Reticulate evolution and incomplete lineage sorting among the ponderosa pines

Ann Willyard a,b,c, Richard Cronn d, Aaron Liston a,*

a Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA
b Department of Biology, University of South Dakota, 414 E. Clark St, Vermillion, SD 57069, USA
c Biology Department, Hendrix College, 1600 Washington Ave., Conway, AR 72032, USA
d Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:
Received 6 January 2009
Revised 10 February 2009
Accepted 13 February 2009
Available online 26 February 2009

Keywords:
Pinus
Pine
Hybridization
Low copy nuclear loci
Reticulate evolution
LEA
WD-40
trnG

ABSTRACT

Interspecific gene flow via hybridization may play a major role in evolution by creating reticulate rather than hierarchical lineages in plant species. Occasional diploid pine hybrids indicate the potential for introgression, but reticulation is hard to detect because ancestral polymorphism is still shared across many groups of pine species. Nucleotide sequences for 53 accessions from 17 species in subsection Ponderosa (Pinus) provide evidence for reticulate evolution. Two discordant patterns among independent low copy nuclear gene trees and a chloroplast haplotype are better explained by introgression than incomplete lineage sorting or other causes of incongruence. Conflicting resolution of three monophyletic Pinus coulteri accessions is best explained by ancient introgression followed by a genetic bottleneck. More recent hybridization transferred a chloroplast from P. jeffreyi to a sympatric P. washoensis individual. We conclude that incomplete lineage sorting could account for other examples of non-monophyly, and caution against any analysis based on single-accession or single-locus sampling in Pinus.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Many species in the genus Pinus are characterized by large effective population sizes, limited interspecific divergence, and century-long fertile life spans (Syring et al., 2007). These factors appear to have created a remarkable evolutionary web that is profoundly affected by incomplete lineage sorting. Hybridization between distinct Pinus species may also have created reticulate rather than strictly hierarchical patterns of descent (e.g. Matos and Schaal, 2000). Despite persistent attempts to classify pine species based on morphological, biochemical, cytological, and molecular characters, the taxonomic relationships among many species remain unsolved. There is a growing understanding of the causes.

First, Pinus is an ancient genus, diverging from other extant genera at least 100 million years ago in the Cretaceous (Alvin, 1960). Nevertheless, integration of genetic and fossil evidence indicates that many of the roughly 100 species of pine arose rather recently, especially when measured in generations rather than years. For example, the stem lineage of the 17 species in subsection Ponderosa (Pinaceae, Pinus, subgenus Pinus, section Tifoliae) diverged within the last 15 million years, and the crown divergence began around 5 million years ago (Willyard et al., 2007). Using an average generation time of 50 years (Syring et al., 2007), this 17-species clade began diverging some 300,000 generations ago and the crown divergence occurred on the order of 100,000 generations ago. Thus, the first factor confounding their evolutionary relationships is that pine species may be relatively young, even though the genus is ancient and individuals are generally long-lived.

Second, Pinus classification is bedeviled by plastic, homoplastic, and highly variable morphological character states. For example, serotinous cones are variable within and among populations (Borchert, 1985) and have evolved independently in several lineages (Gernandt et al., 2005). Abundant within-population genetic variation appears to be the norm. Using evidence from quantitative traits, allozymes, or molecular markers, more than 90% of variation is generally contained within vs. among populations, although a few pine species exhibit higher among-population differentiation (Leddig, 1998; Sorenson et al., 2001). The interplay between homoplasy and intraspecific variability makes delimitations between some groups of pine species challenging, encouraging some to propose hybrid origins to explain the enigma. An important example may be P. densata (subgenus Pinus, section Pinus), often cited as a diploid hybrid species (Ma et al., 2006 and references therein).

A third factor affecting pine evolutionary relationships is that lineage sorting between pine species is often incomplete. When molecular sequences for multiple individuals are sampled per species, many conspecific samples lack allelic monophyly. This has been attributed to incomplete lineage sorting (Syring et al.,
2007). Because loci coalesce at different rates and are affected by stochastic processes (Carstens and Knowles, 2007), incomplete lineage sorting creates conflicting topologies for pine species from unlinked loci (Syring et al., 2005). Many of the factors that delay allelic coalescence are present in woody trees (and specifically in Pinus): predominantly outcrossed mating, high within-species (and within-population and within-individual) mean heterozygosity, long generation time, and large effective population sizes where alleles are rarely purged (Rosenberg, 2003). Incomplete lineage sorting may be especially troublesome in Pinus because speciation has been rapid relative to effective population sizes (Willyard et al., 2007).

Fourth, hybridization between diverged lineages is thought to be an important factor in the evolution of many plants (Arnold, 1997; Grant, 1981; Stebbins, 1950); an idea supported by empirical data (Arnold, 1993; Cronn and Wendel, 2004; Doyle et al., 2004). Interspecific gene flow has been proposed to facilitate adaptive radiations in plants (Seehausen, 2004), the invention of novel habitats by nascent species (Petit et al., 2003; Rieseberg et al., 2003; Stebbins, 1959) and the invasion of new habitats by exotic plants (Ellstrand and Schierenbeck, 2000). Many researchers have speculated that reticulate evolution, facilitated by introgression, may have influenced the observed complexities among pine species.

Putative natural hybrids have been studied between many sympatric pine species, including P. banksiana and P. contorta (mitochondrial restriction sites; Dong and Wagner, 1993); P. brutia and P. halepensis (chloroplast microsatellite loci; Buccì et al., 1998); P. coulteri and P. jeffreyi (morphology; Libby, 1958; Zobel, 1951); P. jeffreyi and P. ponderosa (morphology; Haller, 1962); P. edulis, and P. monophylla (morphology; Armbruster and Phillips, 1992); P. hartwegii and P. montezumae (morphology and chloroplast; Matos, 1995; Matos and Schaal, 2000); P. montezumae and P. pseudoasobrus (chloroplast microsatellite loci; Delgado et al., 2007); P. muget and P. sylvestris (morphology; Christensen and Dar, 1997); and P. palustris and P. taeda (morphology; Namkoong, 1966).

