



Application of a Simplified Method of Chloroplast Enrichment to Small Amounts of Tissue for Chloroplast Genome Sequencing

Authors: Sakaguchi, Shota, Ueno, Saneyoshi, Tsumura, Yoshihiko, Setoguchi, Hiroaki, Ito, Motomi, et al.

Source: *Applications in Plant Sciences*, 5(5)

Published By: Botanical Society of America

URL: <https://doi.org/10.3732/apps.1700002>

APPLICATION OF A SIMPLIFIED METHOD OF CHLOROPLAST ENRICHMENT TO SMALL AMOUNTS OF TISSUE FOR CHLOROPLAST GENOME SEQUENCING¹

SHOTA SAKAGUCHI^{2,10}, SANEYOSHI UENO³, YOSHIHIKO TSUMURA⁴, HIROAKI SETOGUCHI², MOTOMI ITO⁵, CHIE HATTORI², SHOGO NOZOE², DAIKI TAKAHASHI², RIKU NAKAMASU⁶, TAISHI SAKAGAMI⁶, GUILLAUME LANNUZEL⁷, BRUNO FOGLIANI⁷, ADRIEN S. WULFF^{7,8}, LAURENT L'HUILLIER⁷, AND YUJI ISAGI⁹

²Graduate School of Human and Environmental Studies, Kyoto University, Yoshida-nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan; ³Tree Genetics Laboratory, Department of Forest Genetics, Forestry and Forest Products Research Institute, 1 Matsunosato, Tsukuba, Ibaraki 305-8687, Japan; ⁴Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 3058572, Japan; ⁵Graduate School of Arts and Sciences, University of Tokyo, Tokyo 153-8902, Japan; ⁶Faculty of Integrated Human Studies, Kyoto University, Yoshida-nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan; ⁷Agronomic Institute of New Caledonia (IAC), Diversités biologique et fonctionnelle des écosystèmes terrestres, BP 73, Port Laguerre, Païta 98890, New Caledonia; ⁸SoREco-NC, 57 Route de l'Anse Vata, 98800 Nouméa, New Caledonia; and ⁹Division of Forest and Biomaterials Science, Graduate School of Agriculture, Kyoto University, Kyoto 6068502, Japan

- **Premise of the study:** High-throughput sequencing of genomic DNA can recover complete chloroplast genome sequences, but the sequence data are usually dominated by sequences from nuclear/mitochondrial genomes. To overcome this deficiency, a simple enrichment method for chloroplast DNA from small amounts of plant tissue was tested for eight plant species including a gymnosperm and various angiosperms.
- **Methods:** Chloroplasts were enriched using a high-salt isolation buffer without any step gradient procedures, and enriched chloroplast DNA was sequenced by multiplexed high-throughput sequencing.
- **Results:** Using this simple method, significant enrichment of chloroplast DNA-derived reads was attained, allowing deep sequencing of chloroplast genomes. As an example, the chloroplast genome of the conifer *Callitris sulcata* was assembled, from which polymorphic microsatellite loci were isolated successfully.
- **Discussion:** This chloroplast enrichment method from small amounts of plant tissue will be particularly useful for studies that use sequencers with relatively small throughput and that cannot use large amounts of tissue (e.g., for endangered species).

Key words: chloroplast enrichment; cpSSR; endangered species; high-throughput sequencing.

Chloroplast genomes harbor genetic polymorphisms, such as single nucleotide polymorphisms (SNPs), insertion/deletion polymorphisms (indels), and microsatellite or simple sequence repeat (SSR) markers, which can be widely used in molecular research. Phylogenetic, population genetic, and DNA barcoding studies have benefited from universal primers used to amplify coding or noncoding regions bordered by conserved chloroplast

loci (e.g., Kress et al., 2005; Shaw et al., 2007; Hollingsworth et al., 2009). For intraspecific research, chloroplast microsatellite (cpSSR) markers are important in that they exhibit high levels of allelic variation (Provan et al., 2001). However, a limited number of universal cpSSR primers have been used to date (e.g., Cato and Richardson, 1996; Weising and Gardner, 1999), and novel specific cpSSR primers are required to increase the number of polymorphic markers. Although this can be achieved using cpDNA sequences of target and/or related species deposited in public databases, it may be necessary to perform de novo sequencing of chloroplast genomes in species for which no genomic information is available (Ebert and Peakall, 2009).

High-throughput sequencing of total genomic DNA allows for the recovery and assembly of complete chloroplast genomes (Daniell et al., 2016), although most reads in these situations are derived from the nuclear or mitochondrial genomes and contaminating DNA. For example, cpDNA reads represented only 1.2% of all reads for the conifer *Abies nephrolepis* (Trautv. ex Maxim.) Maxim. (Yi et al., 2016), 2.7% for the pear *Pyrus pyrifolia* (Burm. f.) Nakai (Terakami et al., 2012), and 3.8–11.6%

¹Manuscript received 19 January 2017; revision accepted 9 April 2017.

Funding was provided by a Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (KAKENHI 26850098) and the Environmental Research and Technology Development Fund of the Ministry of the Environment SICORP Program of the Japan Science and Technology Agency (4-1403). The southern province of New Caledonia provided financial support and authorized (by exception) the collection of *C. sulcata* material. The authors thank the traditional owners of the sampling areas for supporting the research program.

