A Set of Plastid Loci for use in Multiplex Fragment Length Genotyping for Intraspecific Variation in *Pinus* (Pinaceae)

Author(s): Austin M. Wofford , Kristen Finch , Adam Bigott , and Ann Willyard
Source: Applications in Plant Sciences, 2(5) 2014.
Published By: Botanical Society of America
DOI: http://dx.doi.org/10.3732/apps.1400002
URL: http://www.bioone.org/doi/full/10.3732/apps.1400002

Applications in Plant Sciences

Applications in Plant Sciences 2014 2(5): 1400002

Application Article

A set of plastid loci for use in multiplex fragment length genotyping for intraspecific variation in *Pinus* (Pinaceae)[1]

Austin M. Wofford[2], Kristen Finch[2,3], Adam Bigott[2], and Ann Willyard[2,4]

[2]Department of Biology, Hendrix College, 1600 Washington Avenue, Conway, Arkansas 72032 USA

- *Premise of the study:* Recently released *Pinus* plastome sequences support characterization of 15 plastid simple sequence repeat (cpSSR) loci originally published for *P. contorta* and *P. thunbergii*. This allows selection of loci for single-tube PCR multiplexed genotyping in any subsection of the genus.
- *Methods:* Unique placement of primers and primer conservation across the genus were investigated, and a set of six loci were selected for single-tube multiplexing. We compared interspecific variation between cpSSRs and nucleotide sequences of *ycf1* and tested intraspecific variation for cpSSRs using 911 samples in the *P. ponderosa* species complex.
- *Results:* The cpSSR loci contain mononucleotide and complex repeats with additional length variation in flanking regions. They are not located in hypervariable regions, and most primers are conserved across the genus. A single PCR per sample multiplexed for six loci yielded 45 alleles in 911 samples.
- *Discussion:* The protocol allows efficient genotyping of many samples. The cpSSR loci are too variable for *Pinus* phylogenies but are useful for the study of genetic structure within and among populations. The multiplex method could easily be extended to other plant groups by choosing primers for cpSSR loci in a plastome alignment for the target group.

Key words: discriminant analysis of principal components; intraspecific taxonomy; microsatellites; *Pinus*; plastid.

Simple sequence repeats (SSRs; microsatellites) have been widely used for nearly two decades to visualize intraspecific genetic variability (Avise, 1994), and SSRs have been used to infer phylogenies in some lineages (Orsini et al., 2004). Because they are highly polymorphic, SSRs are often informative for finer-scale patterns within and among populations (DeFaveri et al., 2013). The fast mutation rate that creates these high levels of polymorphism is widely recognized to increase levels of homoplasy as genetic distance increases (Estoup et al., 2002), although the use of multiple loci may mitigate the impact of homoplasy in a single locus. The genomic location (i.e., coding vs. noncoding) and nucleotide repeat pattern (mononucleotide repeats vs. two or more repeated nucleotides) of a set of SSRs are important factors that affect homoplasy and thus experimental design.

Plastid genomes provide useful markers to infer plant genetic patterns. Plastid data are especially important in *Pinus* L. (Pinaceae) because nuclear genetic markers have been problematic. For example, there has been slow concerted evolution in the nuclear ribosomal DNA internal transcribed spacer (Gernandt et al., 2001), greatly limiting its usefulness. Incomplete lineage sorting in *Pinus* low-copy nuclear loci extends deeply into the tree (Syring et al., 2005). This will likely mean that numerous low-copy nuclear loci will be required to resolve species trees in *Pinus*. Although plastid lineages can suffer from incomplete lineage sorting as well, their faster coalescence times may make them more useful for species-level questions. Because plastids are paternally inherited in Pinaceae (Neale and Sederoff, 1989), plastids potentially track pine pollen flow in contrast to maternal mitochondrial inheritance and biparental nuclear inheritance. Because much of the genetic variation in long-lived forest trees like *Pinus* is contained within rather than among populations (Petit and Hampe, 2006), our long-term goal of delimiting a species complex required a marker that could be efficiently genotyped in many individuals per population so that we could use allele frequencies rather than exemplar sampling. We investigated 15 plastid simple sequence repeat (cpSSR) loci—nine loci based on the *P. thunbergii* Parl. plastid genome (Wakasugi et al., 1994; Vendramin et al., 1996) and six loci designed for *P. contorta* Douglas (Stoehr and Newton, 2002)—that have been genotyped in numerous species and populations of pines and other members of Pinaceae (Echt et al., 1998; Walter and Epperson, 2001; Marshall et al., 2002; Richardson et al., 2002; Robledo-Arnuncio et al., 2005; Godbout et al., 2010; Feng et al., 2011; Jardón-Barbolla et al., 2011).

We clarify the genomic locations of these 15 cpSSRs in a plastome alignment of 107 pine species (Parks et al., 2012), summarize the extent of primer sequence conservation across

[1]Manuscript received 13 January 2014; revision accepted 27 March 2014.

The authors thank A. Duina, J. Finney, D. Gernandt, V. Goremykin, D. Hoose, S. Langer, P. Lea, B. Linz, P. Marquardt, S. Meyers, R. Murray, T. Nguyen, M. Parks, D. Pouncey, C. Rand, B. Schumacher, N. Seagar, J. Smith, K. Spatz, and F. Telewski, and three reviewers. Funding is acknowledged from Hendrix College Odyssey, Arkansas Academy of Sciences (grants to A.M.W., K.F., and A.B.), Beta Beta Beta Foundation (awards to A.M.W. and A.B.), and an Arkansas Department of Education Student Undergraduate Research Fellowship to K.F.

[3]Present address: Conservation Department, Fairchild Tropical Botanic Garden, 10901 Old Cutler Road, Coral Gables, Florida 33156 USA

