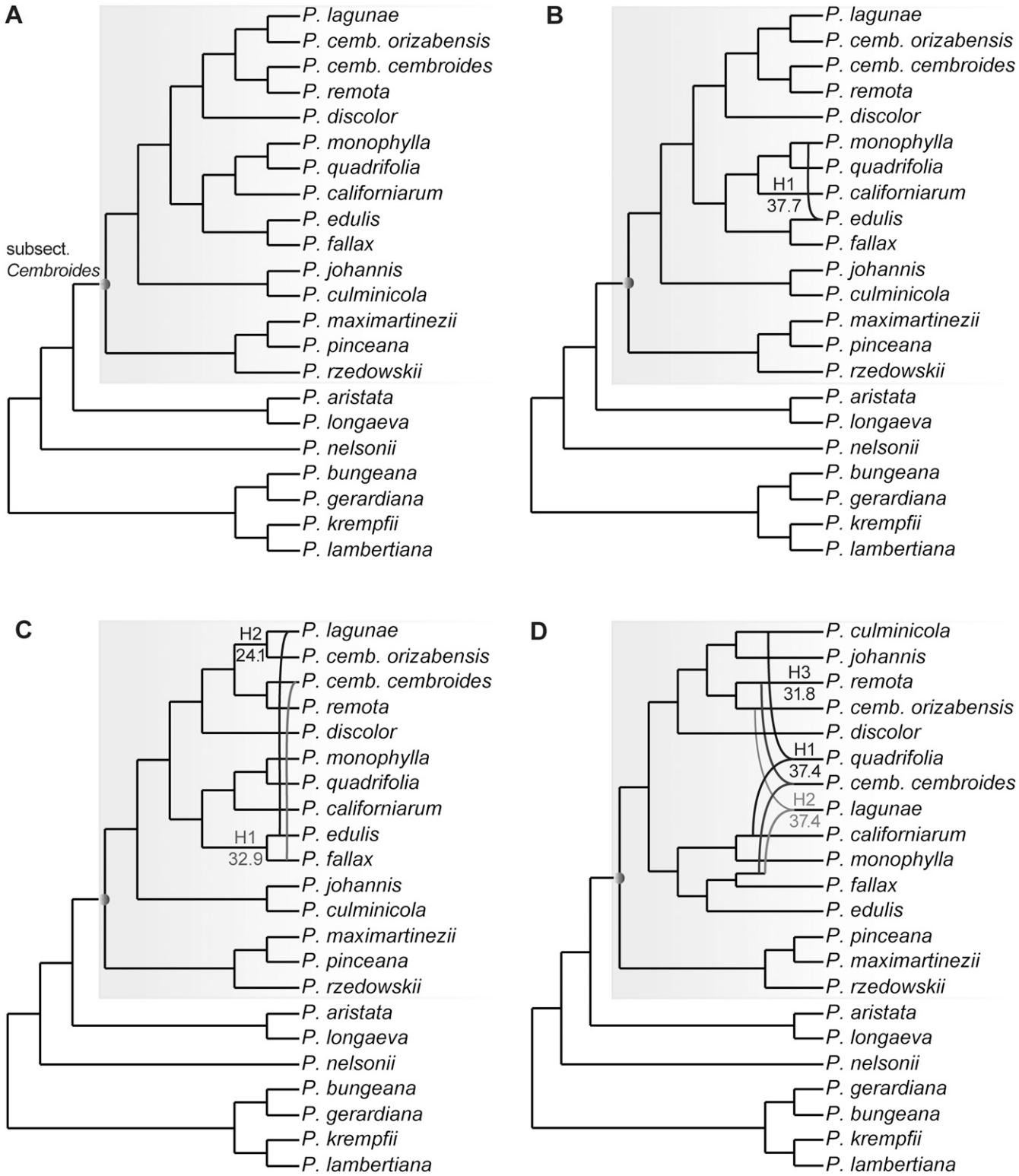

FIG. 4. Results of minimizing deep coalescences (MDC) analyses with and without reticulation. Inheritance probabilities for the minor edge are indicated for the networks (B, C, and D), with the reticulations represented by lines. A. Best MDC tree with no reticulation. B. Best MDC network permitting one reticulation event. C. Best MDC network permitting two reticulation events. D. Best MDC network permitting three reticulation events.

(61% bs) with SNPs. The clade of *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis,* and *P. lagunae* was recovered as monophyletic with the concatenated gene alignment (72% bs), but with SNPs it was paraphyletic due to the placement of *P. cembroides* subsp. *orizabensis* as sister to the clade of *P. discolor*, *P. johannis*, *P. culminicola*, and *P. remota* (61% bs). In the SNPs tree, *P. fallax* was sister to the single-needle pinyons together with *P. quadrifolia* (63% bs) whereas in the concatenated gene tree it was sister to all other small-cone species (100% bs). In both analyses of SVDquartets,

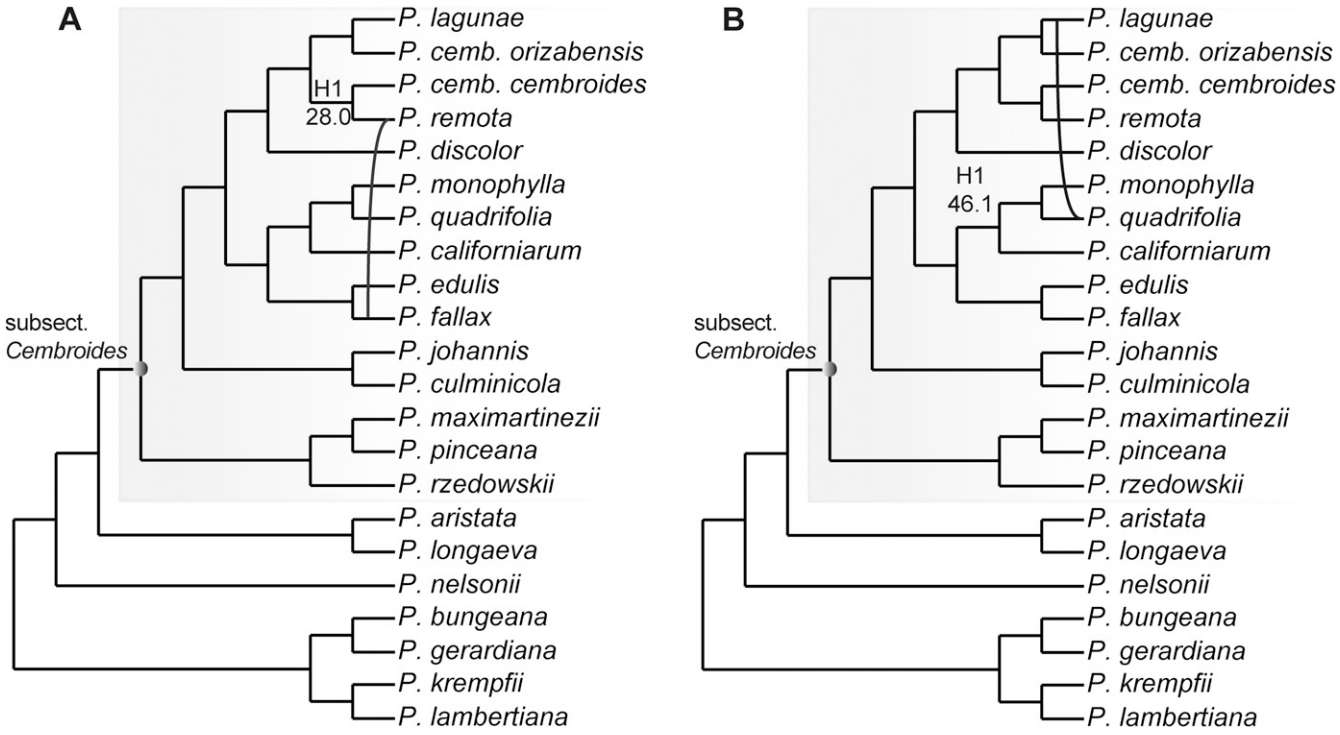

FIG. 5. Tests constraining *P. remota* and *P. quadrifolia* to be hybrids under the minimizing deep coalescences criterion (MDC). A. Best MDC network constraining *P. remota* as a hybrid. B. Best MDC network constraining *P. quadrifolia* as a hybrid. Reticulation events are represented by lines and inheritance probabilities are indicated at reticulations.

P. monophylla was sister to *P. quadrifolia* rather than to *P. californiarum*. With SNPs, *P. nelsonii* + *P. aristata* were sister to subsection *Cembroides* (50% bs), whereas with the nuclear exon alignment only *P. nelsonii* was recovered as sister to subsection *Cembroides* (61% bs).