¹⁰Author for correspondence: sakaguchi.shota.6a@kyoto-u.ac.jp

doi:10.3732/apps.1700002

Applications in Plant Sciences 2017 5(5): 1700002; <http://www.bioone.org/loi/apps> © 2017 Sakaguchi et al. Published by the Botanical Society of America. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY-NC-SA 4.0), which permits unrestricted noncommercial use and redistribution provided that the original author and source are credited and the new work is distributed under the same license as the original.

for five Poaceae grass species (Nock et al., 2011). The low content of chloroplast-derived reads can be more problematic when researchers perform a multiplexed sequencing of genomic DNA samples on desktop sequencers with relatively small throughputs (e.g., using the Ion PGM Sequencer [Thermo Fisher Scientific, Waltham, Massachusetts, USA] or the MiSeq version 2 [Illumina, San Diego, California, USA]). Hence, there are situations when increasing the proportions of cpDNA will allow for more cost-effective sequencing of chloroplast genomes.

Traditionally, chloroplast genome sequencing has been performed on cpDNA prepared from isolated intact chloroplasts. Most methods for cpDNA enrichment involve three steps: separation of chloroplasts from other organelles, chloroplast lysis, and cpDNA purification (Bookjans et al., 1984; Palmer, 1986). All methods employ step gradients for chloroplast isolation that involve repeated centrifugations. In these procedures, it is important to collect sufficient numbers of chloroplasts while minimizing contamination by nuclear genomic DNA. Thus, isolation of intact chloroplasts has been thought to require a substantial amount of leaf tissue to yield an adequate amount of enriched cpDNA; generally tens of grams of fresh leaf materials are enough, while sometimes at least 100 g is recommended (Jansen et al., 2005). In one extreme example, 120 g of seedlings of the conifer *Cryptomeria japonica* D. Don, corresponding to 2500 individuals, was used for chloroplast isolation (Hirao et al., 2008; Hirao and Watanabe, 2012). However, it is sometimes difficult to collect such large amounts of fresh leaves from endangered or small species; it is desirable to start with much smaller amounts of fresh material (i.e., <1 g).

We here applied a method of chloroplast DNA enrichment that involves slight modifications from recently reported methods to isolate intact chloroplasts (Shi et al., 2012; Vieira et al., 2014). To safely recover chloroplasts from small amounts of leaf materials, we employed a high-salt isolation buffer without a step gradient procedure. The enriched cpDNA samples were subjected to multiplexed high-throughput sequencing for reconstruction of complete chloroplast genomes. The method was tested on eight plant species of various lineages and life forms, and enabled significant enrichment of cpDNA-derived reads, facilitating deep chloroplast genome sequencing. As an example of marker development, cpSSRs were explored in one sequenced species: the endangered conifer *Callitris sulcata* Schltr. Sixty-five microsatellite loci were identified in the chloroplast genome. Overall, our method was shown to efficiently generate species-specific cpDNA markers.

MATERIALS AND METHODS

Isolation of chloroplast-enriched DNA—Leaf material was collected from young leaves of eight plant species of seven families (various life forms including conifer, deciduous trees, and herbaceous plants with nuclear genome size ranging from 1.1 to 9.2 C in pg) (Table 1). The input leaf weight ranged from 0.40 g for *Callitris sulcata* to 2.1 g for *Celtis sinensis* Pers. For the conifer *C. sulcata*, the foliage was kept in the dark at 4°C for 5 d to reduce the level of starch. Samples of the other seven angiosperms were used immediately for chloroplast enrichment. After rinsing with deionized water, leaf material was homogenized in 40 mL of 4°C isolation buffer (1.25 M NaCl, 0.25 M ascorbic acid, 10 mM sodium metabisulfite, 0.0125 M borax, 50 mM Tris-HCl [pH 8.0], 7 mM ethylenediamine tetracetic acid [EDTA], 1% polyvinylpyrrolidone [PVP] [w/v], and 0.1% bovine serum albumin [BSA] [w/v]) (Bookjans et al., 1984; Shi et al., 2012; Vieira et al., 2014) in a prechilled blender. Each homogenate was filtered

through two layers of Miracloth (Calbiochem, San Diego, California, USA) and centrifuged at 200 × g for 15 min at 4°C to remove cell debris. This step was repeated twice. The supernatants were transferred to new 10-mL tubes and centrifuged at 2900 × g for 20 min at 4°C to obtain pellets containing chloroplasts. The pellets were not washed with high-salt buffer as this would have reduced the amount of pellet. Rather, the pellet was lysed with 600 μL hexadecyl trimethyl ammonium bromide (CTAB) buffer following Shi et al. (2012) to extract DNA (Murray and Thompson, 1980). The DNA was then purified using a DNeasy Plant Mini Kit according to the manufacturer's instructions (QIAGEN, Hilden, Germany).

High-throughput sequencing of fragmented DNA libraries—Purified DNA (50 ng) was used to construct DNA fragment libraries using the Ion Xpress Plus Fragment Library Kit, following the manufacturer's protocol (Thermo Fisher Scientific). An Ion PGM Hi-Q OT2 Kit was used to prepare template-positive Ion Sphere Particles. Thermal cycling amplification of DNA fragments in microreactors was performed using the Ion OneTouch 2 system (Thermo Fisher Scientific). Positive particles were isolated and purified employing the Ion OneTouch ES system (Thermo Fisher Scientific). The particles were loaded onto Ion 318 chips and sequenced using an Ion PGM Sequencer (Thermo Fisher Scientific). Sequencing of the libraries was separately performed in three independent chips together with other samples for different research projects.