Artificial hybridizations between many pine species within taxonomic subsections yield fertile seeds (Crittfield, 1986), demonstrating that incomplete mating barriers between allopatric Pinus species are common and suggesting that geographic, distance is a major barrier to natural hybridization among pine species. However, not all hypotheses of natural pine hybridization have withstood genetic analysis. For example, trees long suspected of representing a hybrid swarm between P. arizonica and P. ponderosa (Epperson et al., 2003) instead represent a unique third taxon (Epperson et al., 2009).

There is also substantial evidence that pine species are capable of rather rapid migrations (Petit et al., 2004). The combination of incomplete mating barriers and the opportunity for secondary contact via migration may have allowed infrequent, but evolutionarily significant, introgression within Pinus. In fact, the long-term retention of ancestral polymorphism observed in Pinus could be partially driven by migrant alleles from occasional interspecific hybridization. These alleles would increase the diversity within populations and slow the process of allelic coalescence. This may mean that reticulate evolution is an important-driving-force behind pignegetic patterns. Detection of reticulation is not at all straightforward, but is vital because assuming a hierarchical relationship for groups with a net-like rather than tree-like history may lead to erroneous conclusions (Doolittle and Bapteste, 2007; Posada and Crandall, 2001; McDaniel, 1992, 1990). Conflicting relationships for the Ponderosae have been reported based on samples of different characters and exemplars (Eckert and Hall, 2006; Gernandt et al., 2005; Krupkin et al., 1996; Liston et al., 1999). However, phylogenetic incongruence can be due to incomplete lineage sorting, recombination, natural selection, random lineage sorting, homoplasy, errors in phylogenetic inference, and reticulate evolution. Estimates of intraspecific genetic diversity are useful for interpreting causes of incongruence because a large effective population size \( N_e \) suggests that incomplete lineage sorting may provide the entire explanation for lack of monophyly and even for incongruent results, as genomes in diverse species may be "mosaics of conflicting genealogies" (Pollard et al., 2006). On the other hand, incongruence among species with limited diversity may suggest the involvement of other mechanisms, such as reticulate evolution or drift. Despite their large current effective population sizes, it is possible that some pine species have undergone genetic bottlenecks in the past that were severe enough for drift to affect the lineage (Ledig, 2000).

1.1. Ponderosae taxonomy

Nineteen species of Ponderosae are sometimes recognized, but we synonymize P. nubigera J.P. Perry with P. pseudoasobrus, and P. donnell-smitii Masters with P. hartwegii (Farjon and Styles, 1997). Two taxa have sometimes been treated within P. ponderosa (P. arizonica and P. washoensis), but are currently recognized as distinct species. Rehfeldt (1999a) provided evidence for elevating P. arizonica from its varietal status under P. ponderosa. The narrowly endemic P. washoensis is included in the Flora of North America (Kral, 1993). Pinus ponderosa var. scopulorum and P. arizonica var. stormiae are sometimes treated at the species level, our sampling scheme recognizes these two taxa as varieties (Table 1).

The 17 current species are divided into two major groups: the California big-coned pines (Sabinianae; Loudon, 1838) and Ponderosae sensu sticto (s.s.; Table 1). Three species (P. sabiniana, P. coulteri, and P. torreyana) are traditionally grouped in the Sabinianae based on shared morphology (Littie and Critchfield, 1969; Price et al., 1998). Chloroplast results (Gernandt et al., 2009) support proposals based on heptane biochemistry (Mirov, 1961) and seed fatty acids (Wolfelt et al., 2000) that P. jeffreyi, despite its superficial resemblance to P. ponderosa, belongs with the Sabinianae. We use Sabinianae to refer to all four species of California big-coned pines (Table 1).

Floristic treatments (Farjon and Styles, 1997; Martinez, 1948; Perry, 1991; Price et al., 1998) have suggested subdividing Ponderosae s.s. species in very different ways (Table 1), but these groups have not been supported by phylogenetic analyses (Gernandt et al., 2005).

1.2. Experiment design

Given the complexity of previous taxonomic delineations and the potential for a mosaic genome in some pines due to the potential for hybridization, it is vital to sample multiple individuals within each species and to also sample loci that segregate independently. We included a comprehensive sample of Ponderosae species with two to six individuals per species that represent the geographic range of each taxon wherever possible. Three independent gene regions were used: two unlinked low-copy nuclear loci and a chloroplast; Matos, 1995; Matos and Schaal, 2000); P. montezumae and P. pseudoasobrus (chloroplast microsatellite loci; Delgado et al., 2007); P. muget and P. sylvestris (morphology; Christensen and Dar, 1997); and P. palustris and P. taeda (morphology; Namkoong, 1966).

Artificial hybridizations between many pine species within taxonomic subsections yield fertile seeds (Crittfield, 1986), demonstrating that incomplete mating barriers between allopatric Pinus species are common and suggesting that geographic, distance is a major barrier to natural hybridization among pine species. However, not all hypotheses of natural pine hybridization have withstood genetic analysis. For example, trees long suspected of representing a hybrid swarm between P. arizonica and P. ponderosa (Epperson et al., 2003) instead represent a unique third taxon (Epperson et al., 2009).

The 17 current species are divided into two major groups: the California big-coned pines (Sabinianae; Loudon, 1838) and Ponderosae sensu sticto (s.s.; Table 1). Three species (P. sabiniana, P. coulteri, and P. torreyana) are traditionally grouped in the Sabinianae based on shared morphology (Littie and Critchfield, 1969; Price et al., 1998). Chloroplast results (Gernandt et al., 2009) support proposals based on heptane biochemistry (Mirov, 1961) and seed fatty acids (Wolfelt et al., 2000) that P. jeffreyi, despite its superficial resemblance to P. ponderosa, belongs with the Sabinianae. We use Sabinianae to refer to all four species of California big-coned pines (Table 1).