Discussion

Phylogeny of Pinus Subsection Cembroides—The phylogenies inferred with low-copy nuclear genes for *Pinus* subsection *Cembroides* recovered two main lineages (Figs. 2, 8). Nonetheless, some interrelationships vary among analysis and with previous studies. In studies with the nrDNA ITS region (Gernandt et al. 2001) and cpDNA (Gernandt et al. 2005; Parks et al. 2012; Ortiz-Medrano et al. 2016), *P. rzedowskii* is sister to the remaining species of *Pinus* subsection *Cembroides*. Here, *P. rzedowskii* is sister to *P. pinceana* and *P. maximartinezii* (100% bs). The same relationship of these large-cone pinyon pines was recovered with plastid DNA sequences by Gernandt et al. (2007) but did not receive high support (72% bs). In both the concatenated and coalescence-based analyses, the *P. rzedowskii* + *P. pinceana* + *P. maximartinezii* clade is always sister to the rest of the pinyon pines with strong support (100% bs). If *P. rzedowskii* forms a clade with *P. maximartinezii* + *P. pinceana*, instead of being the sister group of all the other species in subsection *Cembroides,* this implies either multiple gains of enlarged, functionally wingless seeds in subsection *Cembroides*, or a reversion to small winged seeds in *P. rzedowskii* (Gernandt et al. 2007). Maximum likelihood analyses of plastid DNA support the sister relationships between *P. discolor* + *P. johannis* and *P. culminicola* (Gernandt et al. 2007; Parks et al. 2012). The results with low-copy nuclear genes support that relationship with the concatenated alignment in RAxML and the SVDquartets species tree (Figs. 2, 7). In contrast, in the ASTRAL-III and SNPs trees (Figs. 3A, 8), *P. discolor* is recovered as sister to *P. culminicola* + *P. johannis*; this relationship was also supported with plastomes (Parks et al. 2012). Although Farjon and Styles (1997) treated *P. discolor* and *P. johannis* as a single variety of *P. cembroides* (as *P. cembroides* var. *bicolor* Little), the stomata of *P. discolor*, *P. johannis*, and *P. culminicola* are limited to the adaxial surfaces and the seed megagametophyte is white rather than pink. Our results with low-copy nuclear genes from the concatenated alignment, SVDquartets, and ASTRAL-III, agree with analyses of plastid DNA (Gernandt et al. 2007) in recovering *P. johannis* + *P. discolor* + *P. culminicola* as the sister group of *P. remota*, rather than grouping these species with *P. cembroides*. This relationship was not recovered with cladistic analyses of morphological characters (Malusa 1992). The coalescence analyses at the species level strongly support the clade of pinyon pines that are distributed in the Sierra Madre Oriental (SMO), *P. discolor* (mainly distributed in the Sierra Madre Occidental and the Sky Islands of the United States, but also present in the southern part of the Sierra Madre Oriental), *P. johannis*, and *P. culminicola* (100% bs in SVDquartets and ASTRAL-III; Figs. 7–8). The monophyly of the three taxa has been attributed to the evolution from an ancestor that was resistant to the calcareous soils that predominate in the SMO (Malusa 1992). Therefore, limestone soils tolerance may be a synapomorphy for *P. discolor*, *P. johannis*, and *P. culminicola*.

Infraspecific taxonomies of *P. cembroides* and their relationship to *P. lagunae* have disagreed (Zavarin and Snajberk 1985; Passini 1987; Farjon and Styles 1997; Farjon and Filer 2013). Phylogenetic results based on three plastid DNA regions recovered *P. cembroides* subsp. *cembroides*, *P. cembroides*

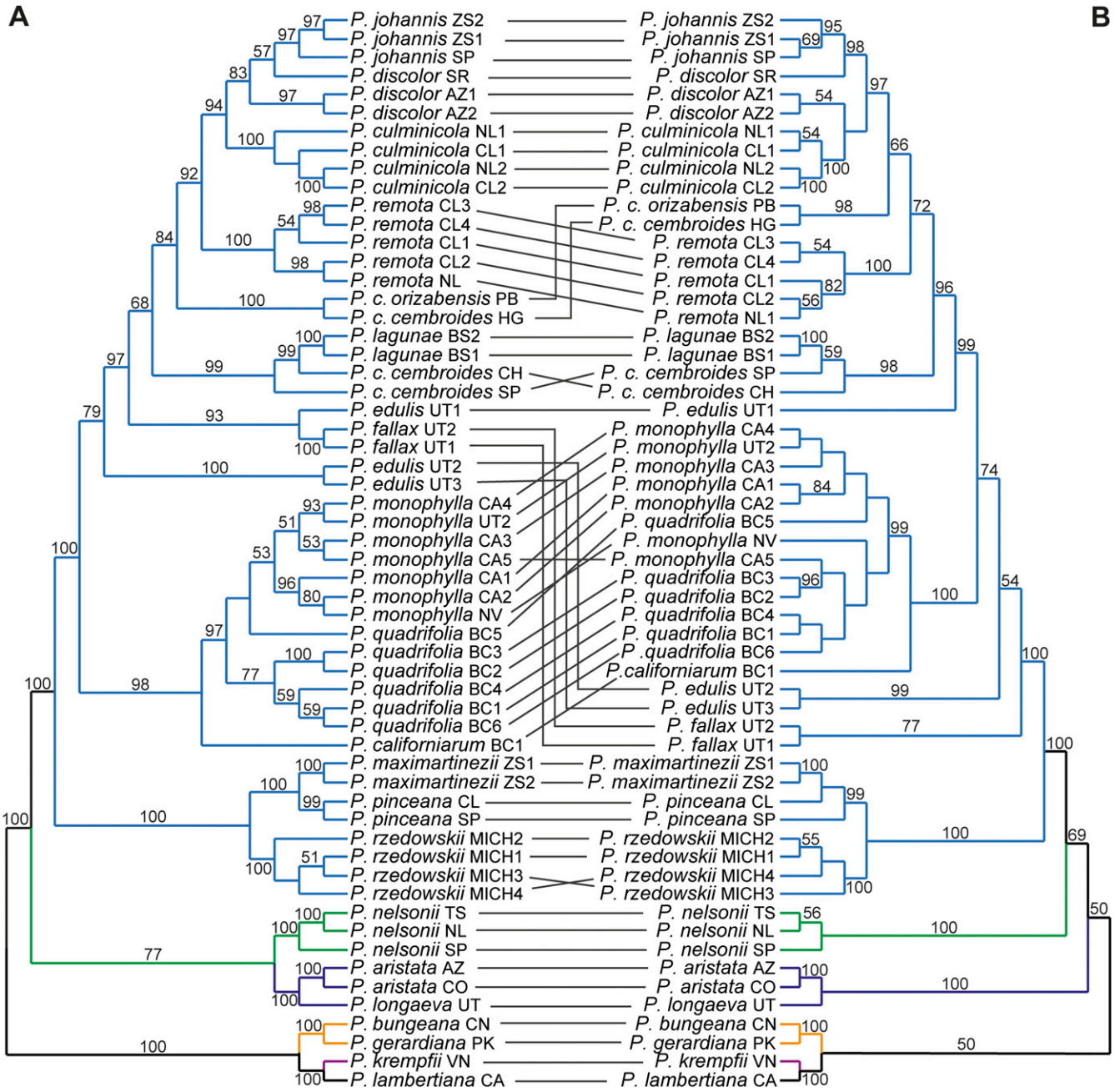

FIG. 6. Tanglegram of RAxML tree and SVDquartets lineages tree. A. Maximum likelihood tree based on concatenated alignment with 304 nuclear genes. B. Coalescent-based tree based on 1,000,000 quartets. Trees were estimated with a matrix of 222,129 bp and 58 terminals. Bootstrap values > 50% are shown above the branches. The colors of the branches follow Fig. 2.

subsp. *orizabensis*, and *P. lagunae* as a single exclusive lineage (Gernandt et al. 2003). Our results from nuclear gene assemblies (but not from SNPs; Fig. 9A) coincide with Gernandt et al. (2003) in recovering both varieties of *P. cembroides* and *P. lagunae* as an exclusive lineage (Figs. 7–9). Previously, Whang et al. (2001) reported differences of the leaf internal cuticle among *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. lagunae*. For instance, *P. cembroides* subsp. *orizabensis* differs from *P. cembroides* subsp. *cembroides* and *P. lagunae* by the width of the epidermal cell apex (thick in *P. cembroides* subsp. *orizabensis*), continuity of cell walls, stomatal apparatus shape, and cuticular flange-guard cell. *Pinus cembroides* subsp. *cembroides* and *P. lagunae* share more characters of the internal cuticle with each other than with *P. cembroides* subsp. *orizabensis*.

