AJB Advance Article published on June 17, 2011, as 10.3732/ajb.1100088.
The latest version is at http://www.amjbot.org/cgi/doi/10.3732/ajb.1100088

American Journal of Botany

American Journal of Botany: e000–e000. 2011.

AJB Primer Notes & Protocols in the Plant Sciences

Newly developed primers for complete *ycf*1 amplification in *Pinus* (Pinaceae) chloroplasts with possible family-wide utility[1]

Matthew Parks[2,4], Aaron Liston[2], and Rich Cronn[3]

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331-2902 USA; and [3]Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon 97331 USA

- *Premise of the study:* Primers were designed to amplify the highly variable locus *ycf*1 from all 11 subsections of *Pinus* to facilitate plastome assemblies based on short sequence reads as well as future phylogenetic and population genetic analyses.
- *Methods and Results:* Primer design was based on alignment of 33 *Pinus* and four Pinaceae plastomes with mostly incomplete *ycf*1 sequences. Sanger sequencing of 12 *Pinus* accessions resulted in open reading frames ranging in size from 5.2 to 6.1 kbp. Highest sequence diversity was identified in two regions totaling 5.5 kbp aligned length, which can be targeted in all pine subsections with three primer combinations. Preliminary results suggest the primers described also amplify homologous targets in the broader Pinaceae.
- *Conclusions:* The successful design and implementation of PCR primers spanning the large, variable locus *ycf*1 in *Pinus* represents the development of a valuable tool in pine genetic studies, and should facilitate studies throughout Pinaceae.

Key words: chloroplast; Pinaceae; *Pinus*; *ycf*1.

"Next-generation" sequencing technologies, which feature nucleotide sequence output measured in millions to billions of base pairs, are revolutionizing DNA- and RNA-based research. However, the short-read length characteristic of these technologies (currently tens to several hundreds of base pairs) often makes it difficult or impossible to accurately assemble and characterize regions highly divergent from available reference sequences. For phylogenetic or population genetic pursuits, such regions are often of great utility. In the genus *Pinus*, the locus *ycf*1 was identified as highly variable based on nearly complete plastome sequences assembled from Illumina short-read sequence data (Parks et al., 2009), and subsequently used to study the phylogeny of the species-rich subsection *Ponderosae* (Gernandt et al., 2009). Nonetheless, a fuller accounting of this locus is desirable, as the *ycf*1 sequences assembled by Parks et al. (2009) and Cronn et al. (2008) contained a substantial portion of undetermined positions (largely due to difficulties in assembly of short reads), and the analyses of Gernandt et al. (2009) used only ca. 20% of total *ycf*1 sequence length.

We designed 14 primers to allow amplification of the entire *ycf*1 locus from all 11 of the currently recognized *Pinus* subsections (Gernandt et al., 2005). Our sequencing efforts focused on 10 of the 11 subsections, as a complete *ycf*1 sequence is already available for subsection *Pinus* (GenBank record NC_001631.1, Wakasugi et al., 1994). In addition, although a complete *ycf*1 sequence is also available for subsection *Strobus* (*Pinus koraiensis*, GenBank record NC_004677.2), we sequenced two additional members of this subsection to verify a repetitive region reported in *P. koraiensis*. Amplicons were subsequently sequenced using Sanger technology and combined to assemble complete or nearly complete reading frames of *ycf*1 for each accession. Potential applicability of these primers to the broader Pinaceae was also investigated by alignment of *ycf*1 sequences to four non-*Pinus* members of Pinaceae and PCR amplifications using two primer pairs.

[1]Manuscript received 23 February 2011; revision accepted 7 March 2011.

The authors would like to thank David Gernandt for sharing partial ycf1 sequences from *Pinus ponderosa*. This work was supported by NSF ATOL-0629508.

[4]Corresponding author: parksma@science.oregonstate.edu

doi:10.3732/ajb.1100088

METHODS AND RESULTS

Total genomic DNA was extracted from frozen leaf or mega-gametophyte tissues of 12 accessions representing 10 of the 11 *Pinus* subsections sensu Gernandt et al. (2005) (Appendix 1) using the standard FastDNA extraction protocol (MP Biomedicals, Solon, Ohio, USA).

