Length polymorphism scanning is an efficient approach for revealing chloroplast DNA variation

Matthew E. Horning and Richard C. Cronn

Abstract: Phylogeographic and population genetic screens of chloroplast DNA (cpDNA) provide insights into seed-based gene flow in angiosperms, yet studies are frequently hampered by the low mutation rate of this genome. Detection methods for intraspecific variation can be either direct (DNA sequencing) or indirect (PCR–RFLP), although no single method incorporates the best features of both approaches. We show that screening universal chloroplast amplicons for length polymorphism provides an accurate and efficient method for identifying cpDNA variation. By sequencing 4500 bp of cpDNA from 17 accessions of *Purshia tridentata* (bitterbrush), we detected 9 haplotypes, 8 of which were identifiable by unique multilocus length combinations resolvable by automated fragment analysis. In silico estimates of PCR–RFLP for these loci show that 5 haplotypes would be resolved by agarose electrophoresis. A survey of 4 intraspecific data sets from diverse angiosperms revealed that length variation in cpDNA amplicons is nearly ubiquitous, and 61 of 67 haplotypes identified by direct sequencing could be identified by screening length variation. Combined with automated fluorescent detection, length polymorphism screening of universal cpDNA regions offers a simple screen for intraspecific variation that can be used across angiosperms with minimal optimization, providing detection limits that rival direct sequencing at a fraction of the cost.

Key words: cpDNA, intraspecific polymorphism, population genetics, phylogeography, indels.

Résumé : Des analyses phylogéographiques et de génétique des populations sur l'ADN chloroplastique (ADNcp) nous éclairent sur les flux géniques via les graines chez les angiospermes, mais de telles études sont souvent limitées par le faible taux de mutation au sein de ce génome. Les méthodes de détection pour la variation intraspécifique peuvent être directes (séquençage de l'ADN) ou indirectes (PCR–RFLP), mais aucune méthode ne combine les caractéristiques les plus souhaitables de ces deux approches. Les auteurs montrent ici que le criblage d'amplicons chloroplastiques universels pour le polymorphisme de longueur constitue une méthode précise et efficace pour identifier la variation au sein de l'ADNcp. En séquençant 4500 pb d'ADNcp chez 17 accessions du *Purshia tridentata* (purchie tridentée), les auteurs ont détecté 9 haplotypes dont 8 étaient identifiables en raison d'une combinaison unique de tailles des amplicons, lesquelles étaient évaluées par analyse automatisée. Des prédictions in silico du polymorphisme PCR–RFLP ont révélé que 5 haplotypes seraient identifiables sur gel d'agarose. Une analyse de 4 jeux de données sur la variation intraspécifique chez diverses angiospermes a révélé que la variation pour la taille des amplicons d'ADNcp est quasi universelle et que 61 de 67 haplotypes identifiés par séquençage direct auraient pu être détectés en examinant la variation pour la taille des amplicons. Combinée à la détection automatisée en fluorescence, l'analyse du polymorphisme de taille des amplicons au sein de régions chloroplastiques universelles s'avère une méthode simple pour la détection de variation intraspécifique. De plus, elle peut être employée chez diverses angiospermes à la suite d'une optimisation minime et offre des limites de détection qui sont proches de celles offertes par le séquençage direct mais à une fraction des coûts.

Mots clés : ADNcp, polymorphisme intraspécifique, génétique des populations, phylogéographie, indels.

[Traduit par la Rédaction]

Introduction

The genomes of cytoplasmic organelles make attractive markers for population genetic and phylogeographic studies because they are haploid, non-recombinant, uniparentally inherited, and usually achieve high rates of fixation relative to nuclear markers (Avise 1994; McCauley 1995; Provan et al. 2001). Because the effective population size of organellar genomes is one-half to one-third the size of a diploid nuclear genome (Birky et al. 1983; Hamilton and Miller 2002), organellar markers respond more rapidly to processes such as migration and drift, and may show increased differentiation relative to nuclear markers across an equivalent geographic scale (McCauley 1995; Sunnucks 2000). In plants, the fact that organellar markers are haploid is a significant feature that can simplify comparisons among taxa exhibiting chromosome number or autopolyploid variation (Trewick et al. 2002).

Many "universal" chloroplast DNA (cpDNA) PCR primers are available for amplifying introns and intergenic spacer

Received 7 March 2005. Accepted 29 August 2005. Published on the NRC Research Press Web site at http://genome.nrc.ca on 8 February 2006.

Corresponding Editor: B. Golding.

M.E. Horning and R.C. Cronn.[1] USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331, U.S.A.

[1]Corresponding author (e-mail: rcronn@fs.fed.us).

 doi:10.1139/G05-093

regions, and enjoy widespread use in studies of intraspecific diversity and divergence (Taberlet et al. 1991; Demesure et al. 1995; Dumolin-Lapègue et al. 1997*b*; Hamilton 1999; Shaw et al. 2005). Despite the popularity of these markers, chloroplast sequences diverge at 1/10 the rate of mammalian mitochondrial genomes ($<1 \times 10^{-9}$ substitutions·(site·yr)$^{-1}$; Wolfe et al. 1987; Clegg et al. 1994) and show considerable rate heterogeneity across different non-coding regions in the genome (Schaal et al. 1998; Hamilton et al. 2003; Shaw et al. 2005). Consequently, studies focusing on intraspecific sequence divergence are often plagued by a paucity of variation relative to the effort expended sequencing loci (McCauley 1995; Schaal et al. 1998). A relatively abundant and evolutionarily labile class of mutations is insertions–deletions (indels; Provan et al. 2001) involving one to several nucleotides. Indels are often associated with repeated motifs (e.g., mono- and di-nucleotide repeats) through the action of polymerase- or strand-slippage, and can accumulate rapidly ($\sim 1 \times 10^{-3}$ to 1×10^{-5} mutations·(site·yr)$^{-1}$; Provan et al. 1999; Marshall et al. 2002). Secondary structure of group II introns may also play a role in larger recombination-mediated deletions, as regions participating in hairpin structures appear prone to deletion events (Kelchner and Wendel 1996; Kelchner 2002; Hamilton et al. 2003).