Floristic treatments (Farjon and Styles, 1997; Martinez, 1948; Perry, 1991; Price et al., 1998) have suggested subdividing Ponderosae s.s. species in very different ways (Table 1), but these groups have not been supported by phylogenetic analyses (Gernandt et al., 2005).
et al., 2007), suggesting that without considering recombination, phylogenetic informativeness for some nuclear loci may be comparable to chloroplast loci in *Pinus*. We explore this possibility by selecting nuclear loci with the highest level of divergence that successfully amplify and align across this taxonomic subsection. Our loci contain large regions of potentially neutral introns using primers anchored in or near exons (Liston et al., 2001). Based on our translation of the *P. taeda* cDNA for WD-40, we infer that our amplicon contains two exons (137 bps) with the remainder intronic. Our translation has highest similarity (BLASTP; nonredundant proteins) to *Arabidopsis thaliana* plasma membrane intrinsic protein (WD-40; GenBank Accession No. NP_175413). We used published primers (Temesgen et al., 2001) 8898F (ATG GGG GTG CAG AA C) and 8898R (GGG ATG GCA ACA ACA AAA A).

For both nuclear loci, 40 µl PCR reactions contained ca. 50 ng of DNA template, 0.4 mM each primer, 0.2 mM dNTPs, 0.1 µg/µl BSA, and 2 units of *Taq* polymerase in supplied buffer (Fisher Scientific, Pittsburgh, PA, USA). Reactions contained 1.5 mM MgCl₂ for LEA-like and 2.0 mM MgCl₂ for WD-40. The thermal cycler program preheated at 80°C for 2 min and then ran 35 cycles of denaturing for 1 min at 94°C, annealing for 1 min, and extending for 1.5 min at 72°C, with a 10 min final extension at 72°C. Optimized PCR annealing temperature was 55°C for LEA-like and 60°C for WD-40.

### 2.3. Chloroplast locus

The *tmGuc* intron was PCR-amplified using published primers 3'trnG (GTA GCG GGA ATC GAA CCC GCA TC) and 5'trnG2 (GCC GGT ATA GAT TAG TGG TAA AA) (Shaw et al., 2005). The *Pinus* amplicon aligns with positions 8857 through 9610 of the *P. thunbergii* chloroplast sequence (GenBank Accession No. D17510; Liston et al., 2007). PCR used 20 µl reactions with ca. 30 ng of DNA template, 0.2 mM each primer, 0.2 mM dNTPs, 1.5 mM MgCl₂, 0.1 µg/µl BSA, and 1 unit of *Taq* polymerase in supplied buffer (Fisher Scientific, Pittsburgh, PA, USA). The thermocycler program preheated at 80°C for 2 min and then ran 35 cycles of denaturing for 1 min at 95°C, annealing for 1 min, and extending for 1 min at 72°C, with a 10 min final extension at 72°C. Optimized PCR annealing temperature was 55°C for LEA-like and 60°C for WD-40.

### 2.4. Sequencing

Products were cleaned with an ethanol precipitation, sequenced using BigDye® v. 3.1 (Applied Biosystems, Foster City, CA, USA) and visualized on an Applied Biosystems 3730 Genetic Analyzer. DNA from single seeds was insufficient for direct PCR in two accessions. These samples were pre-amplified using whole-genome multiple displacement following the protocol of Hosono et al. (2003) with
Table 2
Plants sampled: geographic source, generalized map location (see Fig. 1), specimen voucher. and GenBank accessions. Herbarium abbreviations follow Holmgren and Holmgren (1998).

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Collecting locality</th>
<th>Map No.</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Accession/specimen voucher</th>
<th>GenBank Accession Nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. arizonica var. arizonica</em></td>
<td>USA: NM: Silver City</td>
<td>17</td>
<td>33.13°N</td>
<td>-108.00°W</td>
<td>ARIZ01/Reheford s. (OSC)</td>
<td>FJ395915</td>
</tr>
<tr>
<td></td>
<td>USA: AZ: Saguaro National Park</td>
<td>18</td>
<td>32.20°N</td>
<td>-110.53°W</td>
<td>ARIZ02/Reheford s. (OSC)</td>
<td>FJ395916</td>
</tr>
</tbody>
</table>

| *P. arizonica var. stormiae* | México: COAH: Pala | 21      | 25.50°N  | -102.50°W | ARIZ03/Henrichson 24294 (OSC) | FJ395917 |


| *P. coulteri* | USA: CA: Sania Barbara | 12      | 34.92°N  | -116.92°W | COUL01/Winsa s.n (RSA) | FJ395920 |
|              | USA: CA: San Gabriel Mts. | 11      | 34.35°N  | -117.98°W | COUL02/Winsa s.n (RSA) | FJ395921 |

| *P. devoniana* | México: HGO: Huehuac | 23      | 20.20°N  | -98.60°W  | DEV001/German s. (OSC) | FJ395923 |


| *P. durangensis* | México: SIN: Concordia | 22      | 23.49°N  | -105.85°W | COOPOP1/Inst. Forest Genetics s.n. (OSC) | FJ395920 |

| *P. engelmannii* | México: CHIH: San Juanito | 20      | 28.00°N  | -107.58°W | DEV001/Hernandez s. (OSC) | FJ395927 |

| *P. engelmannii* | USA: AZ: Cave Creek | 20      | 31.72°N  | -110.70°W | ENG01/Hernandez s. (OSC) | FJ395928 |

| *P. engelmannii* | USA: AZ: Florida Canyon | 18      | 31.73°N  | -110.83°W | ENG02/Hernandez s. (OSC) | FJ395929 |

| *P. hartwegii* | Guatemala: Quetzaltenango | 32      | 14.08°N  | -81.84°W  | DON001/Sander s. (OSC) | FJ395930 |

| *P. hartwegii* | México: MEX: Tamascaltepec | 26      | 19.12°N  | -89.76°W  | HART01/German & Sherrwood 445 (OSC) | FJ395931 |

| *P. jeffreyi* | México: GRO: Yerba Santa | 28      | 17.52°N  | -99.96°W  | HART07/Sherwood 1016 (OSC) | FJ395913 |

| *P. jeffreyi* | USA: CA: San Gabriel Mts. | 14      | 34.03°N  | -117.92°W | JEFF01/Williams s. et al. (OSC) | FJ395934 |

| *P. jeffreyi* | USA: CA: Bishop | 10      | 37.37°N  | -118.39°W | JEFF04/Reheford s. (OSC) | FJ395935 |

| *P. jeffreyi* | USA: CA: Susanville | 6       | 40.30°N  | -120.87°W | JEFF06/Law s. (OSC) | FJ395936 |