Reconstruction of the chloroplast genome—The raw reads were imported into CLC Genomics Workbench version 7.5.1 (CLC bio, Aarhus, Denmark) for adapter and quality-based trimming. Low-quality bases were removed (quality limit = 0.03). Cleaned reads were assembled using MITObim version 1.8 (Hahn et al., 2013), guided by the complete chloroplast genome sequences of related species (Table 1). Cleaned reads were then mapped back to the MITObim-derived contig to assess the proportions of chloroplast DNA-derived reads and read depths across the genomes using CLC Genomics Workbench 7.5.1.

Characterization of the chloroplast genome sequence of *C. sulcata* and SSR marker development—To verify the quality of cpDNA assembly by MITObim, gene annotation and synteny analysis were performed for *C. sulcata* as a case analysis. Gene annotation was performed with the aid of the CPGAVAS AnnoGenome module (Liu et al., 2012) with a cutoff BLASTN *E*-value of 1e-10. The reference species for annotation was *Cryptomeria japonica* (GenBank accession no. AP009377). The annotated chloroplast genome was visualized as a circular genome using GenomeVx (Conant and Wolfe, 2008). PipMaker (Schwartz et al., 2000) was used to create dot plots and gene identity plots between *C. sulcata* and another Cupressaceae species, *Calocedrus formosana* (Florin) Florin (GenBank accession no. AB831010).

We used MSATCOMMANDER (Faircloth, 2008) to screen chloroplast regions, including microsatellites, for ≥10 mononucleotide repeats, ≥6 dinucleotide repeats, and ≥4 tri- to hexanucleotide repeats. We designed PCR primer pairs for these regions using Primer3 (Rozen and Skaletsky, 1999); we chose an optimal annealing temperature of 60(±1)°C, a GC content of 30–70%, and product size ranges of 100–500 bp. Nineteen primer pairs, including seven amplifying mononucleotide, six dinucleotide, four trinucleotide, and one tetranucleotide repeats, were employed to explore PCR amplification and polymorphism in two *C. sulcata* populations from the northern catchments of the Dumbéa River (*n* = 21; 22°06'S, 166°30'E) and the Koéalagouamba River (*n* = 20; 22°01'S, 166°23'E). Voucher specimen accession numbers were as follows: KYO 00019999 for the Dumbéa River populations and MO 727065 for the Koéalagouamba River population. For all loci, the forward primer contained one of three different M13 sequences (5'-CACGACGTTGTAACCGAC-3', 5'-TGTGGAATTGTGAGCGG-3', or 5'-CTATAGGGCAGCGGTGGT-3') and the reverse primer was tagged with a PIG-tail sequence (5'-GTTTCTT-3') (Brownstein et al., 1996). PCR reactions were performed according to the protocol of the Multiplex PCR kit (QIAGEN) in final volumes of 10 μL, containing approximately 5 ng of DNA, 5 μL 2× Multiplex PCR Master Mix, 0.01 μM of forward primer, 0.2 μM of reverse primer, and 0.1 μM of fluorescently labeled M13 primer. The PCR thermal profile was as follows: denaturation at 95°C for 3 min; followed by 35 cycles at 95°C for 30 s, 60°C for 3 min, and 68°C for 1 min; and a final 20-min extension step at 68°C. PCR products were sequenced using an ABI 3130xl sequencer (Applied Biosystems, Foster City, California, USA), and the fragment size was determined by GeneMapper (Applied Biosystems).

TABLE 1. Chloroplast genome sequencing of eight plant species.

Species (Family)	Life form	Genome size (1C in pg) ^a	Reference sequence (GenBank no.)	Input tissue weight (g)	No. of cleaned reads	Recovered chloroplast genome length (bp)	No. of mapped reads	Percentage of mapped reads	Average read depth	GenBank accession no.	Voucher accession no. ^b
<i>Parinaria triloba</i> (Miq.) Miq. var. <i>tateuchiana</i> (Makino) Ohwi (Valerianaceae)	Deciduous herb	1.11 (0.75–1.29)	<i>Kolkwitzia amabilis</i> Graebn. (KT966716)	2.00	413,236	158,524	246,049	59.54	237	AP017910	KYO 00019985
<i>Tricyrtis macropoda</i> Miq. (Liliaceae)	Deciduous herb	4.24 (3.73–4.75)	<i>Lilium isingtaense</i> Gilg (KM103365)	0.95	197,233	153,981	48,516	24.6	52	AP017907	KYO 00019984
<i>Rumex nepalensis</i> Spreng. subsp. <i>andreaeanus</i> (Makino) Yonek. (Polygonaceae)	Deciduous herb	2.58 (0.48–6.24)	<i>Rumex acetosa</i> L. (KC817303)	1.40	280,233	158,781	63,443	22.64	74	AP017909	KYO 00019986
<i>Asarum sakawanum</i> Makino var. <i>stellatum</i> (F. Maek. ex Akasawa) T. Sugaw. (Aristolochiaceae)	Evergreen herb	6.75 (4.65–8.95)	<i>Piper kadsura</i> (Choisy) Ohwi (KT23569)	0.84	81,578	161,922	16,816	20.61	20	AP017908	KYO 00019902
<i>Celtis sinensis</i> Pers. (Cannabaceae)	Deciduous tree	1.35	<i>Humulus lupulus</i> L. (KT266264)	2.10	303,264	154,809	53,858	17.76	64	AP017911	KYO 00019989
<i>Zelkova serrata</i> (Thunb.) Makino (Ulmaceae)	Deciduous tree	Unknown	<i>Humulus lupulus</i> L. (KT266264)	2.00	516,820	157,692	83,609	16.18	93	AP017905	KYO 00019987
<i>Aphananthe aspera</i> (Thunb.) Planch. (Cannabaceae)	Deciduous tree	Unknown	<i>Humulus lupulus</i> L. (KT266264)	2.00	1,249,307	156,963	181,915	14.56	222	AP017906	KYO 00019988
<i>Callitris sulcata</i> Schltr. (Cupressaceae)	Evergreen conifer	9.22 (8.25–10.40)	<i>Calocedrus formosana</i> (Florin) Florin (AB831010)	0.40	1,939,765	131,609	196,403	10.13	220	AP017904	KYO 00019999