In this report, we show how indel accumulation in universal cpDNA amplification products can be used to efficiently detect chloroplast variants. Length polymorphism in 3 non-coding regions of a dryland restoration shrub (*Purshia tridentata* (Rosaceae)) has the potential to reveal nearly all of the haplotype diversity detected by direct nucleotide sequencing of 6 loci exceeding 4500 bp. To determine whether this trend is evident in other taxa, we surveyed cpDNA sequences from published population-level surveys from 4 diverse species, *Calamagrostis breweri* (Poaceae), *Lilium philadelphicum* (Liliaceae), *Silene latifolia* (Caryophyllaceae), and *Arabis holboellii* (Brassicaceae). This comparison shows that indels are a frequent and variable component of "universal" cpDNA amplicons from these species. Critically, 61 of 67 haplotypes detected by direct sequencing in these 5 species are marked by unique multilocus length combinations. The intrinsic property of chloroplasts to rapidly accumulate indels makes length polymorphism screening a simple but powerful approach for identifying chloroplast variants in population-level studies, particularly if combined with fluorescent amplification and automated detection methods.

Materials and methods

Studied species

Chloroplast DNA sequences from *Purshia tridentata* were generated as part of an ongoing effort to characterize ecologically important restoration species. Sequences for *Calamagrostis breweri, Lilium philadelphicum, Silene latifolia,* and *Arabis holboellii* have been described elsewhere, but are outlined below.

Purshia tridentata

Antelope bitterbrush (*Purshia tridentata* Pursh DC; Rosaceae) is a dryland shrub that occupies about 3×10^6 acres of rangeland and dry forest between the Cascade, Sierra Nevada, and Rocky Mountain ranges, USA (Young and Clements 2002). Leaves from this shrub provide palatable winter forage for large mammals, and the seed is an important food for granivores (Young and Clements 2002). Bitterbrush is slow growing, taking up to 10 years to reach flowering age, and varies widely in its ability to regenerate from roots after fire. These features combine to make bitterbrush a widely planted shrub in habitat restoration projects (Young and Clements 2002). In this study, 16 plants were sampled across a wide geographic range to gain insight into the magnitude of cytoplasmic variation (Table 1).

Comparative studies

Chloroplast DNA sequences were evaluated from population studies of 4 different angiosperms. Shorthair reedgrass (*Calamagrostis breweri* Thurber) is native to subalpine meadows in the central Sierra Range of California and 2 isolated populations exist on Cascade volcanoes in Oregon. Owing to its rarity, interrelationships between 3 northern hexaploid populations and 1 southern tetraploid population of *C. breweri* were evaluated at 4 cpDNA loci (the *rpl16* intron, *trnH*$^{(GUG)}$–*psbA*, *trnL*$^{(UAA)}$ 3′ exon – *trnF*$^{(GAA)}$, and *trnT*$^{(UGU)}$ – *trnL*$^{(UAA)}$ 5′ exon) as part of a conservation assessment for the USDA Forest Service (Cronn et al. 2004; Table 1). The *Lilium philadelphicum* cpDNA data set consists of 3 intergenic spacer regions (*trnD*$^{(GUC)}$–*trnY*$^{(GUA)}$ (Shiraishi et al. 2001), *trnH*$^{(GUG)}$–*psbA*, and *trnT*$^{(UGU)}$ – *trnL*$^{(UAA)}$ 5′ exon) from 22 samples collected across the range of this North American species (Table 1). DNA isolation, PCR amplification of cpDNA markers, and sequence identification are described in Horning (2003). Sequences from *S. latifolia* were described in two previous studies focussing on the influence of cytoplasmic male sterility on organellar diversity (Ingvarsson and Taylor 2002) and phylogeographic interrelationships of *S. latifolia* and *Silene vulgaris* across Europe (Ingvarsson et al. 2003). This data set includes 4 intergenic spacer regions (*trnG*$^{(UCC)}$–*trnS*$^{(GCU)}$, the *trnL*$^{(UAA)}$ intron, *trnL*$^{(UAA)}$–*trnF*$^{(GAA)}$, and *trnH*$^{(GUG)}$–*psbA*) from 25 individuals. Sequences from *Arabis holboellii* were included in a phylogeographic study of North American *A. holboellii*, *A. drummondii*, and their hybrid derivative *A.* × *divaricarpa* (Dobeš et al. 2004). Because we are interested in intraspecific diversity, we only analyzed the 25 accessions identified as *A. holboellii* that were examined at 1 intergenic spacer region (*trnL*$^{(UAA)}$–*trnF*$^{(GAA)}$) and 2 introns (*trnL*$^{(UAA)}$ and *rpoC1*; Dobeš et al. 2004). All previously published sequences are available from GenBank (http://www.ncbi.nlm.nih.gov/; Table 1).

Chloroplast DNA amplification, sequencing, and sequence analysis

Chloroplast loci from bitterbrush were amplified and sequenced with universal cpDNA primers designed for the following regions: the *trnH*$^{(GUG)}$–*psbA* spacer (Hamilton 1999; Tate and Simpson 2003; for *Purshia* only), the *trnT*$^{(UGU)}$–*trnL*$^{(UAA)}$ and *trnL*$^{(UAA)}$–*trnF*$^{(GAA)}$ spacers (Taberlet et al. 1991), the *rpl16* intron (Jordan et al. 1996), the *trnK* spacer (Cronn et al. 2002), and the *accD–psaI* spacer (Small et al. 1998). Genomic DNA from *Purshia* specimens was isolated from 50–150 mg of fresh or dried leaf tissue using the

Table 1. Collection information for taxa included in this study.