[4]Author for correspondence: willyard@hendrix.edu

doi:10.3732/apps.1400002

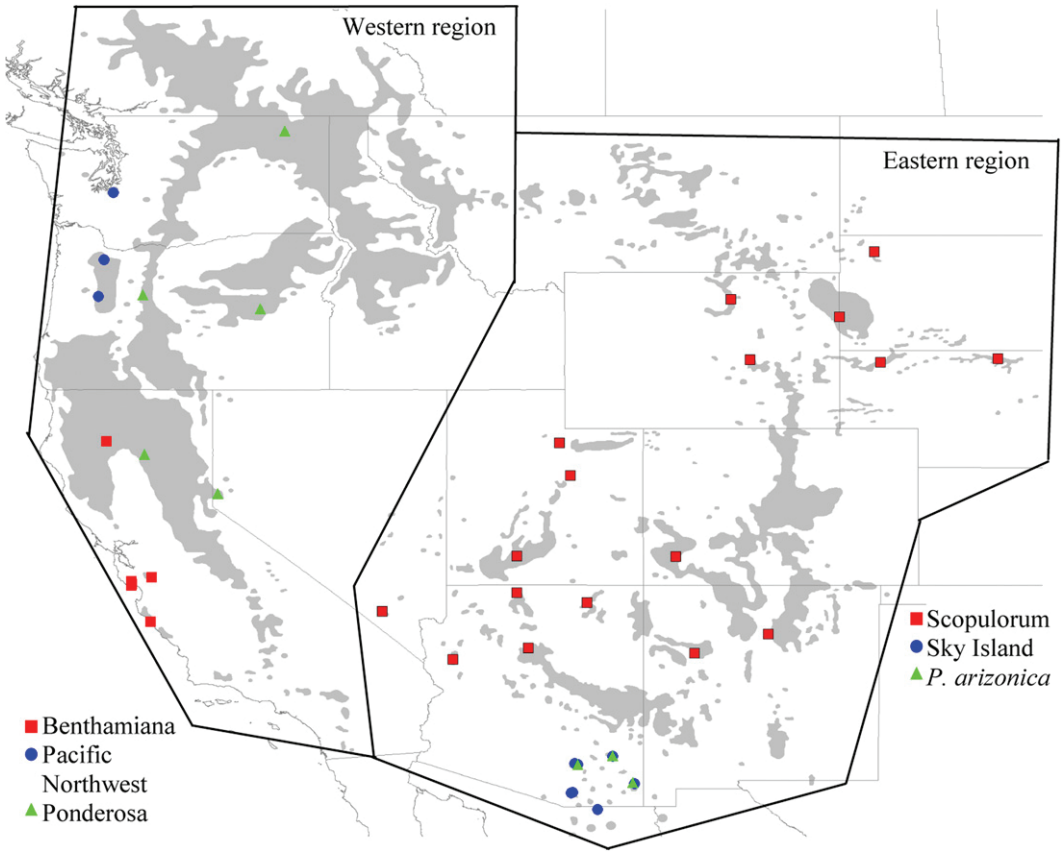

Fig. 1. Geographic locations of 41 populations (Appendix 2) genotyped for Fig. 4 showing their assignment to one of three operational taxonomic units (OTUs) in each region. Areas shaded in light gray represent the distribution of the *P. ponderosa* species complex. In the western region, the Pacific Northwest OTU includes one population from Fort Lewis, Washington, and two from Willamette Valley, Oregon; the Benthamiana OTU includes six populations from western California; the Ponderosa OTU includes five inland populations. In the eastern region, the Sky Island OTU includes seven populations in southeastern Arizona, the *P. arizonica* OTU includes three populations that are partially sympatric with the Sky Island OTU, and the Scopulorum OTU includes 17 allopatric populations.

the genus to allow selection of loci that can be used for any pine species, and investigate whether these loci are in hypervariable regions of the plastome.

To evaluate the most distant relationships for which the cpSSR loci might be useful, we investigated interspecific comparisons. Subsect. *Ponderosae* (sect. *Trifoliae*) was chosen for this test because of our ongoing study of that group. The highly variable *ycf1* locus has been suggested as a useful region for species-level phylogenies (Parks et al., 2012). We compared interspecific information content of the multilocus cpSSR fragment lengths with *ycf1* sequences from the same 15 individuals in subsect. *Ponderosae*.

A set of six nonredundant loci were chosen that can be economically and efficiently amplified in a single-tube multiplexed PCR, and we demonstrate the use of these six cpSSR loci in 911 samples. We also evaluate the impact of homoplasy on these loci using multilocus linkage disequilibrium as another criterion to evaluate their usefulness. For loci on the nonrecombining plastid genome, a finding of significant linkage disequilibrium suggests that the multilocus haplotypes are unlikely to be created by homoplasy (Angioi et al., 2009). Based on preliminary evidence in our own laboratory and on other reports, we hypothesize that two distinctive geographic regions represent divergent lineages within the *P. ponderosa* P. Lawson & C. Lawson species complex (Fig. 1). We hypothesize for the western region (corresponding to *P. ponderosa* var. *ponderosa*) that the Willamette Valley, Oregon, and Fort Lewis, Washington, populations (Pacific Northwest operational taxonomic unit [OTU]) are genetically distinct from the populations of western California (Benthamiana OTU) and/or inland populations (Ponderosa OTU) (Bouffier et al., 2003; Potter et al., 2013). For the eastern region (corresponding to *P. ponderosa* var. *scopulorum* Engelm. in S. Watson), we hypothesize that populations in southeastern Arizona (Sky Island OTU) are distinct from other, mostly allopatric populations of *P. ponderosa* var. *scopulorum* (Scopulorum OTU) and/or from partly sympatric *P. arizonica* Engelm. (Rehfeldt, 1999; Epperson et al., 2009). Patterns among these populations were observed using a method that does not require an assumption of uncorrelated alleles and allows for a priori definition of groups to emphasize among-group rather than within-group variation.

MATERIALS AND METHODS

The published nucleotide sequences for 15 cpSSR primer pairs (Table 1) were located within the aligned plastomes of 107 species of *Pinus* and six Pinaceae outgroups (TreeBase S12640) (Parks et al., 2012). Unique primer matches were confirmed by conducting a BLAST search for each primer sequence within the *P. ponderosa* var. *ponderosa* plastome (GenBank FJ899555).

TABLE 1. Characteristics of 15 *Pinus* cpSSR loci assessed in this study.

Locus	Source species	Primer design	Forward primer	Fluorescent marker used in this study	Forward primer with tag added in this study	Reverse primer
Pc10	*P. contorta*	Stoehr and Newton, 2002	CAGAAGCCCAAGCTTATGGC	VIC	CTAGTTATTGCTCAGCGGTCAGAAGCCCAAGCTTATGGC	CGGATTGATCCTAACCATAC
Pc69	*P. contorta*	Stoehr and Newton, 2002	TTTCGGGCTCCACTGTTATC	FAM	TGTAAAACGACGGCCAGTTTTCGGGCTCCACTGTTATC	CGTACTCAATTTGTTACTAC
Pc987	*P. contorta*	Stoehr and Newton, 2002	ACTGCAAGGAACAGTAGAAC	NED	TACGAGTGCAGCAAGCATACTGCAAGGAACAGTAGAAC	CGGAACGTTTTCTGATGCAC
PcG2R1	*P. contorta*	Stoehr and Newton, 2002	AGATCGGGACAATGTATGCC	PET	CACTGCTTAGAGCGATGCAGATCGGGACAATGTATGCC	TGTCCTATCCATTAGACGAT
PcI1A2	*P. contorta*	Stoehr and Newton, 2002	TTCAAGTCCAGGATAGCCCA	PET	CACTGCTTAGAGCGATGCTTCAAGTCCAGGATAGCCCA	CTACCAACTGAGCTATATCC
PcL2T1	*P. contorta*	Stoehr and Newton, 2002	ACCAATTCCGCCATATCCCC	PET	CACTGCTTAGAGCGATGCACCAATTCCGCCATATCCCC	CTAGGGGAGGATAATAACATTGC
Pt100783	*P. thunbergii*	Vendramin et al., 1996	ATACGTATGTATCCCCTAACTGTCA	FAM	TGTAAAACGACGGCCAGTATACGTATGTATCCCCTAACTGTCA	TCAATTTTTGCCATATCCTGA
Pt107517	*P. thunbergii*	Vendramin et al., 1996	AAAGCTTTATTGCGGCC	VIC	CTAGTTATTGCTCAGCGGTAAAGCTTTATTGCGGCC	ATGGCAGTTCCAAAAAAGC
Pt110048	*P. thunbergii*	Vendramin et al., 1996	TAAGGGGACTAGAGCAGGCTA	NED	TACGAGTGCAGCAAGCATTAAGGGGACTAGAGCAGGCTA	TTCGATATTGAACCTTGGACA
Pt1254	*P. thunbergii*	Vendramin et al., 1996	CAATTGGAATGAGAACAGATACG	FAM	TGTAAAACGACGGCCAGTCAATTGGAATGAGAACAGATACG	TGCGTTGCACTTCGTTATAG
Pt15169	*P. thunbergii*	Vendramin et al., 1996	CTTGGATGGAATAGCAGCC	VIC	CTAGTTATTGCTCAGCGGTCTTGGATGGAATAGCAGCC	GGAAGGGCATTAAGCTCATTA
Pt30204	*P. thunbergii*	Vendramin et al., 1996	TCATAGCGGAAGATCCTCTTT	NED	TACGAGTGCAGCAAGCATTCATAGCGGAAGATCCTCTTT	CGGATTGATCCTAACCATACC
Pt36480	*P. thunbergii*	Vendramin et al., 1996	TTTTGGCTTACAAAATAAAAGAGG	FAM	TGTAAAACGACGGCCAGTTTTTGGCTTACAAAATAAAAGAGG	AAATTCCTAAAGAAGGAAGAGCA
Pt71936	*P. thunbergii*	Vendramin et al., 1996	TTCATTGGAAATACACTAGCCC	VIC	CTAGTTATTGCTCAGCGGTTTCATTGGAAATACACTAGCCC	AAAACCGTACATGAGATTCCC
Pt87268	*P. thunbergii*	Vendramin et al., 1996	GCCAGGGAAAATCGTAGG	NED	TACGAGTGCAGCAAGCATGCCAGGGAAAATCGTAGG	AGACGATTAGACATCCAACCC

