Appendix 1.1: Supplementary materials and methods: DNA extraction, library preparation, and high-throughput sequencing

The novel genomes sequences included in this study are all from Australian Rainforest tree species (Table A1.1). Leaf material for each species was collected from Dorrigo National Park, Nightcap National Park, or Washpool National Park in northern New South Wales, Australia. Total genomic DNA was extracted from silica-dried leaf material at the Australian Genome Research Facility (Adelaide, Australia) using the NucleoSpin® Plant II Maxi kits (Machery-Nagel). Total DNA concentrations used for library preparation ranged between 14 ng/μl and 146 ng/μl. Nextera libraries were prepared according to the Nextera® DNA Sample Preparation Guide from Illumina. Libraries for the 11 species presented in this manuscript were multiplexed along with 44 libraries from another study (to a total of 55 libraries per lane) and paired-end (2×100 bp) shotgun sequencing was performed on the Illumina HiSeq 2000. Library preparation and sequencing were conducted by the Ramaciotti Centre for Genomics (University of New South Wales, Australia).

Assembly of chloroplast DNA data from the whole genome shotgun libraries was conducted using CLC Bio genomics Workbench 6.5.1 (available from http://www.clcbio.com) following the approach previously outlined by McPherson et al. (2013) and van der Merwe et al. (2014). The paired-end Illumina reads with insert sizes from 100 to 600 bp were imported using the Illumina pipeline 1.8 or later, with default settings. Reads below 50 bp were discarded and the remaining reads were quality trimmed using default settings. For each species, two de novo assemblies were conducted, one with a minimum contig length of 200 bp and another with a minimum contig length of 1000 bp, both using the default settings. The contigs were combined and subjected to a BLASTn similarity search in CLC against a database of 213 angiosperm whole chloroplast genomes (downloaded from the NCBI Organelle Genome Resources page at http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=plastid&taxid=3398, on 4 June 2013). Default settings were used except word size was 20 bp,

and maximum number of hit sequences set to 10. From this BLASTn result contigs with e-values of zero (very high similarity to the chloroplast genome database) were selected.

For each species, the chloroplast contigs were assembled together in Geneious Pro v6.1.7 (Biomatters, http://www.geneious.com) using a two-step approach. During the first step, the contigs generated in CLC were assembled using the de novo assembly available in Geneious Pro. In the second step, the new contigs assembled in Geneious Pro and any contigs generated in CLC that failed to assemble in the first step were concatenated. The contigs were aligned to a reference(s) and concatenated according to the order of this alignment. The chloroplast genome available in GenBank that was taxonomically closest to the new species was selected as the reference (except in the cases of Wilkiea huegeliana and Cinnamomum oliveri where the relevant species-specific cpDNA reference sequences from van der Merwe et al. (2014) were also used as a reference) and alignment implemented the Mauve algorithm (Darling et al. 2010) in Geneious Pro.

For each species, the resulting chloroplast sequence was imported into CLC and the quality-trimmed Illumina reads for each species were mapped onto the relevant sequence using a length fraction of 0.9 and with otherwise default settings. The consensus sequences were submitted to CPGAVAS (http://www.herbalgenomics.org/0506/cpgavas/analyzer/annotate) for annotation using all available plant reference sequences and all other settings as default. The resultant GFF3 files were used to extract the genes used in this study for each species.

References

Darling A.E., Mau B., Perna N.T. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLOS ONE. 5:e11147.

McPherson H., van der Merwe M., Delaney S.K., Edwards M.A., Henry R.J., McIntosh E., Rymer P.D., Milner M.L., Siow J., Rossetto M. 2013. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. BMC Ecol. 13:8.

van der Merwe M., McPherson H., Siow J., Rossetto M. 2014. Next-Gen phylogeography of rainforest trees: exploring landscape-level cpDNA variation from whole-genome sequencing. Mol. Ecol. Resour. 14:199–208.