| *P. jeffreyi* | USA: CA: Warner Mts. | 4       | 41.17°N  | -120.28°W | JEFF12/Willyard 1018 (OSC) | FJ395937 |

| *P. jeffreyi* | USA: NV: Mt. Rose | 8       | 39.33°N  | -119.88°W | JEFF13/Willyard 1019 (OSC) | FJ395938 |

| *P. jeffreyi* | USA: NV: Reno | 8       | 39.24°N  | -119.84°W | JEFF14/Willyard 1020 (OSC) | FJ395939 |

| *P. jeffreyi* | USA: OR: Abert Rim | 4       | 42.38°N  | -120.23°W | WASH02/Willyard 1007 (OSC) | FJ395940 |

| *P. jeffreyi* | USA: OR: Blue Mtns. | 4       | 42.07°N  | -118.78°W | WASH03/Willyard 1023 (OSC) | FJ395941 |

| *P. jeffreyi* | USA: UT: Price Canyon | 7       | 39.77°N  | -110.92°W | WASH04/McArthur s. (OSC) | FJ395942 |

| *P. jeffreyi* | Honduras: COM: Minas de Oro | 27      | 17.92°N  | -57.83°W  | MAX001/Hernandez s. (OSC) | FJ395933 |

| *P. montezumae* | Guatemala: Huatulco | 30      | 15.47°N  | -90.37°W  | MAX05/Sobrados s. (OSC) | FJ395930 |

| *P. montezumae* | México: HGO: Mapicac | 23      | 20.11°N  | -86.01°W  | MON02/German 416 (MEXU) | FJ395941 |

| *P. montezumae* | Guatemala: Hue | 31      | 15.22°N  | -91.87°W  | MON02/Sobrados s. (OSC) | FJ395942 |

| *P. montezumae* | Bolivia: Chaco | 28      | 17.65°N  | -99.84°W  | PSEU04/Sherwood 1010 (OSC) | FJ395903 |

| *P. montezumae* | México: GRO: Filo de Caballos | 8       | 40.73°N  | -122.94°W | SAB01/Sherwood 1001 (OSC) | FJ395940 |

| *P. psuedoaroxylepis* | USA: CA: Weaverville | 5       | 39.18°N  | -122.70°W | SAB02/Biren s. (OSC) | FJ395930 |

| *P. psuedoaroxylepis* | USA: CA: Redding | 5       | 40.55°N  | -122.46°W | SAB03/Willyard 984 (OSC) | FJ395931 |

| *P. psuedoaroxylepis* | USA: California: Santa Rosa | 12      | 23.95°N  | -120.10°W | TOBO01/Tobin 1236 (OSC) | FJ395932 |

| *P. psuedoaroxylepis* | USA: California: Santa Rosa | 12      | 23.95°N  | -120.10°W | TOBO01/Tobin 1236 (OSC) | FJ395932 |

| *P. psuedoaroxylepis* | USA: California: Santa Rosa | 12      | 23.95°N  | -120.10°W | TOBO01/Tobin 1236 (OSC) | FJ395932 |

| *P. psuedoaroxylepis* | USA: California: Santa Rosa | 12      | 23.95°N  | -120.10°W | TOBO01/Tobin 1236 (OSC) | FJ395932 |

| *P. psuedoaroxylepis* | USA: California: Santa Rosa | 12      | 23.95°N  | -120.10°W | TOBO01/Tobin 1236 (OSC) | FJ395932 |

| *P. psuedoaroxylepis* | USA: California: Santa Rosa | 12      | 23.95°N  | -120.10°W | TOBO01/Tobin 1236 (OSC) | FJ395932 |

| *P. psuedoaroxylepis* | USA: California: Santa Rosa | 12      | 23.95°N  | -120.10°W | TOBO01/Tobin 1236 (OSC) | FJ395932 |
phi29 DNA polymerase, pyrophosphatase (New England Biolabs, Ipswitch, MA, USA), and random hexamer primers (Operon, Huntsville, AL, USA). The whole-genome product was used as template for PCR. For five samples, DNA was isolated from an excised gel band with Ultra Clean\textsuperscript{TM} DNA purification kit (MoBio Laboratories, Carlsbad, CA) prior to sequencing. Nucleotide sequences are available in GenBank (Table 2).

2.5. Analysis of nuclear loci

Forward and reverse nucleotide reads were assembled using CodonCode (vers. 1, 4, 6; CodonCode Corporation, Dedham, MA) and edited by hand to create a consensus sequence. Alignments were made by eye to minimize the number of inferred indels. Each locus was analyzed independently. For the Bayesian analysis, a nucleotide substitution model was selected with the Akaike Information Criterion (AIC) using MrModeltest (vers. 2.0; Nylander et al., 2004). Gaps were treated as missing data in the nucleotide partition and coded as present/absent with the simple indel coding method (Simmons and Ochoterena, 2000) using SeqState (vers. 1.4; Muller, 2005). Indel characters were analyzed using an equal-rate binary model. Three partitions were defined: nonsynonymous (approximated with 1st and 2nd codon positions of inferred exons); synonymous (3rd codon positions plus noncoding sequences); and indel characters. We assessed the usefulness of partitions with AIC scores (Akaike, 1974) and a comparison of the number of supported nodes.

We performed two runs using MrBayes (vers. 3.1; Ronquist and Huelsenbeck, 2003) for each locus. Each run used four simultaneous chains and 10 million generations of Metropolis-coupled Monte Carlo simulations, sampling every 1000 generations to save 10,000 trees per run, with default settings for chain heating and rates allowed to vary by partition. We assessed convergence and chose the number of samples to discard as burn-in based on stationarity of a plot of the generation versus log-likelihood for each run. We also compared plots of tree distances and split frequencies.
between two runs using the Comparetree function in MrBayes. One majority-rule consensus tree was built for each locus by combining trees generated by two runs, discarding the first 1000 trees from each run. Branch lengths were estimated by averaging across all retained trees. Nodes with less than 0.95 posterior probabilities were collapsed. Alignments and trees are available at TreeBase (study accession number S2297; matrix accession number M4362).