We made slight manual adjustments to improve the alignment in areas where cpSSRs were located, and then used annotations for FJ899555 to determine whether the primers, SSR regions, and flanking sequences were coding or noncoding. Using the same plastome alignment, primer conservation was determined for each taxonomic subsection (Gernandt et al., 2009). Primers were regarded as being highly conserved if they had no more than one base position mismatch. Alignments are available on the Dryad Digital Repository (http://doi.org/10.5061/dryad.5nc25; Wofford et al., 2014).

To test if these loci were in hypervariable regions of the plastome, we measured nucleotide variation in the regions immediately surrounding each locus using the same plastome alignment by extracting a 1-kb segment centered on the repeat region. Using the script *sorter.pl* (Goremykin et al., 2010) on the iPlant Discovery Environment (http://www.iplantcollaborative.org), we calculated the observed variability (OV) for each base position. OV calculates a mean of all possible pairwise comparisons, excluding gaps. Mean OV was also calculated for the full plastome alignment. For comparison, we also counted the number of unique amplicon lengths for each locus in the alignment (measuring from the outside of each primer pair) and conducted a Spearman's rank correlation test between the mean OV of the 1-kb segments and the amplicon lengths.

To evaluate interspecific information content, we selected the 14 samples that represent subsect. *Ponderosae* s.s. (Gernandt et al., 2009; Willyard et al., 2009) from the plastome alignment described above and *P. jeffreyi* Balf. (subsect. *Sabinianae*) to serve as the outgroup (Appendix 1). These 15 plastomes were used to compare the information content in the fragment lengths of the 15 cpSSR loci vs. the nucleotide sequences for the highly variable *ycf1* region using median joining haplotype networks with star contraction preprocessing (Network version 4.6; Fluxus Engineering, Suffolk, England).

To improve the PCR multiplex, we removed loci that were monomorphic in early testing (Pt107517), had numerous failures (Pc69, Pc987, Pt1254, Pt15169, Pt36480), or amplified the same region as other cpSSR loci (Pcl1A2, Pt30204; see Results). Six loci (Pc10, PcG2R1, Pcl2T1, Pt100783, Pt71936, and Pt87268) amplified consistently in subsect. *Ponderosae*, were polymorphic, and had lengths that allowed confident four-color genotyping. For intraspecific comparisons, these six cpSSR loci were genotyped for 911 individuals from 41 populations of subsect. *Ponderosae* (Fig. 1; Appendix 2) using a PCR multiplex protocol that integrates fluorescent labels during PCR (Culley et al., 2008; Culley et al., 2013). New forward primers were purchased (Integrated DNA Technologies, Coralville, Iowa, USA) with a unique nucleotide sequence for one of four fluorescent dyes added to the 5′ end of the published primer (Table 2). The same four unique sequences (Table 2) were purchased as fluorescently labeled primers (Life Technologies, Carlsbad, California, USA). A 1-μM primer master mix was created with six forward, six reverse, and four labeled primers in a 1 : 4 : 4 (forward : reverse : labeled) volume ratio to limit forward primers as recommended by the manufacturer's protocol for the Multiplex PCR kit (part number 206143, QIAGEN, Germantown, Maryland, USA). We isolated DNA using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA) according to the manufacturer's protocol except that leaves dried in silica gel were homogenized in QIAGEN AP1 buffer and RNase A using the FastPrep homogenizer (MP Biomedicals, Santa Ana, California, USA) with a ceramic bead and garnet sand in FastPrep tubes, processing three times for 20 s each at 6 m/s. PCR reactions were 10 μL, using 1 μL of 1 mM primer master mix and 1 μL of DNA eluted from the DNeasy procedure. Thermocycler parameters were 15 min at 95°C; 35 cycles of 30 s denaturing at 94°C, 90 s annealing at 58°C, and 90 s extension at 72°C; and a final extension of 10 min at 72°C. PCR success (expecting multiple fuzzy bands because of the six-plex) was determined on 0.8% agarose gels using 2 μL of PCR product with 1 : 1000 SYBR Green loading dye (Sigma-Aldrich, St. Louis, Missouri, USA). PCR products were diluted 1 : 10 and genotyped (University of Missouri DNA Core Facility; ABI 3730xl DNA Analyzer, Life Technologies) with a GS600 LIZ (Life Technologies) size standard.

Table 2. Fluorescently labeled primers for 15 *Pinus* cpSSR loci.

Fluorescent dye	Tag source	Tag sequence[a]
PET	M13B	CACTGCTTAGAGCGATGC
6-FAM	M13	TGTAAAACGACGGCCAGT
NED	M13A	TACGAGTGCAGCAAGCAT
VIC	T7	CTAGTTATTGCTCAGCGGT

[a]Each tag sequence is used as a fluorescently labeled primer and added to the 5′ end of the forward primer.

Linkage disequilibrium was estimated using MultiLocus (version 1.2; http://www.bio.ic.ac.uk/evolve/software/multilocus) and significance was estimated using 100 randomizations. Patterns among predefined OTUs were observed using discriminant analysis of principal components (DAPC) (Jombart et al., 2010). Two separate DAPC analyses were run, one for the western and one for the eastern region, with a priori grouping into three OTUs each (Fig. 1). DAPC and scatter plots of the first two principal components were run using *adegenet* (version 1.3-9.2; Jombart et al., 2010) in R (version 3.0.2; http://www.R-project.org). We used default parameters to place inertia ellipses for each OTU.

