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

SHOTA SAKAGUCHI1,2, SANEYOSHI UENO3, YOSHIHIKO TSUMURA4, HIROAKI SETOGUCHI2, MOTOMI ITO5, CHIE HATTORI2, SHOGO NOZOE2, DAIKI TAKAHASHI2, RIKU NAKAMAST6, TAISHI SAKAGAMI6, GUILLAUME LANNUZEL7, BRUNO FOGLIANI7, ADRIEN S. WULFF7,8, LAURENT L’HUILIER2, and YUI ISAGI9

• 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%...
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 throughput (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 capable of producing complete chloroplast genomes. The method was then validated on a case study. Gene annotation was performed with the aid of the msatcommander (Faircloth, 2008) to screen chloroplast regions, including microsatellites, for ≥1 mononucleotide repeats, 26 dinucleotide repeats, and ≥4 tri- to hexanucleotide repeats. We designed PCR primer pairs for these regions using Primer3 (Rozen and Skaltsky, 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 Koalagouanga River (n = 20; 22°01'S, 166°23'E). Voucher specimen accession numbers were as follows: KY00019999 for the Dumbéa River populations and MO72065 for the Koalagouanga River population. For all loci, the forward primer contained one of three different M13 sequences (5′-CAGGCGTGTGAAAACGAC-3′, 5′-TTGGAATTGTGACCGG-3′, or 5′-CTATAGGCGACCGTTGTG-3′) and the reverse primer was tagged with a PIG-tail sequence (5′-GGTTTCTT-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 program 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).

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 C. 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 metabsulfite, 0.0125 M borax, 50 mM Tris-HCl [pH 8.0], 7 mM ethylenediamine tetraacetic 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 precooled 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 2000 × 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 purified DNA was then 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 the 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 GenomeVs (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. ABB831010).

We used MSATCOMMANDER (Faircloth, 2008) to screen chloroplast regions, including microsatellites, for ≥1 mononucleotide repeats, 26 dinucleotide repeats, and ≥4 tri- to hexanucleotide repeats. We designed PCR primer pairs for these regions using Primer3 (Rozen and Skaltsky, 1999); we chose an optimal annealing temperature of 60°C±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 Koalagouanga River population (n = 20; 22°01’S, 166°23’E). Voucher specimen accession numbers were as follows: KY00019999 for the Dumbéa River populations and MO72065 for the Koalagouanga River population. For all loci, the forward primer contained one of three different M13 sequences (5′-CAGGCGTGTGAAAACGAC-3′, 5′-TGTTGAATTGTGACCGG-3′, or 5′-CTATAGGCGACCGTTGTG-3′) and the reverse primer was tagged with a PIG-tail sequence (5′-GGTTTCTT-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 program 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 |
|------------------|-----------|------------------------|----------------------------------|-------------------------|----------------------|------------------------------------------|---------------------|--------------------------|---------------------|-----------------------|-----------------------|
| *Patrinia triloba* (Miq.) Miq. var. takeuchiana (Makino) Ohwi (Valerianaceae) Deciduous herb | 1.11 (0.75–1.29) | *Kolkwitzia amabilis* Greb. (KT966716) | 2.00 | 413,236 | 158,524 | 246,049 | 59.54 | 237 | AP017910 KYO 00019985 |
| *Tricyrtis macropoda* Miq. (Liliaceae) Deciduous herb | 4.24 (3.73–4.75) | *Lilium tsingtauense* Gilg (KM103365) | 0.95 | 197,233 | 153,981 | 48,516 | 24.6 | 52 | AP017907 KYO 00019984 |
| *Rumex nepalensis* subsp. andreaeanus (Makino) Yonek. (Polygonaceae) Deciduous herb | 2.58 (0.48–6.24) | *Rumex acetosa* L. (KC817303) | 1.40 | 280,233 | 158,781 | 63,443 | 22.64 | 74 | AP017909 KYO 00019986 |
| *Asarum sakawanum* Makino var. stellatum (F. Maek. ex Akasawa) T. Sugaw. (Aristolochiaceae) Evergreen herb | 6.75 (4.65–8.95) | *Piper kadsura* (Choisy) Ohwi (KT223569) | 0.84 | 81,578 | 161,922 | 16,816 | 20.61 | 20 | AP017908 KYO 00019902 |
| *Celtis sinensis* Pers. (Cannabaceae) Deciduous tree | 1.35 | *Humulus lupulus* L. (KT266264) | 2.10 | 303,264 | 154,809 | 53,858 | 17.76 | 64 | AP017911 KYO 00019989 |
| *Zelkova serrata* Thunb. Makino (Ulmaceae) Deciduous tree | Unknown | *Humulus lupulus* L. (KT266264) | 2.00 | 516,820 | 157,692 | 83,609 | 16.18 | 93 | AP017905 KYO 00019987 |
| *Aphananthe aspera* (Thunb.) Planch. (Cannabaceae) Deciduous tree | Unknown | *Humulus lupulus* L. (KT266264) | 2.00 | 1,249,307 | 156,963 | 181,915 | 14.56 | 222 | AP017906 KYO 00019988 |
| *Callitris sulphata* Schltr. (Cupressaceae) Evergreen conifer | 9.22 (8.25–10.40) | *Calocedrus formosana* (Florin) Florin (AB831010) | 0.40 | 1,939,765 | 131,609 | 196,403 | 10.13 | 220 | AP017904 KYO 00019999 |

*a The 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).  
*b Vouchers 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 × (no. of cpDNA reads) – 0.0015 × (genome size) + 1.115 × (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× and 100×, 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× 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 mature (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 genomic 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 autoscaler. 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; Doorduin et al., 2011; Njugu 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 nepholproles* (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...
herbaceous species (20.6–59.5%), comparable to those obtained using RCA approaches (Atherton et al., 2010; Du et al., 2015). The protocol used in this study is simple (i.e., no elaborate chloroplast isolation steps or long-range PCR, which could introduce PCR errors), widely applicable, and yields relatively high proportions of chloroplast reads compared with those of genomic libraries and RCA-based methods.

