A Long PCR–Based Approach for DNA Enrichment Prior to Next-Generation Sequencing for Systematic Studies

Authors: Uribe-Convers, Simon, Duke, Justin R., Moore, Michael J., and Tank, David C.

Source: Applications in Plant Sciences, 2(1)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1300063

Applications in Plant Sciences

Applications in Plant Sciences 2014 2(1): 1300063

Protocol Note

A long PCR–based approach for DNA enrichment prior to next-generation sequencing for systematic studies[1]

Simon Uribe-Convers[2,3,5], Justin R. Duke[3], Michael J. Moore[4], and David C. Tank[2,3]

[2]Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, Idaho 83844-3051 USA; [3]College of Natural Resources, University of Idaho, 875 Perimeter Drive MS 1133, Moscow, Idaho 83844-1133 USA; and [4]Department of Biology, Oberlin College, Science Center K111, 119 Woodland St., Oberlin, Ohio 44074-1097 USA

- *Premise of the study:* We present an alternative approach for molecular systematic studies that combines long PCR and next-generation sequencing. Our approach can be used to generate templates from any DNA source for next-generation sequencing. Here we test our approach by amplifying complete chloroplast genomes, and we present a set of 58 potentially universal primers for angiosperms to do so. Additionally, this approach is likely to be particularly useful for nuclear and mitochondrial regions.
- *Methods and Results:* Chloroplast genomes of 30 species across angiosperms were amplified to test our approach. Amplification success varied depending on whether PCR conditions were optimized for a given taxon. To further test our approach, some amplicons were sequenced on an Illumina HiSeq 2000.
- *Conclusions:* Although here we tested this approach by sequencing plastomes, long PCR amplicons could be generated using DNA from any genome, expanding the possibilities of this approach for molecular systematic studies.

Key words: angiosperms; chloroplast enrichment; long PCR; next-generation sequencing; plastome; universal chloroplast PCR primers.

Advancements in next-generation sequencing (NGS) technologies have permitted the assembly of large, genome-scale data sets that have shed light on the evolutionary history of many taxa (e.g., Parks et al., 2009; Moore et al., 2010; Xi et al., 2012; Eaton and Ree, 2013; Tennessen et al., 2013). For plant phylogenetics, there has been a major focus on methods for chloroplast phylogenomics (e.g., Parks et al., 2009; Moore et al., 2010), although methods for collecting phylogenomic data sets from the nuclear and mitochondrial genomes have also been developed (e.g., Straub et al., 2012; Eaton and Ree, 2013). Stull et al. (2013) developed a custom RNA probe set designed to capture angiosperm plastomes via solution-based hybridization. While their capture system was broadly successful, Stull et al. (2013) found that the most variable spacer regions were often captured at much-reduced coverage compared to more conserved regions, and were sometimes missed entirely if the target taxon was phylogenetically divergent from one of the 22 plastomes used in the bait design. Moreover, the current cost of the capture probes makes this method most efficient for projects dealing with hundreds of species. Another commonly employed method for plant phylogenomic studies is genome skimming (Straub et al., 2012), which takes advantage of the fact that organellar DNA and nuclear ribosomal DNA are present at high copy numbers in genomic DNA. However, a significant limitation of this method for systematic studies is that only high-copy number regions are recovered consistently across all samples, whereas regions with lower representation are only recovered in some samples and missed completely in others (Straub et al., 2011). This can be problematic for molecular systematic studies where missing data may result in misleading phylogenetic results (Lemmon et al., 2009). Moreover, being limited to high-copy regions in the genome becomes restrictive for experimental design as it excludes putatively highly informative regions in the genome such as single-copy nuclear genes (e.g., the single-copy orthologous genes [COSII] and the pentatricopeptide repeat [PPR] gene family; Wu et al., 2006, and Yuan et al., 2009, respectively).

As an alternative, we present an NGS approach that combines long PCR and Illumina sequencing to strategically compile phylogenomic data sets for molecular systematic studies. Long PCR, or long-range PCR, uses a combination of two polymerases—a nonproofreading polymerase at high concentration and a proofreading polymerase at a lower concentration—to amplify DNA fragments that range between 3 and 15 kilobases (kb), although cases of extremely large fragments (22–42 kb) have been reported (e.g., Cheng et al., 1994). Long PCR has been used extensively in human genome projects (e.g., Craig et al., 2008) and to sequence complete mitochondrial genomes (e.g., Knaus et al., 2011; Alexander et al., 2013), using both

[1]Manuscript received 26 July 2013; revision accepted 3 December 2013.

The authors thank R. Cronn, T. C. Peterson, D. F. Morales-Briones, S. J. Jacobs, H. E. Marx, and four anonymous reviewers for helpful comments on the manuscript, and the University of Idaho Institute for Bioinformatics and Evolutionary Studies (NIH/NCRR P20RR16448 and P20RR016454) and the iPlant Collaborative (NSF DBI-0735191) for bioinformatic resources. Funding for this work was provided by the National Science Foundation (DEB-1210895 to D.C.T. for S.U.C., and DEB-1253463 to D.C.T.) and by the University of Idaho Student Grant Program and Seed Grant Program to S.U.C. and D.C.T., respectively.

[5]Author for correspondence: uribe.convers@gmail.com

doi:10.3732/apps.1300063

TABLE 1. List of species included in this study, with voucher information, tissue sources, and NGS assembly statistics when available.[a]