Zavarin and Snajberk (1985) found that the populations of *P. cembroides* subsp. *orizabensis* from southern Puebla and northeastern Veracruz differ from *Pinus cembroides* subsp. *cembroides* and *P. lagunae* in their chemical composition of monoterpenes but are very similar morphologically. They suggested that the divergence of southern populations of *P. cembroides* (*P. cembroides* subsp. *orizabensis*) as an isolated taxon most likely resulted from climatic and geographic isolation (middle Miocene). They also suggested that the isolation of *P. lagunae* from Baja California was related to movement of the coastal region from California during the

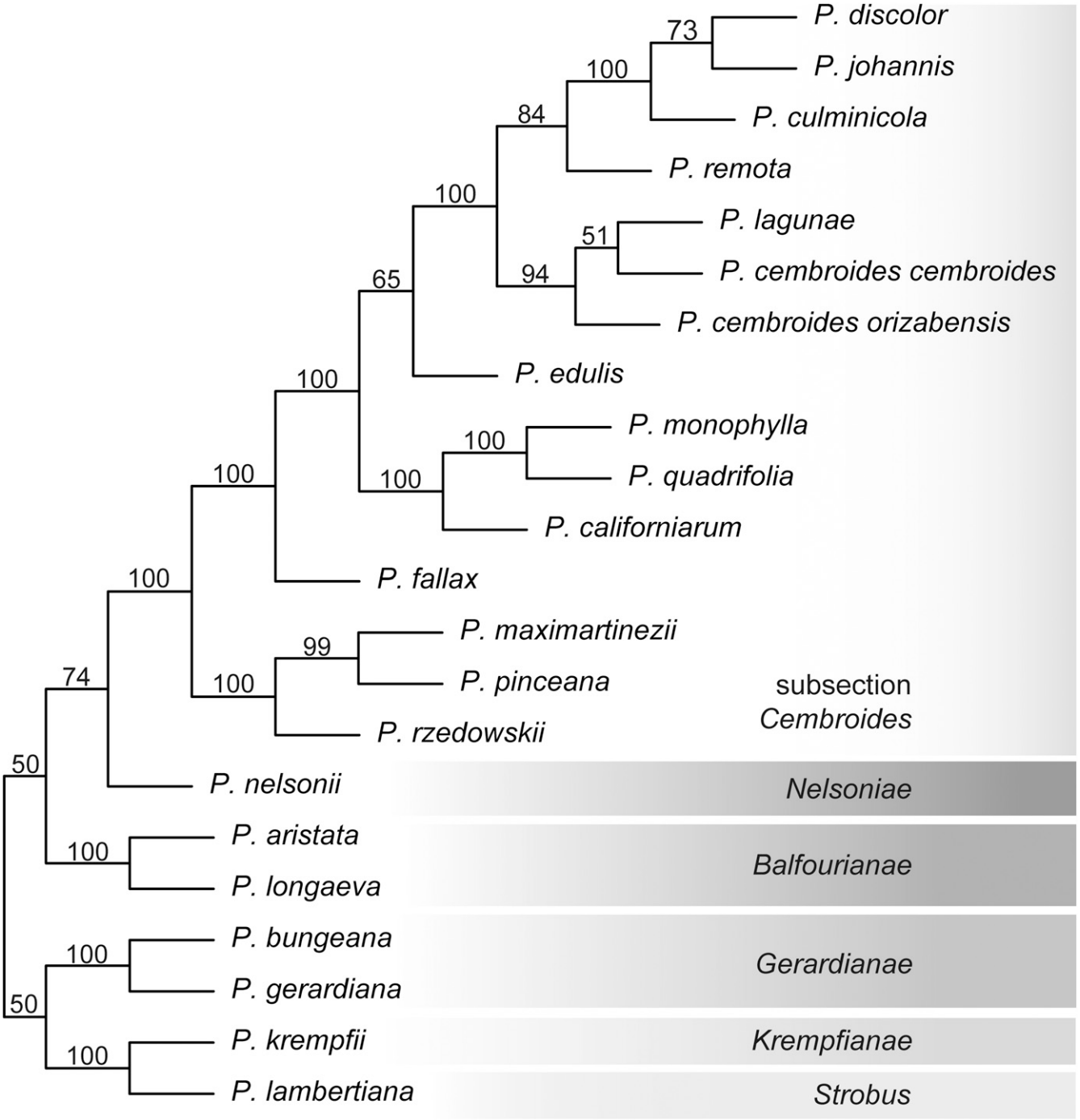

FIG. 7. Species tree inferred with SVDquartets. Coalescent-based tree estimated from 1,000,000 quartets (222,129 bp, 58 taxa and 304 genes). Total weight of incompatible quartets = 12.89%, and total weight of compatible quartets = 87.11%. Bootstrap values > 50% are shown above the branches.

Miocene and the formation of the Sierra Madre Occidental, resulting in the separation of this population from the rest of the continental populations by the breach of the Gulf of California (Zavarin and Snajberk 1985). However, this vicariance event is probably too old to explain divergence between *P. lagunae* and *P. cembroides*. The entire small-cone pinyon pine clade was estimated to have diversified during the Miocene, ~11 MYA (Gernandt et al. 2008; Saladin et al. 2017). More studies are needed to determine whether these three taxa represent independent evolutionary entities or the same species.

The coalescence-based analyses of concatenated genes and SNPs support the sister relationship between *P. monophylla* and *P. quadrifolia* and place both taxa as sister to *P. californiarum*. This relationship was not recovered with morphology, where *P. quadrifolia* is sister to *P. johannis* + *P. discolor* + *P. culminicola*, united by sharing resinous cones (Malusa 1992). In this study we observed two different topologies with respect to the origin of a single needle. In both the SVDquartets and ML analysis of the concatenated alignment, single-needle taxa occurred in independent lineages (Figs. 2, 7). In contrast, ASTRAL-III and analyses of SNPs

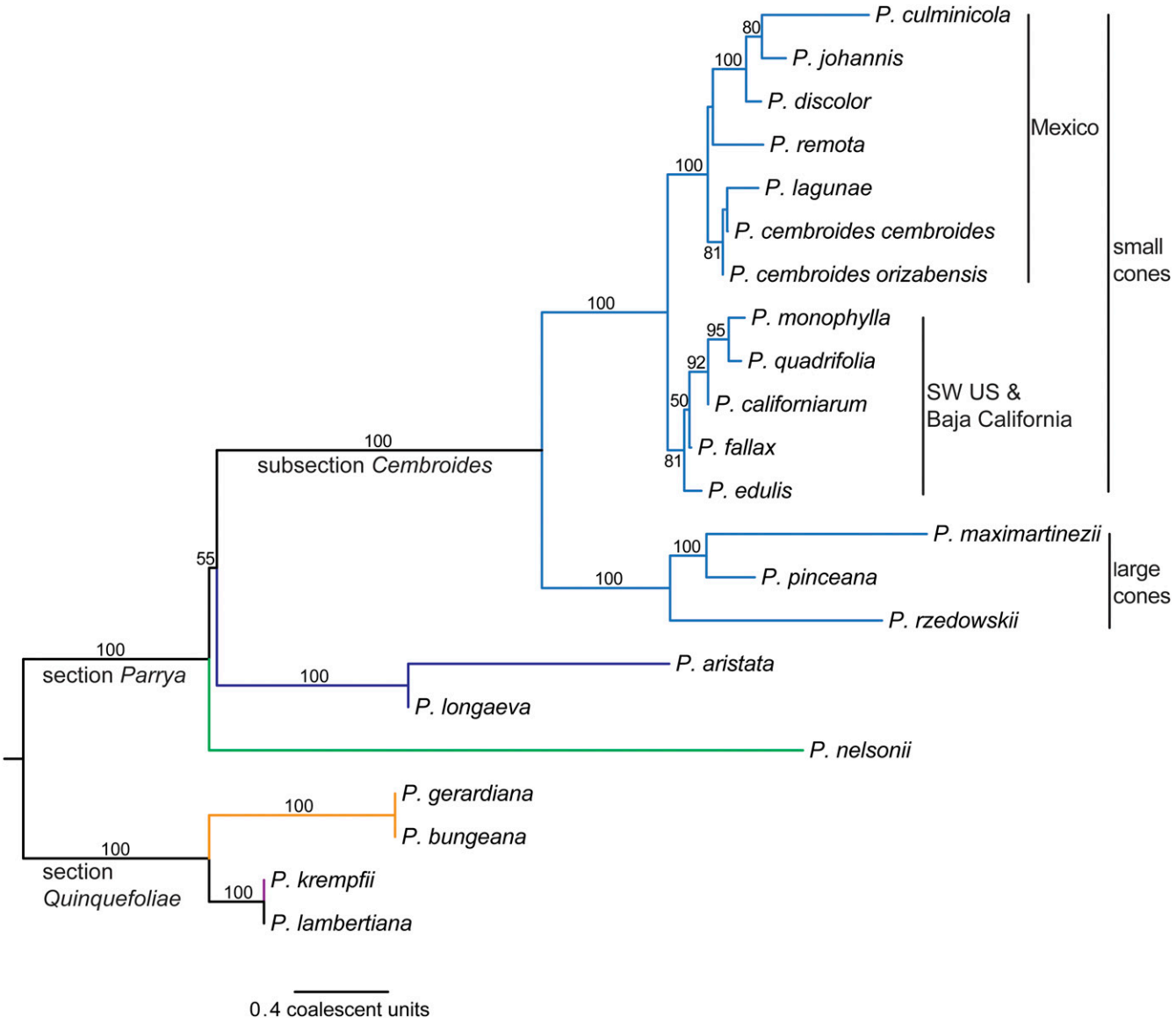

FIG. 8. Species tree inferred with ASTRAL-III. The tree was estimated under a species coalescent model using 304 nuclear gene trees with 58 individuals as input. Final quartet score = 73,455,073 and final normalized quartet score = 0.58 (very high incomplete lineage sorting). The bootstrap support values are provided above branches. The branch length represents coalescent units. The colors of the branches follow Fig. 2.