Fourteen primers (Table 1, Fig. 1) were designed to cover the *ycf*1 locus based on conserved regions in an alignment of 33 *Pinus* species and four non-pine outgroups from the Pinaceae (Parks et al., 2009) (GenBank FJ899555-FJ899583, EU998739-EU998746). PCR amplifications using various combinations of these primers were performed with either Phusion DNA polymerase and Phusion buffer HF or Taq polymerase and Thermopol buffer (New England Biolabs, Ipswich, Massachusetts, USA). Generally, PCR reactions were carried out as: 30 s 98°C (one cycle), 8 s 98°C, 30 s 55–59°C, 30 s/kb 72°C (25–30 cycles), 5 min 72°C (1 cycle), final hold temperature of 4°C. For problematic amplifications, several strategies were pursued, including: (1) varying annealing temperatures (max 60°C, min 52°C); (2) use of alternative buffer or additives (for example, Phusion GC buffer with 0.25 µl 100% DMSO / 50 µl reaction volume or addition of 0.1 µl BSA (10 mg/ml) / 50 µl reaction volume when using Taq polymerase); (3) pairing primers herein with alternative primers designed to target the *Pinus* plastome in the vicinity of *ycf*1 (Cronn et al., 2008). Amplification success was determined by gel electrophoresis using 1% agarose gels stained with GelRed (ca. 1:30 000 v/v) (Phenix Research Products, Candler North Carolina, USA) and run for 30 min at 110 V. Successful amplifications (i.e., strong, single bands) were submitted for Sanger sequencing to the University of Washington High-Throughput Genomics Unit (www.htseq.org) using the above-described PCR primers for sequencing.