^aThe genome sizes are shown as the mean (with minimum and maximum) for each genus. Genome size data were taken from the RBG Kew Angiosperm DNA C-values database, version 8.0 (Bennett and Leitch, 2012).

^bVouchers deposited at the Kyoto University Herbarium, Kyoto, Japan.

RESULTS

Sequencing of cpDNA-enriched libraries—Sequencing of the fragment DNA libraries yielded 4,981,436 cleaned reads in total. The number of reads varied from 81,578 for *A. sakawanum* Makino var. *stellatum* (F. Maek. ex Akasawa) T. Sugaw. to 1,939,765 for *C. sulcata* (Table 1; DNA Data Bank of Japan [DDBJ] Sequence Read Archive accession numbers: DRA005207–DRA005214). The baiting and iterative mapping approach implemented in MITObim reconstructed the chloroplast genome sequences of the eight species (accession numbers: AP017904–AP017911). The lengths ranged from 131,609 bp in *C. sulcata* to 161,922 bp in *A. sakawanum* var. *stellatum*. Chloroplast DNA-derived reads were more abundant among cleaned reads of herbaceous species (average = 31.9%). DNA libraries from deciduous tree species had lower levels of cpDNA reads (average = 14.7%); this was particularly evident in the evergreen conifer species *C. sulcata* (10.1%) (Table 1). The read depth (averaged across the chloroplast genome) was modeled by the linear regression formula: (read depth) = $0.0011 \times (\text{no. of cpDNA reads}) - 0.0015 \times (\text{genome size}) + 1.115 \times (\text{average read length}) + 44.6$, with a coefficient of determination (an R^2 value) = 0.993, and all P values for each term were significant ($P < 0.01$). Given a genome size of 155 kb and an average read length of 170 bp, the formula indicates that 14,000 and 96,000 of cpDNA reads should generate read depths of 10 \times and 100 \times , respectively. Although the estimates vary with the proportions of cpDNA reads in read pools, a variety of next-generation sequencing platforms with throughputs of millions of reads can generate sufficient numbers of reads to recover multiple whole chloroplast genomes. Read mapping back to the recovered genomes revealed that sequencing of cpDNA was relatively even for five species with less than 10 \times variation across genomic regions, with a few short regions with much higher coverage (one region in *Aphananthe aspera* (Thunb.) Planch., four in *Tricyrtis macropoda* Miq., and one in *Patrinia triloba* (Miq.) Miq. var. *takeuchiana* (Makino) Ohwi) (Appendix S1).

Structure of the chloroplast genome of *C. sulcata*—CPGAVAS annotated 117 genes in the chloroplast genome of *C. sulcata* (Fig. 1, Table 2). These included 61 self-replication genes (52.1%) including 31 transfer RNA genes; 49 photosynthesis genes (41.9%); genes encoding a maturase (*matK*), an envelope membrane protein (*cemA*), a subunit of acetyl-CoA-carboxylase (*accD*), a c-type cytochrome synthesis gene (*ccsA*); and three genes of unknown function (*ycf* genes) (Table 2). Comparisons with other Cupressaceae conifers revealed strong conservation in terms of genome synteny (Fig. 2); however, no significant homolog of a large *ycf1* gene (7011 bp in *C. formosana*) was evident in the cpDNA sequence of *C. sulcata*. The percentage identities in genomic sequences between *C. sulcata* and *C. formosana* were relatively low (ca. 75%) in most of the region corresponding to *ycf1*, in contrast with the surrounding genic regions (Fig. 3).

Characterization of chloroplast microsatellites in *C. sulcata*—A microsatellite motif search performed using MSATCOMMANDER identified 65 microsatellite regions: one tetranucleotide, eight trinucleotide, nine dinucleotide, and 47 mononucleotide repeats. For these, 19 primer pairs were designed to bracket the microsatellite regions, four of which included multiple microsatellite repeats distributed less than 50 bp apart.

PCR amplification of 14 microsatellite loci was successful, with clear microsatellite peaks evident upon fragment analysis using an autosequencer. Allelic variation was detected at five loci evaluated in 41 individuals from two populations (Table 3). The number of alleles per locus ranged from two to seven, and 10 resultant multilocus haplotypes were evident across the populations.

DISCUSSION

Alternative methods for chloroplast genome sequencing—Chloroplast genomes have been sequenced after isolation of essentially intact chloroplasts, but this is labor-intensive and requires substantial amounts of fresh tissue (Jansen et al., 2005). The newer, massive high-throughput sequencing techniques have encouraged the development of other approaches toward determination of chloroplast genome sequences of nonmodel plant species; the isolation step is unnecessary. One powerful approach is to sequence genomic DNA without any enrichment procedures, which enables the use of very small amounts of plant materials. It does not require any labor-intensive experiments and can recover whole chloroplast genome sequences for nonmodel species, although the portion of chloroplast-derived reads is usually low (Terakami et al., 2012; Yi et al., 2016). This approach is suitable when one can use high-throughput sequencing platforms that generate hundreds of millions of reads, but can be challenging in performing highly multiplexed sequencing with desktop sequencers with relatively small throughputs.