Each nuclear locus was also analyzed with the parsimony criterion using PAUP (vers. 4.0b10; Swofford, 2002). The heuristic search used stepwise addition of starting trees, 1000 random additions, with 50,000 trees retained, and tree-bisection-reconnection branch-swapping. Nonparametric bootstrap was performed with 100 replicates, holding 50 trees each for a random addition of 1000 replicates. Strict consensus trees were compared with Bayesian consensus trees and Bayesian-supported nodes with parsimony bootstraps less than 50 were collapsed.

Because phylogenetic models assume a hierarchical, bifurcating tree that may not apply to these lineages, we explored an alternative network method that allows reticulate evolution. Nuclear networks were created with neighbor-net (Huson and Bryant, 2006) using SplitsTree (vers. 4.8). For distance calculations, we chose the nucleotide substitution model favored by A. Willyard et al. (2009) using the formula: 1.665 x 2^N; (N = number of years for each species until allelic monophyly is more than paraphyly using the formula: 1.665 x 2^N).

As another coarse estimate of the expectation that gene tree topologies could arise through random lineage sorting, we simulated 1000 trees for each of three levels of effective population size (30,100, and 200 x 10^3) using the option in Mesquite (vers.2, beta 2; Maddison and Maddison, 2006) to generate gene trees within a species tree using a simple coalescence model (i.e. for a neutral gene and a constant population size). For this test, we used a species tree that unites two polytomies (i.e. four Sabiniae and 13 Ponderosae 5.5). The symmetric, distance (Penny and Hendy, 1985) was calculated in PAUP for three 'clouds' of 1000 simulated trees, and for each consensus gene tree (with poorly-supported nodes collapsed as described above) against each cloud of simulated trees. We compared the distribution of symmetric distances for each gene tree to each cloud versus the distribution within each cloud.

2.6. Analysis of chloroplast locus

Nucleotide sequences of trnG were aligned by hand and haplotype networks were created using median joining (Bandelt et al., 1999) in Network version 4.5.0.0, www.fluxus-engineering.com.

3. Results

3.1. Nuclear alignments

Sequences from 45 individuals representing 17 species for LEA-like were aligned with a length of 1630 bps and 3.05% missing data. The inferred intron varied from 837 to 1515 bps. The simple indel coding method inferred 71 indel characters, 33 shared and 38 singlтонs. Sequences from 51 individuals representing 16 species for WD-40 aligned across 1182 bps with 4.43% missing data. The length of the inferred intronic regions varied from 1000 to 1116 bps. We coded 18 indel characters for WD-40, 14 shared and four singlтонs. Alignment of sequences from 49 individuals representing 16 species for the trnG intron required one indel in a mononucleotide repeat, for an aligned length of 722 bps.

The LEA-like sequence for P. maximinoi (Oaxaca) is highly divergent, resolving as sister to the remaining Patijerosaceae on the gene tree and yielding a 8_6 in relation to the Honduras accession nearly three SO from the mean (data not shown). Both accessions of P. maximinoi were verified by repeated amplification and resequencing. High intraspecific divergence in both nuclear loci for P. montezumae (data not shown) was also verified by repeated amplification and resequencing. The LEA-like sequence for P. sabiniana (Clearlake) contains several small indels in the inferred exon, suggesting that our PCR amplicon may be a pseudogene. We chose to retain this sample in our analyses.

3.2. Bayesian phylogeny

The AIC favored GTR+ G for LEA-like and HKY+ G for WD-40. For both loci, two independent Bayesian runs yielded majority-rule consensus trees with identical topologies (Figs.2 and 3). For two runs, the average standard deviations of split frequencies were 0.003840 and 0.003971 for LEA-like and WD-40 loci, respectively.

Analyses with and without indel coding resulted in identical topologies, differing only in posterior probabilities; some nodes were moved above or below our threshold of 0.95 for well-supported nodes. Inclusion of indel coding resolved more deep nodes for LEA-like and more highly-supported nodes for WD-40. Topologies were identical and branch lengths were nearly identical for both partitioning schemes for each locus. For LEA-like (53 exon-in-bps), the two-partition model (nucleotides and indel-codes) was preferred, but for WD-40 (using 103 bps from the first exon), three partitions (synonymous, non synonymous, and indel coding) per-
formed better. We present only trees inferred with the inclusion of indel coding and the preferred partition.

3.3. Parsimony phylogeny

The parsimony strict consensus gene trees inferred similar topologies to their Bayesian counterparts except that in WD-40, eight nodes with 0.95 or greater posterior probability received bootstrap support below 50. We collapsed these nodes as well. Based on this criterion, two nodes would collapse in LEA-like. However, one node with a low bootstrap proportion (highlighted with a star in Fig. 2) received a Bayesian posterior probability of 0.98. Because the low support from parsimony may be an artifact, we elected to show this node in Fig. 2.
3.4. Nuclear gene trees and networks

The nuclear gene trees (Figs. 2 and 3) contain numerous polytomies. If the basal reticulations on the LEA-like and WD-40 networks (Figs. 4 and 5) are viewed as ambiguities rather than lateral transfers, then these networks appear similar to their respective gene trees. A LEA-like network that excludes \textit{P. maximinoi} (Oaxaca) displayed a similar level of resolution (results not shown). LEA-like resolves \textit{P. maximinoi} (Oaxaca) as an outlier to two clades: one contains \textit{Ponderosae s.s.}, the other with the four \textit{Sabinianae} (including \textit{P. jeffreyi} as expected; Figs. 2 and 4). In contrast, WD-40 resolves three clades (Figs. 3 and 5). One unites all \textit{P. arizonica} samples with \textit{P. ponderosa var. scopulorum}, \textit{P. montezumae} (Hidalgo), and \textit{P. maximinoi} (Honduras). The second clade contains the remaining \textit{Ponderosae s.s.} plus a derived, monophyletic grouping of all three samples of \textit{P. coulteri}. The third clade contains all Sabin-
Fig. 4. LEA-like neighbor-net network created using maximum likelihood distances. Oval highlights clade that contains four Sabiniae species.

Fig. 5. WD-40 neighbor-net network created using maximum likelihood distances. Ovals highlight two clades that contain the four Sabiniae species.
ianae except *P. coulteri*, and also contains five accessions classified in *Ponderosa s.s.*: two *P. ponderosa*, and three *P. washoensis*.

