Botanical Sciences 103 (Special Issue): S95-S112. 2025
DOI: 10.17129/botsci.3721

Received: May 5, 2025, Accepted: August 10, 2025
On line first: September 1, 2025

Systematics / Sistemática

Multilocus phylogenetics of *Pinus* subsection *Ponderosae* using the Hyb-Seq method

David S. Gernandt[1]*, Ann Willyard[2], Alejandra Vázquez-Lobo[3], Alejandra Moreno-Letelier[4], Patricia Delgado[5], Dante S. Figueroa[6], and M. Socorro González-Elizondo[7]

[1] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico.
[2] Hendrix College, Biology Department, Conway, Arkansas, USA.
[3] Centro de Investigación en Biodiversidad y Conservación, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico.
[4] Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico.
[5] Facultad de Agrobiología "Presidente Juárez", Universidad Michoacana de San Nicolás de Hidalgo, Uruapan, Michoacán, Mexico.
[6] Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico.
[7] CIIDIR Unidad Durango, Instituto Politécnico Nacional, Durango, Mexico.
*Corresponding author: dgernandt@ib.unam.mx

Abstract

Background: Hyb-Seq combines solution hybridization based target enrichment and genome skimming for the study of plant phylogenetics and evolution, including the detection of incomplete lineage sorting and hybridization.
Questions: What are the relationships among the principal lineages of *Pinus* subsection *Ponderosae* in Mexico?
Studied species: *Pinus* subsection *Ponderosae*.
Study sites and dates: Mexico, Guatemala, and the United States, 2015-2025.
Methods: We assembled sequences of low copy nuclear loci and plastomes to compare phylogenetic results between the two genomic compartments. Parsimony and maximum likelihood were used to infer trees for both datasets, and the coalescent method ASTRAL was used for the nuclear sequences.
Results: Strict alignment selection criteria, including exclusion of putative paralogs and manual editing of alignments, resulted in less homoplasy and slightly better correspondence to morphology compared to a dataset subject to relaxed filtering. ASTRAL and concatenated analyses recovered the same main clades for the nuclear loci. There were some striking discordances between the nuclear and plastid trees in the placement of specific individuals.
Conclusions: The nuclear and plastid trees corroborated previous studies in recovering subsection *Ponderosae* as monophyletic with high branch support. Branch lengths separating Mexican taxa were relatively short, and support values in the nuclear coalescent trees were low, consistent with rapid speciation. Instances of cytonuclear discordance in the Hyb-Seq data were consistent with plastid capture, but more work is needed to disentangle homoplasy, incomplete lineage sorting, and reticulation.
Keywords: coalescence, introgression, pines, low copy nuclear genes, *Pinus* section *Trifoliae*, plastomes.

Resumen

Antecedentes: Hyb-Seq combina enriquecimiento híbrido en solución y secuenciación poco profunda para el estudio de la filogenética y evolución de plantas, facilitando la identificación de repartición incompleta de linajes e hibridización.
Preguntas: ¿Cuáles son las relaciones entre los principales linajes de *Pinus* subsección *Ponderosae* en México?
Especies de estudio: *Pinus* subsección *Ponderosae*.
Sitio y años del estudio: México, Guatemala y Estados Unidos, 2015-2025.
Métodos: Ensamblamos secuencias de loci nucleares y plastomas para comparar los resultados filogenéticos entre los dos compartimentos genómicos. Se utilizó la máxima verosimilitud para inferir árboles en ambos conjuntos de datos, y el método coalescente ASTRAL para los datos nucleares.
Resultados: Criterios de selección más estrictos, incluyendo la exclusión de parálogos, redujeron la homoplasia, y mejoran ligeramente la correspondencia entre morfología en comparación con un conjunto de datos sometido a un filtrado relajado. ASTRAL y máxima verosimilitud recuperaron los mismos clados principales para los loci nucleares. Se observaron discrepancias notables entre los árboles nucleares y de plástidos en la ubicación de individuos.
Conclusiones: Los árboles nucleares y de plástidos corroboraron estudios previos que recuperaron la subsección *Ponderosae* como monofilética con alto apoyo de ramas. La longitud de las ramas que separan los taxones mexicanos fue relativamente corta, y los valores de apoyo fueron bajos en los análisis coalescentes, consistente con la especiación rápida. Los casos de discordancia citonuclear en los datos Hyb-Seq soportaron la captura de plastos, pero se necesita más investigación para desentrañar la homoplasia, la repartición incompleta de linajes y la reticulación.
Palabras claves: coalescencia, introgresión, pinos, genes nucleares de bajo número de copia, *Pinus* sección *Trifoliae*, plastomas.

Massively parallel sequencing has greatly expanded the breadth of questions that can be addressed in life sciences. More specifically, the use of short read sequencing in combination with target enrichment using solution hybridization with biotinylated RNA probes has allowed for the generation of datasets of hundreds to thousands of loci (Gnirke *et al.* 2009). A variant adapted for plants, the Hyb-Seq method, combines genome skimming to recover high copy number DNA (*e.g.*, organellar and nuclear ribosomal DNA) and target enrichment of low copy nuclear loci (Weitemier *et al.* 2014). It is especially useful for plants with genome sizes that are too large for whole genome sequencing with current methods.

Pinus subsection *Ponderosae* is an ecologically and economically important lineage of pines distributed from western Canada to Nicaragua (Critchfield & Little 1966, Farjon & Styles 1997, Debreczy & Rácz 2011). It is part of the North American hard pines, *Pinus* section *Trifoliae*, which also includes *Pinus* subsects. *Contortae*, *Australes*, *Attenuatae* and *Sabinianae* (Krupkin *et al.* 1996, Gernandt *et al.* 2005, Hernández-León *et al.* 2013, Jin *et al.* 2021, Willyard *et al.* 2021a). Species level relationships and delimitation in subsect. *Ponderosae* remain uncertain, despite previous phylogenetic studies using plastid DNA sequences (*e.g.*, Gernandt *et al.* 2009, Hernández-León *et al.* 2013), low copy nuclear genes and plastomes generated with the Hyb-Seq method (Willyard *et al.* 2021a), and transcriptomes (Jin *et al.* 2021). Here we follow Willyard *et al*. (2021a) in separating the four species of California big cone pines into subsection *Sabinianae*. *Pinus* subsect. *Ponderosae* may represent an extreme case of genealogical discordance across loci and of alleles of a single locus caused by both incomplete lineage sorting and interspecific gene flow (Matos & Schaal 2000, Delgado *et al.* 2007, Willyard *et al.* 2021b). Jin *et al.* (2021) estimated that its Miocene crown clade age is more recent than other Mexican pine lineages such as the Oocarpae, Cembroides, and Strobus clades.

Early low copy nuclear gene studies of pines found evidence of genealogical discordance across loci (Syring *et al.* 2005) and within alleles of the same locus (Syring *et al.* 2007) for a subset of species from *Pinus* subgenus *Strobus* (the soft pines). This pattern observed in low copy nuclear genes was hypothesized to be primarily a result of incomplete lineage sorting, with interspecific gene flow as a secondary cause. A subsequent study from a subset of the group, *Pinus* subsection *Strobus* (the white pines) based on 121 loci in multiple individuals per species documented topological differences in results from concatenation and several coalescent methods available at the time (DeGiorgio *et al.* 2014). Studies of hundreds of low copy nuclear loci in North American hard pines (*Pinus* subgenus *Pinus* sect. *Trifoliae*; Gernandt *et al.* 2018, Willyard *et al.* 2009, 2021a) and pinyon pines (*Pinus* subgenus *Strobus*, section *Parrya*; Montes *et al.* 2019) have further documented the prevalence of genealogical nonmonophyly and demonstrated multiple cases of discordance with plastid DNA based trees. These studies presented limited evidence that the main cause of genealogical discord is incomplete lineage sorting; there is also evidence supporting at least some interspecific gene flow, which has been reported widely in *Pinus* (*e.g.*, Ma *et al.* 2006). The genus is also characterized by exceptionally high levels of genetic diversity (Ledig 1998) and contrasting uniparental inheritance of plastid (paternal) and mitochondrial (maternal) DNA (Mogensen 1996).

Here we present results from an expanded Hyb-Seq dataset for subsect. *Ponderosae* that focus on taxa distributed primarily in Mexico. We included previously unsampled populations of *P. stormiae* (Martínez) Frankis (treated by Willyard *et al*. (2021a) as *P. arizonica* var. *stormiae* Martínez), which was described from Mexico, and representatives of two taxa, *P. martinezii* Larsen and *P. rudis* Endl., placed in synonymy with *P. durangensis* Martínez and *P. hartwegii* Lindl., respectively, by Farjon & Styles (1997). We also included representatives of the main lineages of the North American hard pines (*Pinus* sect. *Trifoliae*) as outgroups. The objectives of this study are to use improved plastid and nuclear DNA sequence datasets for the group to identify its principal lineages, examine species relationships, compare results between the plastid and nuclear genomic compartments, and use the global results to evaluate species concepts based principally on morphology. We also address discrepancies in phylogenetic relationships inferred with a target enrichment dataset of Willyard *et al.* (2021a) and a transcriptome dataset of Jin *et al.* (2021).

Materials and methods

Sampling, enrichment, and sequencing. We included 87 samples, of which 73 belonged to *Pinus* subsect. *Ponderosae*, 11 to subsect. *Sabinianae* (California big-cones pines), one each from subsects. *Australes*, and *Contortae* (these 86 hard pine taxa belong to *Pinus* sect. *Trifoliae*) and one representative of *Pinus* sect. *Pinus* subsect. *Pinus* (Table 1; Table S1). Thirty-three of the samples were included in the study of Willyard *et al.* (2021a).