Species	Accession and collection locality or reference	Latitude (°N), longitude (°W)	GenBank accession Nos.
Purshia tridentata	30-3, Lake Roosevelt, Wash.	Unknown[a]	DQ100171–DQ100186 (*accD–psaI*),
	B1-1, Winema National Forest, Ore.	42.41, 121.33	DQ100187–DQ100202 (*trnH–psbA*),
	B11-3, Mt. Hood National Forest, Ore.	45.25, 121.23	DQ100203–DQ100218 (*trnL–trnF*),
	B13, Umatilla National Forest, Ore.	45.10, 118.53	DQ100219–DQ100234 (*trnK*),
	B13-1, Umatilla National Forest, Ore.	45.10, 118.53	DQ100235–DQ100250 (*rpl16*)
	B3-1, Hart Mountain National Wildlife Refuge, Ore.	42.29, 119.38	
	8A, Craters of the Moon National Monument, Idaho	43.4, 113.5	
	10-3, St. Anthony, Idaho	44.0, 111.7	
	B20, Lassen National Forest, Calif.	40.18, 120.32	
	B21, Tahoe National Forest, Calif.	39.22, 120.80	
	B18, Modoc National Forest, Calif.	41.39, 121.27	
	B23, Inyo National Forest, Calif.	37.39, 118.55	
	B24, Elko, Nev.	41.40, 114.35	
	1-1, Granite Seed Company, Nevada source	Unknown[a]	
	23-2, Vernal, Utah	40.4, 109.5	
	20A, Panguitch Lake, Utah	37.7, 112.6	
	S1, Medicine Bow National Forest, Wyo.	Unknown[a]	
Calamagrostis breweri	Data from Cronn et al. (2004)	—	DQ113917–DQ113925 (*trnH–psbA*),
			DQ113926–DQ113934 (*trnL–trnF*),
			DQ113935–DQ113943 (*trnT–trnL*),
			DQ113944–DQ113952 (*rpl16*)
Lilium philadelphicum	Data from Horning (2003)	—	DQ122704–DQ122720 (*trnH–psbA*),
			DQ122721–DQ122737 (*trnD–trnY*),
			DQ122738–DQ122754 (*trnT–trnL*)
Silene latifolia	Data from Ingvarsson and Taylor (2002)	—	AF519013–AF519037 (*trnL*),
			AF518879–AF518903 (*trnL–trnF*),
			AF518904–AF518928 (*trnH–psbA*),
			AF519047–AF519071 (*trnG–trnS*)
Arabis holboellii	Data from Dobeš et al. (2004)	—	AY257693–AY257794 (*trnL–trnF*),
			AY257795–AY257838 (*rpoC1*)

[a]Restoration seed mixes of uncertain geographic origin.

FastDNA® kit (Bio 101). PCRs (40 µL) were performed on an MJ Research PTC-100 thermocycler and contained ~50 ng genomic DNA, 0.2 µmol/L each forward and reverse primers, 1× reaction buffer (50 mmol/L KCl, 10 mmol/L Tris–HCl, pH 9.0), 200 µmol/L dNTPs, 1.5 mmol/L $MgCl_2$, 1.6 U *Taq* polymerase, and 40 µg bovine serum albumin. PCR cycle parameters were 80 °C for 2 min; 34 cycles of 95 °C for 30 s, 50 °C for 30 s, a ramp to 65 °C (temperature increase of 0.1 °C/s), 65 °C for 4 min; and a final extension at 72 °C for 10 min. Reactions were cleaned using polyethylene glycol precipitation (Lis 1980) and sequenced with the ABI PRISM® Big Dye™ terminator cycle sequencing kit (v. 3.1, Applied Biosystems Inc., Foster City, Calif.). Sequences were resolved on an ABI 3100 capillary automated sequencer at the Oregon State University Center for Gene Research and Biotechnology (Corvallis, Ore.; http://www.cgrb.orst.edu/).

Sequence traces were checked using BioEdit v. 5.0.6 (Hall 1999) and alignments were constructed using ClustalW (in BioEdit, default settings), with minor corrections. Alignments are available from GenBank or from the authors. We determined the absolute length of each sequence (aligned length minus the number of gapped positions) and used these to estimate multilocus fragment-length genotypes. We used MEGA v. 2.1 (Kumar et al. 2001) to identify length variants for each locus, and to calculate variable positions and nucleotide diversity (π). Gaps were recoded as binary (1, present; 0, absent) or multistate unordered events following the guidelines of Simmons and Ochoterena (2000), and these characters were appended to nucleotide alignments. These data were evaluated using statistical parsimony (TCS 1.13; Clement et al. 2000) to create most-parsimonious haplotype networks based on both substitutions alone and substitutions plus coded indels.

Fragment length polymorphism screening

To demonstrate the feasibility of fragment screening, we amplified PCR products from 2 loci using fluorescent primers and screened multiplexed products for length polymorphism using capillary electrophoresis. Loci included the *trnH–psbA* spacer (5′[HEX]–*trnH2* primer) and the *accD–psaI* spacer (5′[6FAM]–*psaI–75R* primer). Amplified *trnH–psbA* products were analyzed without modification, and 100–200 ng of *accD–psaI* product was digested with 5 U of *Eco*RI (10 µL total reaction volume) to cut the ~1200 bp amplicon into 2 smaller fragments. Fluoresecent detection

was accomplished using an ABI3100 equipped with a 36 cm capillary and POP4 polymer, and MapMarker® 1000 ROX internal lane standards (BioVentures Inc., Murfreesboro, Tenn.). Bands were visualized using Genotyper v. 3.7NT (Applied Biosystems Inc.) and estimated sizes and standard errors were determined from duplicate capillary runs from each sample (2–4 repititions/haplotype).

Comparisons of fragment length screening to PCR–RFLP were conducted using in silico restriction digests for variable sequences (the *trnH–psbA* spacer, the *trnT–trnL* spacer, the *trnL–trnF* spacer, the *rpl16* intron, the *trnK* spacer, and the *accD–psaI* spacer) using 4-cutter (*Alu*I, *Hae*III, *Hha*I, *Hpa*II, *Mse*I, *Msp*I, *Rsa*I, and *Taq*I), 5-cutter (*Hin*fI), and 6-cutter (*Eco*RI) restriction enzymes identified in published PCR–RFLP surveys. Sequences were digested using the *Escherichia coli* multilocus in silico RFLP tool (http://www.shigatox.net/cgi-bin/mlst7/rflppredictor) and polymorphic fragments were identified using the following criteria: (*i*) polymorphic fragments must differ by ≥2% of the largest fragment length to be detected using high-resolution agarose (such as MetaPhor, Cambrex Biosystems, East Rutherford, N.J.) and (*ii*) individual fragments must be larger than 20 bp to be detected. Even though polyacrylamide is used successfully in PCR–RFLP studies (e.g., Dumolin-Lapègue et al. 1997*a*), agarose electrophoresis is used more frequently. For this reason, agarose electrophoresis is used as the basis for this comparison.