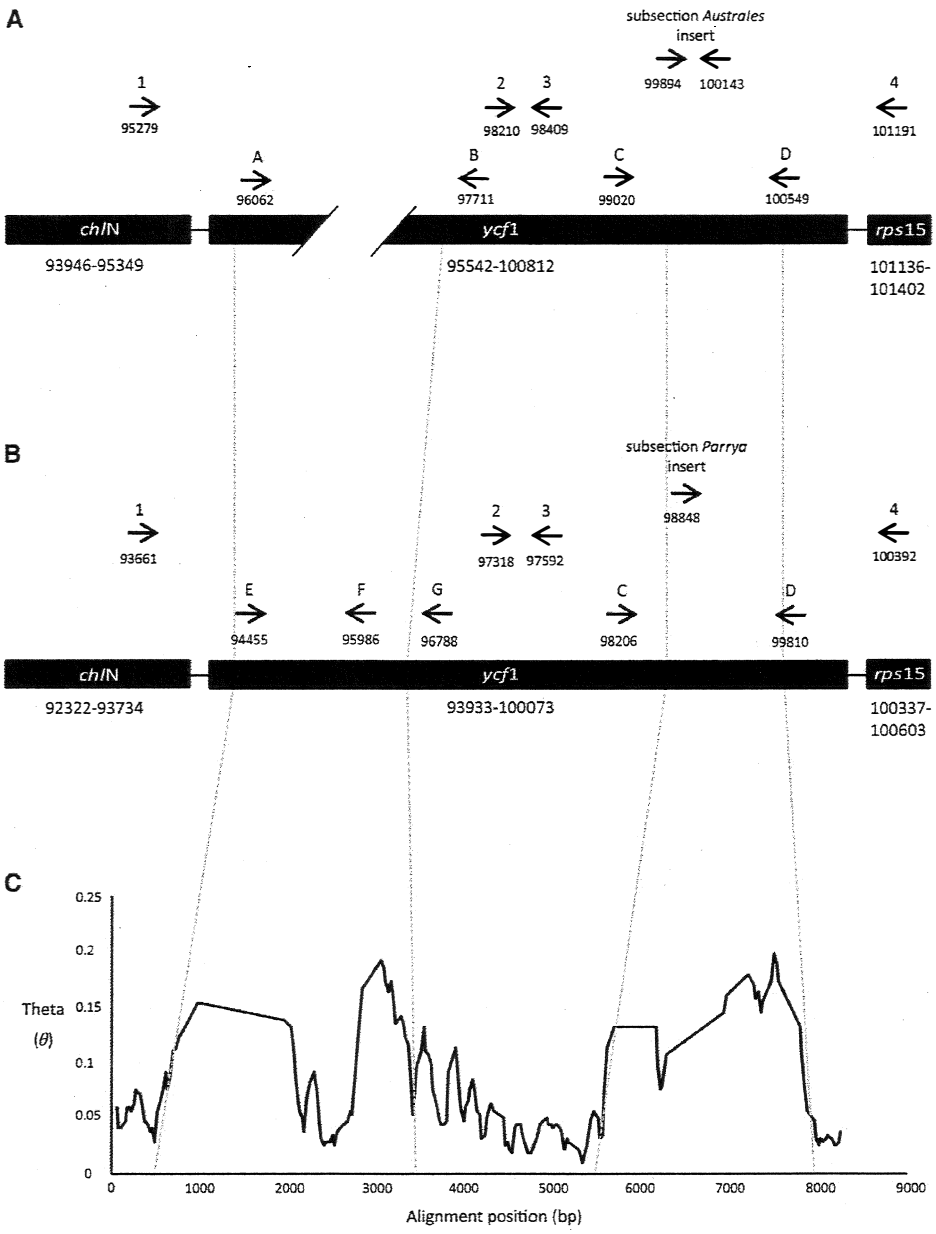

Fig. 1. A, Map of primer locations used in *ycf*1 amplifications in subgenus *Pinus*. Coordinates correspond to locations in the *Pinus thunbergii* chloroplast genome (Wakasugi et al., 1994); B, Map of primer locations used in *ycf*1 amplifications in subgenus *Strobus*. Coordinates correspond to locations in the *Pinus koraiensis* chloroplast genome (GenBank NC_004677.2). In both A and B, primer coordinates represent 5′ end of the primer; gap in *ycf*1 of A) corresponds to a repetitive region of ca. 900 bp aligned length found in subgenus *Strobus* but not present in subgenus *Pinus*. C, Graph of Watterson's theta (θ) as measured in 100bp windows along the aligned length of *ycf*1 for *Pinus* accessions sequenced in this study and *Pinus thunbergii* (Wakasugi et al., 1994) and *Pinus koraiensis* (GenBank NC_004677.2). Light dotted lines indicate approximate locations of regions of high θ values in relation to primer locations in parts A and B.

Quality-trimming of sequence reads was performed by the UW High-Throughput Genomics Unit based on a quality score cutoff of Q20 (corresponding to 99% confidence in a base call), which is a commonly used cutoff for initial quality filtering (Ewing and Green, 1998; Richterich, 1998). Reads were trimmed from their 5′ and 3′ ends until windows of 50 consecutive positions contained fewer than 10 base calls < Q20. At this point the remaining sequence, including all positions in the current window, was considered to have passed quality filtering, although some remaining positions were subject to manual masking in subsequent steps (see below). The resulting filtered sequences were manually aligned to their reference genome in BioEdit v7.0.5 (Hall, 1999); in all cases, the original assembly of Cronn et al. (2008) or Parks et al. (2009) served as the reference for an accession's assembly. Assembly problems were identified by translation of the *ycf*1 reading frame as well as comparison of overlapping aligned reads where possible. Putative internal stop-codons, frame-shift mutations, and discordant base calls between overlapping aligned reads or aligned reads and their reference were investigated through examination of sequence chromatograms and/or resequencing. In some cases positions within the quality-filtered reads were masked by hand based on a combination of low quality

TABLE 1. Information for primers used in *ycf1* amplifications and sequencing. Primers from Cronn et al. (2008) were used mainly as alternative primers for difficult to amplify accessions/regions.