3.5. Chloroplast haplotype network

We excluded position 492 in the *trnG* alignment due to ambiguities in several sequences. Six phylogenetically informative substitutions (positions 91, 180, 548, 563, 568, and 604 in the alignment) yield six haplotypes that differ by one or two substitutions: CTCGAC (OUTGROUP); CTCGTC (MAIN SABINIANAE); ATCTTC (COULTER); ATTTTC (NORTHERN SIERRA MADRE); CTGTC (MEXICAN); and CCTGTA (PONDEROSA) (Fig. 6). The OUTGROUP haplotype differs from MAIN SABINIANAE by one substitution (position 568), and the PONDEROSA haplotype differs from MEXICAN by two unique substitutions (positions 180 and 604), but the main cycle of the network cannot be automatically resolved into a tree because two alternate three-step paths (both involving positions 91, 548, and 563) connect MAIN SABINIANAE and NORTHERN SIERRAMADRE haplotypes with either COULTER or MEXICAN as the intermediate. Plotting *trnG* haplotypes on nuclear gene trees (Figs. 2 and 3) suggests a substantially different chloroplast lineage. However, the *trnG* network cycle can be arbitrarily broken (Fig. 6b) to yield a topology with a Sabinianae-Ponderosa *s.s.* divergence similar to that recovered from cpDNA phylogeny (Gernandt et al., 2009).

For 13 species, only one chloroplast haplotype was observed. However, two accessions of *P. arizonica var.* arizonica carry the NORTHERN SIERRAMADRE, while *P. arizonica var.* stonniae (Coahuila) shares the MEXICAN haplotype. *Pinus maximinoi* (Honduras) and *Pinus washoensis* (Babbitt Peak) share the MAIN SABINIANAE haplotype.

3.6. Recombination

Excluding the divergent *P. maximinoi* (Oaxaca) accession and using alignments with all gaps removed, the Phi test finds significant evidence for recombination for *LEA-like* (*P* = 0.037), but not for *WD-40* (*P* = 0.915). With all missing data removed from *LEA-like*, the Phi test is not significant (*P* = 0.376).

3.7. Neutrality

After correcting for multiple tests, no *intraspecific* test revealed a significant departure from neutrality at either locus. Negative departure is indicated for *LEA-like* at all *interspecific* tests. 

---

Fig. 6. (a) Chloroplast *trnG* haplotype network. Single and double lines represent one and two nucleotide substitutions, respectively, at six aligned positions: 91, 180, 548, 563, 568, and 604. Number of accessions in parenthesis. Species with more than one haplotype are in bold. (b) Arbitrarily resolved network (see Section 3.5).
(\(P < 0.02\)), consistent with positive selection. However, these tests are not significant when all missing data are removed.

3.8. Time to monophyly

For three species with five or more samples, we used intraspecific diversity to infer years until allelic monophyly is more likely than paraphyly. These calculations suggest that \(N\) ranges from ca. 23 x 10^3 to 166 x 10^3 and years to coalesce from ca. 4 to 28 million years (Table 3).

3.9. Coalescent simulations

Distributions of symmetric distances within a 1000-tree cloud of simulated trees are nearly identical at three levels of \(N\): (Fig. 7a). The WD-40 consensus gene tree differs more from a cloud of simulated trees (symmetric distances range from 68 to 70) than does the LEA-like consensus gene tree (symmetric distances range from 64 to 66; Fig. 7b). However, ranges of mean distances from either consensus gene tree to the cloud (Fig. 7b) are far below the range of distances within each cloud of trees (94-102; Fig. 7a).

4. Discussion

Our tests did not reveal evidence for genetic recombination or for a departure from neutrality. Thus, neither of these mechanisms can explain the lack of monophyletic species observed across the nuclear gene trees. Projections based on intraspecific genetic diversity and on coalescent simulations hint that the retention of ancestral polymorphism may explain most incongruence in this Ponderosae data set. Mean coalescence times for two species (P. ponderosa and P. washoensis; Table 3) approach the inferred stem age (about 15 million years) for the entire 17-species subsection and all three species meet or exceed the inferred crown age (about 5 million years) (Willyard et al., 2007), suggesting that allelic monophyly may be unlikely for many loci across their nuclear genomes. Similarly, coalescent simulations for our 17-species phylogenies reveal that tree-to-tree distances within a cloud of simulated trees are large enough to contain the distances of each gene tree to the cloud. Thus, our phylogenetic trees combine two features which increase the probability of incongruence: numerous tips and species with large effective population sizes. For these phylogenies, reticulate evolution need not be invoked to explain most instances of incongruence. In some data sets, the removal of a putative hybrid allele can improve the resolution. For our Ponderosae gene trees, excluding the divergent P. maximinoi (Oaxaca) accession yields a nearly identical topology (results not shown).

A lack of allelic monophyly was also observed in a Ponderosae phylogeny based on different taxonomic sampling that used sequences of chloroplast noncoding regions (Gernandt et al., 2009). Despite their morphological and ecological distinctiveness, a pattern of molecular polyphyly may be expected for these species because effective population sizes are large relative to the number of generations since divergence. In the Ponderosae, this pattern extends rather deeply into the gene trees, and none of the previously proposed subdivisions within Ponderosae s.s. (Table 1) are resolved in either gene tree. This suggests that these groupings are sufficiently young that incomplete lineage sorting could explain conflicting placements within Ponderosae s.s.

However, we found support for the traditional Sabinioidae-Ponderosae s.s. clades (Figs. 2-5 and 6b). Across this deep node, conflicts between three independent genomic regions are less likely to be attributable to incomplete lineage sorting, and can thus be used to identify potential examples of reticulate evolution. The most dramatic example of incongruence in our data set is the resolution of a monophyletic P. coulteri clade within Ponderosae s.s. in WD-40 (Figs. 3 and 5). In addition, five Ponderosae s.s. resolve with Sabinioidae in WD-40 (Figs. 3 and 5) and two Ponderosae s.s. share the MAINSABINIANAE chloroplast haplotype (Fig. 6).