Results

cpDNA fragment length variation in bitterbrush

Amplification products from 6 different cpDNA regions were evaluated from *Purshia tridentata*. One region was invariant (*atpB–rbcL* spacer, 777 bp), and 2 additional regions (*rpl16* intron, 787 bp; *trnL–trnF* spacer, 388 bp) showed single variable sites without length variation. Because these sequences were uninformative with regard to length variation and correspondence to haplotype variation, they are only considered in the next section.

The 3 remaining genes showed absolute length variation that ranged from less than 0.6% (*trnK* intron) to 8.3% (*accD–psaI* spacer) of their aligned length. The intergenic spacer between *accD* and *psaI* genes was the most variable with regard to the range of absolute length (1098–1176 bp; range = 78 bp) and number of length variants detected (n = 5). Length variation at this locus arose entirely through a combination of perfect tandem duplications (ranging from 3 to 55 bp) and unique indels; microsatellites were absent. In contrast, length variation at the *trnH–psbA* spacer and the *trnK* intron was restricted entirely to simple repeats. While indels at *trnH–psbA* and the *trnK* intron resulted in a narrower range of length variation as compared with tandem duplications at *accD–psaI*, length polymorphism was substantial, with 4 variants observed at *trnH–psbA* and 3 at the *trnK* intron. Averaged across loci showing length variation (*accD–psaI, trnH– psbA, trnK*) and loci showing length constancy (*trnL–trnF* spacer, *rpl16* intron), 10 indels were inferred from 4516 aligned bp, or approximately 2.2 indel events/kb.

Multigene length genotypes from these 3 variable loci combined to reveal 8 different haplotypes from the 17 individuals screened (Fig. 1, Table 2). Considering the gene order "*accD–psaI* + *trnH–psbA* + *trnK*", these 8 haplotypes have combined lengths of 1098+455+902 base pairs (for accession B11–3), 1137+460+899 (for B23, 1–1), 1143+460+899 (for 20A), 1153+455+902 (for B1–1, B13, B13–1, B20), 1153+455+904 (for 30–3, B21, B18), 1176+454+902 (for 8A, B24), 1176+455+902 (for B3–1, 10–3), and 1176+456+902 (for 23–2, S1).

Correspondence between length polymorphism and genealogy

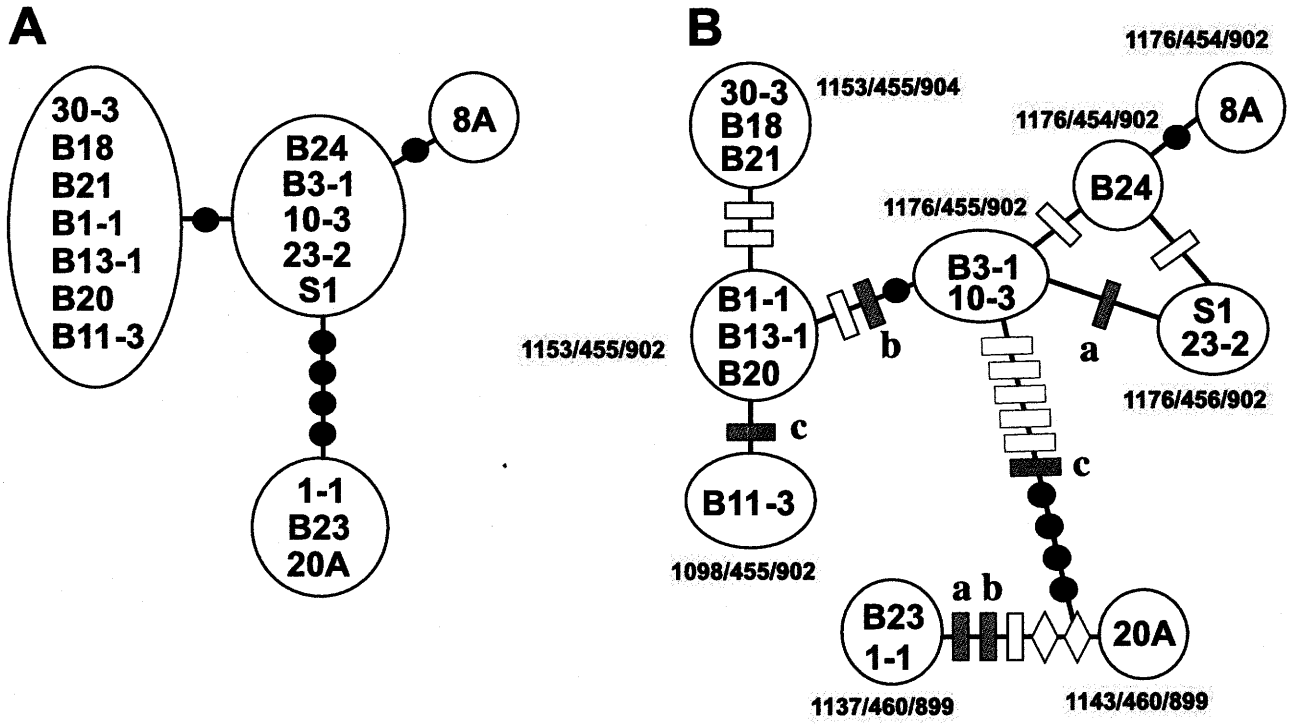
Size homoplasy in cpDNA PCR products can lead to mistaken homology, particularly if recurrent mutations occur in regions of limited variation and (or) constrained length. To determine if dissimilar haplotypes were misclassified as identical owing to homoplasious length mutations, we generated allele networks from nucleotide sequences using statistical parsimony (Clement et al. 2000) and compared sequence-derived networks to fragment length estimates of haplotype identity. Sequences from this sample of geographically diverse germplasm showed high uniformity, with only 6 variable sites detected in 4516 total aligned bases. Consequently, nucleotide diversity among these 17 individuals was exceptionally low ($\pi = 0.0005$), and the 6 substitutions alone yielded 4 haplotypes (Fig. 1A). By including indels (coded as binary characters) in the analysis, 9 haplotypes are revealed (Fig. 1B). Of the 10 scored indels, 3 showed possible evidence of homoplasy (shaded bars, Fig. 1B). One of the indels was located in *trnH–psbA* (lowercase *a* in Fig. 1B; positions 141–151) and involved a T_n microsatellite. At this site, individuals S1, 23-2, B23, and 1-1 shared T_{11} as opposed to T_9 or T_{10}. The other two sites were found in *accD–psaI*, with one caused by a 5 bp deletion (*b*; positions 724–728 of the alignment), and the other caused by a 55 bp deletion (*c*; positions 828–882).