Alternative approaches involve long-range PCR or targeted enrichment of chloroplast fragments using conserved primers (Ebert and Peakall, 2009; Stull et al., 2013; Uribe-Convers et al., 2014; Yang et al., 2014). This approach allows targeted fragments to be effectively enriched (>90% of reads are derived from chloroplast genomes; Yang et al., 2014) and, when combined with high-throughput sequencing, extends the scalability of chloroplast genome sequencing to large-scale phylogenetics and population genomics. Although the conserved primers are very similar among divergent angiosperms, they cannot be used to study certain lineages because of structural changes in the chloroplast genomes. For example, the primers of Yang et al. (2014) cannot be used to study the Asteraceae, Ericaceae, Poaceae, or gymnosperms. For such lineages, novel primers must be developed by reference to conserved regions in the genomes of related species (Cronn et al., 2008; Doorduyn et al., 2011; Njuguna et al., 2013). Recently, Du et al. (2015) introduced an improved method using rolling circle amplification (RCA), which amplifies entire chloroplast genomes from intact chloroplasts prepared using small-scale step gradients in 2-mL tubes. The method is promising for the study of species for which only limited amounts of leaf material are available; only 0.5 g of leaf material was required to obtain a complete chloroplast genome sequence. The reported proportions of cpDNA reads from an RCA-enriched approach were 35.5% for *Quercus spinosa* David ex Franch. (Du et al., 2015) and 19.6% for *Corynocarpus laevigatus* J. R. Forst. & G. Forst. (Atherton et al., 2010).

In this study, we used only 0.4 g of *Callitris sulcata* foliage for DNA extraction and confirmed significant cpDNA enrichment by electrophoresis of the enriched DNA (Appendix S2). In fact, the proportion of cpDNA-derived reads (10.13%) was higher by one order of magnitude than that reported for the conifer *Abies nephrolepis* (1.2%) by Yi et al. (2016), from which a nonenriched genomic DNA library was prepared. The proportions were higher for deciduous trees (14.5–17.8%) and

TABLE 2. Genes detected in the chloroplast genome of *Callitris sulcata*.

Gene category	Gene group	Genes	No.
Self-replication	Ribosomal RNA genes	<i>rrn4.5, rrn5, rrn16, rrn23</i>	4
	Transfer RNA genes	<i>trnC-GCA</i> (×2), <i>trnD-GTC</i> (×2), <i>trnE-TTC</i> , <i>trnF-GAA</i> , <i>trnM-CAT</i> , <i>trnG-GCC</i> , <i>trnH-GTG</i> , <i>trnI-CAT</i> (×2), <i>trnL-CAA</i> (×2), <i>trnL-TAG</i> (×2), <i>trnN-GTT</i> , <i>trnP-GGG</i> , <i>trnP-TGG</i> (×2), <i>trnQ-TTG</i> , <i>trnQ-TTG</i> , <i>trnR-ACG</i> , <i>trnR-TCT</i> , <i>trnS-GCT</i> (×2), <i>trnS-TGA</i> , <i>trnV-GAC</i> , <i>trnV-TAC</i> , <i>trnW-CCA</i> (×2), <i>trnY-GTA</i>	31
Genes for photosynthesis	Large subunit of ribosome	<i>rpl2</i> (×2), <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl36</i>	9
	Small subunit of ribosome	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps8</i> , <i>rps11</i> , <i>rps12</i> (×2), <i>rps14</i> , <i>rps15</i> , <i>rps18</i> , <i>rps19</i>	12
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>	4
	Translational initiation factor	<i>infA</i>	1
	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	5
	Subunits of photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbK</i> , <i>psbL</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>	12
	Subunits of cytochrome	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> (×2), <i>petL</i> (×2), <i>petN</i> (×2)	10
	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> (×2), <i>atpF</i> , <i>atpH</i> , <i>atpI</i>	7
	Large subunit of RuBisCO	<i>rbcL</i>	1
	Chlorophyll biosynthesis	<i>chlB</i> , <i>chlL</i> , <i>chlN</i>	3
	Subunits of NADH dehydrogenase	<i>ndhA</i> , <i>ndhB</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>	11
Other genes	Maturase	<i>matK</i>	1
	Envelope membrane protein	<i>cemA</i>	1
	Subunit of acetyl-CoA-carboxylase	<i>accD</i>	1
	c-Type cytochrome synthesis gene	<i>ccsA</i>	1
Genes of unknown function	Conserved open reading frames	<i>ycf2</i> , <i>ycf3</i> , <i>ycf4</i>	3
Sum			117

apparent (Table 1). Tree species generally yield smaller proportions of cpDNA reads than herbaceous species; evergreen species yield fewer cpDNA reads than deciduous species. These differences are attributable to leaf chloroplast content. It is well-known that leaf nitrogen concentration, which is strongly related to the level of chloroplast photosynthetic machinery

(Evans, 1989), varies greatly among plant functional groups. Generally, herbaceous species have greater leaf nitrogen content (20–60 mg/g) than do deciduous and evergreen tree species (15–40 mg/g and 7–30 mg/g, respectively) (Reich et al., 1997). Nuclear genome sizes are also assumed to influence the yield of cpDNA-derived reads; however, we found no clear relationship (Table 1).