grouped *P. californiarum*, *P. fallax*, and *P. monophylla* together but paraphyletic to *P. quadrifolia* (Figs. 8–9). We performed a character state reconstruction with the ASTRAL-III species tree and the results supported a single reduction to single needles and common origin but with one independent loss in *P. quadrifolia*. Cole et al. (2013) compared variation in needle number to environmental variation and found that the proportions of the number of needles in *P. edulis* and *P. fallax* depend on annual fluctuations in precipitation. In addition, it was shown that *P. edulis* and *P. fallax* occur in an area with monsoon precipitation extremes, whereas *P. monophylla* and *P. californiarum* occur in areas with high levels of winter precipitation (Cole et al. 2008).

In morphology-based views of phylogeny, *P. monophylla*, *P. californiarum*, and *P. fallax* are recovered together by sharing resinous cones, single needles (predominantly), and thinner seed coats (Malusa 1992). With nuclear genes, *Pinus monophylla*, *P. californiarum*, and *P. fallax* are not recovered as monophyletic by the ASTRAL-III and SVDquartets analyses (Figs. 7–8). In fact, in our results *P. fallax* and *P. californiarum* are separate from *Pinus monophylla*, suggesting that *P. fallax* and *P. californiarum* are not taxonomic varieties of *P. monophylla* as proposed by Silba (1990). Taxonomic uncertainty between *P. monophylla* and *P. edulis* can be attributed to the existence of trees with both one and two leaves per fascicle (Tausch and West 1987). However, *P. edulis* is not sister to *P. monophylla*. Furthermore, environmental studies indicate that *P. edulis* is more similar to *P. fallax*, and *P. monophylla* is more similar to *P. californiarum*. Besides, *P. fallax* occurs in an area with moderate summer rains, similar to *P. edulis* (Malusa 1992). Our analyses support the separation of *P. californiarum* from *P. monophylla*. Bailey (1987) segregated *P. californiarum* from *P. monophylla* based principally on the length and amount of curl-back of fascicle sheaths, the number of leaf resin canals, and the number of rows of foliar stomata. The ASTRAL-III analysis recovered a clade with the species

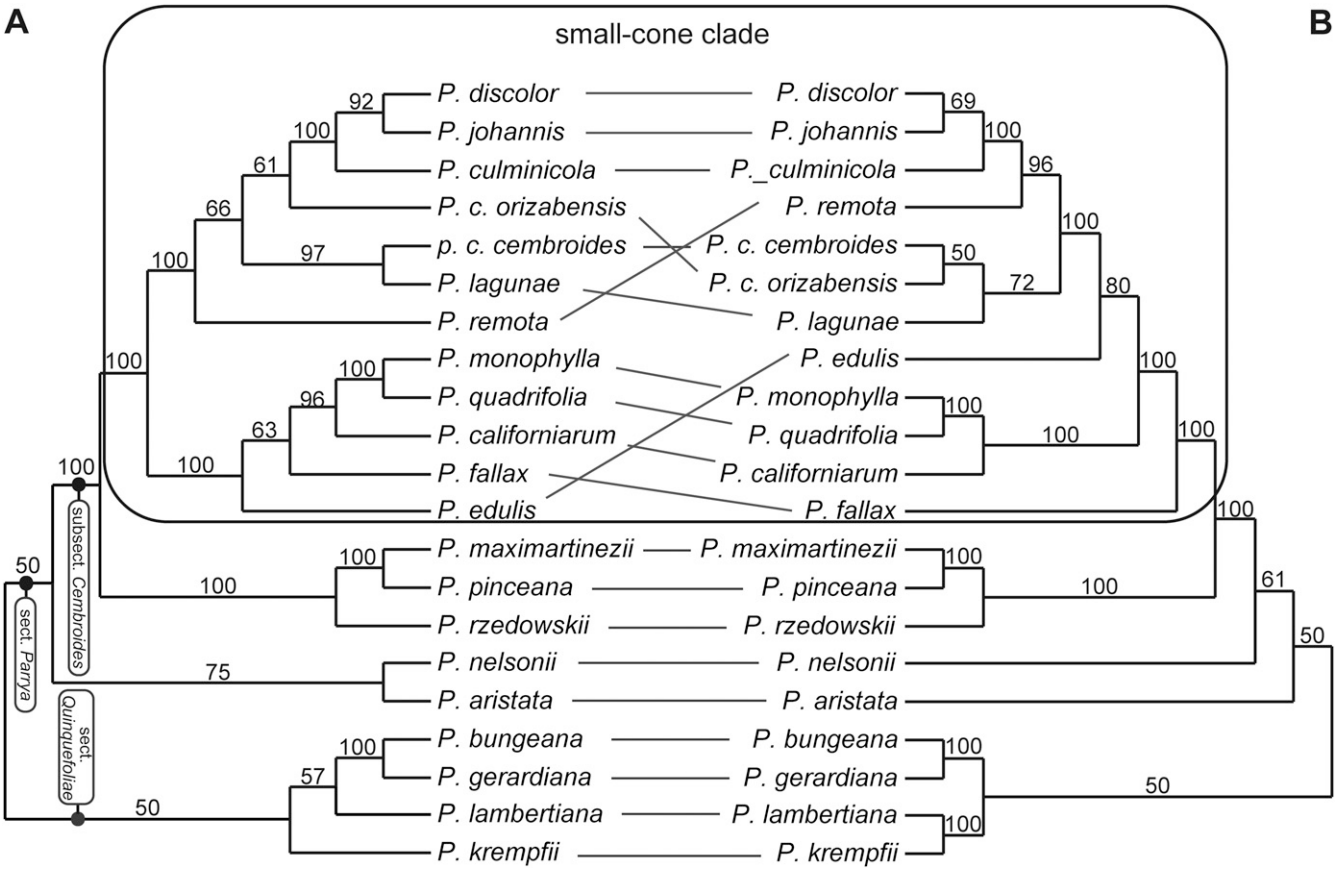

FIG. 9. Comparison between nuclear genes-based and single nucleotide polymorphisms-based trees inferred in SVDquartets. A. Single nucleotide polymorphisms tree based on 26,499 characters with a total weight of incompatible quartets = 5.35%, and total weight of compatible quartets = 94.65%. B. Coalescent-based tree based on 304 nuclear gene trees with a total weight of incompatible quartets = 14.05%, and total weight of compatible quartets = 85.95%. The trees were estimated from 1,000,000 quartets in 51 terminals. Bootstrap values > 50% are shown above the branches.

from the southwestern United States, *P. monophylla*, *P. californiarum*, *P. quadrifolia*, *P. edulis*, and *P. fallax*. The pinyon pines in this region are mainly allopatric or parapatric in distribution (Malusa 1992), but *P. californiarum* and *P. quadrifolia* co-occur in California and Baja California. *Pinus monophylla* occurs in California, extending east (and north) into the Great Basin in the states of Utah, Colorado, Arizona, and Idaho (Critchfield and Little 1966; Farjon 2005; Cole et al. 2008), and populations of *P. fallax* and *P. edulis* co-occur in Arizona and New Mexico (*P. edulis* reaches eastern Nevada and southeastern California). If *P. edulis* and *P. fallax* (adapted to early summer or periodic drought) are paraphyletic to *P. monophylla* (adapted to summer-autumn drought), *P. californiarum*, and *P. quadrifolia* (Figs. 2, 7), this relationship may be explained by vicariance or may have evolved in response to summer drought (Cole et al. 2008). In fact, these taxa are distributed widely in geographical regions with distinct precipitation regimes (Cole et al. 2008). Particularly, *P. monophylla* occurs in regions with different precipitation from *P. fallax* and *P. edulis*, which are characterized by high monsoon precipitation (Cole et al. 2008). The ecological similarities of *P. fallax* to *P. edulis* (Bailey 1987) rather than to *P. monophylla*, and its genetic distinctness from *P. monophylla* (Fig. 8), support our decision to treat *P. fallax* as a separate species from *P. monophylla*.