Of the 9 haplotypes detected, only 2 non-identical sequences (B24 and 8A) show identical fragment-length combinations. These two haplotypes would be considered identical if surveyed by fragment-length screening alone. While this result highlights a limitation of fragment-length screening, it's critical to note that these sequences differ by only 1 substitution in >4500 bp; a less-exhaustive sample of nucleotides by direct sequencing would likely have led to the same inference of identity. Equally as important, B24 and 8A are sister haplotypes that trace to a recent common ancestor, as opposed to more divergent haplotypes that are united through parallel mutation and homoplasy. In the case of bitterbrush, it appears that simple screening of length polymorphism from 3 cpDNA PCR products could reveal nearly as many haplotypes (8 of 9, or 89%) as direct sequencing from 6 genes and 4516 aligned bases.

Fragment length polymorphism screening versus PCR–RFLP in bitterbrush

To determine whether fragment length polymorphism screening could be used to detect the length mutations present in bitterbrush cpDNA, we used capillary fragment analysis to evaluate labeled amplicons from accessions exhibiting small (*trnH–psbA*) and large (*psaI–accD*) length differences. These include accessions 8A (*trnH–psbA* = 454 bp; *psaI–accD* = 1176 bp (sizes do not include primers)), B18 and

Fig. 1. Haplotype networks derived from nucleotides (A) and nucleotides plus indels (B) for *Purshia tridentata*. Multilocus length genotypes of the aligned locus order *accD–psaI/trnH–psbA/trnK* (e.g., 1153/455/904) are highlighted in grey for each unique haplotype identified by direct sequencing. Unobserved mutational events (◇), nucleotide substitutions (●), and indels (▯) are traced on branches. Homoplasious indels are shown with grey rectangles; letters correspond to specific indels described in the text.



B20 (455 and 1153 bp, respectively), and B23-2 and S1 (456 bp and 1176 bp, respectively). Results showed that the length differences in *trnH–psbA* were identified as distinct, with mean fragment sizes estimated at 508.39 bp for accession 8A, 509.04 bp for accessions B18 and B20, and 510.11 for accessions B23-2 and S1 (Table 3). Critically, replicates showed high precision, with 99% confidence intervals ranging from ±0.064 to ±0.137 bp. *Eco*RI-digested *accD–psaI* fragments from the same individuals yielded an invariant *accD* fragment of 515.0 bp, and a variable *psaI* fragment of 637.76 bp (B18, B20) or 660.93 bp (8A, B23-2, S1) in length (Table 3). Error estimates for these larger fragments were proportionately larger, with 99% confidence intervals ranging from ±0.218 to ±0.429 bp. These results show that single-base differences can be easily detected in products 510 bp in length, but that resolution will diminish with larger fragments.

For comparison, in silico PCR–RFLP analysis was performed on variable sequences to estimate the number of haplotypes that could be detected. Substitutions detected in *rpl16* and *trnL–trnF* would not result in a polymorphism for the restriction enzymes surveyed. Four enzymes cut *trnH–psbA* (*Alu*I, *Mse*I, *Rsa*I, *Taq*I), and *Mse*I revealed an RFLP between the larger spacer in accessions 20A, B23 and 1-1 (generating 2 internal bands of 16 and 49 bp) and the smaller spacer from all other accessions (1 internal band of 66 bp). Six enzymes cut the *trnK* intron (*Alu*I, *Mse*I, *Rsa*I, *Taq*I, *Hin*fI, *Eco*RI) although none revealed an RFLP. Digestion with *Mse*I and *Hin*fI could resolve 2 haplotypes owing to length variation in internal restriction fragments. For example, digestion with *Mse*I would yield 1 internal fragment of 137 bp for accessions 20A, B23, and 1-1, whereas the homologous fragment in all other accessions is 140–141 bp. Finally, 7 enzymes cut *accD–psaI* (*Alu*I, *Hae*III, *Mse*I, *Rsa*I, *Taq*I, *Hin*fI, *Eco*RI). *Hin*fI is the only enzyme that reveals an RFLP; this site is associated with the 55 bp indel at positions 828–882. Owing to the extensive length variation at this locus (1098 bp–1176 bp), *Alu*I, *Taq*I, and *Hin*fI also have the potential to resolve 5 haplotypes that show variation in internal fragments (data not shown). If all variable loci are combined, 5 unique multilocus haplotypes could be identified by PCR–RFLP using high-resolution agarose electrophoresis. This number is substantially lower than the 8 haplotypes that could be detected by capillary electrophoresis of cpDNA fragments.

cpDNA fragment length variation in other species

Bitterbrush may represent an unusual case where the correspondence between length polymorphism and haplotype detection is a product of low nucleotide diversity or a disproportionately high rate of indels per substitution. In this regard, comparisons with previously published data are revealing. Population-level sequence data from 13 cpDNA loci and 4 diverse species covers a wide range of nucleotide diversity (Table 2), with a low of 0.0010 (*Lilium*) to a high of 0.0106 (*Arabis*). Length polymorphism was comparably diverse, with variable loci showing between 2 (*trnH–psbA* from *C. breweri*) and 11 (*S. latifolia* and *A. holboellii*) vari-

Table 2. Summary of observed nucleotide and length variation in cpDNA from 5 intraspecific studies.