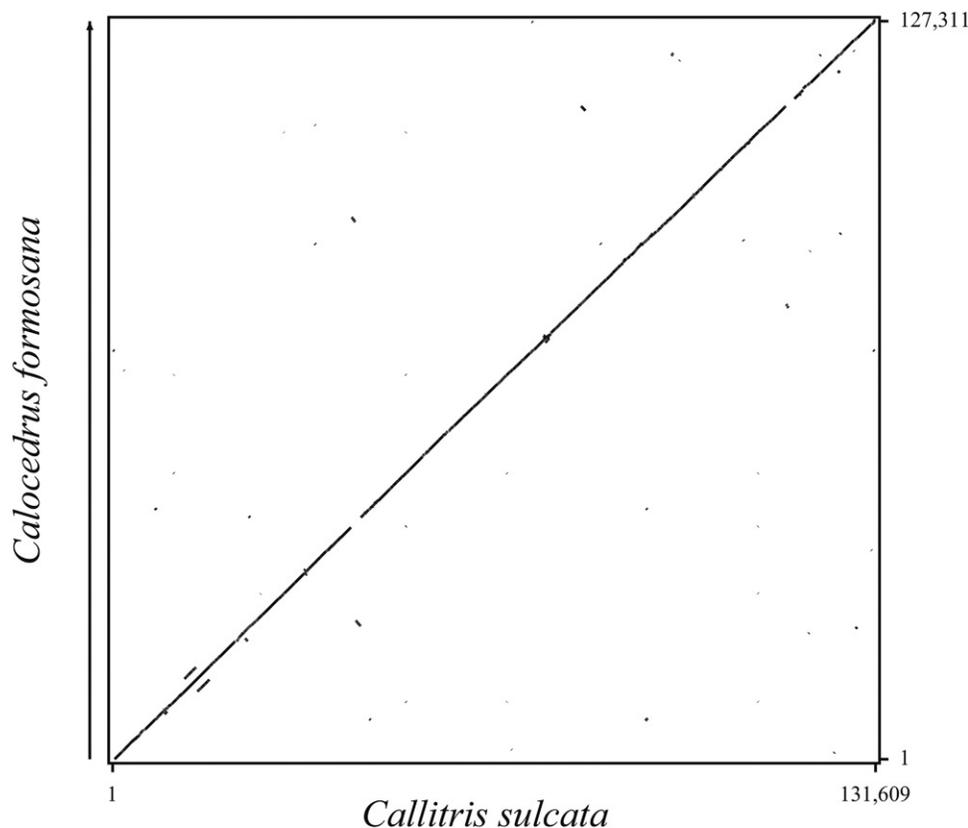


Fig. 2. Dot plot of chloroplast genome sequences shared by *Callitris sulcata* and *Calocedrus formosana*.

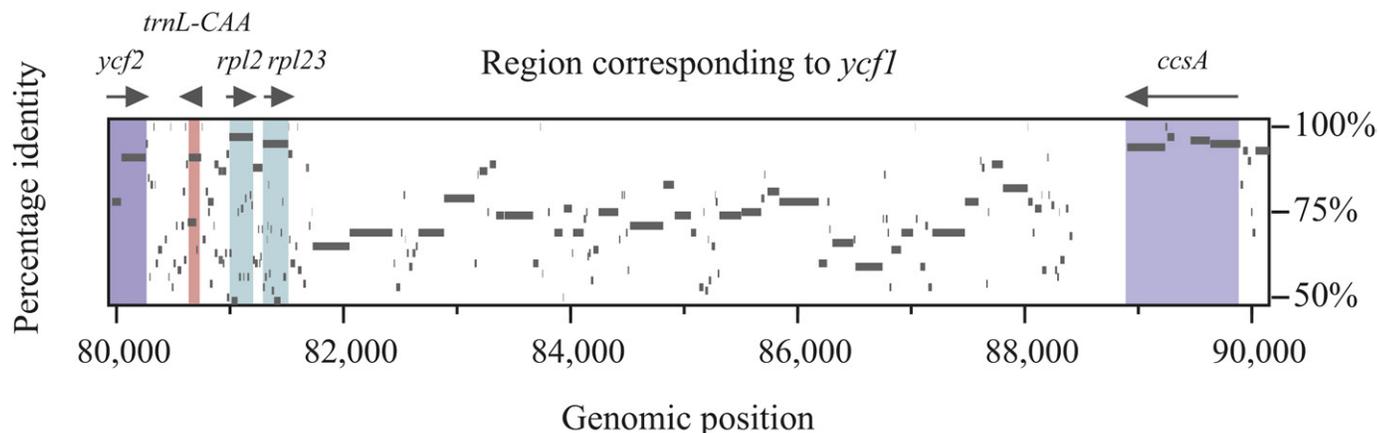


Fig. 3. The percentage identities in gene identity plots of the genomic regions around the *ycf1* gene in *Callitris sulcata* and *Calocedrus formosana*. The color coding of the genic regions corresponds to those used in Fig. 1.