Species	Order/Family	Collection no.	Herbarium	Type of tissue	Collection date	No. of amplified regions	Region no. not amplified[b]	Base pairs sequenced[c]	No. of contigs	CAL bp (min–max)	Ave. assembly depth	No. of masked bp[d]	% of masked bp	N50	% called bases[e]	No. of ambiguous bases	% of ambiguous bases
Bartsia inaequalis Benth.	Lamiales/Orobanchaceae	Uribe-Convers 2010-22	ID	Silica gel–dried	5 July 2010	16	n/a	125,283	25	5011 (204–28,257)	656	2126	1.7	19,294	99.9729	34	0.02714
Castilleja covilleana L. F. Hend.	Lamiales/Orobanchaceae	Tank 1046	ID	Silica gel–dried	13 July 2009	16	n/a	133,595	10	13,360 (1222–48,767)	641	101	0.08	37,107	99.9948	7	0.00524
Castilleja elmeri Fernald	Lamiales/Orobanchaceae	Olmstead 2001-78	WTU	Silica gel–dried	4 July 2001	16	n/a	122,614	11	11,147 (464–34,602)	664	440	0.36	33,049	99.9976	3	0.00245
Castilleja linariifolia Benth.	Lamiales/Orobanchaceae	Tank 2001-49	WTU	Silica gel–dried	21 July 2001	16	n/a	122,046	8	15,256 (819–50,680)	642	260	0.21	28,529	99.9984	2	0.00164
Castilleja miniata Douglas ex Hook.	Lamiales/Orobanchaceae	Tank 1048-b	ID	Silica gel–dried	13 July 2009	16	n/a	134,704	4	33,676 (6157–75,123)	844	35	0.03	75,123	99.9970	4	0.00297
Castilleja pallescens (A. Gray) Greenm.	Lamiales/Orobanchaceae	Tank 2009-8	ID	Silica gel–dried	6 June 2009	16	n/a	125,490	4	31,372 (3039–73,629)	764	29	0.02	73,629	99.9984	2	0.00159
Bartsia stricta (Kunth) Benth.	Lamiales/Orobanchaceae	Uribe-Convers 2010-24	ID	Silica gel–dried	7 July 2010	15	13	119,828	14	8559 (425–67,195)	707	1045	0.87	67,195	99.9967	4	0.00334
Castilleja applegatei Fernald	Lamiales/Orobanchaceae	Tank 2001-35	WTU	Silica gel–dried	24 June 2001	15	10	119,647	14	8546 (204–28,559)	642	394	0.33	18,856	99.9983	2	0.00167
Castilleja virgata (Domb. ex Wedd.) Edwin	Lamiales/Orobanchaceae	Olmstead 2009-22	WTU	Silica gel–dried	5 Mar. 2009	15	7	113,650	21	5412 (178–39,914)	698	1525	1.34	14,541	99.9938	7	0.00616
Castilleja ortegae Standl.	Lamiales/Orobanchaceae	Egger 1213	WTU	Silica gel–dried	22 Feb. 2002	15	13	108,071	3	36,024 (269–97,615)	925	198	0.18	97,615	99.9991	1	0.00093
Castilleja lineariloba (Benth.) T. I. Chuang & Heckard	Lamiales/Orobanchaceae	Tank 2002-04	WTU	Silica gel–dried	27 Apr. 2004	14	9, 10	122,182	23	5312 (179–36,972)	540	810	0.66	11,656	99.9844	19	0.01555
Castilleja victoriae Fairbarns & J. M. Egger	Lamiales/Orobanchaceae	Fairbarns s.n.	WTU	Silica gel–dried	21 July 2005	14	10, 14	111,371	10	11,137 (616–44,011)	688	547	0.49	18,398	99.9982	2	0.00180
Lamourouxia virgata Kunth	Lamiales/Orobanchaceae	Zak & Jaramillo, 3387	F	Herbarium	16 Jan. 1988	14	9, 10	108,767	30	3626 (214–36,850)	652	2255	2.07	11,012	99.9669	36	0.03310
Castilleja oresbia Greenm.	Lamiales/Orobanchaceae	Tank 2001-27	WTU	Silica gel–dried	19 June 2001	10	6, 9, 10, 13, 14, 16	83,384	20	4169 (222–36,830)	717	1544	1.85	9986	99.9676	27	0.03238
Castilleja arvensis Cham. & Schltdl.	Lamiales/Orobanchaceae	Tank 2005-27	WTU	Silica gel–dried	16 Apr. 2005	6	4, 6, 7, 8, 9, 10, 13, 14, 15, 16	73,378	15	4892 (186–36,621)	701	1187	1.62	9803	99.9877	9	0.01227
Penstemon montanus Greene var. *idahoensis* (D. D. Keck) Cronq.	Lamiales/Plantaginaceae	Brunsfeld 4159	ID	Herbarium	14 June 2001	16	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Balsamorhiza sagittata (Pursh) Nutt.	Asterales/Asteraceae	Willard 2013-42	ID	Silica gel–dried	3 July 2013	15	5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Lomatium dissectum (Nutt.) Mathias & Constance	Apiales/Apiaceae	Poor 21	ID	Herbarium	27 May 2004	15	14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Nuphar polysepala Engelm.	Nymphaeales/Nymphaeaceae	Morales-Briones 412	ID	Silica gel–dried	8 July 2013	15	5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Salix scouleriana Barratt ex Hook.	Malpighiales/Salicaeae	Brunsfeld 7213	ID	Herbarium	11 June 2008	15	9	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Crataegus columbiana Howell	Rosales/Rosaceae	Hetrick 1005	ID	Herbarium	10 Apr. 1996	13	9, 14, 17	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Polygonum douglasii Greene	Caryophyllales/Polygonaceae	Smith 8040	ID	Herbarium	23 June 2005	12	5, 6, 9, 15	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Umbellularia californica (Hook. & Arn.) Nutt.	Laurales/Lauraceae	Halse 6901	ID	Herbarium	28 Mar. 2002	12	6, 8, 9, 10	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Bromus tectorum L.	Poales/Poaceae	Clippinger 2	ID	Herbarium	1 May 2004	11	5, 6, 9, 11, 17	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Alnus rhombifolia Nutt.	Fagales/Betulaceae	Gray 52	ID	Herbarium	7Aug. 1989	10	5, 6, 8, 9, 10, 14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Poa bulbosa L.	Poales/Poaceae	Willard 2013-26	ID	Silica gel–dried	3 July 2013	10	5, 6, 9, 12, 13, 14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Senecio integerrimus Nutt. var. *exaltatus* (Nutt.) Cronq.	Asterales/Asteraceae	Willard 2013-21	ID	Silica gel–dried	3 July 2013	10	3, 5, 6, 8, 9, 11	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

TABLE 1. Continued.