RESULTS

We found single locations in the plastome for all 30 primers (Table 3; Fig. 2; http://doi.org/10.5061/dryad.5nc25; Wofford et al., 2014). Two locus pairs were redundant: Pt30204 and Pc10 had the same reverse primer sequence and the forward primer for Pc10 was 50-bp upstream from Pt30204. Thus, they would yield amplicons that encompass the same repeat region. Pt87268 and PcI1A2 also had overlaps that include the same repeat region, despite not having matching primer sequences. Fourteen loci had both primers located in coding regions; primers for Pt71936 were within the *ycf3* intron (Table 3). Fourteen loci had repeat regions located in intergenic spacers or introns. The repeat region for Pt107517 was located entirely within the *rpl32* gene. Repeat motifs for the loci varied. Five were simple mononucleotide repeats, five had two adjacent segments of mononucleotide repeats, and five had complex motifs, including an 11-bp minisatellite in Pc987 and a 10-bp minisatellite in PcL2T1. We also found that indels in flanking regions contribute to the length variation in some loci. Primer conservation was high in all taxonomic subsections for 10 of the 15 primers (Table 3), and opportunities exist to create nearby primers for taxonomic subsections that have diverged (data available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.5nc25; Wofford et al., 2014). As expected from their design in *P. thunbergii* and *P. contorta*, primers were well conserved across subgen. *Pinus* except for Pt71936 in sect. *Trifoliae*. Subgen. *Strobus* had mismatches for primers of four loci: Pc987, Pt100783, Pt1254, and Pt87268.

For the 107 *Pinus* plastomes examined, none of the tested loci were located in hypervariable regions. The mean OV for the entire plastome was 0.0350 (0.1193 SD), while the mean OV for the 15 segments surrounding each locus was 0.0210 (0.0084 SD). For the same 107 *Pinus* plastomes, length variation ranged from four to 37 alleles per locus (Table 3). Pt107517 was monomorphic for the repeat region with only minor length differences in the flanking regions, as expected from its location in a coding region. Pc10, Pt30204, and Pt15169 had the greatest amount of length variation with 37, 35, and 32 unique lengths in 107 species, respectively. Spearman's rank correlation tests (r_s) for the 15 loci showed no significant correlation between mean OV and number of alleles [$r_s(13) = 0.1323$, $P = 0.43$], nor for only the six loci (described below) that we selected for subsect. *Ponderosae* [$r_s(4) = 0.4412$, $P > 0.5$]. Networks based on cpSSRs and *ycf1* nucleotide sequences from the same samples were different (Fig. 3).

Multiplex genotyping in 911 subsect. *Ponderosae* individuals yielded 45 total alleles, with a mean of 7.5 (2.3 SD) alleles per locus (Table 3). Pt71936 was successfully amplified despite minor primer differences. Multilocus linkage disequilibrium was significant ($r_d = 0.95$; $P < 0.01$). Our DAPC analysis included 35 cpSSR alleles in 314 individuals in the western region (Fig. 1). Populations assigned to the Pacific Northwest OTU did not overlap on the scatter plot with the Benthamiana

TABLE 3. Plastid SSR locus characteristics in 107 *Pinus* species and in 911 subsect. *Ponderosae* individuals.

	Characteristics in 107 *Pinus* plastomes											No. of alleles in 911 subsect. *Ponderosae* individuals
	Size (bp)		Genic location[a]					Primer conservation by subsection				
Locus	Min.	Max.	Left flank	Repeat	Right flank	Repeat motif[b]	No. of alleles	*Pinus*	*Trifoliae*	*Parrya*	*Quinquifolia*	
Pc10	147	220	*clpP*	Intergenic →	*rps12*	$(A)_{15}(G)_{14}$	37	yes	yes	yes	yes	8
Pc69	187	198	*rpl14*	Intergenic	*rpI16*	$(A)_{15}$	12	yes	yes	yes	yes	n/a
Pc987	229	327	*psbM*	Intergenic →	*trnD-GCA*	$(TTTGATCTCAT)_5$	15	yes	yes	no	no	n/a
PcG2R1	83	124	*trnG-GCC*	← Intergenic	*trnR-UCU*	$[(A)_{12}(G)_8]_3$	19	yes	yes	yes	yes	5
PcI1A2	89	143	*trnI-GAU*	Intergenic	*trnA-UGC*	$(T)A(T)_{31}$	17	yes	yes	yes	yes	n/a
PcL2T1	229	288	*trnL-UAA*	Intergenic →	*trnT-UGU*	$(CCCCTTCTCT)_9$	18	yes	yes	yes	yes	7
Pt100783	53	120	*ycf1*	Intergenic →	*rps15*	$(T)_7(A)_{17}$	20	yes	yes	no	no	5
Pt107517	95	98	*rpl32*	*rpl32*	*rpl32*	$(T)_{11}$	4	yes	yes	yes	yes	n/a
Pt110048	82	97	*rps12*	← Intergenic →	*rps7*	$(A)_{18}(CA)_4$	14	yes	yes	yes	yes	n/a
Pt1254	61	80	*psbA*	← Intergenic →	*trnK-UUU*	$(T)_{23}$	15	yes	yes	no	no	n/a
Pt15169	92	141 (304[c])	*atpI*	← Intergenic	*rps2*	$(C)_{17}(T)_{33}$	32	yes	yes	yes	yes	n/a
Pt30204	101	163	*clpP*	← Intergenic →	*rps12*	$(A)_{15}(G)_{14}$	35	yes	yes	yes	yes	n/a
Pt36480	138	163	*psbJ*	← Intergenic →	*petA*	$(T)_{26}$	14	yes	yes	yes	yes	n/a
Pt71936	141	153	*ycf3* 5′ intron	*ycf3* 5′ intron	*ycf3* 5′ intron	$(T)_{22}$	10	yes	no	yes	yes	9
Pt87268	119	173	*trnI-GAU*	← Intergenic	*trnA-UGC*	$(T)A(T)_{31}$	18	yes	yes	no	yes	11
Mean							18.7					7.5
SD							9.2					2.3
Total							280					45

Note: n/a = not available; SD = standard deviation.

[a] Arrows indicate that repeat motif extends into adjacent coding region.

[b] Subscript values indicate the highest number of repeats for 107 *Pinus* species.

[c] *P. brutia* has 170-bp insertion.