4.1. Pinus coulteri

The very distinctive Coulter pine is allied with two other California big-cone pines (P. sabinioides and P. torreyana; Price et al., 1998). Unique allozyme alleles are observed in some P. coulteri populations, and hybridization with P. jeffreyi was proposed as a potential source for these alleles (Ledig, 2000). In the present study, three accessions of P. coulteri resolve with Ponderosae s.s. in WD-40 as a monophyletic clade with a moderate branch length. At this level of sampling (three P. coulteri; 48 others), monophyly due to random branching is highly unlikely (\(P < 0.01\); Rosenberg, 2007). Pinus coulteri's intraspecific diversity can be roughly
approximated using the mean of two loci for three accessions in this study as $8n = 0.00729$. This diversity is relatively low, suggesting that P. coulteri is unlikely to have maintained ancient WD-40 alleles across its genome. However, differential retention of ancestral alleles can occur due to locus-specific effects during a genetic bottleneck.

The WD-40 gene tree topology could be explained by hybridization. Artificial and a few putative natural P. coulteri x P Jeffreyi hybrids have been viewed as crossing the Sabinianae-Ponderosae s.s. division (Critchfield, 1966), but our results confirm that P. coulteri and P. Jeffreyi are rather closely related members of the Sabinianae. Pinus coulteri has not been successfully crossed with any extant Ponderosae s.s. species (Critchfield, 1966). Further, the topology argues for 'ancient', rather than 'recent' introgression because the node representing the most recent common ancestor of the "ponderosae s.s.-style" P. coulteri alleles is relatively deep in the WD-40 gene tree.

We propose that P. coulteri might have retained WD-40 alleles from an unknown Ponderosae s.s. parent and that this could account for the unique allozyme alleles as well. Two alternatives can be envisioned: (i) introgression after P. cauten had speciated; or (ii) P. cauten arising as a diplloid hybrid species. In either case, hybridization was likely followed by one or more bottlenecks that purged part of the genome. Pinus coulteri's massive and well-armored cones may have contributed to reproductive isolation. Adaptation to different seed predators (Borchert, 1985) might have allowed P. coulteri to move into hotter and drier foothill habitats than either of its parents (or its introgressing partner), providing reproductive isolation by allopatry. Scenarios of either introgression or hybrid speciation fit a theoretical framework for hybridization as an evolutionary stimulus (Anderson and Stebbins, 1954) and we are conducting further sampling to address the origin of P. coulteri.

4.2. Pinus ponderosa and P. washoensis

Throughout much of its geographic range, P. Jeffreyi meets lower-altitude P. ponderosa or higher-altitude P. washoensis at the margins of each species' zone. Morphological traits support a few natural hybrids between P. Jeffreyi and P. ponderosa (Haller, 1962), and the nature of this putative introgression is more interesting in light of our current understanding that P. Jeffreyi is part of the Sabiniana lineage. In LEA-like, all accessions of P. ponderosa and P. washoensis resolve as expected within Ponderosae s.s. Our trnG network supports a transfer of the MAINSABINIANAchloroplast haplotype into one individual of P. washoensis (Babbitt Peak), which harbors a haplotype that is three substitutions removed from its conspecifics (Fig. 6). WD-40 resolves the Babbitt Peak accession plus two other P. washoensis and two P. ponderosa individuals with the Sabinianae. The remaining 'misplaced' individuals carry the expected PONDEROSA chloroplast haplotype (Fig. 6a). This supports a low level of ongoing introgression between P. Jeffreyi and P. ponderosa and between P. Jeffreyi and P. washoensis.

Studies that place a P. washoensis allele sister to P. Jeffreyi or sister to P. sabini (Patten and Brunsfeld, 2002; Prager et al., 1976), or that place P. Jeffreyi within Ponderosae s.s. (Eckert and Hall, 2006) may be footprints of introgression. We hypothesize that reticulate ancestry might contribute to a preference for high-altitude sites in P. washoensis (Haller, 1965; Mason and Stockwell, 1945). All species of Sabinianae grow in colder climates or on ultramafic soil (P. Jeffreyi; Haller, 1962) or in more arid habitats (P. coulteri, P. sabini, and P. torreyana) than P. ponderosa. Perhaps high-altitude Washo pines represent a lineage that has retained more of the introgressed Sabinianae genome because of traits that are more useful in harsh climates. Additional information from the maternally inherited mitochondrial genome (Godbout et al., 2005) may be useful to detect recent hybridization, and we are currently assessing introgression in P. washoensis using nuclear microsatellite loci for population-level samples.

The lack of reciprocal monophyly for P. ponderosa and P. washoensis and their shared chloroplast haplotype might be interpreted as support for the conclusion that the narrowly endemic P. washoensis is synonymous with the wide-ranging P. ponderosa (Brayshaw, 1997; Lauria, 1997; Niebling and Conkle, 1990; Rehfeldt, 1999b). However, we note that none of the Ponderosae species achieve monophyly in both of our gene trees.

4.3. Incomplete lineage sorting or reticulation?

Anomalous results for P. maximinoi and P. montezumae could be explained by incomplete lineage sorting, but there are indications that future studies designed to detect reticulate ancestry may be fruitful. Pinus maximinoi encompasses an unusual range of genetic diversity across its wide distribution. In our study, P. maximinoi (Oaxaca) is highly divergent from other LEA-like sequences, but this accession is unremarkable in WD-40, and it is the P. maximinoi (Honduras) allele that resolves unexpectedly in WD-40. Artificial hybrids can be created between P. maximinoi and P. taeda (subsect., Australes; Dvorak et al., 2000). These clues suggest that our results could be due to misidentifications, cryptic species, or introgression (perhaps even outside its taxonomic subsection).

Although our two P. montezumae accessions do not sort across the Sabinianae-Ponderosae s.s. divergence, the sequences resolve in very different locations on both nuclear gene trees despite sharing a trnG haplotype (Fig. 6a). Because P. montezumae can be difficult to distinguish from P. devoniata, P. hartwegii, and P. pseudostrobus (Perry, 1991), misidentification cannot be ruled out. Alternatively, the extreme variability may be attributed to the complex patterns of interspecific hybridization reported for P. montezumae with P. pseudostrobus and with P. hartwegii (Delgado et al., 2007; Matos and Schaal, 2000). Recent introgression would not explain our placement of the Hidalgo accession because the geographic ranges of the species that P. montezumae joins (P. durangensis and P. anzonica) do not extend into Hidalgo. Incomplete lineage sorting could create the patterns we observe, but in light of strong evidence for ongoing localized hybridization, our results do not exclude ancient introgression between P. montezumae and previously sympatric species.