Hybridization and Introgression—Both natural and artificial hybridization have been well-documented in conifers (Saylor and Smith 1966; Lanner 1974a, 1974b; Delgado et al. 2007; Wachowiak et al. 2011; Zhou et al. 2017). Overlapping phenology and weak reproductive barriers can influence the direction of pollen-mediated gene flow in natural populations (Hamilton et al. 2013). In conifers, plastid DNA is paternally inherited, and higher gene flow for plastid DNA has been attributed to the high migration capacity of wind-dispersed pollen (Rieseberg and Soltis 1991; Petit et al. 2005).

In pines, plastid introgression has been observed at a low or moderate frequency in sympatric or parapatric populations (Dounavi et al. 2001; Delgado et al. 2007; Zhou et al. 2017). In closely-related *Pinus* species, shared genetic variation can be the result of introgression following secondary contact, with genetic differentiation in parapatric populations lower than in allopatric populations (Zhou et al. 2017). Using the minimizing deep coalescence criterion, we explored reticulation in subsection *Cembroides*. No gene flow was detected between subsection *Cembroides* and other closely related lineages (Figs. 4–5). This result supports the observation by Mirov (1967) that species of subsection *Cembroides* do not form hybrids with species from other pine subsections. In fact, *Pinus* subsection *Cembroides* may have diverged from other subsections (particularly *Balfourianae*) relatively early (Axelrod 1986).

The majority of taxa involved in the reticulation events detected here have geographic distributions that are somewhat close to one another (*P. californiarum*, *P. edulis*, *P. fallax*, *P. monophylla*, *P. quadrifolia*, and *P. remota*; only *P. lagunae* and *P. cembroides* subsp. *orizabensis* are geographically isolated). This coincides with other studies where gene flow has been

reported (Edwards-Burke et al. 1997; Delgado et al. 2007; Zhou et al. 2017). Our results coincide with some hypotheses proposed by Lanner (1974a). We detected gene flow in *P. edulis* in only one reticulation scenario (Fig. 4B). This suggests that *P. edulis* is introgressed with *P. monophylla*. Some populations of *P. edulis* and *P. monophylla* occur in sympatry in the eastern Great Basin where trees of *P. edulis* with both single needles and two needles per fascicle have been observed (Lanner 1974a). *Pinus edulis* also occurs in Arizona and New Mexico (Cole et al. 2013), and sympatric populations of *P. edulis* and *P. monophylla* have been reported in in the Mojave Desert in southeastern California (Munz and Keck 1959; Critchfield and Little 1966), western Utah, and Nevada (Farjon and Filer 2013). For this reason, it would not be unusual if *P. edulis* is introgressed with *P. monophylla* where their populations are in contact. The direction of gene flow inferred from *P. monophylla* to *P. edulis* is consistent with the prevailing winds, which are from west to east. Overlapping phenology could facilitate introgression from *P. monophylla* to *P. edulis*, since both disperse pollen in a short period in the spring (Lanner 1970; Farjon and Styles 1997). Furthermore, the distribution range of *P. edulis* may have been more extensive in the past (Cole et al. 2008), resulting in more widespread contact with *P. monophylla*.

The MDC method detected reticulation in taxa for which gene flow had not been suspected. Currently *Pinus fallax* and *P. cembroides* subsp. *cembroides* have widely separated distributions. One possible explanation for reticulation between the two species is that this inference is incorrect (Fig. 4C). As an alternative explanation, we suggest studying species distribution and demographic history to test whether these two taxa came into contact in the past. The potential distribution and demographic history for *P. fallax* suggests that it may have been more widely distributed in southwestern California, southern Nevada, throughout Arizona and extending beyond into Utah, Colorado, and New Mexico (Cole et al. 2008). Another detected reticulation event that was unexpected was between *P. lagunae* and *P. edulis* (Fig. 4C). These species also are widely separated geographically. The genetic diversity shared by these allopatric taxa seems more likely to be influenced by the retention of ancestral polymorphism. However, ancient introgression events may have been interrupted by migration of the populations of *P. edulis* northward in present-day USA during the Holocene (Cole et al. 2008).

We detected reticulation in *P. quadrifolia* and *P. culminicola* (Fig. 4D) although their populations are allopatric and gene flow had not been suspected. This inference may be incorrect, or long-distance pollen dispersal could have resulted in introgression. It would be interesting to study the past distribution and demographic history of *P. culminicola*, *P. cembroides* subsp. *cembroides*, and *P. lagunae* to test whether they were formerly in contact with other species.

The Phylonet analysis also detected reticulation in taxa such as *P. lagunae* and *P. cembroides* subsp. *cembroides* (Fig. 4D) for which gene flow had not been suspected. The origin of these reticulations implies the existence of an extinct taxon. It is also possible that this inference is incorrect or not significant. Copetti et al. (2017) explained the origin of a reticulation with the existence of an extinct or unsampled taxon in cacti, but the authors do not discuss the result.

Specifying *P. quadrifolia* as a hybrid under MDC (Fig. 5B), our results did not indicate that *P. quadrifolia* is introgressed with *P. californiarum* as reported by Lanner (1974b). *Pinus californiarum* was recovered as the sister to *P. quadrifolia* and *P. monophylla*, but no reticulation was detected between *P. californiarum* or *P. monophylla*. We did detect reticulation in *P. quadrifolia* and *P. lagunae* for which gene flow had not been suspected (Fig. 5B). Long-distance pollen dispersal could have resulted in introgression. Likewise, *Pinus lagunae* may be a relictual population left behind from a time when its ancestors had a range that extended northwest into what is today Baja California and southern California.

Another species reported as a possible hybrid is *P. remota*. Our results do not support the proposal of Little. Little (1968) suggested the possibility that in the past its populations had been in contact with those of *P. edulis*, permitting introgressive hybridization. Nonetheless, reticulation with *P. remota* was inferred from *P. fallax* and not from *P. edulis* (Fig. 5A; Little 1968). No contemporary *P. fallax* populations come into contact with *P. remota*, but past secondary contact could have resulted in introgression. Some populations of *P. fallax* (western New Mexico) occur in proximity to *P. remota* (Texas and northeast Mexico). According to the packrat middens record for the Late Quaternary, *P. remota* appear to have expanded into the south of Edwards Plateau to the west or southwest (Van Devender 1990). Likewise, *P. fallax* in the Sonoran Desert may have expanded from California chaparral to Arizona across the Pleistocene-Holocene boundary (Betancourt et al. 1990).

Although some populations of *P. californiarum* and *P. monophylla* are found in limited sympatry in California (San Bernardino Co.), we did not detect gene flow in any direction, but the admixture analysis provided additional information about the relationship between *P. californiarum* and *P. monophylla* (Fig. 3A). Admixture analyses should be interpreted with care since it is difficult to estimate the real number of clusters, especially for species with long generation times like conifers; however, with the increased use of sequencing technologies and massive generation data, our capacity to detect admixture has substantially improved (Pritchard et al. 2000). The admixture analysis we carried out on the pinyon pines subclade from SW US and Baja California provided a picture of possible interbreeding in this group of pines. Distribution of ancestry fractions indicate that *P. californiarum* is introgressed but this scenario was not recovered in Phylonet. The admixture results indicate that *P. quadrifolia* shares little genetic variation with *P. monophylla* (s. s.). In addition, it stands out that *P. quadrifolia* is clustered in a singular population with $K > 2$. This pattern of ancestral structure suggests that *P. quadrifolia* accumulates particular genetic diversity and could be a valid species and not a hybrid. For $K = 5$ (data not shown) *P. monophylla* was not clustered into a separate population. It is also important to mention that according to clustering patterns *P. fallax* seems to share more genetic variation with *P. edulis* than *P. monophylla*. Clustering of *P. fallax* into a new well-defined population at a particular value of K could reflect enough genetic differentiation from *P. edulis* to support its isolation as a species. Although admixture patterns support in most cases our phylogenetic results, increased sampling is required for a more complete perspective of the admixture events for these complex and widely geographically distributed populations. Despite the reduced sample size, we were able to provide a preliminary insight into the admixture events occurring in this group of pines.

In conclusion, using target enrichment to characterize 304 nuclear genes and 26,499 SNPs, we corroborated the monophyly of *Pinus* subsection *Cembroides* in all analyses. The inferred phylogenies also corroborate other relationships

previously recovered with plastid DNA. The results suggest that *P. fallax* and *P. californiarum* could be considered as valid species rather than as infraspecific taxa of *P. monophylla* or *P. edulis*. Also, our admixture results suggest that *P. quadrifolia* accumulates particular genetic diversity and could be a valid species and not a hybrid. The single-needle pinyons were recovered as a non-monophyletic group, with character reconstructions consistent with a single derivation and subsequent loss of the single-needle condition. The ASTRAL-III tree was consistent with the presence of ILS (very high) in the group of pinyon pines with small cones based on the short length in coalescent units of internal branches. Respecting reticulation events, we identified *P. remota* as having genes introgressed from *P. fallax*, and *P. quadrifolia* having genes introgressed from *P. lagunae*. Some hybridization scenarios were unexpected and not reported in the literature. Finally, further study is needed to determine the relative roles of ILS and introgression in explaining shared genetic diversity in *Pinus* subsection *Cembroides*.