Whole-chloroplast genome sequences as a resource for exploration of species-specific genetic markers—The reconstructed chloroplast genome sequence of *C. sulcata* represents the first plastid genomic resource from the Cupressaceae clade of the Southern Hemisphere, which includes 10 ecologically diverse genera containing more than 30 species (Farjon, 2005). The chloroplast genome of *C. sulcata* was strongly collinear with that of Cupressaceae of the Northern Hemisphere. In addition, PCR using the primers based on the reconstructed chloroplast genome sequence was very successful in terms of SSR marker development. These findings suggested that the cp-DNA of *C. sulcata* assembled in the current study was of a quality adequate to allow development of species-specific genetic markers.

The successful application of our chloroplast enrichment protocol to small amounts of fresh plant material means that it can be used to study endangered and extremely small species; less than 1.0 g of material is required. For example, *C. sulcata* is on the IUCN Red List as endangered (<http://dx.doi.org/10.2305/IUCN.UK.2010.RLTS.T30993A9590761.en>); the known populations are distributed in small regions of the southern province of New Caledonia (Haverkamp et al., 2015). Thus, only small juveniles grown in gardens as back-up populations for ex situ conservation were available for chloroplast genome sequencing. For such species, the method reported here can produce

species-specific chloroplast primers without unnecessary destructive sampling. The polymorphic cpSSR markers of *C. sulcata* developed here can be readily used to explore population structures shaped by paternal contributions to gene dispersal via pollen flow (Sakaguchi et al., 2014), thus providing fundamental information on the history and ecology of the species that will inform conservation strategies.

LITERATURE CITED

- ATHERTON, R. A., B. J. MCCOMISH, L. D. SHEPHERD, L. A. BERRY, N. W. ALBERT, AND P. J. LOCKHART. 2010. Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. *Plant Methods* 6: 22.
- BENNETT, M. D., AND I. J. LEITCH. 2012. Angiosperm DNA C-values database (release 8.0, Dec. 2012). Website <http://data.kew.org/cvalues/CvalServlet?querytype=2> [accessed 21 April 2017].
- BOOKJANS, G., B. M. STUMMANN, AND K. W. HENNINGSEN. 1984. Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. *Analytical Biochemistry* 141: 244–247.
- BROWNSTEIN, M. J., J. D. CARPTEN, AND J. R. SMITH. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques* 20: 1004.
- CATO, S. A., AND T. E. RICHARDSON. 1996. Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. *Theoretical and Applied Genetics* 93: 587–592.

TABLE 3. Characteristics of the polymorphic chloroplast microsatellite markers of *Callitris sulcata*.

Locus	Primer sequences (5'–3')	Repeat motif	Allele size range (bp)	A		
				Dumbéa Northern (n = 21)	Koélagoguamba (n = 20)	Total (n = 41)
Csul_cpSSR_4	F: CACGACGTTGTAAAACGACCCCTCTCTCAGGGCATAACAGC R: GTTCTTTGCCCTCCTACGTTATCCAGAA	(AT) ₆ ...	278–280	2	2	2
Csul_cpSSR_7	F: TGTGGAATTGTGAGCGGCATCCATGGCTGAATGGTAA R: GTTCTTGCATTTGCATCTCACATGAT	(GT) ₆	386–482	5	4	7
Csul_cpSSR_8	F: CACGACGTTGTAAAACGACTGGACGAAGTCGATTTACGAA R: GTTCTTAACAAACAAAGGAGAACCCGTA	(AAAT) ₄	359–363	2	1	2
Csul_cpSSR_16	F: CACGACGTTGTAAAACGACGCAAAATGGATCCAAGAACC R: GTTCTTTTAGATTGCTTGCCGTAGCC	(A) ₁₈ ...	244–284	3	1	3
Csul_1067 ^a	F: CTATAGGGCACGCGTGGTGAACAATACGTCGCCCATG R: GTTCTTTCGGAGTCCCTTCATTGCCC	(AT) ₉	437–441	2	1	2

Note: A = number of alleles.

^aThis cpSSR loci is identical to “Csul_1067” previously reported (Sakaguchi et al., 2015).