Table 1. The taxa of *Pinus* subsections *Ponderosae* and *Sabinianae*. The recognition of five species (*P. apulcensis*, *P. estevezii*, *P. oaxacana*, *P. martinezii*, and *P. stormiae*) is based on results from the present study.

Taxon	Distribution
Pinus subsection *Ponderosae* Loudon	Canada, USA, Guatemala, Honduras, El Salvador
Pinus apulcensis Lindl.	Mexico
Pinus arizonica Engelm.	USA, Mexico
Pinus brachyptera Engelm.	USA, Mexico
Pinus cooperi C.E.Blanco	Mexico
Pinus devoniana Lindl.	Mexico, Guatemala
Pinus durangensis Martínez	Mexico
Pinus estevezii (Martínez) J.P.Perry	Mexico
Pinus engelmannii Carrière	USA, Mexico
Pinus gordoniana Hartw. ex Gordon	Mexico
Pinus hartwegii Lindl.	Mexico, Guatemala, Honduras
Pinus martinezii E.Larsen	Mexico
Pinus maximinoi H.E.Moore	Mexico, Guatemala, Honduras, El Salvador
Pinus montezumae Lamb.	Mexico, Guatemala
Pinus oaxacana Mirov	Mexico, Guatemala
Pinus ponderosa Douglas ex P.Lawson & C.Lawson	Canada, USA
Pinus ponderosa var. *benthamiana* (Hartw.) Vasey	USA
Pinus ponderosa var. *washoensis* (H.Mason & Stockw.) J.R.Haller & Vivrette	USA
Pinus pseudostrobus Lindl.	Mexico, Guatemala
Pinus stormiae (Martínez) Frankis	USA, Mexico
Pinus scopulorum (Engelm.) Lemmon	USA
Pinus yecorensis Debréczy & I.Rácz	Mexico
Pinus subsection *Sabinianae* Loudon	USA, Mexico
Pinus coulteri D.Don	USA, Mexico
Pinus jeffreyi Balf.	USA, Mexico
Pinus sabiniana Douglas	USA
Pinus torreyana Parry ex Carrière	USA

We extracted total genomic DNA from leaf samples using the 2x CTAB protocol of Doyle & Doyle (1987) and quantified DNA for each extraction using a Qubit 3 fluorometer and a dsDNA BR Assay kit (Life Technologies Corp., Carlsbad, California). Purity was measured with A260/A280 ratios using a NanoDrop (ThermoFisher Scientific, Waltham, Massachusetts). Six samples (*P. brachyptera* Engelm. collections AMW1172, DM14, SM14, and PT02, and *P. stormiae* collections AMW1044 and AMW1047) were extracted, prepared as libraries, and enriched with custom probes at Hendrix College (Conway, Arkansas; Willyard *et al.* 2021a) whereas all other samples were extracted at the Universidad Nacional Autónoma de México (Mexico City, Mexico) and shipped to Daicel Arbor Biosciences (Ann Arbor, Michigan) for genomic library preparation, target enrichment, and sequencing.

Custom biotinylated RNA probes ("baits") were used to enrich putative low copy nuclear genes. Initially, we selected for 713 genes based on 11,396 gene models reported in an exon hybridization capture study of *P. taeda* L. and *P. elliottii* Engelm. (Neves *et al.* 2013.) Further details on selection criteria for the loci and the probes were described previously (Gernandt *et al.* 2018). We subsequently augmented the probe set to enrich 1,058 genes (version 2; Table S2), relying primarily on the gene models by Neves *et al.* (2013), and again augmented the probe set to enrich 1,416 genes (version 3), relying primarily on the draft genome of *P. taeda* (Neale *et al.* 2014). There are 703 genes common to all three versions of the probes (Willyard *et al.* 2021a), and the second and third versions of the probes have 1,057 genes in common. Of the 1,437 loci, 1,426 were pine nuclear protein coding genes (although we have since discovered that at least three of these are adjacent to the targeted nuclear exons), five were mitochondrial genes, and three were putative fungal contaminants (see Results). We gave priority to the inclusion of samples enriched with v. 3 of the probes but made exceptions for six samples from the US representing two taxa (*P. stormiae* and *P. brachyptera*). Substantially fewer loci resulted in contigs for these six samples (see Results).

Libraries were prepared at enriched to unenriched ratios of 80 : 20, 70 : 30 or 60 : 40, multiplexed, and sequenced using either Illumina or AVITI paired-end sequencing (additional information on probes, library preparation and sequencing is provided in Table S1).

Sequence assembly and alignment. Sequence reads were processed with Trimmomatic v. 39 (Bolger *et al.* 2014) to remove adapters and low-quality bases (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20). We retained reads with a minimum length of 36 bp. The nuclear gene sequences were assembled with the HybPiper v. 2.3.0 pipeline (Johnson *et al.* 2016). Our reference target file for recruiting and sorting reads with BWA (Li & Durbin 2009) was updated from previous studies (*e.g.*, Gernandt *et al.* 2018, Montes *et al.* 2019, Willyard *et al.* 2021a) to include 1,486 sequences for 1,437 loci (several genes with long introns were divided into multiple loci, and several genes include two or three references). Of the 1,486 reference sequences, > 90 % were from *P. taeda*, including > 900 loci from the *P. taeda* sequences used to generate most of the probes. The remainder were protein coding regions extracted from transcriptome shotgun assembly sequences from other pines and Pinaceae downloaded from GenBank. The pipeline assembled the nuclear loci de novo with SPAdes v. 3.15.5 (Bankevich *et al.* 2012). We created files with a single gene and supercontig sequence for each gene per species, and we used the paralog_retriever script in HybPiper to identify loci with multiple copies and create files with multiple paralogous sequences per sample.

Gene, supercontig, and paralog files were aligned with MAFFT v. 7.505 (Katoh & Standley 2013) with the -auto option enabled to choose a progressive or iterative alignment based on alignment size and complexity. The resulting alignments were trimmed with trimAl v. 1.3 to remove sites with missing data or gaps in 20 % or more of the samples (Capella-Gutiérrez *et al.* 2009).

We created three sets of alignments for phylogenetic analysis. The first, "relaxed" dataset included both single-copy and duplicated (paralogous) loci with 78 or more of the 87 terminals, with a sequence similarity > 88 %, and with a percent identity threshold > 50 %. The latter two criteria were determined in Geneious Prime v. 2024.0.7 (www.geneious.com). This relaxed filter was implemented in an attempt to adhere to the recommendations of the author of ASTRAL (Mirarab 2023), who discouraged excessive data filtering, except in cases of fragmentary data. The second, "strict" dataset only included putative single-copy genes (the number of copies of each gene was determined

based on the SPAdes de novo assemblies with the paralog_retriever script included with HybPiper) and the alignments for these genes were visually inspected, choosing those that either 1) had no evidence of ambiguous alignment or mis-assembly, 2) for which sites could be trimmed so that no ambiguously aligned or mis-assembled positions were included, or 3) six or fewer terminals could be deleted, resulting in a satisfactory alignment. The strict dataset represented an attempt to implement alignment quality control steps that could potentially improve hypotheses of homology and reduce error in phylogenetic inference, as advocated by Simmons *et al.* (2022). The third dataset was assembled with the HybPiper paralog_retriever script. After aligning the locus paralog files with MAFFT (saving the alignments in Clustal format to preserve the numbering of paralogs) and trimming with trimAl as described above, Geneious was used to identify and remove alignments with a sequence similarity < 88 % or with a percent identity threshold < 50 % (the same criteria as in the relaxed dataset). Additionally, we removed paralog alignments < 200 bp in length and individual sequences with lengths < 126 bp.

Plastome assembly and alignment. Plastomes were assembled from off-target reads with HybPiper. We used a target file consisting only of the complete plastome sequence of a single sample (*P. pseudostrobus* DSG1785) assembled with GetOrganelle v. 1.7.7.1 (Jin *et al.* 2020) because we found that this reference resulted in HybPiper retrieving a nearly complete plastome assembly from SPAdes. In contrast, using the plastome protein coding genes as references in the target files resulted in HybPiper functioning as intended, returning only the open reading frames for these genes, which were shorter and less informative compared to the nearly complete plastomes (c. 62 versus 120 kpb). After alignment with MAFFT, we inspected the alignments and noted that one sample (*P. brachyptera* SM14) was not recovered, and another (*P. brachyptera* PT09) was recovered only partially, and with many putative singletons (autapomorphies) which we interpreted as probable assembly errors. We retained this sample but replaced all autapomorphies with the ambiguity code "N" and deleted all sites that were only occupied by this taxon as uncalled ("N"). A total of 2,341 sites were excluded from the alignment corresponding to inversions, putative mis-assemblies, a difficult to align region with many gaps, and one of two remnant inverted repeats (see below). Alignments for the strict nuclear dataset and the plastomes are available online (https://github.com/dgernandt/Pinus_subsection_Ponderosae_data_2025/).

Phylogenetic analyses. Individual loci and the concatenated datasets were analyzed with maximum likelihood in IQ-Tree v. 2.2.2.6 (Minh *et al*. 2020b). ModelFinder (Kalyaanamoorthy *et al.* 2017) was used to choose nucleotide substitution models with the Bayesian Information Criterion, and branch support was estimated with 1,000 fast bootstrap replicates using UFboot (Hoang *et al.* 2018). To complement the likelihood bootstrap values, which can be misleadingly high for large datasets despite conflict among gene trees, gene and site concordance factors were calculated for the likelihood tree of the concatenated strict and relaxed datasets using IQ-Tree (Minh *et al.* 2020a). The site concordance factors were calculated using likelihood probability distributions (Mo *et al.* 2023).