Species	Order/Family	Collection no.	Herbarium	Type of tissue	Collection date	No. of amplified regions	Region no. not amplified[b]	Base pairs sequenced[c]	No. of contigs	CAL bp (min–max)	Ave. assembly depth	No. of masked bp[d]	% of masked bp	N50	% called bases[e]	No. of ambiguous bases	% of ambiguous bases
Abies amabilis Douglas ex J. Forbes	Pinales/Pinaceae	1419-46	WA Park Arb.	Silica gel–dried	24 May 2009	9	4, 6, 7, 9, 10, 11, 12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Capsella bursa-pastoris (L.) Medik.	Brassicales/Brassicaceae	Brunsfeld 6313	ID	Herbarium	1 June 2005	8	4, 6, 8, 9, 10, 13, 14, 17	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Lupinus leucophyllus Douglas ex Lindl.	Fabales/Fabaceae	Willard 2013-03	ID	Silica gel–dried	3 July 2013	8	1, 6, 8, 9, 10, 12, 13, 14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Abies fraseri (Pursh) Poir.	Pinales/Pinaceae	1005-47	WA Park Arb.	Silica gel–dried	24 May 2009	7	4, 5, 6, 7, 8, 9, 10, 11, 12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Balsamorhiza hookeri Nutt.	Asterales/Asteraceae	Smith 9421	ID	Herbarium	4 June 2007	7	4, 5, 6, 7, 8, 9, 10, 11, 13	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Abies grandis (Douglas ex D. Don) Lindl.	Pinales/Pinaceae	1084-49	WA Park Arb.	Silica gel–dried	24 May 2009	6	1, 3, 4, 6, 7, 8, 9, 10, 11, 12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Average								11,493	14.13	13,166.60	698.73	833.07	0.79	35,052.87	99.99	10.6	0.01

Note: CAL = contig average length; F = Field Museum of Natural History Herbarium; ID = University of Idaho Stillinger Herbarium; WA Park Arb. = Washington Park Arboretum; WTU = University of Washington Herbarium.

[a]All data from the 16 chosen primer combinations.

[b]The number of the regions is the same as the order in Fig. 1.

[c]Base pairs (bp) sequenced is the sum of all contigs when including only one copy of the inverted repeat.

[d]Number of bases masked because the minimum sequencing depth of 5× was not achieved.

[e]Percentage of unambiguously called bases.

Sanger sequencing and NGS technologies. Here, we use long PCR to generate chloroplast DNA templates for systematic studies using NGS. While we focus on whole chloroplast amplification, this approach is directly translatable to targeted studies where only particular regions of the plastome are of interest (e.g., the inverted repeat or the small single-copy region). In addition, long PCR could also be very useful for the enrichment of mitochondrial and/or nuclear regions where intron sizes are large or unknown, as well as for regions that are difficult to assemble bioinformatically, such as repetitive regions.

Our focus on the chloroplast genome is driven by its phylogenetic informativeness at essentially all taxonomic scales and its relative ease of amplification (e.g., Downie and Palmer, 1992; Graham and Olmstead, 2000; Moore et al., 2007; Parks et al., 2009; Moore et al., 2010), which have made the chloroplast the workhorse of molecular plant systematics since the beginning of the field. Moreover, the availability of a large number of angiosperm plastome sequences had facilitated the design of potentially universal PCR primers. To test this approach, we amplified the chloroplast genomes of 30 species (17 genera) across angiosperms using a set of 58 chloroplast PCR primers that were designed to potentially be universal in angiosperms and that may work in some gymnosperm lineages.

METHODS AND RESULTS

Representatives of 17 different genera (30 spp.) spanning 12 orders of angiosperms sensu APG III (Angiosperm Phylogeny Group, 2009) were chosen to test this approach (Table 1). Special focus was given to three genera in Orobanchaceae: *Lamourouxia* Kunth (one species), *Bartsia* L. (two species), and *Castilleja* Mutis ex L. f. (12 species). High-quality genomic DNA was extracted from ca. 0.02 g of silica gel–dried or herbarium tissue using a modified 2× cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987), yielding 30–70 ng/μL of DNA per sample. Using the 83 plastid gene angiosperm alignments of Moore et al. (2010; Appendix S1), we developed 58 primers with a goal of maximizing universality across angiosperms (Table 2). Conserved regions for primer design were identified by eye, and the primers were tested with IDT OligoAnalyzer tools (Integrated DNA Technologies, Coralville, Iowa, USA) to ensure that melting temperatures (T_m) were greater than 50°C and that there were no significant hairpins or self-dimerization problems. From these, 16 overlapping primer combinations were chosen to amplify the entire chloroplast genome in appropriately sized, overlapping fragments, making sure to allow at least 100 bp of overlap between regions (Fig. 1, Table 2) to minimize the decrease in sequencing depth usually associated with the ~30 bp immediately adjacent to the primer sites (Cronn et al., 2008; Harismendy and Frazer, 2009; Cronn et al., 2012).

PCRs were performed using a combination of two high-quality *Taq* polymerases—QIAGEN *Taq* DNA Polymerase (5 units/μL) and QIAGEN HotStar HiFidelity DNA Polymerase (2.5 units/μL) (QIAGEN, Valencia, California, USA)—to obtain amplification of fragments between 5 kb and 12 kb. The QIAGEN HotStar HiFidelity DNA Polymerase was diluted to 0.2 units/μL by combining 0.1 μL of 5× QIAGEN HotStar HiFidelity PCR buffer, 0.36 μL of double-deionized water (ddH_2O), and 0.04 μL of QIAGEN HotStar HiFidelity DNA Polymerase (2.5 units/μL). Each PCR had a total volume of 25 μL, was prepared on ice, and contained the following reagents: 2.5 μL of 10× PCR buffer (QIAGEN CoralLoad or colorless, with 15 mM $MgCl_2$), 1.0 μL $MgCl_2$ (QIAGEN 25 mM), 0.75 μL of deoxyribonucleotide triphosphates (dNTPs, each at 10 mM), 5.0 μL of 5× QIAGEN Q solution, 2.5 μL of both forward and reverse primers (each at 5 μM), 0.25 μL (1.25 units) of QIAGEN *Taq* DNA Polymerase, 0.5 μL of the diluted QIAGEN HotStar HiFidelity DNA Polymerase solution, 9 μL of ddH_2O, and 1.0 μL of DNA template. Long PCR profiles were as follows: preheat at 93°C, initial denaturation at 93°C for 3 min followed by 35 cycles of denaturation at 93°C for 15 s, annealing at 48–68°C (depending on the primer pair) for 30 s, and extension at 68°C for 5–12 min (1 min/kb of target). To assess amplification, 2 μL of the final reactions were examined on a 1% agarose gel with appropriate size standards and the final products were kept at 4°C. The complete, step-by-step long PCR protocol can be found in Appendix 1.