The functional and ecological traits of leaf materials are important in terms of the proportions of cpDNA reads. Differences in the cpDNA read percentages among plant life forms are

Fig. 1. Gene distribution in the chloroplast genome of *Callitris sulcata*. The genomic region corresponding to the *ycf1* gene is indicated by a gray arrow.
Table 2. Genes detected in the chloroplast genome of *Callitris sulcata*.

| Gene category                        | Gene group                  | Genes                             | No. |
|--------------------------------------|-----------------------------|-----------------------------------|-----|
| Self-replication                     | Ribosomal RNA genes         | *rrn4.5, rrn5, rrn16, rrn23*      | 4   |
|                                      | Transfer RNA genes          | *trnC-GAG (x2), trnD-GTC (x2), trnE-TTC, trnF-GAA, trnM-CAT, trnG-GCC* | 31  |
|                                      |                             | *trnH-GTG, trnL-CAT (x2), trnL-CCT (x2), trnL-CAA (x2), trnL-TAG (x2), trnN-GTT, trnP-GGG, trnP-TGG (x2), trnQ-TTG, trnQ-TGA, trnR-ACG, trnR-TCT, trnS-GCT (x2), trnS-TGA, trnT-VGC, trnV-TAC, trnV-WCC (x2), trnY-GTA* | 6   |
|                                      | Large subunit of ribosome   | *rpl2 (x2), rpl14, rpl16, rpl20, rpl22, rpl23 (x2), rpl26* | 9   |
|                                      | Small subunit of ribosome   | *rps2, rps3, rps4, rps7, rps8, rps11, rps12 (x2), rps14, rps15, rps18, rps19* | 12  |
|                                      | DNA-dependent RNA polymerase| *rpoA, rpoB, rpoC1, rpoC2*       | 4   |
|                                      | Translational initiation factor| *infA*                           | 1   |
| Genes for photosynthesis             | Subunits of photosystem I   | *psaA, psaB, psaC, psaI*          | 5   |
|                                      | Subunits of photosystem II  | *psaA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbK, psbL, psbN, psbT, psbZ* | 12  |
|                                      | Subunits of cytochrome      | *petA, petB, petD, petG (x2), petL (x2), petN (x2)* | 10  |
|                                      | Subunits of ATP synthase    | *atpA, atpB, atpE (x2), atpF, atpH, atpI* | 7   |
|                                      | Large subunit of RuBisCO    | *rbcL*                            | 1   |
|                                      | Chlorophyll biosynthesis    | *chlB, chlL, chlN*               | 3   |
|                                      | Subunits of NADH dehydrogenase| *ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK* | 11  |
| Other genes                          | Maturase                    | *matK*                            | 1   |
|                                      | Envelope membrane protein   | *cemA*                            | 1   |
|                                      | Subunit of acetyl-CoA-carboxylase| *accD*                           | 1   |
|                                      | c-Type cytochrome synthesis gene| *ccsA*                           | 1   |
| Genes of unknown function            | Conserved open reading frames| *ycf2, ycf3, ycf4*               | 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).
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. McCormish, L. D. Shepherd, L. A. Berry, N. W. Albert, and P. J. Lockhart. 2010. Whole genome sequencing of enriched chloroplast DNA using the Illumina GAI 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. CarPien, 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) | Dumbéa Northern (n = 21) | Koéalagoumba (n = 20) | Total (n = 41) |
|-----------|-------------------------|--------------|------------------------|--------------------------|------------------------|-----------------|
| Csul_cpSSR\_4 | F: CACGAGTTTTGATACCCACCTTCCTCAGGGGATACACGC | (AT)\_h … (AT)\_h | 278–280 | 2 | 2 | 2 |
|           | R: GTTCTTTCGTCCTCTTACGTATTTCCAGAA |              |                       |                          |                        |                 |
| Csul_cpSSR\_7 | F: TGGGAAATTTGAGCGAGCAACCATGGCTGATGTTAA | (GT)\_h | 386–482 | 5 | 4 | 7 |
|           | R: GCTTCTTACGTATTTCCAGAA |              |                       |                          |                        |                 |
| Csul_cpSSR\_8 | F: CACGAGTTTTGATACCCACCTTCCTCAGGGGATACACGC | (AAAT)\_4 | 359–363 | 2 | 1 | 2 |
|           | R: GTTCTTTCGTCCTCTTACGTATTTCCAGAA |              |                       |                          |                        |                 |
| Csul_cpSSR\_16 | F: CACGAGTTTTGATACCCACCTTCCTCAGGGGATACACGC | (A)\_16 … (A)\_16 | 244–284 | 3 | 1 | 3 |
|           | R: GCTTCTTTCGTCCTCTTACGTATTTCCAGAA |              |                       |                          |                        |                 |
| Csul\_1067\* | F: CTTAGGAGCTCCGCTTGGTGAAACATACGTCGGCAGAC | (AT)\_h | 437–441 | 2 | 1 | 2 |
|           | R: GCTTCTTTCGTCCTCTTACGTATTTCCAGAA |              |                       |                          |                        |                 |

*Note: A = number of alleles.

*This cpSSR loci is identical to “Csul\_1067” previously reported (Sakaguchi et al., 2015).*

http://www.bioone.org/loi/apps