Region	Primer Name	Source	Primer sequence (5′–3′)
*chl*N	ycf1.1	This paper	TAGATAACTTGGATCGGACCAC
ycf1	ycf1.2	This paper	TTCCTTTTCGTTTGAAGCCTT
ycf1	ycf1.3	This paper	TCTTATTCCTGTAGATCCCATCAAT
rps15	ycf1.4	This paper	GATCCTCTCTGTTTATCGGGAA
ycf1	ycf1.A	This paper	TGGGCGGTCATATTCTATT
ycf1	ycf1.B	This paper	TTAAGTTCCGACGATAATCTG
ycf1	ycf1.C	This paper	AAGATTTTGAAATTCGTCCTG
ycf1	ycf1.D	This paper	TACGACGTTTTGGAAGC
ycf1	ycf1.E	This paper	GGCGGTCATATTCTATTCAT
ycf1	ycf1.F	This paper	TGCCAATGCTCAGAGATA
ycf1	ycf1.G	This paper	CTCGGCATGATAACGTTT
ycf1	subs. *Australes* insert F	This paper	GAAGGAAACAACAAATGTTCTAG
ycf1	subs. *Australes* insert R	This paper	CATAACCCTGCAAATATTCG
ycf1	subs. *Parrya* insert F	This paper	GATCCGGATTAGATTTAAAATTCTGG
*trn*N-GUU	28F	(Cronn et al., 2008)	TTAACAGCCGACCGCTCTAC
ycf1	28R	(Cronn et al., 2008)	GTAGAGCGGTCGGCTGTTA
ycf1	29F	(Cronn et al., 2008)	TCCCGTATTAACAAGACTGGTG
ycf1	29R	(Cronn et al., 2008)	CCAGTCTTGTTAATACGGGATTT
ycf1	30F	(Cronn et al., 2008)	TTGGATCACGAAAAACCACA
*psa*C	30R	(Cronn et al., 2008)	TGTGGTTTTTCGTGATCCAA

TABLE 2. *ycf1* sequencing and assembly success for accessions representing *Pinus* subsections. (P) and (S) indicate subgenus *Pinus* and *Strobus*, respectively.

Subsection	Species	Estimated length of ycf1 (bp)	Estimated number of undetermined positions
Australes (P)	*Pinus taeda*	5259	350
Contortae (P)	*P. contorta*	5547	0
Pinaster (P)	*P. canariensis*	5448	48
Pinaster (P)	*P. pinaster*	5472	48
Ponderosae (P)	*P. ponderosa*	5784	1
	Average length (SD)	5502 (190)	
Balfourianae (S)	*P. aristata*	5772	2
Cembroides (S)	*P. monophylla*	6054	0
Gerardianae (S)	*P. gerardiana*	5781	497
Krempfianae (S)	*P. krempfii*	6222	0
Nelsoniae (S)	*P. nelsonii*	6036	0
Quinquefoliae (S)	*P. flexilis*	5901	0
Quinquefoliae (S)	*P. lambertiana*	6114	0
	Average length (SD)	5983 (170)	

score (< Q20) and discordance with overlapping reads or their previously published assemblies. Sequence polymorphism of final assemblies (measured as Watterson's θ) was evaluated using the program DnaSP v.5.10.00 (Librado and Rozas, 2009).

Alignment of sequence reads allowed complete or nearly complete assembly of the *ycf1* locus from 12 *Pinus* accessions representing all of the targeted *Pinus* subsections (Table 2). Assembled lengths of *ycf1* ranged from ca. 5.2 to 6.2 kbp, and averaged over 500 bp longer in subgenus *Strobus* than in subgenus *Pinus* (Table 2). Areas of highest sequence diversity in aligned *Pinus ycf1* were found between positions 500–3500 and 5500–8000 of the 8.3 kbp alignment. These regions are mostly bounded by the primers A to B (E to G in subgenus *Strobus*), and C to D, respectively (Fig. 1).