TABLE 2. Universal angiosperm primers used for chloroplast genome amplifications. The 16 primer combinations chosen for this study are in bold with approximate amplicon sizes in kilobases (kb) indicated.[a]

Region no.	Approx. size (kb)	Primer (F/R)	Primer sequence (5′–3′)	Overlap between regions in bp[b]
1	8	**trnH.GUG.6R**	CCTTRATCCACTTGGCTACAT	Regions 1 & 2 = 542
1		**psbK.195R**	ACTTACAGCAGCTTGCCAAAC	Regions 1 & 2a = 542
2/2a	10.3/6.3	**trnQ.UUG.50R**	GGACGGAAGGATTCGAACC	Regions 2a & 2b = 627
2a		**atpH.17F**	CTGCYGCTTCYGTTATTGCT	Regions 2b & 3 = 2059
2b	4	**atpF.65R**	CGGTATTAAACCCGAAACTCC	Regions 2 & 3 = 2059
2/2b		**rpoC2.4805F**	GYCGTATYGATTGGTTRAAAGG	Regions 3 & 4 = 1274
3	7	**atpI.705R**	CRGCTAAAGTTGCAAAAATAAGAGCT	Regions 4 & 5 = 860
3		**rpoC1.1670F**	GRGATCAAATGGCTGTTCAT	Regions 5 & 6 = 618
4	9	**rpoC2.520R**	GTTCGTACAGCAGTATCYACAAC	Regions 6 & 7 = 764
4		**petN.3R**	GCCCAAGCRAGACTTACTATATCC	Regions 7 & 8 = 153
5	10.5	**trnC.GCA.47F**	CCCAGTTCAAATCCGGGT	Regions 8 & 9 = 1216
5		**psaB.2170F**	GCRGCTTTCTTGATTGCYTC	Regions 9 & 10 = 135
6	10	**trnfM.CAU.21R**	GGTTATGAGCCTTGCGAGCTA	Regions 10 & 11 = 771
6		**trnT.UGU.17F**	GGTTAGAGCATCGCATTTGTAATG	Regions 11 & 12 = 2781
7	10.3	**rps4.380R**	GGTTTGCARCGATAACTTGGKATATC	Regions 12 & 13 = 142
7		**rbcL.178R**	GTCCATGTACCAGTAGARGATTC	Regions 13 & 14 = 392
8	9.2	**rbcL.2F**	TGTCACCACAAACAGARACTAAAG	Regions 14 & 15 = 1911
8		**psbJ.3F**	GGCYGATACTACTGGAAGRAT	Regions 16 & 1 = 840
9	9.8	**petA.920F**	CTTCAAGAYCCATTACGTGTHCAAG	
9		**psbB.160R**	TRCCYTGTCTCCACATTGGAT	
10	10.9	**psbB.3F**	GGGTTTRCCTTGGTATCGTGT	
10		**rps3.17F.new**	ATCCACTTGGTTTYMGACTTGG	
11	8.7	**rpl16.3R**	AACCAACGAGTCACACACTAAGC	
11/16		**ycf2.5100R**	CAGATCATGAATGTTTGGAATCCAT	
12	10	**ycf2.2300F**	TCGGGATCCTRATGCATATAGATAC	
12		**rps12.190F**	GTTGCCAGAGTACGMTTAACCT	
13	11	**rps12.360R**	CCCTTGTTGACGATCCTTTACTC	
13		**ycf1.59R**	CCGACCACAACGACCGAAT	
14/15	11.2	**trnN.GUU.7R**	CCGCTCTACCACTGAGCTAC	
14		**ndhA.535F**	GCTGCTCAATCDATTAGTTATGAA	
15	10.5	**ndhI.194R**	CGAACRCATACTTCACAAGCAA	
16	8.2	**psbA.640F**	GCTATGCATGGTTCYTTGGTAAC	
		rps16.50R	CGAACATCAATTGCAACGATTCGATA	
		rps16.50F	TATCGAATCGTTGCAATTGATGTTCG	
		psbK.200F	GGCAAGCTGCTGTAAGTTTTCGA	
		atpF.70F	GGGTTTAATACCGATATTTTAGCAAC	
		trnR.UCU.45F	GGTATAGGTTCAAATCCTATTGGAC	
		trnQ.UUG.47F	CGGAGGTTCGAATCCTTCC	
		trnK.UUU.3R	GAGATGGCAACTCAATCGTTG	
		trnK.UUU.3F	CAACGATTGAGTTGCCATCTC	
		atpA.430F	CGTTCYGTATATGARCCTCTTCAAAC	
		atpA.820F	ATCGMCAAATGTCTCTTCTATTAMG	
		ccsA.890R	TCCAAGTAATAAANGCCCAAGTTTC	
		trnR.ACG.15F	GAGGATTAGAGCACGTGG	
		ycf1.70F	GTGGTCGGACTCTATTATGGAT	
		trnL.UAG.18F	GGTAGACACGCTGCTCTTAGG	
		trnL.UAG.19F	GTAGACACGCTGCTCTTAGGAAG	
		rps12.320R	GGGTTCCTCGAACAATGTGATATC	
		rpl2.550F	GTGCTGTAGCGAAACTGATTG	
		rpl2.640F	TCAGCAACAGTCGGACARGT	
		psbT.3F	TGGAAGCATTGGTTTATACATTYCT	
		atpB.1290R	ARGGTTGTGATAAGAAACGYTCAA	
		trnT.UGU.42F	GATGGTCATCGGTTCGATTC	
		psbC.3R	AGTTCCATTAAAGAGCGTTTCC	
		psbD.860F	CYGGTTTATGGATGAGYGCT	
		rpoB.900R	CGTCGACCAATCYTTCCTAATTC	
		rpoB.470R	CCRGGRCTTTGCAATATTTGATTG	
		rpoC2.430R	ATRGGTAAATCAATCATTTGYCCTTG	

[a]All primers are shown in the 5′ to 3′ direction; the name of each primer consists of three parts: the gene in which the primer is anchored, the approximate position of the primer within that gene, and either an "F" or an "R." It is important to note that the F and R designations do not indicate that the primer should be used as a forward or reverse primer; rather, they indicate the 5′ to 3′ orientation of the primer with respect to the gene—i.e., a primer that is designated as an "F" primer has its 5′ to 3′ orientation in the same orientation as the gene (i.e., on the forward strand), whereas an "R" primer is oriented in the direction opposite to the 5′ to 3′ orientation of the gene (i.e., on the reverse strand).