4.4. Species tree

Because the gene trees inferred by LEA-like and WD-40 are different, we do not present a combined species phylogeny. Our two nuclear gene trees are too incongruent to calculate 'Concordance Factors' for internal nodes (Baum, 2007) using BUCKY (Ane et al., 2007). We note that any quantitative assessment of discordance is obscured by the astronomical number of potential topologies possible for a 53-tip tree (more than $7 \times 10^8$). Felsenstein, 2004). Although networks help visualize the extent to which gene trees are incongruent (Figs. 4 and 5; McNearn and Lockhart, 2006), they do not help distinguish between incomplete-lineage sorting and reticulation. As a further example, the numerous reticulations suggested by T-Rex (data not shown; Makarenkov and Lapointe, 2004) are difficult to interpret because they are inferred on neighbor-joining trees that resolve different nodes than our Bayesian and Parsimony gene trees.

We suggest that the low resolution observed here and in other species-level Ponderosae phylogenies (Gernandt et al., 2005, 2009) is reflective of the recency of species divergence. Despite the wide range of mutation rates among Pinus loci (Willeyard et al., 2007), this low resolution is likely to be mirrored across nuclear and chloroplast genomes.
5. Conclusions

The inference of hybrid ancestry in natural species is a challenging but worthwhile endeavor. Comparing phylogenetic hypotheses based on independent genomic regions is an important method to detect potential reticulations, but requires the exclusion of other causes of incongruence. In this data set, we did not detect genetic recombination or a departure from neutrality. However, our rough calculations suggest that incomplete lineage sorting is a major source of the incongruence observed at all but the deepest nodes of the *Ponderosae* tree. Phylogenetic inferences like the one presented here for reticulate evolution in *P. coulteri* and among *P. jeffreyi, P. ponderosa*, and *P. washoensis* do not meet the standard of resynthesizing a hybrid species (Rieseberg et al., 2003). Further, our imposition of a tree-like hierarchy on obviously not-quite-tree-like relationships is perilous. Nevertheless, if these methods are used and interpreted cautiously, valuable clues can be gleaned about potential examples of reticulate evolution that are worthy of future study. This can be accomplished without the large number of loci that will apparently be required (Maddison and Knowles, 2006) to infer species relationships from multiple gene trees. In particular, a phylogenetic overview like the one created here from independent genomic regions is crucial to the identification of the potential gene-flow 'players', which may be taxonomically distant as well as geographically remote in their current distributions. It is clear from our results that species-level diagnosis based on either single-accession sampling or single-locus sampling in *Pinus* is inadequate.

Acknowledgments

We thank Mariah Parker-De'Eniks, Anthony Shireman, Carla Streng, and Kathleen Farrell for laboratory work. The many collectors who generously contributed plant material are gratefully acknowledged in our GenBank records. We are indebted to David Gernandt for sharing unpublished information as well as helpful comments. We thank Tim Cowman, John R. Haller, Mathew Horning, Susan Huber, Ronald Larmer, Molly Nepokroeff, Bernard Pfeil, and Thomas Rohl for providing plant material who generously contributed plant material are gratefully acknowledged. We thank William Cowman, John R. Haller, Mathew Horning, Susan Huber, Ronald Larmer, Molly Nepokroeff, Bernard Pfeil, and Thomas Rohl for providing plant material who generously contributed plant material.

Appendix A, Supplementary data


References


(Pinaceae). Evolution 30, 637-649.

Lind, W.J., 1958. The backcross hybrid Jeffrey x (jeffrey x Coulter) pine. J, For, 56,
840-842.
phylogenetic, analysis enhances interspecific phylogeographic, inference. A case
study in Pinus lambertiana. Mol, Ecol, 16, 3926-3937.
Pinus (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer
Loudon, J., 1838. Arboretum et fruticetium britannicum. James Ridgway and Sons,
London.

of a diploid hybrid pine Pinus densata inferred from the nucleotide variation at seven
gene loci. Mol, Biol, Evo1, 23, 807-816.
Maddison, W.P., Knowles, L.L. 2006. Inferring phylogeny despite incomplete lineage
inferring phylogenies from complete distance matrices. Bioinformatics 20,
2113-2121.
Martin, D.P., Williamson, C, Posada, D., 2005. RDP2: recombination detection and
analysis from sequence alignments. Bioinformatics 21, 260-262.
Martinez, M., 1948. Los Pinos Mexicanos, Segunda Edicion. Ediciones Betas, Mexico
City, Mexico.
8,61-63.

of a diploid hybrid pine Pinus densata inferred from the nucleotide variation at seven
gene loci. Mol, Biol, Evo1, 23, 807-816.
Maddison, W.P., Knowles, L.L. 2006. Inferring phylogeny despite incomplete lineage
inferring phylogenies from complete distance matrices. Bioinformatics 20,
2113-2121.
Martin, D.P., Williamson, C, Posada, D., 2005. RDP2: recombination detection and
analysis from sequence alignments. Bioinformatics 21, 260-262.
Martinez, M., 1948. Los Pinos Mexicanos, Segunda Edicion. Ediciones Betas, Mexico
City, Mexico.
8,61-63.

of a diploid hybrid pine Pinus densata inferred from the nucleotide variation at seven
gene loci. Mol, Biol, Evo1, 23, 807-816.
Maddison, W.P., Knowles, L.L. 2006. Inferring phylogeny despite incomplete lineage
inferring phylogenies from complete distance matrices. Bioinformatics 20,
2113-2121.
Martin, D.P., Williamson, C, Posada, D., 2005. RDP2: recombination detection and
analysis from sequence alignments. Bioinformatics 21, 260-262.
Martinez, M., 1948. Los Pinos Mexicanos, Segunda Edicion. Ediciones Betas, Mexico
City, Mexico.
8,61-63.