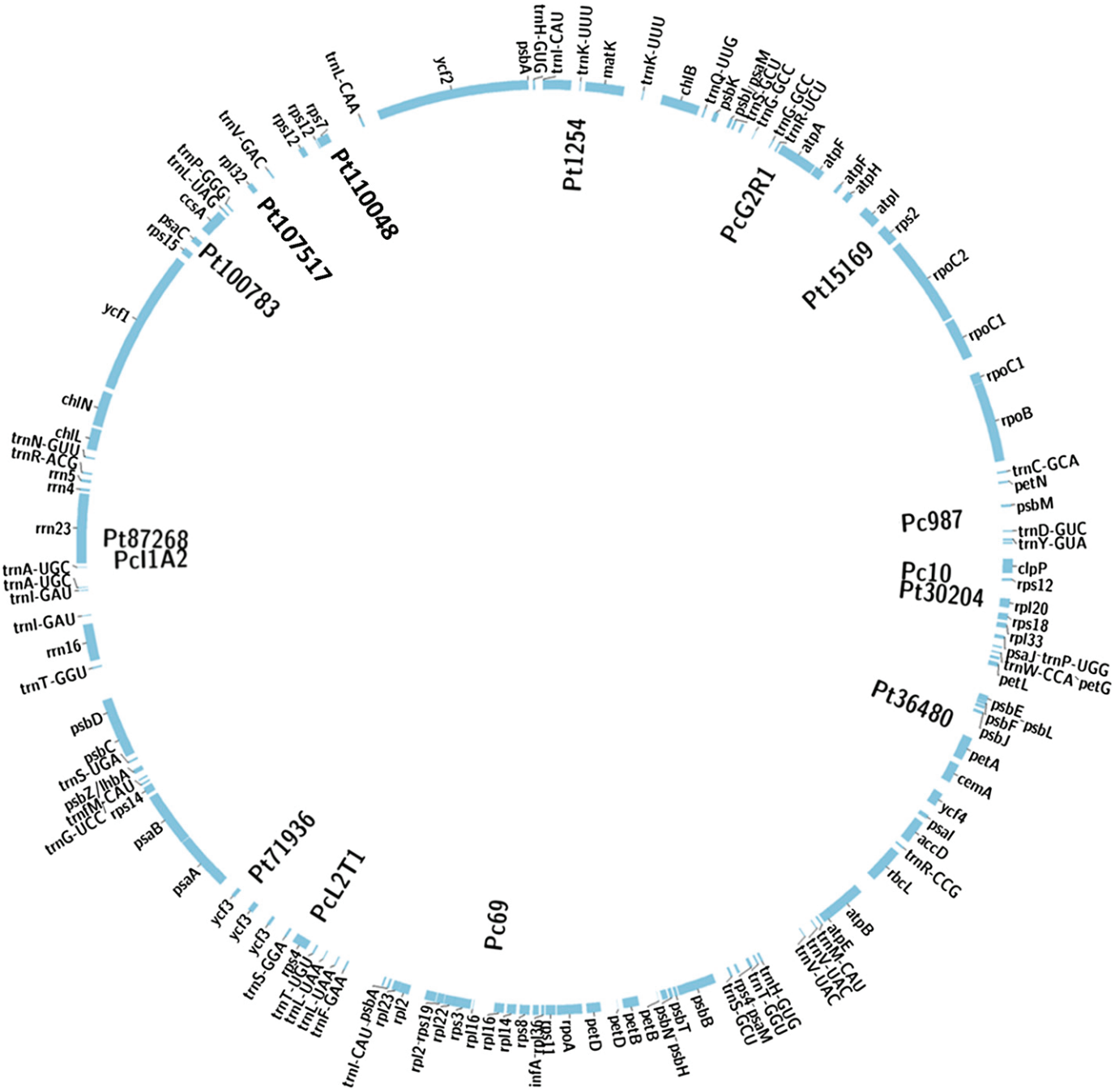

Fig. 2. Locations of 15 cpSSR loci within the *Pinus ponderosa* plastome (GenBank FJ899555).

OTU or with the Ponderosa OTU (Fig. 4A). There were 37 cpSSR alleles in 597 individuals in the eastern region (Fig. 1). The inertia ellipse for the Sky Island OTU did not overlap with the ellipse for the Scopulorum OTU or with the ellipse for the *P. arizonica* OTU (Fig. 4B).

DISCUSSION

Our analysis of plastome alignments confirmed single locations for all 15 primer pairs but found that repeat regions were redundant for two pairs of loci: Pt30204 and Pc10; Pt87268 and PcI1A2. The primers showed generally high levels of sequence conservation across the four taxonomic sections of pine, with some exceptions in subgen. *Strobus* where either the forward or the reverse primer for three loci (Pc987, Pt100783, and Pt1254) had mismatches across the entire subgenus. For these, minor adjustments in primer location to more conserved adjacent regions would potentially increase cross-species transferability (data available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.5nc25; Wofford et al., 2014).

Because the plastid genome is nonrecombining, the significant linkage disequilibrium that we observed in this set of six

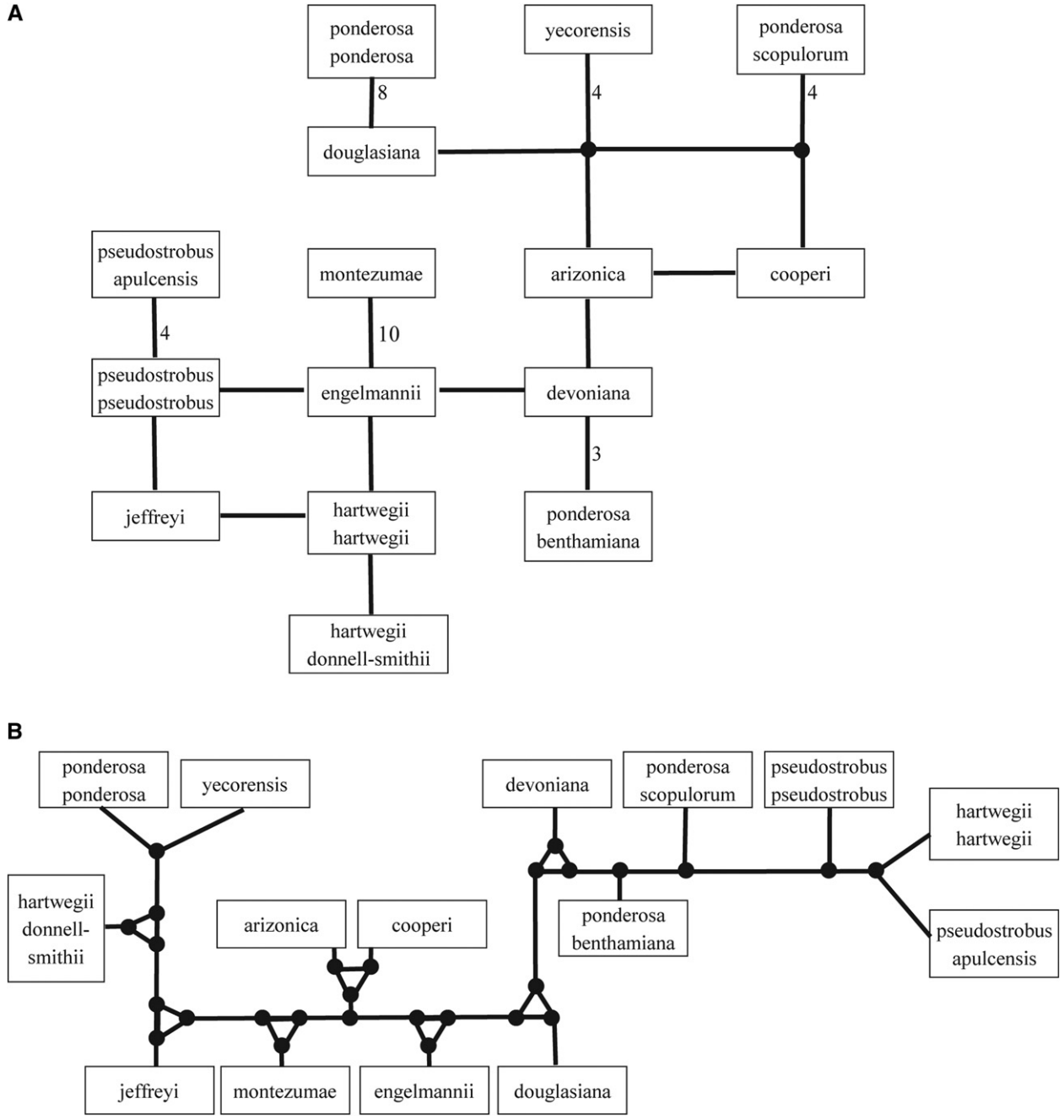

Fig. 3. Comparison of information content of fragment lengths and nucleotide sequences from the same 15 plastomes (Appendix 1) using median joining haplotype networks; black circles indicate median vectors. (A) Based on alleles from six cpSSR loci; numerals indicate additional median vectors. (B) Based on *ycf1* sequences.

cpSSR loci in 911 samples suggests that these multilocus haplotypes are likely to be identical by descent rather than to have been derived by homoplasy.