[b]Overlap between regions is given in number of base pairs (bp), without taking the length of the primers into consideration.

For the three genera of Orobanchaceae in which PCR optimization was performed, amplification of the fragments was straightforward and had an average success rate of 89.7% (range = 73–100%). The most difficult regions to amplify were regions 2 (*trnQ*$^{(UUG)}$-*rpoC2*), 9 (*petA*-*psbB*), 10 (*psbB*-*rps3*), and 14 (*trnN*$^{(GUU)}$-*ndhA*), which are among the largest fragments (10.3 kb, 9.8 kb, 10.9 kb, and 11.2 kb, respectively; Table 2). It was possible to split

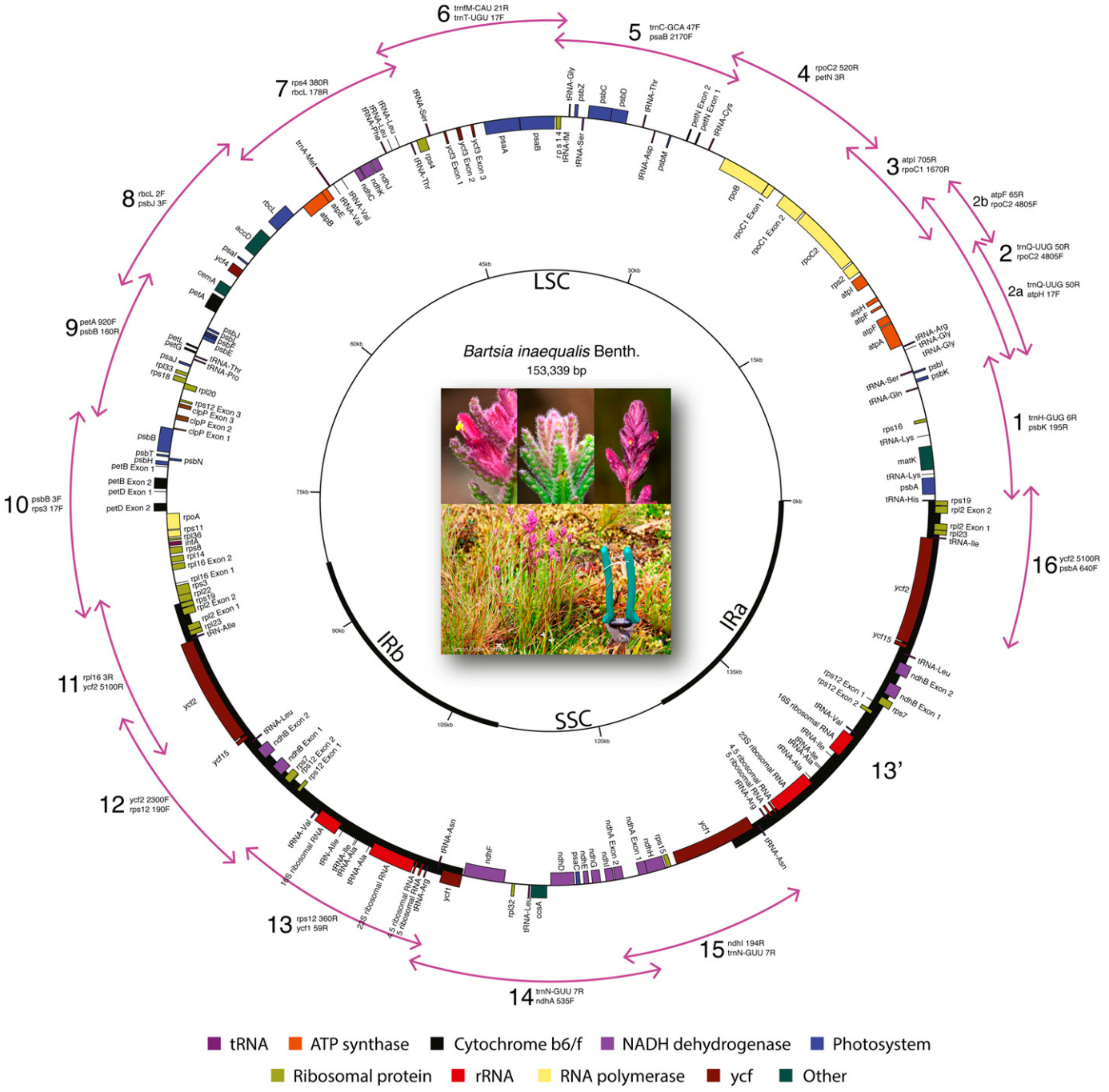

Fig. 1. The final annotated chloroplast genome assembly of *Bartsia inaequalis* with the 16 overlapping primer combinations indicated. Note that the primer combinations for regions 11, 12, 13, and 16 amplify both inverted repeat A and B in a single reaction. Photos by Simon Uribe-Convers.

region 2 into two smaller fragments, 2a (*trnQ*$^{(UUG)}$-*atpH*: 6.3 kb) and 2b (*atpF*-*rpoC2*: 4 kb), which facilitated its amplification in several taxa. This was not the case for regions 9, 10, and 14, for which multiple long PCR experiments using varying amounts of DNA template were necessary to obtain successful amplifications. Amplification outside of Orobanchaceae was highly variable, with an average success rate of 70.8% (range = 22–100%) with regions 5, 6, 9, 10, and 11 showing the lowest success. Importantly, the results for these taxa were obtained after just two rounds of PCR where the annealing temperatures were changed to either 48°C or 55°C. Although we did not optimize the long PCRs for each group, we are confident that optimization on a per group basis (e.g., increasing template volume, altering annealing temperatures, and/or long PCR profiles) and/or the use of fresh tissue for DNA extractions would improve success rates. Furthermore, if genomic rearrangements and/or primer mismatches are present in certain groups, primer combinations other than the 16 that were used here could be tested (Table 2). Nevertheless, we successfully amplified all 16 regions in seven species, whereas in the remaining 23 species it was only possible to amplify between six (1 sp.) and 15 (8 spp.) regions (Table 1). These results translate to 21 species having at least 12 regions amplified (114.7 kb based on potential amplicon size), representing ca. 74% of the chloroplast genome when considering only one copy of the inverted repeat. Even the species with the smallest number of amplified fragments (*Castilleja arvensis* Cham. & Schltdl.) was represented by ~73 kb of data, exemplifying the effectiveness of this approach.