- CONANT, G. C., AND K. H. WOLFE. 2008. GenomeVx: Simple web-based creation of editable circular chromosome maps. *Bioinformatics (Oxford, England)* 24: 861–862.
- 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.
- DANIELL, H., C.-S. LIN, M. YU, AND W.-J. CHANG. 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology* 17: 134.
- DOORDUIN, L., B. GRAVENDEEL, Y. LAMMERS, Y. ARIYUREK, T. CHIN-A-WOENG, AND K. VRIELING. 2011. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Research* 18: 93–105.
- DU, F. K., T. LANG, S. LU, Y. WANG, J. LI, AND K. YIN. 2015. An improved method for chloroplast genome sequencing in non-model forest tree species. *Tree Genetics & Genomes* 11: 114.
- EBERT, D., AND R. PEAKALL. 2009. A new set of universal de novo sequencing primers for extensive coverage of noncoding chloroplast DNA: New opportunities for phylogenetic studies and cpSSR discovery. *Molecular Ecology Resources* 9: 777–783.
- EVANS, J. R. 1989. Photosynthesis and nitrogen relationships in leaves of C₃ plants. *Oecologia* 78: 9–19.
- FAIRCLOTH, B. C. 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* 8: 92–94.
- FARJON, A. 2005. A monograph of Cupressaceae and *Sciadopitys*. Kew Publishing, Richmond, Surrey, United Kingdom.
- HAHN, C., L. BACHMANN, AND B. CHEVREUX. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research* 41: e129.
- HAVERKAMP, C., L. D. PRIOR, B. FOGLIANI, L. L'HUILLIER, M. ANQUEZ, Q. HUA, AND D. M. J. S. BOWMAN. 2015. Effect of landscape fires on the demography of the endangered New Caledonian conifer *Callitris sulcata*. *Biological Conservation* 191: 130–138.
- HIRAO, T., AND A. WATANABE. 2012. Organelle DNA analysis. Bun'ichi Sogo Shuppan, Tokyo, Japan.
- HIRAO, T., A. WATANABE, M. KURITA, T. KONDO, AND K. TAKATA. 2008. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: Diversified genomic structure of coniferous species. *BMC Plant Biology* 8: 70.
- HOLLINGSWORTH, P. M., L. L. FORREST, J. L. SPOUGE, M. HAJIBABAEI, S. RATNASINGHAM, M. VAN DER BANK, M. W. CHASE, ET AL. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA* 106: 12794–12797.
- JANSEN, R. K., L. A. RAUBESON, J. L. BOORE, C. W. DEPAMPHILIS, T. W. CHUMLEY, R. C. HABERLE, S. K. WYMAN, ET AL. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. In E. A. Zimmer and E. H. Roalson [eds.], *Methods in enzymology*, vol. 395: Molecular evolution: Producing the biochemical data, Part B, 348–384. Academic Press, London, United Kingdom.
- KRESS, W. J., K. J. WURDACK, E. A. ZIMMER, L. A. WEIGT, AND D. H. JANZEN. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA* 102: 8369–8374.
- LIU, C., L. SHI, Y. ZHU, H. CHEN, J. ZHANG, X. LIN, AND X. GUAN. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13: 715.
- MURRAY, M. G., AND W. F. THOMPSON. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321–4326.
- NIJUGUNA, W., A. LISTON, R. CRONN, T.-L. ASHMAN, AND N. BASSIL. 2013. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Molecular Phylogenetics and Evolution* 66: 17–29.
- NOCK, C. J., D. L. E. WATERS, M. A. EDWARDS, S. G. BOWEN, N. RICE, G. M. CORDEIRO, AND R. J. HENRY. 2011. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal* 9: 328–333.
- PALMER, J. D. 1986. Isolation and structural analysis of chloroplast DNA. In A. Weissbach and H. Weissbach [eds.], *Methods in enzymology*, vol. 118: Plant molecular biology, 167–186. Academic Press, London, United Kingdom.
- PROVAN, J., W. POWELL, AND P. M. HOLLINGSWORTH. 2001. Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* 16: 142–147.
- REICH, P. B., M. B. WALTERS, AND D. S. ELLSWORTH. 1997. From tundra to tundra: Global convergence in plant functioning. *Proceedings of the National Academy of Sciences, USA* 94: 13730–13734.
- ROZEN, S., AND H. SKALETSKY. 1999. Primer3 on the WWW for general users and for biologist programmers. In S. Misener and S. A. Krawetz [eds.], *Methods in molecular biology*, vol. 132: Bioinformatics: Methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.
- SAKAGUCHI, S., Y. TSUMURA, M. D. CRISP, D. M. J. S. BOWMAN, AND Y. ISAGI. 2014. Genetic evidence for paternal inheritance of the chloroplast in four Australian *Callitris* species (Cupressaceae). *Journal of Forest Research* 19: 244–248.
- SAKAGUCHI, S., G. LANNUZEL, B. FOGLIANI, A. S. WULFF, L. L'HUILLIER, S. KURATA, S. UENO, ET AL. 2015. Development of nuclear and chloroplast microsatellite markers for the endangered conifer *Callitris sulcata* (Cupressaceae). *Applications in Plant Sciences* 3: 1500045.
- SCHWARTZ, S., Z. ZHANG, K. A. FRAZER, A. SMIT, C. RIEMER, J. BOUCK, R. GIBBS, ET AL. 2000. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Research* 10: 577–586.
- SHAW, J., E. B. LICKEY, E. E. SCHILLING, AND R. L. SMALL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany* 94: 275–288.
- SHI, C., N. HU, H. HUANG, J. GAO, Y.-J. ZHAO, AND L.-Z. GAO. 2012. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PLoS ONE* 7: e31468.
- 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.
- TERAKAMI, S., Y. MATSUMURA, K. KURITA, H. KANAMORI, Y. KATAYOSE, T. YAMAMOTO, AND H. KATAYAMA. 2012. Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): Genome structure and comparative analysis. *Tree Genetics & Genomes* 8: 841–854.
- URIBE-CONVERS, S., J. R. DUKE, M. J. MOORE, AND D. C. TANK. 2014. A long PCR-based approach for DNA enrichment prior to next-generation sequencing for systematic studies. *Applications in Plant Sciences* 2: 1300063.
- VIEIRA, L. N., H. FAORO, H. P. F. FRAGA, M. ROGALSKI, E. M. DE SOUZA, F. DE OLIVEIRA PEDROSA, R. O. NODARI, AND M. P. GUERRA. 2014. An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS ONE* 9: e84792.
- WEISING, K., AND R. C. GARDNER. 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.
- YANG, J.-B., D.-Z. LI, AND H.-T. LI. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources* 14: 1024–1031.
- YI, D.-K., K. CHOI, M. JOO, J. C. YANG, F. U. MUSTAFINA, J.-S. HAN, D. C. SON, ET AL. 2016. The complete chloroplast genome sequence of *Abies nephrolepis* (Pinaceae: Abietoideae). *Journal of Asia-Pacific Biodiversity* 9: 245–249.