Alignment of primer sequences to four non-pine members of Pinaceae (*Picea sitchensis* (Bong.) Carr., *Larix occidentalis* Nutt., *Cedrus deodara* (Roxb.) G. Don, and *Abies firma* Siebold and Zucc., GenBank identifiers included above) suggested that most of the primers should be effective throughout the family. For primers 1–4 and A–G, each outgroup accession on average differed at 1.18 ± 1.23 (SD) positions in the primer sequence, while the average distance of these differences from the primer's 3′ end was 10.21 ± 4.98 bp (SD). Using the described PCR strategies, all four non-pine species tested also successfully PCR-amplified with each of two primer combinations (ycf1.1/ycf1.3 and ycf1.2/ycf1.4), resulting in amplicons of expected size.

CONCLUSIONS

Based on our results, the primers described should be useful tools in studies employing *ycf1* as a phylogenetic or population genetic marker in species of *Pinus*, and likely throughout Pinaceae. While full sequencing of this locus may require additional primers, a substantial portion of the most variable regions of this locus can be targeted with a small number of primers. In addition, the complete to nearly complete subsectional *ycf1* sequences generated should aid in reference-guided assembly from next-generation sequencing data in future projects targeting Pinaceae chloroplast genomes or the *ycf1* locus specifically.

LITERATURE CITED

CRONN, R., A. LISTON, M. PARKS, D. S. GERNANDT, R. SHEN, AND T. MOCKLER. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36: e122.

EWING, B., AND P. GREEN. 1998. Base-calling of automated sequencer traces using Phred.II. Error probabilities. *Genome Research* 8: 186–194.

GERNANDT, D. S., S. HERNÁNDEZ-LEÓN, E. SALGADO-HERNÁNDEZ, AND J. A. AND P. L. ROSA. 2009. Phylogenetic relationships of *Pinus* subsection *Ponderosae* inferred from rapidly evolving cpDNA regions. *Systematic Botany* 34: 481–491.

GERNANDT, D. S., G. LOPEZ, S. O. GARCIA, AND A. LISTON. 2005. Phylogeny and classification of *Pinus*. *Taxon* 54: 29–42.

HALL, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.

LIBRADO, P., AND J. ROZAS. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)* 25: 1451–1452.

PARKS, M., R. CRONN, AND A. LISTON. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 84.

RICHTERICH, P. 1998. Estimation of errors in "raw" DNA sequences: A validation study. *Genome Research* 8: 251–259.

WAKASUGI, T., J. TSUDZUKI, S. ITO, K. NAKASHIMA, T. TSUDZUKI, AND M. SUGIURA. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences, USA* 91: 9794–9798.

APPENDIX 1. Voucher information for species of *Pinus* used in *ycf1* amplifications and sequencing. Institutions of voucher deposit are in parentheses.

Taxon; Voucher (Herbarium); GenBank accession of annotated chloroplast sequence.

Pinus aristata Engelmann; *K. Farrell 30* (OSC213681); FJ899567. **Pinus canariensis** C. Smith; *A. Ensoll, M. Maquire, F. Nelson & A. Wright 1* (E); FJ899572. **Pinus contorta** Douglas ex. Loudon; *A. Liston 1315* (OSC218800); EU998740. **Pinus flexilis** James; *K. Farrell 28* (OSC213724); FJ899576. **Pinus gerardiana** Wallich ex. D. Don; *R. Businsky 41123* (RILOG); EU998741. **Pinus krempfii** Lecomte; *RBG Edinburgh First Darwin Expedition 242* (E); EU998742. **Pinus lambertiana** Douglas; *USFS Region 5 Camino Seed Orchard* (OSC213763); EU998743. **Pinus monophylla** Torrey & Fremont; *D. Gernandt 479* (OSC213555); EU998745. **Pinus nelsonii** Shaw; *D. Gernandt 10198-15098* (OSC202126); EU998746 . **Pinus pinaster** Aiton; *R. Businsky 36157a* (RILOG); FJ899583. **Pinus ponderosa** Douglas ex P. & C. Lawson; *USFS R1 Geneticist* (OSC213789); FJ899555. **Pinus taeda** L.; *Southern Institute of Forest Genetics, clone 7-56, LB-A10-05* (OSC226921); FJ899561.