The branches of the maximum likelihood gene trees with UFboot values < 10 % were collapsed and these trees were used as input for multi-individual searches in ASTRAL-IV v. 1.19.4.5 (Zhang *et al.* 2022a), or in the case of the paralogs, with ASTRAL-Pro3 v. 1.22.3.6 (Zhang *et al.* 2020, 2022b). For all three datasets, lineage trees and species trees were inferred using a more thorough search than offered by default by increasing -s and -r (-r 128 -s128). Branch lengths were estimated in substitutions per site (Tabatabaee *et al.* 2023). We analyzed the relaxed and strict datasets with ASTRAL-III v. 5.7.8 (Zhang *et al.* 2018) to calculate normalized quartet scores. The concatenated datasets were also analyzed with parsimony in PAUP* v. 4.0a (Swofford 2003). We performed heuristic searches with 2,000 random addition sequence replicates saving up to 20 optimal trees per replicate, followed by 2,000 bootstrap pseudoreplicates, each with 10 random addition sequence replicates saving up to 10 optimal trees per replicate.

Network analyses. We tested for reticulation by subsampling terminals from a strict selection of nuclear locus alignments (265 putative orthologs) using the script BeforePhylo (https://github.com/qiyunzhu/BeforePhylo) so that the number of terminals (12 or 13) could be analyzed in reasonable times. The alignments were inspected visually to

only include those with informative sites that did not appear to include any assembly or alignment problems. These alignments were analyzed with maximum likelihood using IQ-Tree as described above, except that *P. jeffreyi* Balf. was designated as the outgroup, choosing the best nucleotide substitution model, and collapsing branches with less than 10 % bootstrap values. The resulting trees were converted into NEXUS format and used as input for Phylonet v. 3.8.2 (Than *et al.* 2008, Wen *et al.* 2018). We calculated the best unrooted tree using minimizing deep coalescences and the best networks with exactly one or two reticulations under parsimony and maximum pseudolikelihood optimality criteria.

Results

Assembly and alignment statistics. Sequence read assembly statistics for each sample are provided in Table S3. The mean number of paired reads processed for HybPiper was 21,716,602 (range: 3,964,116-49,343,752). Six samples (*P. brachyptera* samples AMW1172, DM14, PT09, and SM14, and *P. stormiae* samples AMW1044 and AMW1047) were low outliers, ranging from 3,964,116 to 13,432,116 reads. The mean number of mapped reads was 6,731,749 (range: 190,001-18,603,065), and the mean number of genes mapped was 1,425.9 (range: 1,170-1,435) out of a total of 1,435 genes in the target file, with the same six samples representing low outliers (range: 1,170-1,423). The mean number of genes with contigs at least 75 % of the length of the reference sequences was 1,246.0 (range: 289-1,339). A single sample (*P. stormiae* AMW1047) had only 289 genes assembled at 75 % of the reference length. This was the only sample included that was enriched with the first version of the probes, which targeted 711 loci, compared to approximately double that in the third version of the probes (Table S2). We observed in the alignments that all six of the outlier samples often had fragmentary assemblies, but this sample stood out as most often being completely absent from individual gene alignments.

For the enriched nuclear and five enriched mitochondrial sequences, a total of 1,431 loci assembled as genes were recovered from the 87 samples. Four loci (0-15757, and the three putative fungal genes: 0-15365, and 0-15842, and 0-17556) failed to assemble (reported for the first time here with their best BLASTX hits: 0-15365, *Thelephora terrestris* ribosomal protein L2 GenBank Accession No. KAF9777530.1; 0-15842, *Thelephora terrestris* ARF/SAR GenBank Accession No. KAF9779517.1; and 0-17556, *Fusarium oxysporum* 40S ribosomal protein S14 GenBank Accession No. XP_018248369.1).

The total aligned length of the gene assemblies (in which the assemblies are trimmed to the length of the target file locus, which usually corresponds to an open reading frame) after the automated trimming step was 1,378,340, with an average locus length of 963.2 bp (range: 63-6,611 bp). The supercontig assembly (open reading frames, introns, and flanking regions) also recovered 1,431 loci (Figure S1). The total aligned length of the supercontig loci was 2,018,062 bp with an average length of 1,410.2 bp (range: 152-7,045 bp). After applying a relaxed filter by removing alignments missing eight or more terminals and with high divergence (see Methods), and merging the genes and supercontigs, retaining the longer supercontigs whenever both met our criteria for inclusion, the total aligned length of the resulting dataset of 1,258 loci was 1,567,589 bp (average length per locus of 1,246.1; range: 63 to 6,611). None of the five mitochondrial alignments passed the filter.

The numbers of paralogous copies assembled with HybPiper per sample and per gene are provided as a heatmap (Figure S1) and a table (Table S4). For the 86 samples of sect. *Trifoliae* (excluding the sect. *Pinus* outgroup), 816 of 1,435 loci were inferred to be single copy. The paralog assembly recovered sequence alignments for 1,430 loci (one locus, scaffold208126, retrieved 83 assemblies but could not be aligned). The number of terminals in this dataset often exceeded the number of samples (87) because of the inclusion of paralogs, resulting in alignments with 2 to 2,191 terminals. The total aligned length of the paralog assembly was 1,321,282 bp, with an average locus length of 924.0 bp (range: 63-6,611 bp).

The strict filter resulted in 308 alignments with a total length of 390,242 bp (average length per locus of 1,267.0; range: 261-5,874). The relaxed dataset of 1,258 loci had 150,305 parsimony informative characters and 138,526 variables but uninformative characters. Filtering the paralog alignments resulted in a dataset of 985 loci with a total

aligned length of 995,684 and an average locus length of 1,010.8 (range: 201-6,611). The number of terminal sequences per gene ranged from 50-448. One of the five mitochondrial alignments (*matR*) passed the filter.

Phylogenetic analyses of nuclear loci. Phylogenetic relationships of *Pinus* subsect. *Ponderosae* lineages as estimated from 308 putative nuclear orthologs with ASTRAL-IV are shown in Figure 1 (branch support values are provided in Figure S2). The subsection was recovered as monophyletic with strong support as measured by local posterior probability branch support values (lpp = 1), and sister to subsect. *Sabinianae* (lpp = 0.84). The deepest division in subsect. *Ponderosae* was between *P. ponderosa* P. & C.Lawson together with its varieties (lpp = 0.98) and the remaining taxa (lpp = 1), which in turn comprise a clade of *P. brachyptera* samples from west-central North America and more southerly distributed taxa, most with centers of distribution in Mexico (lpp = 0.64). Taxon specific clades were identified in Mexico, but only one received high support: a sister relationship between two collections of *P. stormiae* from Nuevo León.

The species tree for the 308 loci also recovered *Pinus* subsect. *Ponderosae* as monophyletic with high support (lpp = 1.0) and sister to subsect. *Sabinianae* (Figure 2A). Within subsect. *Ponderosae*, *P. ponderosa* was sister to all remaining species of the subsection (lpp = 1) and *P. brachyptera* was successively sister to a clade centered in Mexico (lpp = 1). Support for the monophyly of the Mexican clade was low (lpp = 0.66), as were most of the relationships among its component taxa. In this Mexican clade, collections of *P. stormiae* from Texas and Nuevo León were sister to all the remaining taxa. A sister relationship between *P. gordoniana* and *P. yecorensis* Debreczy & I.Rácz received moderate support (lpp = 0.75), as did a sister relationship between these two species and *P. maximinoi* (lpp = 0.93). Our initial circumscription of *P. pseudostrobus* Lindl. resulted in a paraphyletic relationship between the typical variety and two taxa recognized in alternate taxonomic treatments: *P. oaxacana* Mirov (and another possible segregate, *P. nubicola* Perry) and *P. apulcensis* Lindl. (and another possible segregate, *P. estevezii* (Martínez) J.P.Perry). *Pinus montezumae* Lamb., and *P. hartwegii* Lindl. were sister to *P. pseudostrobus* (lpp = 0.50), with the former two species in a poorly supported (lpp = 0.65) sister relationship. Sister to all these taxa in the predominantly Mexican and Central American clade was a clade that included *P. arizonica* Engelm., *P. durangensis*, *P. cooperi* C.E.Blanco, *P. engelmannii*, *P. martinezii,* and *P. devoniana* (lpp = 0.36). *Pinus arizonica* and *P. durangensis* were weakly supported as sisters (lpp = 0.49), and in turn these two species were sister to *P. cooperi* (lpp = 0.59), and *P. engelmannii* was sister to these three species (lpp = 0.43).

The strict dataset of 308 putatively orthologous loci had 9,289 parsimony informative characters and 17,289 variable but uninformative characters. The parsimony heuristic search recovered a single tree (length = 61,149 steps, consistency index = 0.4485, consistency index excluding uninformative characters = 0.2252, and retention index = 0.5246; Figure S3). Both the parsimony and the concatenated maximum likelihood analyses of the strict dataset recovered subsect. *Ponderosae* as the sister group to subsect. *Sabinianae* (Figure S4). Principal relationships among the subsect. *Ponderosae* taxa were mostly the same as for the ASTRAL tree, except that in the parsimony and likelihood trees a clade of *P. hartwegii*, *P. montezumae*, and *P. pseudostrobus* was sister to a clade of *P. gordoniana*, *P. maximinoi*, and *P. yecorensis*. The sister group to these species was *P. devoniana*. Support values were higher (*e.g.*, > 80 %) in the likelihood tree than in the ASTRAL and parsimony trees for most relationships, despite these relationships receiving low (often 0) gene concordance factor support (Figure S4). Site concordance factors were relatively higher.