It is notable that many of the DNAs that were tested were extracted from herbarium tissues that ranged from five to 25 yr old when isolated. In addition, we tested these primers in several species of *Abies* Mill. (Pinaceae; Table 1) with surprising success, amplifying between six and nine regions without any PCR optimization. We caution that our long PCR protocol works best using recent DNA extractions that have not been through multiple freeze-thaw cycles. Ideally, long PCR should be conducted using new DNA extractions that are stored at 4°C while performing experiments. Additionally, discrete PCR bands were only obtained using high-quality *Taq* polymerases. When conventional polymerases were used (e.g., Go*Taq* [Promega Corporation, Madison, Wisconsin, USA] or Top*Taq* [QIAGEN]), the resulting PCR products were smears rather than discrete bands and were not used for sequencing.

To confirm that our long PCR approach was compatible with NGS and that our primers would yield complete chloroplast genomes, the amplicons from each of the 15 Orobanchaceae taxa were purified by precipitation in a 20% polyethylene glycol 8000 (PEG)/2.5 M NaCl solution and washed in 70% ethanol. The amplicons were sheared by nebulization at 30 psi for 70 s, yielding an average shear size of 500 bp as measured by a Bioanalyzer High-Sensitivity Chip (Agilent Technologies, Santa Clara, California, USA). DNA normalization is a critical step when pooling samples for multiplexing in NGS; however, due to the large number of plastomes per cell and the very few samples that were being sequenced in such a high-throughput sequencing platform, no DNA quantification was made and the sheared amplicons were pooled by species at equal volume ratios. Sequencing libraries were constructed using the Illumina TruSeq library preparation kit and protocol (Illumina, San Diego, California, USA) and were standardized at 2 nM prior to sequencing. Library concentrations were determined using the KAPA qPCR kit (KK4835; Kapa Biosystems, Woburn, Massachusetts, USA) on an ABI StepOnePlus Real-Time PCR System (Life Technologies, Grand Island, New York, USA). The resulting libraries were multiplexed in one Illumina HiSeq 2000 lane (~187.5 million reads per lane [Glenn, 2011]) at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, yielding ~12.5 million 100-bp single-end reads for each taxon (GenBank Sequence Read Archive accessions: SRR1023085, SRR1023089, SRR1023095, SRR1023112, SRR1023113, SRR1023126, SRR1023128–SRR1023136). Average depth of coverage of our sequencing experiment was ~8333× (taking 150 kb as the average plastome size). The results obtained here clearly do not maximize the potential of the Illumina HiSeq 2000 for plastome sequencing. To take full advantage of the large amount of data produced by a HiSeq 2000 for plastome sequencing, it would be theoretically possible to sequence ~4170 samples per lane and still reach the 30× minimum threshold generally regarded as ideal for plastome sequencing (Straub et al., 2012). However, high-level multiplexing in NGS with this or any other high-throughput method requires careful normalization of DNA concentrations across samples and sufficient adapter barcodes; commonly used commercial kits currently offer either 96 (NEXTflex DNA Barcode kit; Bioo Scientific, Austin, Texas, USA) or 386 (Fluidigm, San Francisco, California, USA). Alternatively, one could choose to perform this type of experiment on an NGS platform that yielded a lesser amount of data, e.g., 1 million 250-bp paired-end reads on an Illumina MiSeq Reagent Nano Kit version 2, which would allow a 30× sequencing depth for 96 samples (or 50× sequencing depth for 64 samples).

Because of the high depth of coverage of our sequencing experiment, reads were cleaned at high stringency (minimum quality = 30/40, maximum number of low-quality bases per read = 5, maximum number of duplicate reads = 10, minimum number of duplicate reads = 2) and assembled against a reference genome (*Sesamum indicum* L., GenBank accession no. JN637766) using the Alignreads pipeline version 2.25 (Straub et al., 2011) with the following options: percent identity = medium, minimum coverage depth = 5, and single nucleotide polymorphism (SNP) minimum coverage depth = 25 with 80% of those reads supporting the SNP. The resulting assemblies had an average depth of ~700×, an average of 0.79% bases that were masked for not reaching the minimum sequencing depth of 5×, and an average N50 of 35,053 bp (Table 1; contigs and ACE files deposited in the Dryad Digital Repository: http://doi.org/10.5061/dryad.kc75n; Uribe-Convers et al., 2014). We noticed a small decrease in sequencing depth in regions immediately adjacent to some primer sites, which is a phenomenon that has been reported in the past (Whittall et al., 2010; Knaus et al., 2011; reviewed in Cronn et al., 2012). Given that our shortest overlap between amplicons is 135 bp (between regions 9 and 10; Table 2), with the rest spanning hundreds of base pairs (Table 2), and that our experiment yielded a high sequencing depth, we had no problems calling bases unambiguously (99.99% on average, Table 1). The *Bartsia inaequalis* Benth. assembly (Fig. 1; GenBank accession no. KF922718) was annotated using DOGMA (Wyman et al., 2004) and visualized in GenomeVx (Conant and Wolfe, 2008).

CONCLUSIONS

We present an alternative approach for systematic studies that combines long PCR and NGS to strategically compile phylogenomic data sets for molecular systematic studies. This approach is on par with genome skimming in terms of costs, but it has the advantage of being a targeted approach and has the potential to produce data more uniformly across samples, i.e., minimizing missing data across taxa. Although this approach was only tested with chloroplast data, we emphasize that the long PCR amplicons can be generated using DNA from any genome, expanding the possibilities of long PCR and NGS for molecular systematic studies. This last point is important for studies targeting the mitochondrion or low-copy regions of the genome that otherwise might be missed or not shared across all samples using genome skimming approaches. For example, this approach may be particularly useful for the enrichment of nuclear regions, where intron sizes are large or unknown.