Species	Variable loci	Aligned length (bp)	Variable nucleotide positions (π)[a]	No. of inferred indels arising from simple repeats[b]	No. of inferred indels arising from tandem duplications[c]	No. of inferred indels arising from other indels	Observed lengths (bp)	No. of length haplotypes / no. of sequence haplotypes	% total haplotypes detected by FLP[d]
P. tridentata (*n*=16)	*accD–psaI*	1198	0	0	3	1	1098, 1137, 1143, 1153, 1176		
	trnH– psbA	462	2	4	0	0	454, 455, 456, 460		
	trnK intron	904	2	2	0	0	899, 902, 904		
	trnL–trnF	388	1	0	0	0	388		
	rpl16 intron	787	1	0	0	0	787		
Cumulative			6 (0.0005)	6	3	1		8/9	89
C. breweri (*n* = 9)	*rpl16* intron	796	5	0	0	0	796		
	trnH–psbA	598	3	1	0	0	597, 598		
	trnL–trnF	414	3	0	1	1	407, 413, 414		
	trnT–trnL	842	7	2	0	1	836, 837, 838, 841		
Cumulative			18 (0.0019)	3	1	2		6/6	100
L. philadelphicum (*n* = 22)	trnD–trnY	363	1	0	1	1	338, 362, 363		
	trnH–psbA	427	1	1	0	2	369, 370, 426		
	trnT–trnL	521	1	0	0	0	521		
Cumulative			3 (0.0010)	1	1	3		5/7	71
S. latifolia (*n* = 25)	*trnL* intron	576	4	2	3	2	526, 527, 528, 532, 534, 535, 539, 540, 542, 545, 548		
	trnL–trnF	391	7	0	2	1	366, 385, 386, 391		
	trnH–psbA	198	13	3	0	3	185, 186, 187, 188, 198		
	trnG–trnS	428	6	2	3	3	398, 408, 409, 410, 411, 416, 418		
Cumulative			30 (0.0027)	7	8	9		22/24	92
A. holboellii (*n* = 25)	*trnL–trnF*	953	40	0	0	3	860, 864, 874, 877, 880, 881, 887, 890, 893, 895, 896		
	rpoC1	746	22	0	1	4	717, 719, 720, 721, 726, 732, 734, 735		
Cumulative			62 (0.0106)	0	1	4		20/22	91

[a]Average P distance (nucleotide diversity, or π) calculated across all loci and samples.
[b]Simple repeats include variable length single nucleotide runs (e.g., A_n), as well as dinucleotide repeats.
[c]Tandem duplications include perfect or near perfect duplications from 3–55 bp in length.
[d]FLP, fragment length polymorphism.

Table 3. Resolution of three *Purshia tridentata* chloroplast haplotypes using length polymorphism analysis of fluorescently labeled cpDNA fragments.

Accession	*N* per haplotype	[HEX]*trnH–psbA*		accD–psaI[6FAM][a]	
		Size±SD (bp)	99% CI	Size±SD (bp)	99% CI
8A	2	508.39±0.035	508.32–508.45	660.93±0.408	660.50–661.35
B23-2, S1	4	509.04±0.090	508.92–509.16	637.67±0.167	637.45–637.89
B18, B20	4	510.11±0.107	509.97–510.25	660.93±0.408	660.50–661.35

Note: PCR products from [HEX]*trnH–psbA* are full length, whereas *accD–psaI*[6FAM] products were digested with *Eco*RI to reduce the length. All samples were run in duplicate. SD, standard deviation; CI, confidence interval; *N*, the number of replicate fragment analysis runs per haplotype.

[a]Size information is not provided for the accD fragment, which was invariant across these samples.

ants per locus, and a median value of 3.5. Only 2 locus–taxon combinations showed length constancy (*rpl16* intron, *C. breweri*; *trnT–trnL* intron, *L. philadelphicum*).

As found with *Purshia*, the number of haplotypes inferred from length polymorphism alone is comparable to the number revealed by direct sequencing, ranging from a high of 100% to a low of 71%. Most efficiently, *Calamagrostis breweri* showed 18 variable nucleotide positions and 6 inferred indels across 2650 aligned bases from 4 loci that combined to identify 6 haplotypes. Notably, length variants at *trnH–psbA* (2), *trnL–trnF* (3), and *trnT–trnL* (4) can be combined to reveal the same 6 haplotypes. Length variation at 4 cpDNA loci encompassing 1593 aligned bp was estimated for *S. latifolia*, and the number of length-inferred haplotypes was nearly identical to the number inferred from direct sequencing (22 of 24 haplotypes recovered, 92%). Length variation at 2 cpDNA regions encompassing 1699 aligned bp from *A. holboellii* showed a similar correspondence, with 20 of 22 (91%) haplotypes marked by unique length combinations. Data from *L. philadelphicum* showed the largest discrepancy between length polymorphism screening and direct sequencing, as only 5 of 7 (71%) haplotypes could be identified. In the *L. philadelphicum* example, low detection efficiency by length polymorphism could be attributable to the inclusion of only 2 variable loci, the extremely low nucleotide diversity ($\pi = 0.0010$), or the relatively small sample of nucleotides (1311 aligned bp) compared with other taxa (Table 2). Averaging across these 5 data sets, length polymorphism alone reveals 61 of 67 total haplotypes.

Discussion

Many approaches have been used to infer cpDNA haplotype diversity, including RFLPs (Mason-Gamer et al. 1995; Maskas and Cruzan 2000), PCR–RFLPs (Bacles et al. 2004), cpSSRs (Mengoni et al. 2001), PCR – single strand conformational polymorphism (PCR–SSCP; Isoda et al. 2000; Shiraishi et al. 2001), direct sequencing (Gutiérrez Larena et al. 2002), and combinations thereof (Grivet and Petit 2003). Of these approaches, traditional RFLP and PCR–RFLP studies frequently use agarose gel electrophoresis, a medium that is insensitive to single-nucleotide differences. Because changes in simple sequence repeat length rarely result in restriction site changes, RFLP-based methods would only be capable of detecting abundant polymorphisms created by simple sequence repeats if capillary or acrylamide electrophoresis were used.