The cpSSR haplotype network for one exemplar each of 15 species has two unresolvable cycles, and outliers are attached to the network by as many as 10 median vectors (Fig. 3A). This suggests that when using a single sample per taxonomic unit these six cpSSR loci are too saturated to make useful interspecific comparisons in subsect. *Ponderosae*. The *ycf1* network (Fig. 3B) differs from the cpSSR network. It has 27 median nodes, seven cycles, and fails to group most of the clades that were resolved from a whole-plastome phylogeny using the same samples (Parks et al., 2012). This suggests that nucleotide sequences of *ycf1* are also inadequate for interspecific comparisons across this taxonomic subsection.

DAPC scatter plots for our two intraspecific analyses each support our hypothesized OTU. The Pacific Northwest OTU is clearly distinct from the Benthamiana and Ponderosa OTUs (Fig. 4A), with no intermingled sample points. Although some Scopulorum sample points are intermingled with Sky Island

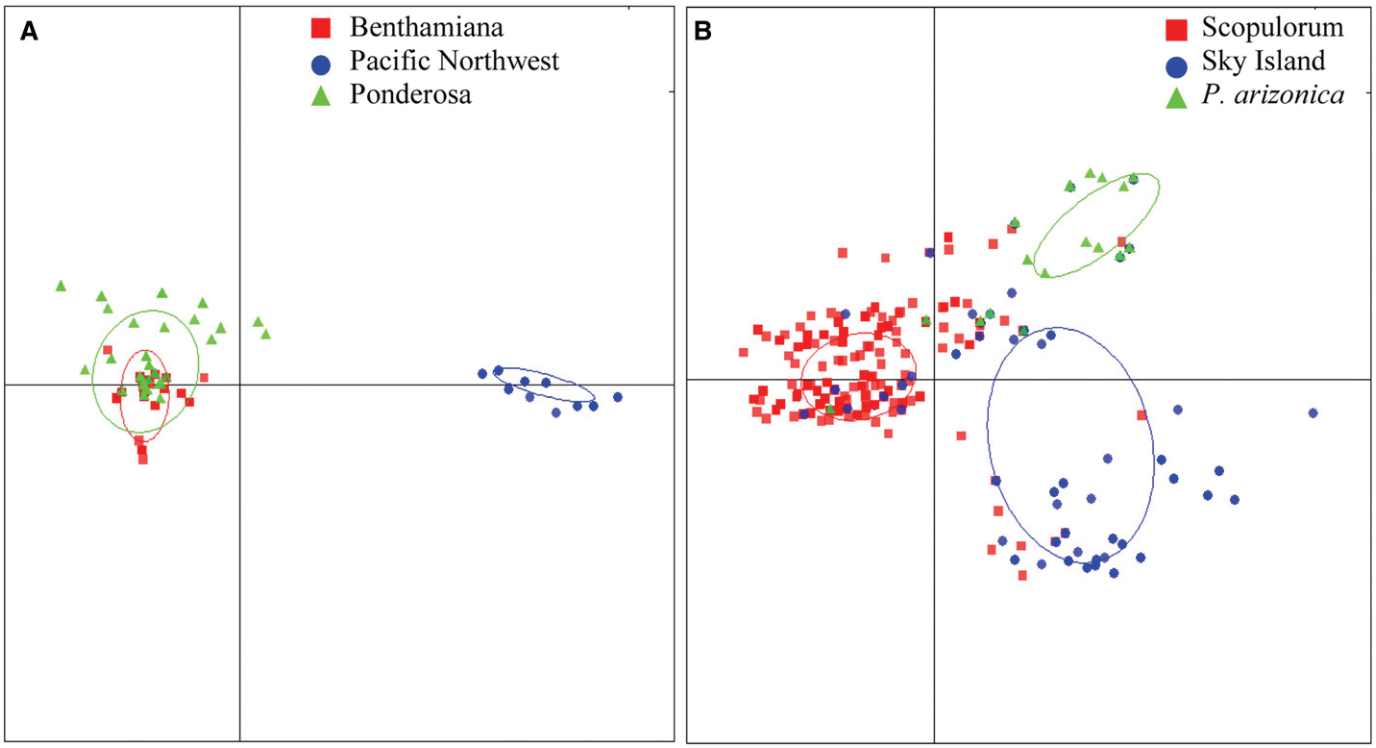

Fig. 4. Scatter plots of discriminant analysis of principal components (DAPC) of cpSSR data for 911 pine samples (Fig. 1; Appendix 2). For each run, a priori groups were assigned to the operational taxonomic units (OTUs) shown in the legend. Ovals are inertia ellipses inferred by *adegenet*. (A) Three hundred fourteen individuals from 14 populations that contained 35 alleles for six cpSSR loci, retaining 10 principal components and two discriminant functions. (B) Five hundred ninety-seven individuals from 27 populations that contained 37 alleles for six cpSSR loci, retaining 30 principal components and two discriminant functions.

OTU samples, the inertia ellipse for the Sky Island OTU does not overlap with the inertia ellipses for the Scopulorum OTU or for *P. arizonica* (Fig. 4B). Using data from six loci for 911 individuals, we were not able to infer an optimal number of clusters (k) using the Bayesian Information Criterion implemented in the *find.clusters* algorithm of *adegenet*. However, this feature may be useful to assign individuals to populations to identify potentially admixed populations.

As we finish our data set for all subsect. *Ponderosae* populations of interest, DAPC of cpSSRs will certainly play an important role. We will test a range of nested OTU groupings to observe relative distinctiveness of these subdivisions. An important caveat is that these cpSSR loci are all linked on the plastid genome and are uniparentally inherited. DAPC offers a way to use these cpSSR data that avoids the discriminant analysis assumption that variables are uncorrelated yet takes advantage of the a priori group assignment feature of discriminant analysis, a feature that is lacking in principal components (Jombart et al., 2010), and is likely to be important in cases like ours where much of the variation is contained within populations.