LITERATURE CITED

Alexander, A., D. Steel, B. Slikas, K. Hoekzema, C. Carraher, M. Parks, R. Cronn, and C. S. Baker. 2013. Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. *Genome Biology and Evolution* 5: 113–129.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

Cheng, S., C. Fockler, W. M. Barnes, and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proceedings of the National Academy of Sciences, USA* 91: 5695–5699.

Conant, G. C., and K. H. Wolfe. 2008. GenomeVx: Simple web-based creation of editable circular chromosome maps. *Bioinformatics (Oxford, England)* 24: 861–862.

Craig, D. W., J. V. Pearson, S. Szelinger, A. Sekar, M. Redman, J. J. Corneveaux, T. L. Pawlowski, et al. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. *Nature Methods* 5: 887–893.

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. doi:10.1093/nar/gkn502

Cronn, R., B. J. Knaus, A. Liston, P. J. Maughan, M. Parks, J. V. Syring, and J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291–311.

Downie, S. R., and J. D. Palmer. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants, 14–35. Chapman and Hall, New York, New York, USA.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Eaton, D. A. R., and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62: 689–706.

Glenn, T. C. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11: 759–769.

Graham, S. W., and R. G. Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

Harismendy, O., and K. Frazer. 2009. Method for improving sequence coverage uniformity of targeted genomic intervals amplified by LR-PCR using Illumina GA sequencing-by-synthesis technology. *BioTechniques* 46: 229–231.

Knaus, B. J., R. Cronn, A. Liston, K. Pilgrim, and M. K. Schwartz. 2011. Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology* 11: 10. doi:10.1186/1472-6785-11-10

Lemmon, A. R., J. M. Brown, K. Stanger-Hall, and E. M. Lemmon. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Systematic Biology* 58: 130–145.

Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.

Moore, M. J., P. S. Soltis, C. D. Bell, J. G. Burleigh, and D. E. Soltis. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.

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. doi:10.1186/1741-7007-7-84

Straub, S. C. K., M. Fishbein, T. Livshultz, Z. Foster, M. Parks, K. Weitemier, R. C. Cronn, and A. Liston. 2011. Building a model: Developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12: 211. doi:10.1186/1471-2164-12-211

Straub, S. C. K., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Stull, G. W., M. J. Moore, V. S. Mandala, N. A. Douglas, H.-R. Kates, X. Qi, S. F. Brockington, et al. 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. *Applications in Plant Sciences* 1: 1200497. doi:10.3732/apps.1200497

Tennessen, J. A., R. Govindarajulu, A. Liston, and T.-L. Ashman. 2013. Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, *Fragaria vesca* ssp. *bracteata* (Rosaceae). *G3; Genes|Genomes|Genetics* 3: 1341–1351.

Uribe-Convers, S., J. R. Duke, M. J. Moore, and D. C. Tank. 2014. Data from: A long PCR–based approach for DNA enrichment prior to next-generation sequencing for systematic studies. Dryad Digital Repository. http://doi.org/10.5061/dryad.kc75n.

Whittall, J. B., J. Syring, M. Parks, J. Buenrostro, C. Dick, A. Liston, and R. Cronn. 2010. Finding a (pine) needle in a haystack: Chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology* 19(Suppl 1): 100–114.

Wu, F., L. A. Mueller, D. Crouzillat, V. Petiard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: A test case in the euasterid plant clade. *Genetics* 174: 1407–1420.

Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics (Oxford, England)* 20: 3252–3255.

Xi, Z., B. R. Ruhfel, H. Schaefer, A. M. Amorim, M. Sugumaran, K. J. Wurdack, P. K. Endress, et al. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences, USA* 109: 17519–17524.

Yuan, Y., C. Liu, H. Marx, and R. Olmstead. 2009. The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. *New Phytologist* 182: 272–283.

Appendix 1. Protocol for long PCR for amplification of 4–20-kb targets. Developed by the Tank Laboratory, University of Idaho; published January 2014.

Product	Contents	Catalog no.
QIAGEN *Taq* DNA Polymerase[1]	250 units *Taq* DNA Polymerase, 10× PCR Buffer,[†] 5× Q-Solution, 25 mM $MgCl_2$	201205
QIAGEN HotStar HiFidelity DNA Polymerase[2]	100 units HotStar HiFidelity DNA Polymerase,[2] 10× HotStar PCR Buffer, 5× Q-solution, 25 mM $MgSO_4$	202602

[1] Almost any high-quality *Taq* polymerase should work; however, cheap *Taq* polymerases (e.g., QIAGEN Top*Taq* or Promega Go*Taq*) do not work and result in large smears, rather than discrete bands.

[2] QIAGEN HotStar HiFidelity DNA Polymerase was the only high-fidelity polymerase used in this study.

[†] Q-solution does seem to be an important additive, thus the use of QIAGEN *Taq*. However, this does work using Q-solution with other high-quality *Taq* polymerases such as Promega's or New England Biolab's standard *Taq* (i.e., if you have a stock of Q-solution, but no QIAGEN *Taq*).

Genomic DNA must be high quality. Run a 0.8% or 1% gel to check. Standard CTAB extractions from silica gel–dried or herbarium material work well if they (1) are recent (extraction and tissue), and (2) contain high-molecular-weight DNA. Most important, we have found that recent DNA extractions that have not been through numerous freeze-thaw cycles work best. **For best results, long PCR should be done using new DNA extractions stored at 4°C while performing long PCR experiments**.

All preparations should be done on ice.