Acknowledgments

The authors thank Alejandra Vázquez-Lobo and Patricia Ornelas for valuable suggestions. We also thank Aaron Liston for his collaboration in the study design, and Xitlali Aguirre-Dugua, Mario Montes, and Angélica Castolo for their participation in fieldwork, and José Delgadillo Rodríguez for logistical support. We also thank Patricia Rosas Escobar for laboratory assistance, and Alison Devault and Jake Enk at Arbor Biosciences for valuable technical assistance. We also thank Dra. Lidia I. Cabrera and the LANABIO of the Instituto de Biología, UNAM. Funding was provided by PAPIIT-DGAPA, UNAM Grant IN209816 and CONACyT Grant 221694. This work is part of the Ph.D. thesis of J.R. Montes in the Posgrado de Ciencias Biológicas, UNAM.

Author Contributions

JRM performed field- and labwork, assembled the DNA sequences, performed the phylogenetic analyses, and was the primary author for the manuscript. JRM, AML, AW, DP, and DSG designed the study. PP provided SNP data and performed admixture analyses. AML and DSG participated in fieldwork and analyses. All authors reviewed and edited the manuscript.

Literature Cited

Aberer, A. J., D. Krompass, and A. Stamatakis. 2013. Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. *Systematic Biology* 62: 162–166.

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

Axelrod, D. I. 1986. Cenozoic history of some western American pines. *Annals of the Missouri Botanical Garden* 73: 565–641.

Bailey, D. K. 1987. A study of *Pinus* subsection *Cembroides* I: The single-needle pinyons of the Californias and the Great Basin. *Notes from the Royal Botanic Garden Edinburgh* 44: 275–310.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Leslin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Betancourt, J. L., T. R. Van Devender, and P. S. Martin. 1990. *Packrat Middens: The Last 40,000 Years of Biotic Change*. Tucson, Arizona: University of Arizona Press.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Chifman, J. and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

Critchfield, W. B. and E. L. Little. 1966. *Geographic Distribution of the Pines of the World*. Washington, D.C.: Department of Agriculture Miscellaneous Publication 991.

Cole, K. L., J. F. Fisher, S. T. Arundel, J. Cannella, and S. Swift. 2008. Geographical and climatic limits of needle types of one-and two-needled pinyon pines. *Journal of Biogeography* 35: 257–269.

Cole, K. L., J. F. Fisher, K. Ironside, J. I. Mead, and P. Koehler. 2013. The biogeographic histories of *Pinus edulis* and *Pinus monophylla* over the last 50,000 years. *Quaternary International* 310: 96e110.

Copetti, D., A. Búrquez, E. Bustamante, J. L. M. Charboneau, K. L. Childs, L. E. Eguiarte, S. Lee, T. L. Liu, M. M. McMahon, N. K. Whiteman, R. A. Wing, M. F. Wojciechowski, and M. J. Sanderson. 2017. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences USA* 114: 12003–12008.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

Debreczy, Z. and I. Rácz. 2011. *Conifers Around the World*. Budapest: DendroPress Ltd.

DeGiorgio, M., J. Syring, A. J. Eckert, A. Liston, R. Cronn, D. B. Neale, and N. A. Rosenberg. 2014. An empirical evaluation of two-stage species tree inference strategies using a multilocus dataset from North American pines. *BMC Evolutionary Biology* 14: 67.

Delgado, P., R. Salas-Lizana, A. Vázquez-Lobo, A. Wegier, M. Anzidei, E. R. Alvarez-Buylla, G. G. Vendramin, and D. Piñero. 2007. Introgressive hybridization in *Pinus montezumae* Lamb. and *Pinus pseudostrobus* Lindl. (Pinaceae): Morphological and molecular (cpSSR) evidence. *International Journal of Plant Sciences* 168: 861–875.

Dounavi, K. D., N. Koutsias, and K. P. Panetsos. 2001. Natural interspecific hybridization between *Pinus brutia* (Ten.) and *Pinus halepensis* (Mill.), verified by using the logistic regression modeling on morphological characters. *Forest Genetics* 8: 151–158.

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

Eckenwalder, J. E. 2009. *Conifers of the World: The Complete Reference*. Portland, Oregon: Timber Press.

Edwards-Burke, M. A., J. L. Hamrick, and R. A. Price. 1997. Frequency and direction of hybridization in sympatric populations of *Pinus taeda* and *P. echinata* (Pinaceae). *American Journal of Botany* 84: 879–886.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048.

Farjon, A. 2005. *Pines: Drawings and Description of the Genus* Pinus. Leiden, Netherlands: Brill.

Farjon, A. and B. Filer. 2013. *An Atlas of the World's Conifers: An Analysis of their Distribution, Biogeography, Diversity and Conservation Status*. Leiden, Netherlands: Brill.

Farjon, A. and B. Styles. 1997. *Flora Neotropica Monograph*: Pinus *(Pinaceae)*. New York: The New York Botanical Garden.

Flores-Rentería, L., A. Wegier, D. Ortega Del Vecchyo, A. Ortíz-Medrano, D. Piñero, A. V. Whipple, F. Molina-Freaner, and C. A. Domínguez. 2013. Genetic, morphological, geographical and ecological approaches reveal phylogenetic relationships in complex groups, an example of recently diverged pinyon pine species (subsection *Cembroides*). *Molecular Phylogenetics and Evolution* 69: 940–949.

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. Phylogenetic of *Pinus* subsections *Cembroides* and *Nelsoniae* inferred from cpDNA sequences. *Systematic Botany* 28: 657–673.

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

Gernandt, D. S., O. Zerón, and I. Goyenechea. 2007. Inferencia filogenética mediante secuencias de DNA: Un ejemplo con los pinos piñoneros. Pp. 55–65 In *La Sistemática Base del Conocimiento de la Biodiversidad*, vol. 2, eds. A. Contreras-Ramos, C. Cuevas Cardona, I. Goyenechea, and U. Iturbide. Hidalgo: Universidad Autónoma del Estado de Hidalgo.

Gernandt, D. S., S. Magallón, G. Geada López, O. Zerón Flores, A. Willyard, and A. Liston. 2008. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. *International Journal of Plant Sciences* 169: 1086–1099.

Gernandt, D. S., X. Aguirre-Dugua, A. Vázquez-Lobo, A. Willyard, A. Moreno Letelier, J. A. Pérez de la Rosa, D. Piñero, and A. Liston. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. *American Journal of Botany* 105: 711–725.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, and C. Nusbaum. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.

Hamilton, J. A., C. Lexer, and S. N. Aitken. 2013. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*). *Molecular Ecology* 22: 827–841.

Huson, D. H. and C. Scornavacca. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* 61: 1061–1067.

IUCN. 2017. IUCN red list of threatened species version 2017–3. Gland, Switzerland and Cambridge, UK: IUCN.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Keng, H. and E. L. Little. 1961. Needle characteristics of hybrid pines. *Silvae Genetica* 10: 131–146.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179–1191.

Lanner, R. M. 1970. Origin of the summer shoot of pinyon pines. *Canadian Journal of Botany* 48: 1759–1765.

Lanner, R. M. 1974a. Natural hybridization between *Pinus edulis* and *Pinus monophylla* in the American Southwest. *Silvae Genetica* 23: 108–116.

Lanner, R. M. 1974b. A new pine from Baja California and the hybrid origin of *Pinus quadrifolia*. *The Southwestern Naturalist* 19: 75–95.

Lanner, R. M. 1981. *The Piñon Pine: A Natural and Cultural History*. Reno, Nevada: University of Nevada Press.

Lanner, R. M. and A. M. Phillips. 1992. Natural hybridization and introgression of pinyon pines in northwestern Arizona. *International Journal of Plant Sciences* 153: 250–257.

Leaché, A. D. and J. R. Oaks. 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology Evolution and Systematics* 48: 69–84.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Little, E. L. 1966. A new pinyon variety from Texas. *Wrightia* 3: 181–187.

Little, E. L. 1968. Two new pinyon varieties from Arizona. *Phytologia* 17: 329–342.

Little, E. L. and W. B. Critchfield. 1969. *Subdivisions of the Genus* Pinus *(Pines)*. Washington D.C.: US Forest Service Miscellaneous Publication 1144.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Maddison, W. P. and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.

Maddison, W. P. and D. R. Maddison. 2018. Mesquite: A modular system for evolutionary analysis, version 3.6. Technical report. Software available from <http://mesquiteproject.org>.

Madrigal, S. X. and D. M. Caballero. 1969. Una nueva especie mexicana de *Pinus*. *Boletín Técnico del Instituto Nacional de Investigaciones Forestales* 26: 1–11.

Malusa, J. 1992. Phylogeny and biogeography of the pinyon pines (*Pinus* subsect. *Cembroides*). *Systematic Botany* 17: 42–66.