The lineage tree for the relaxed dataset was similar in topology, except the sister group of subsects. *Ponderosae* was *P. contorta* Dougl. ex Loudon and *P. taeda*, and *P. devoniana* Lindl. was sister to *P. gordoniana* Hartw. ex Gordon, *P. maximinoi* H.E.Moore, and *P. yecorensis* (Figure S5). It had marginally higher support values for most but not all branches. The ASTRAL-III analysis recovered a normalized quartet score of 0.4079 for the relaxed dataset, indicating a relatively low number of compatible gene trees. In comparison, the ASTRAL-III analysis of the 308 loci recovered a normalized quartet score of 0.4321, a slight increase in compatible quartets relative to the relaxed dataset. The species tree for the relaxed dataset (Figure 2B) was topologically similar to its lineage tree (Figure S5).

The parsimony heuristic search recovered a single tree (length = 1,307,638 steps, consistency index = 0.2504, consistency index excluding uninformative characters = 0.1558, and retention index = 0.2907; Figure S6). Subsections

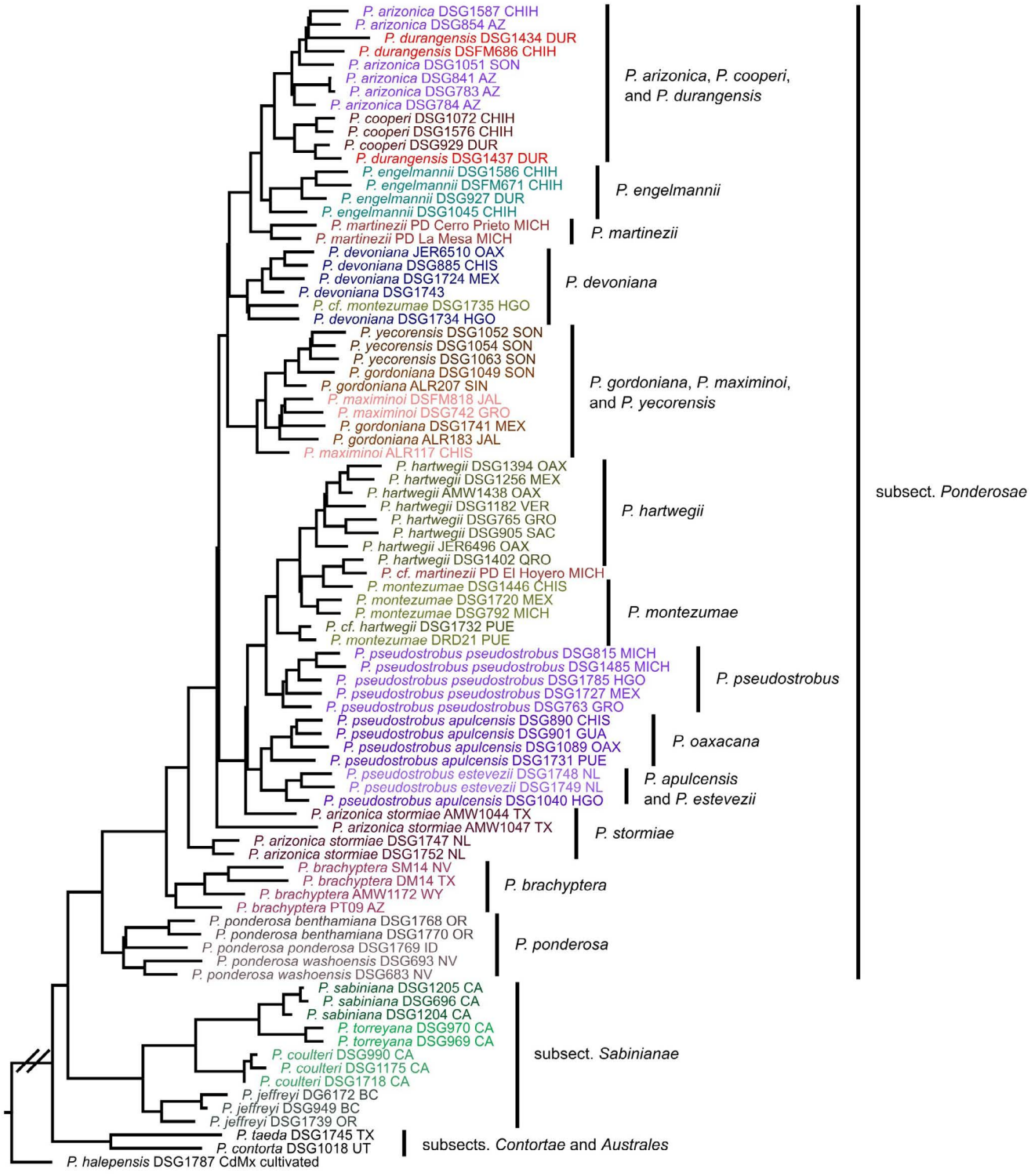

Figure 1. Lineage tree of *Pinus* subsection *Ponderosae* and outgroups based on an ASTRAL-IV analysis of 87 terminals and 308 loci. The scale bar indicates branch lengths in substitutions per site. The branch leading to the outgroup has been truncated.

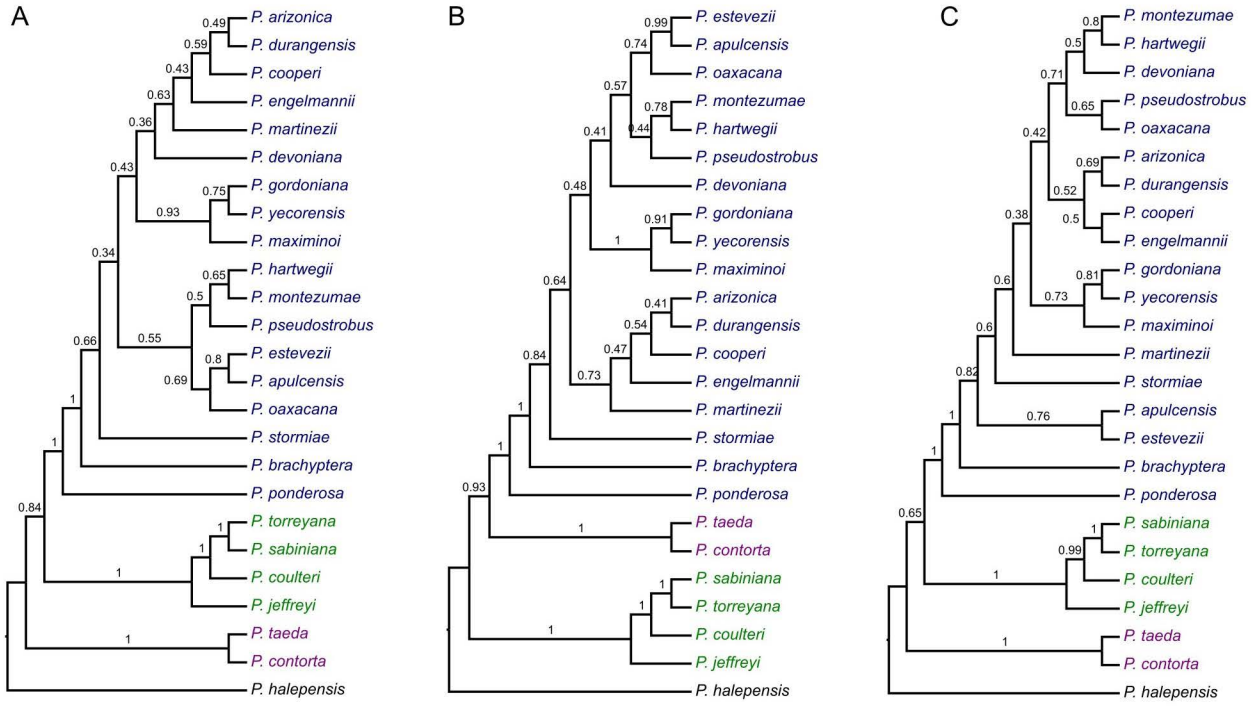

Figure 2. Species trees for *Pinus* subsection *Ponderosae*. A. Tree based on an ASTRAL-IV analysis of 87 terminals and 308 genes or supercontigs. B. Tree based on an ASTRAL-IV analysis of 87 terminals and 1258 genes or supercontigs. C. Tree based on an ASTRAL-Pro3 analysis of 985 loci including paralogs. Blue text = *Pinus* subsection *Ponderosae*, green text = *Pinus* subsection *Sabinianae*, and purple text = *Pinus* subsects. *Contortae* and *Australes*. Values above branches are local posterior probabilities.

Ponderosae and *Sabinianae* were reciprocally monophyletic, but within subsect. *Ponderosae*, several species groups recovered with the strict dataset (*e.g.,* samples of *P. engelmannii*, the *P. hartwegii* and *P. montezumae* clade, and the *P. gordoniana* + *P. maximinoi* + *P. yecorensis* clade) were non-monophyletic with the relaxed dataset. In the maximum likelihood tree, subsect. *Ponderosae* was sister to the clade of subsects. *Australes* and *Contortae* rather than to subsect. *Sabinianae* (Figure S7). The maximum likelihood tree had higher bootstrap values than the parsimony tree based on the same dataset (Figure S6), and than the maximum likelihood tree based on the strict dataset (Figure S4). As in the parsimony tree, several species recovered as monophyletic with the strict dataset were non-monophyletic.