Incomplete lineage sorting is an important factor in pine molecular studies (Syring et al., 2005) and can lead to incongruence among nuclear and plastid phylogenies (Willyard et al., 2009). Plastid lineages in pines might also be incongruent with nuclear lineages in areas of secondary contact via the widespread phenomenon that has been called "chloroplast capture" (Matos and Schaal, 2000; Liston et al., 2007). This plastid-nuclear conflict has been attributed to hybridization in many plant families, although other mechanisms play a role (Stegemann et al., 2012). For the *P. ponderosa* species complex, mitochondrial haplotypes in some cases support further subdivision of OTUs indicated by our plastid evidence, support fewer subdivisions, or suggest different geographic delineations between OTUs. For example, although data presented here show that the Willamette Valley, Oregon, and Fort Lewis, Washington, populations have similar plastids (Pacific Northwest OTU in Fig. 4A), they do not have similar mitochondria. A Fort Lewis, Washington, population shares a mitochondrial haplotype with populations represented by our Benthamiana OTU, and a Willamette Valley, Oregon, population shares a mitochondrial haplotype with populations represented by our Ponderosa OTU (Potter et al., 2013). We also have preliminary evidence from nuclear SSRs for some incongruent groupings, suggesting a genetic mosaic for the *P. ponderosa* species complex. Although the patterns are certainly affected by incomplete lineage sorting, we expect that pollen flow (revealed by paternal plastid inheritance) and seed dispersal (revealed by maternal mitochondrial inheritance) have shaped the genotype of divergent pine populations in contact zones. The extent of organelle transfer and nuclear introgression across contact zones of long-separated subsect. *Ponderosae* populations seems to be rather limited (Latta and Mitton, 1999), but emerging patterns suggest that there are other major contact zones that are yet to be explored. Thus, our taxonomic conclusions in subsect. *Ponderosae* will await nuclear and mitochondrial data, as well as morphological characters and ecological niche models. DAPC will be an important tool to combine these independent data sets because it can accommodate correlated variables and

provides group weightings to compensate for unequal contributions from each partition.

Considering the current possible alternatives for measuring genetic diversity in wild plants, multiplex genotyping of cpSSRs in *Pinus* provided an efficient and relatively informative view of genomic diversity for use in estimating genetic distance in the plastid lineage. Although we demonstrated the utility of these six loci within subsect. *Ponderosae*, the conservation of primers across the genus suggests that many of the 13 nonredundant cpSSR loci will provide useful data for other *Pinus* taxonomic subsections. In conjunction with other criteria for population genetic structure and species delimitation, these fragment length characters can provide useful insights into pine relationships. We suggest that the method would be easy to extend to other plants using readily available plastome alignments to design primers specific for the target group (Angioi et al., 2009).

LITERATURE CITED

Angioi, S. A., F. Desiderio, D. Rau, E. Bitocchi, G. Attene, and R. Papa. 2009. Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biology* 11: 598–612.

Avise, J. C. 1994. Molecular markers, natural history and evolution. Springer, New York, New York, USA.

Bouffier, L. A., B. L. Gartner, and J.-C. Domec. 2003. Wood density and hydraulic properties of ponderosa pine from the Willamette Valley vs. the Cascade mountains. *Wood and Fiber Science* 35: 217–233.

Culley, T. M., S. G. Weller, A. K. Sakai, and K. Putnam. 2008. Characterization of microsatellite loci in the Hawaiian endemic shrub *Schiedea adamantis* (Caryophyllaceae) and amplification in related species and genera. *Molecular Ecology Resources* 8: 1081–1084.

Culley, T. M., T. I. Stamper, R. L. Stokes, J. R. Brzyski, N. A. Hardiman, M. R. Klooster, and B. J. Merritt. 2013. An efficient technique for primer development and application that integrates fluorescent labeling and multiplex PCR. *Applications in Plant Sciences* 1: 1300027.

DeFaveri, J., H. Viitaniemi, E. Leder, and J. Merilä. 2013. Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Molecular Ecology Resources* 13: 377–392.

Echt, C., L. DeVerno, M. Anzidei, and G. Vendramin. 1998. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology* 7: 307–316.

Epperson, B., F. Telewski, and A. Willyard. 2009. Chloroplast diversity in a putative hybrid swarm of *Ponderosae. American Journal of Botany* 96: 707–712.

Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11: 1591–1604.

Feng, F. J., D. Zhao, X. Sui, and X. Y. Sun. 2011. Study on mating system of *Pinus koraiensis* in natural population based on cpSSR technology. *Advanced Materials Research* 183–185: 700–704.

Gernandt, D., A. Liston, and D. Pinero. 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: Implications for molecular systematics studies of pine species complexes. *Molecular Phylogenetics and Evolution* 21: 449–467.

Gernandt, D. S., S. Hernández-León, E. Salgado-Hernández, and J. Pérez de la Rosa. 2009. Phylogenetic relationships of *Pinus* subsection *Ponderosae* inferred from rapidly evolving cpDNA regions. *Systematic Botany* 34: 481–491.

Godbout, J., J. Beaulieu, and J. Bousquet. 2010. Phylogeographic structure of jack pine (*Pinus banksiana*; Pinaceae) supports the existence of a coastal glacial refugium in northeastern North America. *American Journal of Botany* 97: 1903–1912.

Goremykin, V. V., S. V. Nikiforova, and O. R. P. Bininda-Emonds. 2010. Automated removal of noisy data in phylogenomic analyses. *Journal of Molecular Evolution* 71: 319–331.

Jardón-Barbolla, L., P. Delgado-Valerio, G. Geada-López, A. Vázquez-Lobo, and D. Piñero. 2011. Phylogeography of *Pinus* subsection *Australes* in the Caribbean Basin. *Annals of Botany* 107: 229–241.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.

Latta, R. G., and J. B. Mitton. 1999. Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution* 53: 769–774.

Liston, A., M. Parker-Defeniks, J. V. Syring, A. Willyard, and R. Cronn. 2007. Interspecific phylogenetic analysis enhances intraspecific phylogeographic inference: A case study in *Pinus lambertiana. Molecular Ecology* 16: 3926–3937.

Marshall, H. D., C. Newton, and K. Ritland. 2002. Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). *Theoretical and Applied Genetics* 104: 367–378.

Matos, J. A., and B. A. Schaal. 2000. Chloroplast evolution in the *Pinus montezumae* complex: A coalescent approach to hybridization. *Evolution* 54: 1218–1233.

Neale, D., and R. Sederoff. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics* 77: 212–216.

Orsini, L., S. Huttunen, and C. Schlotterer. 2004. A multilocus microsatellite phylogeny of the *Drosophila virilis* group. *Heredity* 93: 161–165.

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

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

Potter, K. M., V. D. Hipkins, M. F. Mahalovich, and R. E. Means. 2013. Mitochondrial DNA haplotype distribution patterns in *Pinus ponderosa* (Pinaceae): Range-wide evolutionary history and implications for conservation. *American Journal of Botany* 100: 1562–1579.

Rehfeldt, G. E. 1999. Systematics and genetic structure of *Ponderosae* taxa (Pinaceae) inhabiting the mountain islands of the Southwest. *American Journal of Botany* 86: 741–752.

Richardson, B. A., S. J. Brunsfeld, and N. B. Klopfenstein. 2002. DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). *Molecular Ecology* 11: 215–227.

Robledo-Arnuncio, J. J., C. Collada, R. Alia, and L. Gil. 2005. Genetic structure of montane isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. *Journal of Biogeography* 32: 595–605.

Stegemann, S., M. Keuth, S. Greiner, and R. Bock. 2012. Horizontal transfer of chloroplast genomes between plant species. *Proceedings of the National Academy of Sciences, USA* 109: 2434–2438.

Stoehr, M. U., and C. H. Newton. 2002. Evaluation of mating dynamics in a lodgepole pine seed orchard using chloroplast DNA markers. *Canadian Journal of Forest Research* 32: 469–476.

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

Vendramin, G. G., L. Lelli, P. Rossi, and M. Morgante. 1996. A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology* 5: 595–598.