Mirarab, S. and T. Warnow. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: i44–i52.

Mirov, N. T. 1967. *The Genus* Pinus. New York: Ronald Press.

Montes, J. R., P. Peláez, A. Willyard, A. Moreno-Letelier, D. Piñero, and D. S. Gernandt. 2019. Data from: Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. Dryad Digital Repository. https://doi.org/10.5061/dryad.1f8p5d6.

Morales-Briones, D. F., K. Romoleroux, F. Kolár, and D. C. Tank. 2018. Phylogeny and evolution of the neotropical radiation of *Lachemilla* (Rosaceae): Uncovering a history of reticulate evolution and implications for infrageneric classification. *Systematic Botany* 43: 17–35.

Munz, P. A. and D. D. Keck. 1959. *A California Flora*. Berkeley, California: University of California Press.

Neale, D. B., J. L. Wegrzyn, K. A. Stevens, A. V. Zimin, D. Puiu, M. W. Crepeau, C. Cardeno, M. Koriabine, A. E. Holtz-Morris, J. D. Liechty, P. J. Martínez-García, H. A. Vasquez-Gross, B. Y. Lin, J. J. Zieve, W. M. Dougherty, S. Fuentes-Soriano, L. S. Wu, D. Gilbert, G. Marçais, M. Roberts, C. Holt, M. Yandell, J. M. Davis, K. E. Smith, J. F. Dean, W. W. Lorenz, R. W. Whetten, R. Sederoff, N. Wheeler, P. E. McGuire, D. Main, C. A. Loopstra, K. Mockaitis, P. J. de Jong, J. A. Yorke, S. L. Salzberg, and C. H. Langley. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59.

Neves, L. G., J. M. Davis, W. B. Barbazuk, and M. Kirst. 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* 75: 146–156.

Ortiz-Medrano, A., D. P. Scantlebury, A. Vázquez-Lobo, A. Mastretta-Yanes, and D. Piñero. 2016. Morphological and niche divergence of pinyon pines. *Ecology and Evolution* 6: 2886–2896.

Parks, M., R. Cronn, and A. Liston. 2012. Separating the wheat from the chaff: Mitigating the effects of noise in plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evolutionary Biology* 12: 100.

Passini, M. F. 1987. The endemic pinyon of lower California: *Pinus lagunae*. *Phytologia* 63: 337–338.

Pérez de la Rosa, J. A., S. A. Harris, and A. Farjon. 1995. Noncoding chloroplast DNA variation in Mexican pines. *Theoretical and Applied Genetics* 91: 1101–1106.

Petit, R. J. and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics* 37: 187–214.

Petit, R. J., J. Duminil, S. Fineschi, A. Hampe, D. Salvini, and G. G. Vendramin. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14: 689–701.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Price, R. A., A. Liston, and S. H. Strauss. 1998. Phylogeny and systematics of *Pinus*. Pp. 49–68 in *Ecology and Biogeography of Pinus*, ed. D. M. Richardson. Cambridge, UK: Cambridge University Press.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559–575.

Rambaut, A. 2012. FigTree. Tree figure drawing tool version 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/.690.

Rieseberg, L. H. and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.

Saladin, B., A. B. Leslie, R. O. Wüest, G. Litsios, E. Conti, N. Salamin, and N. E. Zimmermann. 2017. Fossils matter: Improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology* 17: 95.

Saylor, L. C. and B. W. Smith. 1966. Meiotic irregularity in species and interspecific hybrids of *Pinus*. *American Journal of Botany* 53: 453–468.

Sayyari, E. and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668.

SEMARNAT. 2010. Norma Oficial Mexicana NOM-059-ECOL-2001. Protección ambiental. Especies nativas de México de flora y fauna silvestres. Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio. Lista de especies en riesgo. Diario Oficial de la Federación, 30 de diciembre de 2010.

Silba, J. 1990. A supplement to the international census of the coniferae, II. *Phytologia* 68: 7–78.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Swofford, D. L. 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods), v. 4.0b10. Sunderland, Massachusetts: Sinauer Associates.

Syring, J., A. Willyard, R. Cronn, and A. Liston. 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *American Journal of Botany* 92: 2086–2100.

Tausch, R. J. and N. E. West. 1987. Morphological variation/precipitation relationships of Great Basin single-needled pinyon. Pp. 86–91 In *Proceedings of the Pinyon-Juniper Conference*, ed. R. L. Everett. Reno, Nevada: General Technical Report INT-215. U.S.D.A., Forest Service.

Than, C. and L. Nakhleh. 2009. Species tree inference by minimizing deep coalescences. *PLoS Computational Biology* 5: e1000501.

Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9: 322.

Van Devender, T. R. 1990. Late Quaternary vegetation and climate of the Chihuahuan Desert, United States and Mexico. Pp. 104–133 In *Packrat Middens: The Last 40,000 Years of Biotic Change*, eds. J. L Betancourt, T. R. Van Devender, and P. S. Martin. Tucson, Arizona: The University of Arizona Press.

Wachowiak, W., A. E. Palmé, and O. Savolainen. 2011. Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). *Molecular Ecology* 20: 1729–1743.

Wang, J., R. J. Abbott, Y. L. Peng, F. K. Du, and J.-Q. Liu. 2011. Species delimitation and biogeography of two fir species (*Abies*) in central China: Cytoplasmic DNA variation. *Heredity* 107: 362–370.

Wegrzyn, J. L., J. D. Liechty, K. A. Stevens, L.-S. Wu, C. A. Loopstra, H. A. Vasquez-Gross, W. M. Dougherty, B. Y. Lin, J. J. Zieve, P. J. Martínez-García, C. Holt, M. Yandell, A. V. Zimin, J. A. Yorke, M. W. Crepeau, D. Puiu, S. L. Salzberg, P. J. de Jong, K. Mockaitis, D. Main, C. H. Langley, and D. B. Neale. 2014. Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196: 891–909.

Weitemier, K., S. C. K. Straub, R. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. HybSeq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 9: 1400042.

Whang, S. S., J.-H. Pak, R. S. Hill, and K. Kim. 2001. Cuticle micromorphology of the leaves of *Pinus* (Pinaceae) from Mexico and Central America. *Botanical Journal of the Linnean Society* 135: 349–373.

Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* 24: 90–101.

Zavarin, E. and K. Snajberk. 1985. Monoterpenoid and morphological differentiation within *Pinus cembroides*. *Biochemical Systematics and Ecology* 13: 89–104.

Zhang, C., E. Sayyari, and S. Mirarab. 2017. ASTRAL-III: Increased scalability and impacts of contracting low support branches. Pp. 53–75 in *Comparative Genomics, 15th International Workshop, RECOMB CG*, eds. J. Meidanis and L. Nakhleh. Barcelona, Spain: Springer Press. doi: 10.1007/978-3-319-67979-2_4.

Zhou, Y., L. Duvaux, G. Ren, L. Zhang, O. Savolainen, and J. Liu. 2017. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* 118: 211–220.

Appendix 1. Collection information for individuals included in the study. Voucher information for this study, presented in the following order: Taxon; voucher specimen: collector and number, (herbarium acronym), locality.

Ingroup: ***Pinus californiarum*** D.K.Bailey; *D.S. Gernandt 403*, *1561*, (MEXU), Mexico, Baja California. ***Pinus cembroides*** **subsp.** ***cembroides*** Zucc.; *D.S. Gernandt 444*, *593*, *1042*, (MEXU), Mexico. ***Pinus cembroides*** **subsp.** ***orizabensis*** D.K.Bailey; *D.S. Gernandt 7399*, (MEXU), Mexico, Puebla. ***Pinus culminicola*** Andresen & Beaman; *D.S. Gernandt 1135*, *1137*, *01S6*, *D.O. Burge 1212*, (MEXU), Mexico. ***Pinus discolor*** D.K.Bailey & Hawksw.; *D.S. Gernandt 1067*, (MEXU), Mexico, Sonora, *D.S. Gernandt 785*, (MEXU), *F. Hammond 02S2*, (OSC), United States, Arizona. ***Pinus edulis*** Engelm.; *D.S. Gernandt 485*, *1020*, *1028*, (MEXU), United States, Utah. ***Pinus fallax*** (Little) Businský; *D.S. Gernandt 492*, *494*, (MEXU), United States, Utah. ***Pinus johannis*** M.F.Robert; *D.S. Gernandt 501*, *7999*, *8199*, (MEXU), Mexico. ***Pinus lagunae*** (Robert-Passini) D.K.Bailey; *A.M. González 9279*, *6399*, (MEXU), Mexico, Baja California Sur. ***Pinus maximartinezii*** Rzed. *D.S. Gernandt 1010*, *7799*, (MEXU), Mexico, Zacatecas. ***Pinus monophylla*** Torr. & Frém.; *D.S. Gernandt 478*, *480*, *1214*, *1509*, *1512*, *1513*, *A. Liston 1298*, *R. Halse 6668*, (MEXU), United States. ***Pinus pinceana*** Gordon; *D.S. Gernandt 1163*, *8999*, (MEXU), Mexico. ***Pinus quadrifolia*** Parl. ex Sudw.; *D.S. Gernandt 961*, *1099*, *1499*, *1560*, *1599*, (MEXU), *D. Gernandt, A. Liston & Ann Willyard 035*, (OSC), Mexico, Baja California. ***Pinus remota*** (Little) D.K.Bailey & Hawksw.; *D.S. Gernandt 801*, *1301 19498*, *22498*, *23298*, (MEXU), Mexico. ***Pinus rzedowskii*** Madrigal & M. Caball. *D.S. Gernandt 635*, *636*, *637*, (MEXU), *R. Businský 47131*, (OSC), Mexico, Michoacán.