The lineage tree for the paralog alignments (Figure S8) resembled the strict and relaxed datasets in recovering subsect. *Ponderosae* as monophyletic (Figure 2C). As with the strict dataset, subsects. *Ponderosae* and *Sabinianae* were sister groups, and *P. ponderosa* samples were sister to all other subsect. *Ponderosae* taxa. In both the lineage tree and species tree, relationships differed from the strict dataset in several respects and received low local posterior probability support values.

Plastome results. The full plastome alignment had a length of 122,443 bp. We then excluded eight parts corresponding to 2,341 bp: a 10 bp inversion in the *psbA-trnI* spacer of the three accessions of *P. jeffreyi*, an 18 bp inversion between *ycf3* and *psaA* of the three accessions of *P. coulteri*, 106 bp with a putative mis-assembly in rrn16 of *P. hartwegii* DSG1394, a 79 bp putative mis-assembly in rrn23 of *P. coulteri* DSG990, a 1,629 bp region in 3' *ycf1* around a perfect repeat that assembled with many gaps, a 6 bp inversion between *ycf1* and *rps15* in all three accessions of *P. coulteri*, a 9 bp inversion between *ndhD* and *psaC* in seven outgroup accessions, and 485 bp corresponding to the IRB, which includes *trnI* and a duplicated piece of *psbA*. The final alignment had a total length of 120,102 bp. A total of 1,046 sites were parsimony informative, and 2,386 were variable but uninformative for parsimony.

Plastid relationships inferred with maximum likelihood are shown in Figure 3. *Pinus* subsect. *Ponderosae* was recovered as monophyletic (bootstrap = 100 %) and sister to subsect. *Sabinianae* (bootstrap = 100 %). Within subsect. *Ponderosae*, *P. ponderosa* vars. *ponderosa*, *benthamiana*, and *washoensis* were monophyletic (bootstrap = 100 %) and sister to all other individuals of *Ponderosae*, which were monophyletic (bootstrap = 100 %). The Mexican clade comprised subclades of *P. pseudostrobus*, *P. hartwegii* + *P. montezumae*, *P. devoniana*, *P. arizonica* + *P. cooperi* + *P. durangensis* + *P. brachyptera* + *P. scopulorum*, *P. gordoniana* + *P. maximinoi* + *P. yecorensis*, and *P. engelmannii*.

The *P. pseudostrobus* + *P. oaxacana* + *P. apulcensis* clade (100 % bootstrap) also included individuals of other species that may represent examples of plastid capture: *P. montezumae* (DSG1446 CHIS), *P. hartwegii* (DSG905 SAC), *P. maximinoi* (ALR117 CHIS), and *P. stormiae* (AMW1044 TX and AMW1047 TX). Two of three individuals from Michoacán of *P. martinezii* (PD Cerro Prieto and PD La Mesa) also occurred within the clade. *Pinus apulcensis* did not show clear separation from *P. pseudostrobus*, but *P. estevezii* did, and furthermore was sister to the two individuals of *P. stormiae* from Texas.

The *P. hartwegii* + *P. montezumae* clade included four subclades, all receiving high support (100 % bootstrap). One subclade included samples of both species, either from Veracruz, Puebla, Mexico State, or Guerrero. A second subclade of two individuals comprised a sample originally identified as *P. martinezii* from a relatively high elevation site in Michoacán sister to a *P. montezumae* collected from near the type locality of *P. martinezii* near Uruapan, Michoacán. The third subclade included *P. hartwegii* collections from Oaxaca, Puebla, and Querétaro, and the fourth, which was sister to the rest, included two *P. hartwegii* individuals from Oaxaca.

The *P. devoniana* clade (100 % bootstrap) included all collections of *P. devoniana*, including one with shorter leaves and cones from the type locality of the same species, but with leaf and cone dimensions closer to *P. montezumae* (possibly corresponding to *P. russelliana* Lindl., treated as a synonym of *P. montezumae* by Farjon & Styles 1997). A single collection of *P. engelmannii*, which grouped with other *P. engelmannii* in the nuclear tree, was strongly supported as a member of the *P. devoniana* clade in the plastid tree.

The plastid tree included a large, well-structured clade (100 % bootstrap) that included *P. arizonica*, *P. cooperi*, *P. durangensis*, and *P. brachyptera*. None of the species represented by multiple individuals formed its own clade, but there were two well supported subclades within it. Both subclades included collections from both the United States and Mexico. One included all but one of the Arizona taxa and presumably corresponds to the typical *P. arizonica* haplotype, although it also included one collection of *P. brachyptera* and two taxa from Mexico determined as *P. durangensis*. An early split in this same predominantly *P. arizonica* clade resulted in *P. stormiae* from Nuevo León near its type locality (represented by two collections) as sister. The second clade included all collections of *P. cooperi* together with a single *P. durangensis* individual from Durango, a single individual of *P. brachyptera* from Texas, and a single individual of *P. arizonica* from Arizona. *Pinus brachyptera* did not group with these taxa in the nuclear trees (Figures 1, 2).

The *P. gordoniana*, *P. maximinoi*, and *P. yecorensis* clade (100 % bootstrap; Figure 3) included three subclades (all with 100 % bootstrap support). Four individuals from Sonora, three corresponding morphologically to *P. yecorensis*, formed one clade. *Pinus gordoniana* and *P. maximinoi* shared haplotypes in the other two clades.

Three of four individuals of *P. engelmannii* formed a well-supported clade (100 % bootstrap) in the plastid tree that was supported as sister to the *P. gordoniana*, *P. maximinoi*, *P. yecorensis* clade (95 % bootstrap). This position contrasted with the nuclear results where *P. engelmannii* was sister to a clade that included *P. arizonica*, *P. cooperi*, and *P. durangensis* (Figures 1, 2A).

Phylogenetic networks. Analyses of 12 individuals representing species with Phylonet, using maximum pseudolikelihood without permitting reticulation, resulted in a tree with a similar topology to the nuclear tree when rooting subsect. *Ponderosae* with *P. jeffreyi* (DSG1739). A single collection of *P. ponderosa* (DSG1769) was sister to the remaining taxa, and clades corresponding to *P. arizonica* (DSG1587) + *P. cooperi* (DSG1072), *P. devoniana* (DSG1734) + *P. engelmannii* (DSG1586), *P. gordononiana* (ALR207) + *P. maximinoi* (DSG742) + *P. yecorensis* (DSG1052), and *P. hartwegii* (DSG1256) + *P. montezumae* (DSG1720) sister to *P. pseudostrobus* (DSG1040; total log probability = -0.4585554285; Figure 4A). Allowing a single reticulation resulted in the inference of gene flow between the

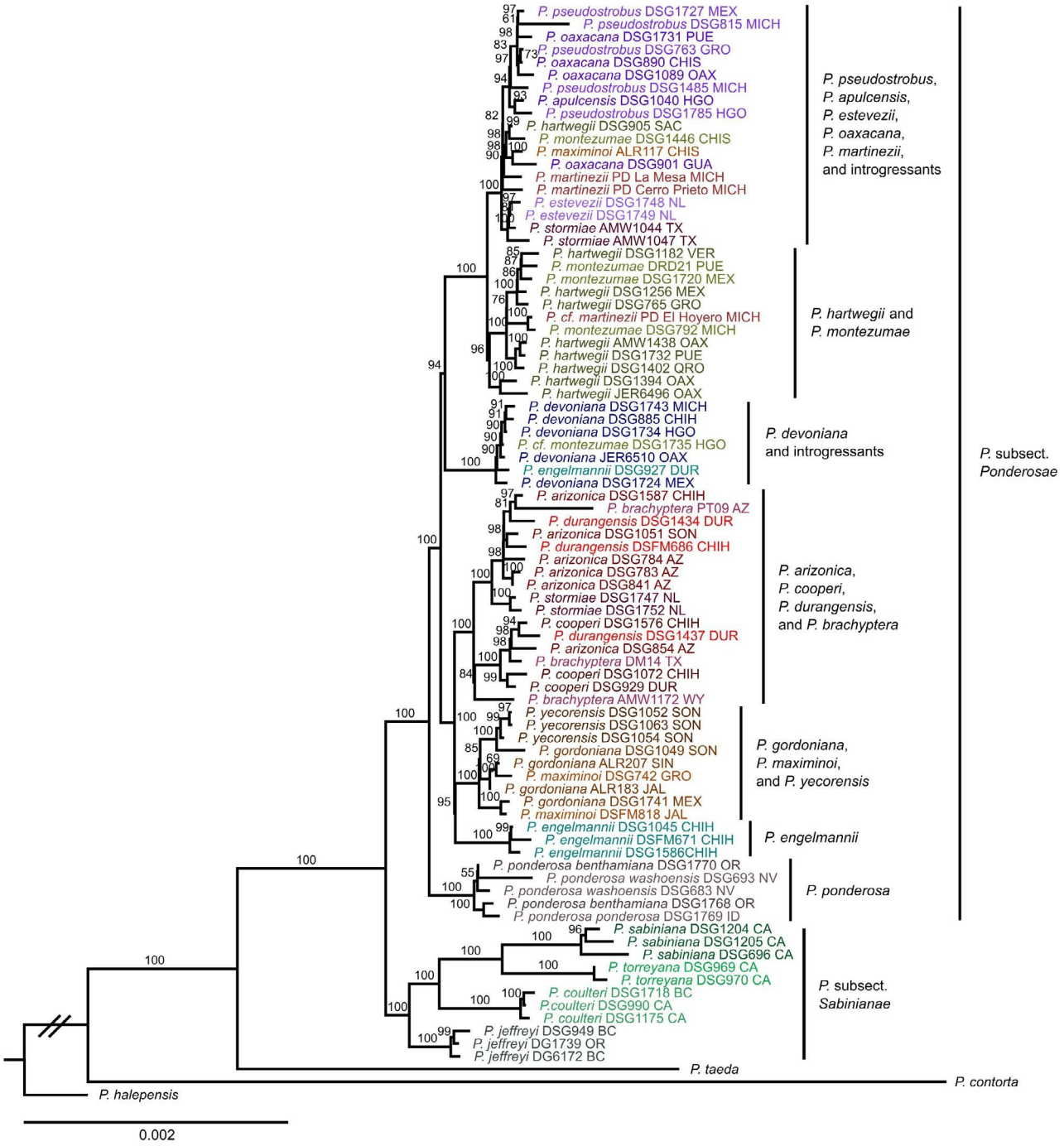

Figure 3. Maximum likelihood tree inferred from plastomes. Eighty-six of the 87 samples analyzed in the nuclear locus analysis are included. Bootstrap values > 50 % are shown above branches. The scale bar indicates branch lengths in substitutions per site. The branch leading to the outgroup has been truncated.