The remaining methods can detect minute length variation in cpDNA amplification products but each has its own drawbacks. SSCP analysis reveals length and nucleotide differences, but the method requires optimization and is complicated by multi-banded phenotypes. A major drawback of SSCP is the inherent size limitation (≤300 bp) of fragments that can be accurately resolved. Chloroplast microsatellite analyses have proven powerful in revealing indel variation (Weising and Gardner 1999; Provan et al. 2001; Grivet and Petit 2003), but attempts to develop "universal" microsatellite primers in angiosperms has met with limited success (Provan et al. 2004). Also, substantial length variation may lie outside regions containing single-nucleotide repeats. This is demonstrated clearly in *Purshia tridentata*, where a larger number of haplotypes are marked by complex indels (5 in *accD–psaI*) than are marked by single-nucleotide repeats (4 in the combined *trnH–psbA* and *trnK* intron regions). Direct sequencing is the most informative method, but the high cost and typically low levels of sequence variation (<2 variable sites/kb in *Purshia tridentata*) makes this an expensive option.

The method described herein capitalizes on the simplicity of fragment length screening and the tendency for cpDNA to accumulate length variation at a rate equivalent to nucleotide substitutions. Our evaluation of intraspecific cpDNA sequence variation from 5 species shows that the number of haplotypes inferred from multilocus length combinations closely approximates the actual number of haplotypes detected by direct sequencing, especially when 3 or more variable regions are examined. Once identified, length variants can be treated as classical haplotypic markers (for calculating F_{ST} or for use in AMOVA) in population genetic studies, or they can be sequenced as required by phylogenetically explicit methods.

This screening method is only useful to the extent that the rate of indel formation is comparable to the rate of nucleotide substitution. For this reason, the method cannot be applied to coding regions or other regions of the chloroplast genome with length constraints. In addition, the method may show less promise if species accumulate indels at a markedly lower rate than substitutions. This phenomenon has been described for *Oryza sativa*, which averages 2.7 substitutions per indel event (72 substitutions, 27 indels) across the genomes of 4 cultivars (Tang et al. 2004). It should be noted that our comparison includes sequences that span a wide range in the ratio of substitutions per indel (7.9 for *A. holboellii*; 0.6 for *L. philadelphicum*), and the absolute

number of indels detected per kilobase of sequence (15.1 in *S. latifolia*; 2.2 in *P. tridentata*). In this context, we feel the method should prove useful across flowering plants.

cpDNA length polymorphism scanning: putting principles into practice

To maximize the efficiency of this method, one would ideally amplify and survey a large number of universal cpDNA products for length variation. In practice, our survey of 5 species shows that as few as 3 polymorphic regions could reveal a large proportion of haplotypic variation. By labeling universal cpDNA amplification primers with fluorescent tags (e.g., 6FAM, HEX), amplification products can be multiplexed and variation can be evaluated at several regions in a single run. Alternatively, PCR products could be labeled by incorporating fluorescently labeled deoxynucleotide triphosphate into the reaction (Wu et al. 2003). The only significant limitation of this approach is that single-nucleotide resolution is limited to DNA fragments smaller than most cpDNA introns and spacers (which often exceed 1kb; Shaw et al. 2005). For example, Applied Biosystems' product literature suggests that an ABI3100 equipped with a 36 cm capillary can resolve 400 bp fragments with a standard deviation of 0.15 bp. Our results show that similar precision can be obtained for larger fragments, but single nucleotide resolution is probably not feasible above 650 bp (Table 3). Because of this limitation, larger introns and spacers should be amplified as smaller fragments using internal primers, or reduced in size with restriction enzymes that cut once per fragment. Using this generalized approach, we routinely screen samples at 4 or more loci (>2500 bp) in a single capillary run. In our laboratory, this efficient approach provides a substantial cost savings, as the per-sample cost of fragment analyses for screening 2500 bp is currently less than 2% that of DNA sequencing at our core facility (Oregon State University Center for Gene Research and Biotechnology; http://www.cgrb.orst.edu/).

Acknowledgements

This research was funded by the U.S. Forest Service and the USDA NRI competitive grants program (05-35101-15341). The authors thank R. Huot and S. Huber for their assistance in the laboratory. Our use of trade names is to provide information to the reader and does not constitute endorsement or preferential treatment of the named products.

References

Avise, J.C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York, N.Y.

Bacles, C.F.E., Lowe, A.J., and Ennos, R.A. 2004. Genetic effects of chronic habitat fragmentation on tree species: the case of *Sorbus aucuparia* in a deforested Scottish landscape. Mol. Ecol. **13**: 573–584.

Birky, C., Maruyama, T., and Fuerst, P. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics, **103**: 513–527.

Clegg, M., Gaut, B., Learn, G., Jr., and Morton, B. 1994. Rates and patterns of chloroplast DNA evolution. Proc. Natl. Acad. Sci. U.S.A. **91**: 6795–6801.

Clement, M., Posada, D., and Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9**: 1657–1659.

Cronn, R.C., Small, R.L., Haselkorn, T., and Wendel, J.F. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am. J. Bot. **89**: 707–725.

Cronn, R., Horning, M., and Huot, R. 2004. Patterns of genetic diversity in hexaploid *Calamagrostis breweri* (Poaceae), a sensitive plant of USFS region 6. Unpublished annual report for the Willamette and Mt. Hood National Forests, Pacific Northwest Research Station, Corvallis, Ore.

Demesure, B., Sodzi, N., and Petit, R.J. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. **4**: 129–131.

Dobeš, C.H., Mitchell-Olds, T., and Koch, M.A. 2004. Extensive chloroplast haplotype variation indicates pleistocene hybridization and radiation of North American *Arabis drummondii*, *A.* × *divaricarpa*, and *A. holboellii* (Brassicaceae). Mol. Ecol. **13**: 349–370.

Dumolin-Lapègue, S., Demesure, B., Fineschi, S., Le Corre, V., and Petit, R. 1997*a*. Phylogeographic structure of white oaks throughout the European continent. Genetics, **146**: 1475–1487.

Dumolin-Lapègue, S., Pemonge, M.-H., and Petit, R.J. 1997*b*. An enlarged set of consensus primers for the study of organelle DNA and plants. Mol. Ecol. **6**: 393–397.