Outgroup: ***Pinus aristata*** Engelm.; *K. Ferrell 30*, *37*, (OSC), United States. ***Pinus bungeana*** Zucc. ex Endl.; *J.E.R 03S3A*, (OSC), China. ***Pinus gerardiana*** Wall. ex D.Don; *R. Businský 41105*, (OSC), Pakistan, Gilgit-Baltistan. ***Pinus krempfii*** Lecomte; *P. Thomas 242*, (E), Vietnam, Lam Dong. ***Pinus lambertiana*** Douglas; *D.S. Gernandt 1195*, (MEXU), United States, California. ***Pinus longaeva*** D.K.Bailey; *D.S. Gernandt 1027*, (MEXU), United States, Utah. ***Pinus nelsonii*** Shaw. *D.S. Gernandt 1096*, *10198*, *31798*, (MEXU), Mexico.

Appendix 2. Sequence statistics for the 60 *Pinus* samples assembled in HybPiper.For each sample, species name is followed by sample ID; sequencing run; yield (Mb); reads; reads mapped; genes mapped; and percent recovered gene.

Subsection ***Cembroides***. ***Pinus californiarum***: DSG1509; 2; 894; 7,601,929; 3,800,860; 969; 97.3. ***P. californiarum***: DSG1561; 1; 932; 8,243,891; 4,121,790; 969; 97.3. ***P. californiarum:*** DSG1512; 1; 999; 8,764,857; 4,382,336; 969; 97.3. ***P. californiarum***: DSG1513; 2; 1108; 9,405,258; 4,702,662; 969; 97.3. ***P. californiarum***: DSG403; 2; 1150; 9,862,288; 4,931,170; 968; 97.2. ***P. californiarum***: AL1298; 2; 1283; 11,713,774; 5,856,883; 969; 97.3. ***P. cembroides*** **subsp.** ***cembroides***: DSG593; 2; 1003; 8,658,650; 4,329,349; 969; 97.3. ***P. cembroides*** **subsp.** ***cembroides***: DSG444; 2; 1203; 10,785,141; 5,392,515; 969; 97.3. ***P. cembroides*** **subsp.** ***cembroides***: DSG1042; 2; 1275; 11,310,686; 5,655,408; 969; 97.3. ***P. cembroides*** **subsp.** ***orizabensis***: DSG7399; 2; 923; 8,474,182; 4,237,030; 968; 97.2. ***P. culminicola***: DOB1212; 1; 733; 6,396,584; 3,198,382; 968; 97.2. ***P. culminicola***: DSG1135; 2; 1068; 9,108,291; 4,553,996; 969; 97.3. ***P. culminicola***: 01S6; 2; 1120; 9,549,507; 4,774,712; 968; 97.2. ***P. culminicola***: DSG1137; 2; 1223; 10,962,417; 5,481,235; 969; 97.3. ***P. discolor***: 02s2; 2; 1036; 9,070,577; 4,535,221; 968; 97.2. ***P. discolor***: DSG1067; 2; 1191; 10,417,878; 5,208,859; 968; 97.2. ***P. discolor***: DSG785; 2; 1180; 10,617,625; 5,308,819; 969; 97.3. ***P. edulis***: DSG485; 2; 989; 9,054,447; 4,527,033; 969; 97.3. ***P. edulis***: DSG1028; 1; 994; 9,061,862; 4,530,904; 969; 97.3. ***P. edulis***: DSG1020; 2; 1126; 9,546,477; 4,773,241; 969; 97.3. ***P. fallax***: DSG494; 2; 484; 4,261,767; 2,130,955; 968; 97.2. ***P. fallax***: DSG492; 2; 904; 7,384,002; 3,692,088; 969; 97.3. ***P. johannis***: DSG501; 2; 907; 7,844,117; 3,922,046; 969; 97.3. ***P. johannis***: DSG08199; 2; 1313; 11,297,858; 5,648,973; 969; 97.3. ***P. lagunae***: AGM9263; 2; 1069; 9,553,367; 4,776,613; 969; 97.3. ***P. lagunae***: AGM9279; 2; 1233; 10,814,782; 5,407,465; 969; 97.3. ***P. maximartinezii***: DSG07799; 2; 937; 8,321,810; 4,160,907; 969; 97.3. ***P. maximartinezii***: DSG1010; 2; 704; 10,502,750; 5,251,251; 969; 97.3. ***P. maximartinezii***: DSG6499; 2; 1174; 10,502,750; 5,251,251; 969; 97.3. ***P. monophylla***: RH6668; 1; 618; 5,605,098; 2,802,565; 969; 97.3. ***P. monophylla***: DSG1214; 1; 942; 8,427,317; 4,213,596; 969; 97.3. ***P. monophylla***: DSG478; 2; 999; 8,718,219; 4,359,102; 969; 97.3. ***P. monophylla***: DSG480; 2; 1241; 11,108,138; 5,554,149; 969; 97.3. ***P. pinceana***: DSG1163; 2; 909; 8,001,565; 4,000,671; 969; 97.3. ***P. pinceana***: DSG8999; 2; 1320; 11,789,410; 5,894,652; 968; 97.2. ***P. pinceana***: DSG7999; 2; 1576; 14,115,122; 7,057,552; 969; 97.3. ***P. quadrifolia:*** DSG1599; 2; 862; 7,306,710; 3,653,325; 968; 97.2. ***P. quadrifolia***: DSG1560; 1; 890; 8,193,597; 4,096,686; 968; 97.2. ***P. quadrifolia***: DSG01499; 2; 1096; 9,892,499; 4,946,163; 969; 97.3. ***P. quadrifolia***: DSG961; 2; 1195; 10,111,721; 5,055,873; 969; 97.3. ***P. quadrifolia***: DSG01099; 2; 1152; 10,445,399; 5,222,697; 968; 97.2. ***P. quadrifolia***: quad035; 2; 1336; 10,879,895; 5,440,165; 968; 97.2. ***P. remota***: DSG1301; 2; 981; 8,112,584; 4,056,345; 969; 97.3. ***P. remota***: DSG19498; 2; 972; 8,686,551; 4,343,208; 969; 97.3. ***P. remota***: DSG22498; 2; 1010; 8,887,062; 4,443,557; 969; 97.3. ***P. remota***: DSG801; 2; 1123; 9,854,338; 4,926,998; 969; 97.3. ***P. remota***: DSG23298; 2; 1117; 10,125,019; 5,062,440; 969; 97.3. ***P. rzedowskii***: DSG637; 2; 1199; 10,663,690; 5,331,879; 968; 97.2. ***P. rzedowskii***: RB47131; 2; 1219; 10,740,126; 5,370,129; 968; 97.2. ***P. rzedowskii***: DSG635; 2; 1310; 11,448,532; 5,724,214; 968; 97.2. ***P. rzedowskii***: DSG636; 2; 1287; 11,475,267; 5,737,519; 969; 97.3. **Subsection** ***Balfourianae***. ***P. aristata***: KF37; 2; 680; 6,039,708; 3,020,016; 969; 97.3. ***P. aristata***: KF30; 2; 813; 7,027,768; 3,513,851; 969; 97.3. ***P. longaeva***: DSG1027; 1; 681; 6,029,708; 3,010,016; 969; 97.3. **Subsection** ***Gerardianae***. ***P. bungeana***: 03s3A; 2; 842; 7,036,428; 3,518,309; 969; 97.3. ***P. gerardiana***: RB41105; 2; 464; 4,111,029; 2,055,501; 968; 97.2. **Subsection** ***Krempfianae***. ***P. krempfii***: PT242; 2; 864; 7,248,613; 3,624,217; 969; 97.3. **Subsection** ***Nelsoniae***. ***P. nelsonii***: DSG1096; 2; 954; 8,461,342; 4,230,564; 969; 97.3. ***P. nelsonii***: DSG10198; 2; 1141; 10,225,576; 5,112,715; 968; 97.2. ***P. nelsonii***: DSG31798; 2; 1261; 11,150,773; 5,575,312; 968; 97.2. **Subsection** ***Strobus***. ***P. lambertiana***: DSG1195; 1; 1122; 10,161,689; 5,080,591; 969; 97.3.