P. ponderosa lineage and the stem of the clade that included *P. arizonica*, *P. cooperi*, *P. devoniana*, *P. engelmannii*, and *P. pseudostrobus* (total log probability = -0.4579931681; Figure 4B). Allowing a second reticulation inferred gene flow between the stem of *P. hartwegii* and *P. montezumae* and the ancestor of all subsect. *Ponderosae* taxa except *P. ponderosa* (total log probability = -0.45774812466; Figure 4C). Inclusion of a thirteenth collection (*P.* cf. *montezumae* DSG1720), which we had identified in the field based on morphology as a possible hybrid between *P. mon-*

tezumae and *P. devoniana*, resulted in inferred gene flow between it and the stem lineage of *P. montezumae* and *P. hartwegii* (Figure 4D), lending some support to the possibility of reticulation occurring at the site. Additional analyses that included a thirteenth collection relative to Figure 4A with evidence of chloroplast capture between clades (*P. maximinoi* ALR117 and *P. martinezii* PD Cerro Prieto) did not recover inferred gene flow involving these taxa (not shown).

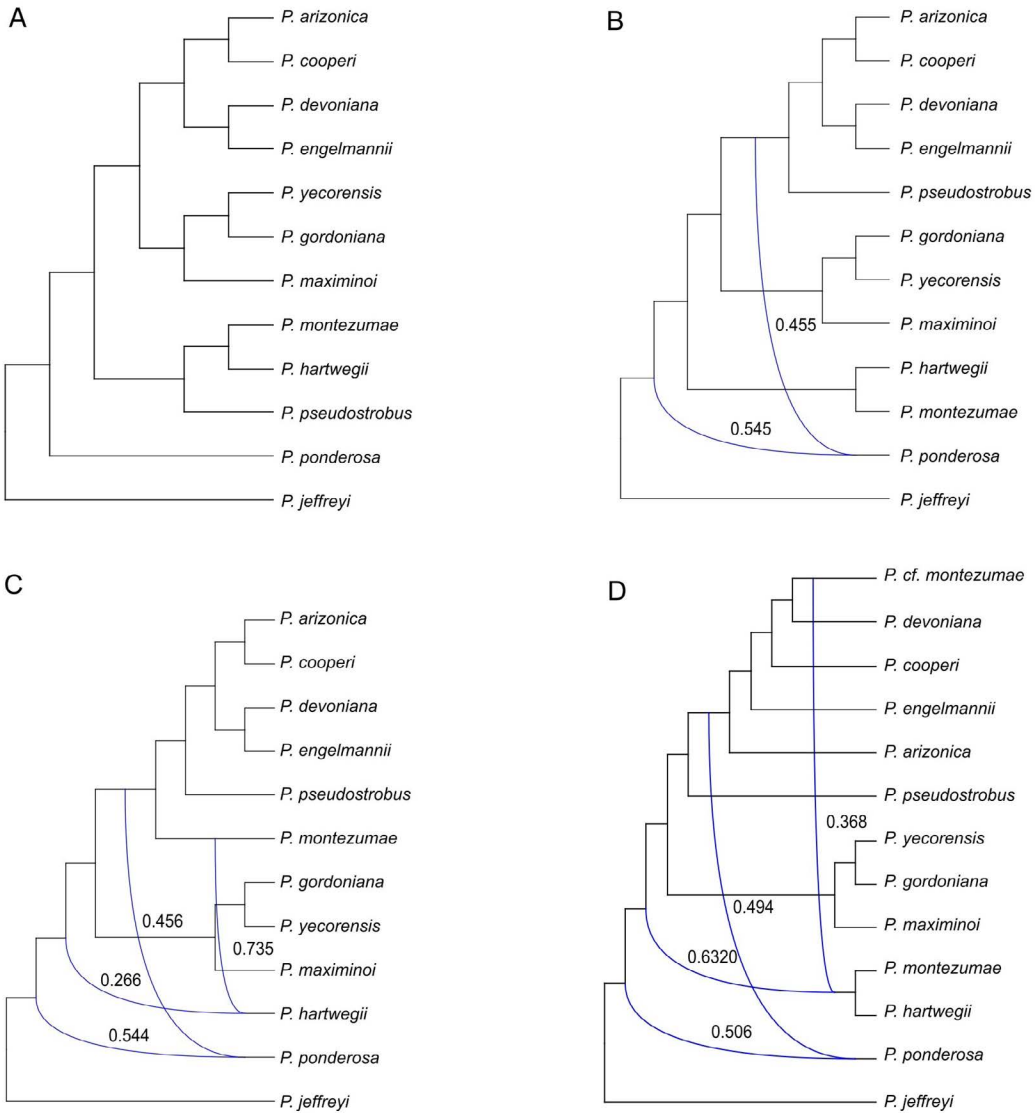

Figure 4. Reticulation results from Phylonet using maximum pseudolikelihood based on informative putative orthologs. A. Tree resulting from an analysis of 12 individuals, without permitting a single reticulation. B. Network resulting from an analysis of 12 individuals, permitting a single reticulation. C. Network resulting from an analysis of 12 individuals, permitting two reticulations. D. Network resulting from an analysis of 13 individuals, including a collection morphologically intermediate between *P. devoniana* and *P. montezumae*. Numbers next to curved lines are inheritance probabilities.

Discussion

The total alignment length of the genes recovered by our probe set is greater than that of previous studies of pines. The supercontig assembly alignment exceeded 2.1 Mbp for 1,431 loci, which was reduced to 1,567,589 Mbp for 1,258 nuclear loci after we applied the relaxed filter. However, the strict filter, which excluded paralogs based on

information from HybPiper, resulted in 308 loci with a total alignment length of 390,242 bp. This number of loci was less than was analyzed in other Hyb-Seq studies, but the overall length per locus was greater. Gernandt *et al.* (2018) reported a total alignment of 268,477 bp for 339 nuclear loci of *Pinus* subsection *Australes* after relatively stringent filtering, Montes *et al.* (2019) reported a total alignment length of 221,129 bp for 304 nuclear loci of *Pinus* subsection *Cembroides*, and Willyard *et al.* (2021a), who did not remove paralogs, reported a total alignment length of 603,747 bp for 600 loci of *Pinus* subsect. *Ponderosae*.

Some of the overall increase in alignment length reported here is due to the addition of more genes in a third version of the probe set and improved bioinformatic methods. However, identifying the open reading frame in most of our previous references, and using these open reading frames as references in the present study allowed us to assemble longer sequences with the HybPiper supercontig assemblies. Many of these longer sequences passed our quality filters. Further work to separate hundreds of putative paralogs and to refine the exon-intron limits in the reference sequences used in HybPiper should result in improved alignment quality and greater discriminating power for studies of phylogenetic relationships, reticulation, and other aspects of molecular evolution.

Both the nuclear and plastid analyses reported here recovered *Pinus* subsect. *Ponderosae* as a well-supported monophyletic group that is separate from subsect. *Sabinianae*. The nuclear relationships that we recovered for the main lineages of sect. *Trifoliae* with the relaxed dataset were similar to those recovered by Jin *et al.* (2021) in that subsection *Ponderosae* was sister to a clade that included representatives of subsects. *Contortae* and *Australes* (Figure S2), whereas subsect. *Sabinianae* was sister to all remaining lineages of sect. *Trifoliae*. In contrast, for the strict filter of nuclear loci (Figures 1, 2B), the paralog dataset (Figure 2C) and the plastid data (Figure 3) subsect. *Sabinianae* was sister to subsect. *Ponderosae*, consistent with previous plastid studies (Gernandt *et al.* 2005, Hernández-León *et al.* 2013). The agreement of the strict dataset tree with the plastid tree should not be attributed generally to the use of strict filtering, however, as other strict datasets that we evaluated with different filtering criteria recovered the same sister group to subsect. *Ponderosae* as the relaxed filter (results not shown).