Grivet, D., and Petit, R.J. 2003. Chloroplast DNA phylogeography of the hornbeam in Europe: evidence for a bottleneck at the outset of postglacial colonization. Con. Genet. **4**: 47–56.

Gutiérrez Larena, B., Fuertes Aguilar, J., and Nieto Feliner, G. 2002. Glacial-induced altitudinal migrations in *Armeria* (Plumbaginaceae) inferred from patterns of chloroplast DNA haplotype sharing. Mol. Ecol. **11**: 1965–1974.

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

Hamilton, M.B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol. Ecol. **8**: 513–525.

Hamilton, M., and Miller, J. 2002. Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. Genetics, **162**: 1897–1909.

Hamilton, M.B., Braveman, J.M., and Soria-Hernanz, D.F. 2003. Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in the new world species of the Lecythidaceae. Mol. Bio. Evol. **20**: 1710–1721.

Horning, M.E. 2003. A genetic approach to the conservation and management of *Lilium philadelphicum* in fragmented habitats. Ph.D. dissertation, Washington State University, Pullman, Wash.

Ingvarsson, P.K., and Taylor, D.R. 2002. Genealogical evidence for epidemics of selfish genes. Proc. Natl. Acad. Sci. U.S.A. **99**: 11265–11269.

Ingvarsson, P.K., Ribstein, S., and Taylor, D.R. 2003. Molecular evolution of insertions and deletion in the chloroplast genome of *Silene*. Mol. Bio. Evol. **20**: 1737–1740.

Isoda, K., Shiraishi, S., Watanabe, S., and Kitamura, K. 2000. Molecular evidence of natural hybridization between *Abies veitchii* and *A. homolepis* (Pinaceae) revealed by chloroplast, mitochondrial and nuclear DNA markers. Mol. Ecol. **9**: 1965–1974.

Jordan, W., Courtney, M., and Neigel, J. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweds (Lemnaceae). Am. J. Bot. **83**: 430–439.

Kelchner, S.A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. Am. J. Bot. **89**: 1651–1669.

Kelchner, S.A., and Wendel, J.F. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. Curr. Genet. **30**: 259–262.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. 2001. MEGA2: molecular evolutionary genetics software. Bioinformatics, **17**: 1244–1245.

Lis, J.T. 1980. Fractionation of DNA fragments by polyethylene glycol induced precipitation. Meth. Enzym. **65**: 347–353.

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

Maskas, S.D., and Cruzan, M.B. 2000. Patterns of intraspecific diversification in the *Piriqueta Caroliniana* complex in southeastern North America and the Bahamas. Evolution, **54**: 815–827.

Mason-Gamer, R., Holsinger, K., and Jansen, R. 1995. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). Mol. Bio. Evol. **12**: 371–381.

McCauley, D.E. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends Ecol. Evol. **10**: 198–202.

Mengoni, A., Barabesi, C., Gonnelli, C., Galardi, F., Gabbrielli, R., and Bazzicalupo, M. 2001. Genetic diversity of heavy metal-tolerant populations in *Silene paradoxa* L. (Caryophyllaceae): a chloroplast microsatellite analysis. Mol. Ecol. **10**: 1909–1916.

Provan, J., Soranzo, N., Wilson, N., Goldstein, D., and Powell, W. 1999. A low mutation rate for chloroplast microsatellites. Genetics, **153**: 943–947.

Provan, J., Powell, W., and Hollingsworth, P.M. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends. Ecol. Evol. **16**: 142–147.

Provan, J., Biss, P.M., McMeel, D., and Mathews, S. 2004. Universal primers for the amplification of chloroplast microsatellites in grasses (Poaceae). Mol. Ecol. Notes, **4**: 262–264.

Schaal, B.A., Hayworth, D.A., Olsen, K.M., Rauscher, J.T., and Smith, W.A. 1998. Phylogeographic studies in plants: problems and prospects. Mol. Ecol. **7**: 465–474.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E., and Small, R.L. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. **92**: 142–166.

Shiraishi, S., Maeda, H., Toda, T., Seido, K., and Sasaki, Y. 2001. Incomplete paternal inheritance of chloroplast DNA recognized in *Chamaecyparis obtusa* using an instraspecific polymorphism of the *trnD–trnY* intergenic spacer region. Theor. Appl. Genet. **102**: 935–941.

Simmons, M.P., and Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analysis. Syst. Biol. **49**: 369–381.

Small, R., Ryburn, J., Cronn, R., Seelanan, T., and Wendel, J. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. **85**: 1301–1315.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends Ecol. Evol. **15**: 199–203.

Taberlet, P., Ludovic, G., Pautou, G., and Bouvet, J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. **17**: 1105–1109.

Tang, J., Xia, H., Cao, M., Zhang, X., Zeng, W., Hu, S., Tong, W., Wang, J., Yu, J., Yang, H., and Zhu, L. 2004. A comparison of rice chloroplast genomes. Plant Physiol. **135**: 412–420.

Tate, J., and Simpson, B. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Syst. Bot. **28**: 723–737.

Trewick, S.A., Morgan-Richards, M., Russell, S.J., Henderson, S., Rumsey, S.J., Pinter, I., Barrett, J.A., Gibby, M., and Vogel, J.C. 2002. Polyploidy, phylogeography and pleistocene refugia of the rockfern *Asplenium ceterach*: evidence from chloroplast DNA. Mol. Ecol. **11**: 2003–2012.

Weising, K., and Gardner, R.C. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome, **42**: 9–19.

Wolfe, K., Li, W.-H., and Sharp, P. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. U.S.A. **84**: 9054–9058.

Wu, X., Larson, S., Hu, Z., Palazzo, A.J., Jones, T.A., Wang, R.R.-C., Jensen, K.B., and Chatterton, N.J. 2003. Molecular genetic linkage maps for allotetraploid *Leymus* wildryes (Gramineae: Triticeae). Genome, **46**: 627–646.

Young, J., and Clements, C. 2002. *Purshia*: the wild and bitter rose. University of Nevada Press, Reno, Nev.