1. Number tubes or prepare plate. Make sure to include appropriate negative controls.

2. Prepare QIAGEN HotStar HiFidelity DNA polymerase dilution:

Reagents to prepare the HotStar *Taq* dilution	Volumes for 25 reactions (total 12.5 µL)	Volumes for 50 reactions (total 25 µL)	Volumes for 100 reactions (total 50 µL)
5× HotStar HiFidelity PCR buffer	2.5 µL	5.0 µL	10 µL
H_2O	9.0 µL	18 µL	36 µL
QIAGEN HotStar *Taq*	1.0 µL	2.0 µL	4.0 µL

3. Prepare cocktail:

Cocktail	×1 (25 µL reaction)
10× PCR buffer (QIAGEN CoralLoad PCR Buffer or colorless, 15 mM $MgCl_2$)	2.5 µL
$MgCl_2$ (25 mM)	1.0 µL (3 mM final conc.; adjustable)
dNTP (10 mM each)	0.75 µL (3 µL of 2.5 mM each)
Q solution (5×)	5.0 µL
5′ primer (5 µM)	2.5 µL (0.5 µM final conc.)
3′ primer (5 µM)	2.5 µL (0.5 µM final conc.)
Taq DNA polymerase (QIAGEN)	0.25 µL (1.25 units)[1]
QIAGEN HotStar DNA polymerase (diluted)	0.50 µL
H_2O	to 25 µL (9 µL if using 1.0 µL DNA)

[1] The success rate was lower when a smaller quantity was used, but the best DNAs work with ≥0.125 µL.

4. Add 1–2 µL of template to each of the tubes.

5. While the tubes/plate with template are on ice, add 24 µL of cocktail to each tube, being careful not to cross contaminate. Spin down to bring all liquid to the bottom of the tube.

6. Run appropriate long PCR profile. Generic temperatures and times are:
   i. 93°C infinity (important to go directly from ice to hot block)
   ii. 93°C for 3 min (initial denaturation)
   iii. 93°C for 15 s
   iv. 48–68°C for 30 s (T_a should be ~5°C below T_m of primers)
   v. 68°C for 5–20 min (1 min/kb of target)
   vi. go to step 3, 34×
   vii. 4°C infinity

7. Check reactions by running 2 µL on 1% agarose gel with appropriate size standards.

Primer combinations for long PCR amplification of the chloroplast genome.[1,2]

Region no.	Approx. size (kb)	Primers (F/R)[3]	Primer sequence (5′–3′)
1	8	trnH.GUG.6R	CCTTRATCCACTTGGCTACAT
		psbK.195R	ACTTACAGCAGCTTGCCAAAC
2	10.3	trnQ.UUG.50R	GGACGGAAGGATTCGAACC
		rpoC2.4805F	GYCGTATYGATTGGTTRAAAGG
2a[4]	6.3	trnQ.UUG.50R	GGACGGAAGGATTCGAACC
		atpH.17F	CTGCYGCTTCYGTTATTGCT
2b[4]	4	atpF.65R	CGGTATTAAACCCGAAACTCC
		rpoC2.4805F	GYCGTATYGATTGGTTRAAAGG
3	7	atpI.705R	CRGCTAAAGTTGCAAAAATAAGAGCT
		rpoC1.1670F	GRGATCAAATGGCTGTTCAT
4	9	rpoC2.520R	GTTCGTACAGCAGTATCYACAAC
		petN.3R	GCCCAAGCRAGACTTACTATATCC
5	10.5	trnC.GCA.47F	CCCAGTTCAAATCCGGGT
		psaB.2170F	GCRGCTTTCTTGATTGCYTC
6	10	trnfM.CAU.21R	GGTTATGAGCCTTGCGAGCTA
		trnT.UGU.17F	GGTTAGAGCATCGCATTTGTAATG
7	10.3	rps4.380R	GGTTTGCARCGATAACTTGGKATATC
		rbcL.178R	GTCCATGTACCAGTAGARGATTC
8	9.2	rbcL.2F	TGTCACCACAAACAGARACTAAAG
		psbJ.3F	GGCYGATACTACTGGAAGRAT
9	9.8	petA.920F	CTTCAAGAYCCATTACGTGTHCAAG
		psbB.160R	TRCCYTGTCTCCACATTGGAT
10	10.9	psbB.3F	GGGTTTRCCTTGGTATCGTGT
		rps3.17F.new	ATCCACTTGGTTTYMGACTTGG
11	8.7	rpl16.3R	AACCAACGAGTCACACACTAAGC
		ycf2.5100R	CAGATCATGAATGTTTGGAATCCAT
12	10	ycf2.2300F	TCGGGATCCTRATGCATATAGATAC
		rps12.190F	GTTGCCAGAGTACGMTTAACCT
13[5]	11	rps12.360R	CCCTTGTTGACGATCCTTTACTC
		ycf1.59R	CCGACCACAACGACCGAAT
14	11.2	trnN.GUU.7R	CCGCTCTACCACTGAGCTAC
		ndhA.535F	GCTGCTCAATCDATTAGTTATGAA
14′[6]	7	trnR.ACG.15F	GAGGATTAGAGCACGTGG
		ccsA.890R	TCCAAGTAATAAANGCCCAAGTTTC
15	10.5	ndhI.194R	CGAACRCATACTTCACAAGCAA
		trnN.GUU.7R	CCGCTCTACCACTGAGCTAC
16	8.2	psbA.640F	GCTATGCATGGTTCYTTGGTAAC
		ycf2.5100R	CAGATCATGAATGTTTGGAATCCAT

[1] Universal primers designed by M.J.M.; compiled and tested by D.C.T. and S.U.C.

[2] T_a should be ~5°C below T_m of primers; however, temperatures of 55°C have worked for all primer combinations.

[3] The name of each primer consists of three parts: (1) the gene in which the primer is anchored in, (2) the approximate position of the primer within that gene (based on all-angiosperm alignment per Moore et al., 2007), and (3) either an "F" or an "R." The F and R designations do not indicate that the primer should be used as a forward or reverse primer; rather, they indicate the 5′ to 3′ orientation of the primer with respect to the gene. In other words, a primer that is designated as an F primer has its 5′ to 3′ orientation in the same orientation as the gene (i.e., on the forward strand, or from start to stop), whereas an R primer is oriented in the direction opposite to the 5′ to 3′ orientation of the gene (i.e., on the reverse strand).

[4] Regions 2a and 2b can be used to amplify region 2 in two pieces.

[5] Regions 11, 12, and 13 represent a large portion of the inverted repeat (IR), thus, one amplification for both IRa and IRb.

[6] Region 14′ amplifies ca. 2/3 of region 14.