Regarding the earliest diverging taxa, which have distributions in the United States or Canada, subsect. *Ponderosae* had a clear split at its base between *P. ponderosa* and its varieties (distributed along the Pacific Coast, and the Sierra Nevada, west of the Great Basin) and the remaining species. This relationship was recovered in both nuclear and plastid trees, and in previous multilocus nuclear analyses using target enrichment (Willyard *et al.* 2021a) and transcriptome orthologs (Jin *et al.* 2021). Above this split, we recovered a clade of *P. brachyptera* (Figures 1, S2, S3), which occur primarily in the Rocky Mountains and Sky Islands of the southwestern United States and northern Mexico. Branch support (measured as lpp) for this clade dropped in the ASTRAL analysis when we used the strict dataset (Figures 1, S2) but was higher for the larger relaxed dataset (Figure S5), and for the maximum likelihood of the concatenated data, as measured by the bootstrap (Figures S4, S7). These *P. brachyptera* samples did not group together in the analysis of plastomes (Figure 3); instead, they were dispersed in the *P. arizonica*, *P. durangensis*, and *P. cooperi* clade. These same samples showed similar nuclear-cytoplasmic discordance in the study of Willyard *et al.* (2021a) and were not included in the transcriptome study of Jin *et al.* (2021). Furthermore, the four samples representing *P. brachyptera* had high amounts of missing data and fragmentary alignments compared to most of the other samples. These kinds of data flaws can be a source of error in phylogenetic analyses with ASTRAL (Mirarab 2023). Thus, the position and circumscription of *P. brachyptera* requires further study.

The other multispecies relationship that was recovered with moderate or high support in the nuclear tree was the *P. gordoniana*, *P. maximinoi*, *P. yecorensis* clade (Figures 1-2). In the plastid tree, one individual from this clade (*P. maximinoi* from Chiapas) occurred in a predominantly *P. pseudostrobus* clade, consistent with plastid capture. This individual was morphologically assignable to *P. maximinoi*, and was sister to all other members of the *P. gordoniana*, *P. maximinoi*, and *P. yecorensis* clade in the nuclear tree (Figure 1). All other representatives were recovered together in the plastid tree with high bootstrap support (100 %). López-Reyes *et al.* (2015) documented plastid haplotype sharing of the three species in this nuclear clade and *P. pseudostrobus* using a larger sample size. The nuclear sequences generated with Hyb-Seq readily separated them from the morphologically similar species, *P. pseudostrobus* (Figures 1-2).

Another difference between this study and Willyard *et al.* (2021a) is that here, *P. arizonica*, *P. cooperi*, and *P. durangensis* were recovered as a clade in the species tree for the strict and relaxed datasets (Figure 2A-B). The samples of *P. stormiae* that were collected closer to the type locality in the Sierra Madre Oriental of Mexico did not group with collections from Texas that were tentatively assigned to *P. stormiae* by Willyard *et al.* (2021a) but formed their own lineage in the nuclear tree (Figure 1). The putative *P. stormiae* samples from Texas occurred with *P. estevezii* in the plastid tree (Figure 3) and were paraphyletic to more inclusive Mexican clades in the nuclear tree (Figure 1).

Larsen (1964) described *P. martinezii* for specimens with six needles per fascicle from near Uruapan, Michoacán. Farjon & Styles (1997) placed *P. martinezii* in synonymy with *P. durangensis*, which has slender, not glaucous nor rigid needles that vary in the number of leaves per fascicle. Two collections of this putative taxon grouped sister to the *P. arizonica*, *P. cooperi*, *P. durangensis,* and *P. engelmannii* in the nuclear tree based on the strict alignment filter (Figure 1), but not in the plastid trees, where they instead grouped together in a clade composed of *P. pseudostrobus* and *P. oaxacana*, suggesting that pines in Michoacán with typically five or six leaves per fascicle could be a relative of *P. durangensis* that captured the plastid of *P. pseudostrobus*. The other individual identified based on morphology as *P. martinezii* occurred in the *P. hartwegii*-*P. montezumae* clade in both the nuclear and plastid trees. This individual morphologically resembled *P. montezumae* more than the other two, although reticulation with another taxon might explain its morphological differences with *P. montezumae*. Further investigation is needed on the extent of gene flow in Michoacán, and more generally throughout the Mexican Transverse Volcanic Belt, among *P. martinezii*, *P. montezumae*, *P. pseudostrobus*, and *P. hartwegii*.

Several taxonomists have hesitated to separate *P. hartwegii* from *P. montezumae*. Engelmann (1880) and Shaw (1909, 1914) treated *P. hartwegii* as a variety of *P. montezumae* but both were recognized as species by subsequent specialists (*e.g.*, Martínez 1992, Matos 1995, Farjon & Styles 1997, Debreczy & Rácz 2011). The close relationship between *P. hartwegii* and *P. montezumae* is reflected in both the nuclear (Figure 1) and plastome trees (Figure 3), where individuals assigned to one or other of these species intermix. We were unable to identify any phylogenetic structure within our nine collections of *P. hartwegii* that could justify the recognition of *P. rudis* Endl., which agrees with Matos (1995), who was unable to statistically distinguish *P. rudis* from *P. hartwegii* on two elevational transects in central Mexico and concluded that the considerable variation in needles per fascicle, needle length, fascicle color, cone length, cone color, cone scale thickness, among others corresponded to a single species, *P. hartwegii*. Farjon & Styles (1997) also treated *P. rudis* as a synonym of *P. hartwegii*. Our nine individuals determined as *P. hartwegii* spanned an elevational range of 2,450 to 4,300 m and included individuals with long and short needles that varied in their number per fascicle. There is some uncertainty as to the types of *P. hartwegii* and *P. rudis*. Lindley (1839) described *P. hartwegii* from “Campanario” in Michoacán, and the lectotype of *P. hartwegii*, designated by Farjon & Styles (1997), is original material of Hartweg that lacks locality information, but has relatively long (erect) needles and therefore appears to be a collection from an intermediate, rather than high elevation population.

The plastome tree included a clade with all samples of *P. pseudostrobus*, *P. apulcensis*, *P. estevezii*, and *P. oaxacana*, but also individuals representing several other taxa, including *P. hartwegii* from Guatemala, *P. montezumae* from Chiapas, two individuals of *P. stormiae* from Texas, and two individuals assigned to *P. martinezii*. The geographic distribution of the clade is extensive, and includes the United States (Texas), where neither *P. pseudostrobus* nor *P. estevezii* occurs. It would be worth investigating whether this haplotype lineage is positively favored by natural selection.

Phylogenetic tests for reticulation arguably are in an early stage of development. Using Phylonet, we consistently recovered an ancient reticulation event between the *P. ponderosa* lineage and a clade of Mexican pines. We also examined the possibility that specific collections were hybrids. These latter analyses had only limited success: we identified an individual identified as *P.* cf. *montezumae* as a possible hybrid between *P. montezumae* and *P. devoniana*, the most abundant species at the site. Other attempts to find reticulation, limiting the analyses to a maximum of two events, were unsuccessful. These involved individuals that morphologically did not appear to be hybrids, but their position in the plastid tree was consistent with plastid capture.

Hyb-Seq can yield hundreds to thousands of nuclear loci and complete or nearly complete plastomes. We attempted to show here using a clade of Mexican pines that the ability to infer phylogenies from this quantity and type of data is immediately useful for studying evolutionary relationships of recalcitrant groups. Coalescent analysis of the nuclear data resulted in topologies that had low branch support (Figure 2), but we noticed that maximum likelihood of the strict and relaxed datasets resulted in similar trees with relatively high bootstraps (Figures S4, S6), but not as high as the plastid tree (Figure 3). Instances of conflicting signal between nuclear and plastid DNA have the potential to inform us on possible cases of chloroplast capture, and there is enough information in the nuclear loci to detect hybridization in some cases. Although our analyses here were limited to the use of phylogenetic network analysis of sequence alignments, other approaches, such as those that work with single nucleotide polymorphisms could also be implemented with Hyb-Seq data.

Supplementary material

Supplemental data for this article can be accessed here: https://doi.org/10.17129/botsci.3721

Acknowledgements

DNA extractions for this study were performed in the Laboratorio Nacional de Biodiversidad, managed by L. Cabrera Martínez. We thank Y. García Bermudez for assistance in the laboratory and herbarium, A. López Reyes, and J. Reyes, for providing collections, P. Pelaez for selection of probes, and J. Nicolas Cruz for assisting in the identification of open reading frames of the sequence references. We also thank N. Turland for nomenclatural advice, M. Johnson for his feedback on plastome assembly with HybPiper, and S. Lara-Cabrera and another anonymous reviewer for their valuable comments on a previous version of the manuscript.

Literature Cited

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, *et al.* 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455-477. DOI: https://doi.org/10.1089/cmb.2012.0021

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120. DOI: https://doi.org/10.1093/bioinformatics/btu170

Capella-Gutiérrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973. DOI: https://doi.org/10.1093/bioinformatics/btp348

Critchfield WB, Little EL. 1966. *Geographic distribution of the pines of the world*. Washington, DC- US: Department of Agriculture, Forest Service.

Debreczy Z, Rácz I. 2011. *Conifers around the world*. Budapest: DendroPress Ltd, 1089 pp. ISBN: 978-963-219-061-7 Ö

DeGiorgio M, Syring J, Eckert AJ, Liston A, Cronn R, Neale DB, Rosenberg NA. 2014. An empirical evaluation of two-stage species tree inference strategies using a multilocus dataset from North American pines. *BMC Evolutionary Biology* **14**: 67. DOI: https://doi.org/10.1186/1471-2148-14-67

Delgado P, Salas-Lizana R, Vázquez-Lobo A, Wegier A, Anzidei M, Alvarez-Buylla ER, Vendramin GG, Piñero D. 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. DOI: https://doi.org/10.1086/518260

Doyle JJ, Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11-15.