Wakasugi, T., J. Tsudzuki, S. Ito, M. Shibata, and M. Sugiura. 1994. A physical map and clone bank of the black pine (*Pinus thunbergii*) chloroplast genome. *Plant Molecular Biology Reporter* 12: 227–241.

Walter, R., and B. Epperson. 2001. Geographic pattern of genetic variation in *Pinus resinosa*: Area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology* 10: 103–111.

Willyard, A., R. Cronn, and A. Liston. 2009. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution* 52: 498–511.

Wofford, A. M., K. Finch, A. Bigott, and A. Willyard. 2014. Data from: A set of plastid loci for use in multiplex fragment length genotyping for intraspecific variation in *Pinus* (Pinaceae). Dryad Digital Repository. http://doi.org/10.5061/dryad.5nc25

Appendix 1. Taxon name and sample, GenBank number, country, state, and geographic coordinates of 15 samples used in Fig. 3. NA = not available.

arizonica: *Pinus arizonica* Engelm. isolate ARIZ01; JN854225; USA: New Mexico (33.13, −108.00)

cooperi: *P. arizonica* Engelm. var. *cooperi* (C. E. Blanco) Farjon & Styles isolate COOP01; JN854216; Mexico: Durango (NA)

devoniana: *Pinus devoniana* Lindl. isolate DEVO02; JN854208; Mexico: Michoacán (19.42, −101.82)

douglasiana: *Pinus douglasiana* Martínez isolate DOUG01; JN854205; Mexico: Jalisco (19.53, −103.52)

engelmannii: *Pinus engelmannii* Carrière isolate ENGE02; JN854201; USA: Arizona (31.73, −110.83)

hartwegii hartwegii: *Pinus hartwegii* Lindl. isolate HART07; JN854196; Mexico: Guerrero (17.52, −99.96)

hartwegii donnell-smithii: *Pinus hartwegii* Lindl. subsp. *donnell-smithii* (Mast.) Silba isolate DONN02; JN854206; Guatemala (14.80, −91.52). Note: shown as *P. donnell-smithii* in whole-plastome tree (Parks et al., 2012).

jeffreyi: *Pinus jeffreyi* Grev. & Balf. isolate JEFF04; JN854193; USA: California (37.37, −118.39)

montezumae: *Pinus montezumae* D. Don in Lamb. isolate MONZ01; JN854183; Mexico: Hidalgo (20.11, −98.61)

ponderosa benthamiana: *Pinus ponderosa* P. Lawson & C. Lawson var. *benthamiana* (Hartw.) Vasey isolate POND21; JN854172; USA: California (39.69, −121.69)

ponderosa ponderosa: *Pinus ponderosa* P. Lawson & C. Lawson; FJ899555; USA: Montana (47.969, −115.985)

ponderosa scopulorum: *Pinus ponderosa* P. Lawson & C. Lawson var. *scopulorum* Engelm. in S. Watson isolate POND59; JN854171 USA: South Dakota (44.295, −103.828)

pseudostrobus pseudostrobus: *Pinus pseudostrobus* Lindl. isolate PSEU03; JN854169; Guatemala (15.38, −91.43)

pseudostrobus apulcensis: *Pinus pseudostrobus* Lindl. var. *apulcensis* (Lindl). Shaw isolate OAXA02; JN854178; Mexico: Oaxaca (17.32, −96.43). Note: shown as *P. oaxacana* in whole-plastome tree (Parks et al., 2012).

yecorensis: *Pinus yecorensis* Debreczy & I. Rácz isolate YECO02; JN854152; Mexico: Sonora (28.38, −108.87)

Appendix 2. Operational taxonomic unit (OTU), population, collector(s), collector number or herbarium voucher, U.S. state, and GPS coordinates of 41 populations shown in Fig. 1 and used in Fig. 4. A herbarium voucher for each population has been deposited at Hendrix College Herbarium (HXC in *Index Herbariorum*).

P. arizonica: AM; Willyard, Liston, Gernandt; DSG874; Arizona; 32.63323, −109.82387. CH; Willyard, Marquardt, Epperson; AMW1080; Arizona; 31.94951, −109.30785. PA; Telewski, Marquardt; HXC5826; Arizona; 32.41369, −110.70954.

Benthamiana: HE; Finney, Willyard; AMW1107; California; 37.19180, −121.54477. HL; Finney, Willyard; AMW1103; California; 37.02610, −122.04603. PS; Willyard; AMW1017; California; 40.66833, −122.69750. QH; Finney, Willyard; AMW1102; California; 37.08527, −122.05990. UZ; Finney, Willyard; AMW1104; California; 36.98420, −122.06352. V1; Finney, Willyard; AMW1105; California; 36.05983, −121.57147.

Ponderosa: IH; Finch, Willyard; AMW1111; Washington; 48.61683, −11816853. KF; Finch, Willyard; AMW1115; Oregon; 44.41843, −121.76797. PB; Willyard; AMW1015; Oregon; 44.06538, −118.78980. PC; Willyard; AMW1098; California; 40.34333, −121.73333; RW; Willyard; AMW999; Nevada; 39.33073, −104.17405.

Scopulorum: A1; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1136; Arizona; 36.54290, −110.46870. A2; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1137; Arizona; 35.38140, −110.96400. A3; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1138; Arizona; 35.09800, −113.88110. A4; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1139; Arizona; 36.79550, −112.25500. C1; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1135; Colorado; 37.72930, −108.20920. C2; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1143; Wyoming; 44.31150, −106.81270. DG; Gernandt; DSG1016; Utah; 39.80000, −110.90000. N1; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1146; Nebraska; 42.69560, −103.00520. N2; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1147; Nebraska; 42.79000, −100.0200. RG; Langer HXC5827; South Dakota; 45.52388, −103.16423. S1; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1145; South Dakota; 43.85300, −104.05320. SF; Finch, Nguyen, Segear, Willyard; AMW1073; New Mexico; 35.73355, −105.56277. SM; Langer, Rand HXC5833; Nevada; 36.32038, −115.67522. TA; Langer HXC5828; New Mexico; 35.24800, −107.72900. U1; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1140; Utah; 37.72950, −112.25050. U2; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1141; Utah; 40.63240, −111.17180. W1; Bigott, Lea, Pouncey, Schumacher, Spatz, Willyard; AMW1142; Wyoming; 42.76420, −106.32690.

Sky Island: BF; Willyard, Marquardt, Epperson; AMW1078; Arizona; 31.91702, −109.27539. HS; Willyard, Marquardt, Epperson; AMW1081; Arizona; 31.25815, −110.20852. LA; Telewski, Marquardt; HXC5829; Arizona; 32.43837, −110.79093. LB; Willyard, Marquardt, Epperson; AMW1077; Arizona; 32.62208, −109.82723. PT; Telewski, Marquardt; HXC5830; Arizona; 34.04893, −111.09373. SL; Langer; HXC5831; Arizona; 31.70020, −110.84718. SR; Willyard, Marquardt, Epperson; AMW1082; Arizona; 31.68636, −110.87816.

Pacific Northwest: F1; Willyard; AMW1099; Washington; 47.03832, −122.52643. VB; Meyers; AMW1112; Oregon; 44.37922, −122.91157. VW; Meyers; HXC5832; Oregon; 45.32350, −122.76570.