Engelmann G. 1880. Revision of the genus *Pinus*, and description of *Pinus elliottii*. *Transactions of the Academy of Science of St. Louis* **IV**: 161-189.

Farjon A, Styles BT. 1997. *Pinus* (Pinaceae). *Flora Neotropica monograph* 75. Bronx: The New York Botanical Garden, ISBN: 978-0893274115

Gernandt DS, Aguirre Dugua X, Vázquez-Lobo A, Willyard A, Moreno Letelier A, Pérez de la Rosa JA, Piñero D, Liston A. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. *American Journal of Botany* **105**: 711-725. DOI: https://doi.org/10.1002/ajb2.1052

Gernandt DS, Geada López G, Ortiz García S, Liston A. 2005. Phylogeny and classification of *Pinus*. *Taxon* **54**: 29-42. DOI: https://doi.org/10.2307/25065300

Gernandt DS, Hernández-León S, Salgado-Hernández E, Pérez de la Rosa JA. 2009. Phylogenetic relationships of *Pinus* subsection *Ponderosae* inferred from rapidly evolving cpDNA regions. *Systematic Botany* **34**: 481-491. DOI: https://doi.org/10.1600/036364409789271290

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* **27**: 182-189. DOI: https://doi.org/10.1038/nbt.1523

Hernández-León S, Gernandt DS, Pérez de la Rosa, JA, Jardón-Barbolla L. 2013. Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae* inferred from plastid DNA. *Plos One* **8:** e70501. DOI: https://doi.org/10.1371/journal.pone.0070501

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**: 518-522. DOI: https://doi.org/10.1093/molbev/msx281

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**: 241. DOI: https://doi.org/10.1186/s13059-020-02154-5

Jin W-T, Gernandt DS, Wehenkel C, Xia X-M, Wei X-X, Wang X-Q. 2021. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *Proceedings of the National Academy of Sciences* **118**: e2022302118. DOI: https://doi.org/10.1073/pnas.2022302118

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4**: 1600016. DOI: https://doi.org/10.3732/apps.1600016

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587-589. DOI: https://doi.org/10.1038/nmeth.4285

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772-780. DOI: https://doi.org/10.1093/molbev/mst010

Krupkin AB, Liston A, Strauss SH. 1996. Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. *American Journal of Botany* **83:** 489-498. DOI: https://doi.org/10.1002/j.1537-2197.1996.tb12730.x

Larsen E. 1964. A new species of pine from Mexico. *Madroño* **17**: 217-218. Ledig FT. 1998. Genetic variation in *Pinus*. *In:* Richardson DM. ed. *Ecology and Biogeography of Pinus*. Cambridge: Cambridge University Press. 251-280. ISBN: 0-521-78910-9

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760. DOI: https://doi.org/10.1093/bioinformatics/btp324

Lindley J. 1839. Miscellaneous notices: Mexican pines. *Edwards's Botanical Register* **25**: 62-64.

López-Reyes A, Pérez de la Rosa J, Ortiz E, Gernandt DS. 2015. Morphological, molecular, and ecologi-

cal divergence in *Pinus douglasiana* and *P. maximinoi*. *Systematic Botany* **40**: 658-670. DOI: https://doi.org/10.1600/036364415X689384

Ma X-F, Szmidt AE, Wang X-R. 2006. Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. *Molecular Biology and Evolution* **23**: 807-816. DOI: https://doi.org/10.1093/molbev/msj100

Martínez M. 1992. *Los pinos mexicanos*. 3rd ed. DF, México. Librería y Ediciones Botas, SA. de CV.

Matos JA. 1995. *Pinus hartwegii* and *P. rudis*: A critical assessment. *Systematic Botany* **20**: 6-21. DOI: https://doi.org/10.2307/2419628

Matos JA, Schaal BA. 2000. Chloroplast evolution in the *Pinus montezumae* complex: A coalescent approach to hybridization. *Evolution* **54**: 1218-1233. DOI: https://doi.org/10.1111/j.0014-3820.2000.tb00556.x

Minh BQ, Hahn MW, Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* **37**: 2727-2733. DOI: https://doi.org/10.1093/molbev/msaa106

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530-1534. DOI: https://doi.org/10.1093/molbev/msaa015

Mirarab S. 2023. Species tree estimation using ASTRAL: Practical considerations. *In*: Kubatko L, Knowles L, eds. *Species Tree Inference: A Guide to Methods and Applications*. Princeton: Princeton University Press, 2023, pp. 43-67 DOI: https://doi.org/10.1515/9780691245157-007

Mo YK, Lanfear R, Hahn MW, Minh BQ. 2023. Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics* **39**: btac741. DOI: https://doi.org/10.1093/bioinformatics/btac741

Mogensen HL. 1996. The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* **83**: 383-404. DOI: https://doi.org/10.1002/j.1537-2197.1996.tb12718.x

Montes JR, Peláez P, Willyard A, Moreno-Letelier A, Piñero D, Gernandt DS. 2019. Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. *Systematic Botany* **44**: 501-518. DOI: https://doi.org/10.1600/036364419X15620113920563

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martínez-García PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu L-S, Gilbert D, Marçais G, Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JFD, Lorenz WW, Whetten RW, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, deJong PJ, Yorke JA, Salzberg SL, Langley, CH. 2014.Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* **15**: R59. DOI: https://doi.org/10.1186/gb-2014-15-3-r59

Neves LG, Davis JM, Barbazuk WB, Kirst M. 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* **75**: 146-156. DOI: https://doi.org/10.1111/tpj.12193

Shaw GR. 1909. The pines of Mexico. *Publications of the Arnold Arboretum No. 1*. Forge Village, Massachusetts: The Murray Printing Company.

Shaw GR. 1914. The genus *Pinus*. *Publications of the Arnold Arboretum No. 5*. Cambridge, Massachusetts: Riverside Press.

Simmons MP, Maurin O, Bailey P, Brewer GE, Roy S, Lombardi JA, Forest F, Baker WJ. 2022. Benefits of alignment quality-control processing steps and an Angiosperms 353 phylogenomics pipeline applied to the Celastrales. *Cladistics* **38**: 595-611. DOI: https://doi.org/10.1111/cla.12507

Swofford DL 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.

Syring J, Farrell K, Businský R, Cronn R, Liston A. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. *Systematic Biology* **56**: 163-181. DOI: https://doi.org/10.1080/10635150701258787

Syring J, Willyard A, Cronn R, Liston A. 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *American Journal of Botany* **92**: 2086-2100. DOI: https://doi.org/10.3732/ajb.92.12.2086

Tabatabaee Y, Zhang C, Warnow T, Mirarab S. 2023. Phylogenomic branch length estimation using quartets. *Bioinformatics* **39**: i185-i193. DOI: https://doi.org/10.1093/bioinformatics/btad221

Than C, Ruths D, Nakhleh L. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**: 322. DOI: https://doi.org/10.1186/1471-2105-9-322

Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, Liston A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* **2**: 1400042. DOI: https://doi.org/10.3732/apps.1400042

Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. *Systematic Biology* **67**: 735-740. DOI: https://doi.org/10.1093/sysbio/syy015

Willyard A, Cronn R, Liston A. 2009. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution* **52**: 498-511. DOI: https://doi.org/10.1016/j.ympev.2009.02.011

Willyard A, Gernandt DS, Cooper B, Douglas C, Finch K, Karemera H, Lindberg E, Langer SK, Lefler J, Marquardt P, Pouncey DL, Telewski F. 2021a.Phylogenomics in the hard pines (*Pinus* subsection *Ponderosae*; Pinaceae) confirms paraphyly in *Pinus ponderosa,* and places *Pinus jeffreyi* with the California big cone pines. *Systematic Botany* **46**: 538-561. DOI: https://doi.org/10.1600/036364421X16312067913435

Willyard A, Gernandt DS, López-Reyes A, Potter KM. 2021b. Mitochondrial phylogeography of the ponderosa pines: Widespread gene capture, interspecific sharing, and two unique lineages. *Tree Genetics & Genomes* **17**: 47. DOI: https://doi.org/10.1007/s11295-021-01529-4

Zhang C, Mirarab S. 2022a. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Molecular Biology and Evolution* **39**: msac215. DOI: https://doi.org/10.1093/molbev/msac215

Zhang C, Mirarab S. 2022b. ASTRAL-Pro 2: Ultrafast species tree reconstruction from multi-copy gene family trees. *Bioinformatics* **38:** 4949-4950. DOI: https://doi.org/10.1093/bioinformatics/btac620

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153. DOI: https://doi.org/10.1186/s12859-018-2129-y

Zhang C, Scornavacca C, Molloy EK, Mirarab S. 2020. ASTRAL-Pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution* **37**: 3292-3307. DOI: https://doi.org/10.1093/molbev/msaa139

Associate editor: Eduardo Ruiz Sanchez

Author contributions: DSG, study design, field and laboratory work, data analysis, writing; AW, study design, field work, writing; AVL, study design, analysis, writing; AML, study design, field work, writing; PD, field work, writing; DSF, field and laboratory work, writing; MSGE, writing.

Supporting agencies: Financial support was provided by the Dirección General de Asuntos del Personal Académico (DGAPA), Universidad Nacional Autónoma de México, through the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (IN210422), awarded to DSG and the Secretaría de Educación Pública y Consejo Nacional de Ciencias y Tecnología (CB-2013 / 221694), awarded to DSG. DGAPA also supported a sabbatical to Oregon State University in 2020 by DSG through its Programa de Apoyos para la Superación Académica.

Conflict of interests: The authors declare that there is no conflict of interest, financial or personal, in the information, presentation of data and results of this article.