Muelleria

41: 44–49

Published online in advance of the print edition, 30 June 2023

ROYAL BOTANIC GARDENS VICTORIA

Genome size estimates for Genomics for Australian Plants sequencing projects

Alexander N. Schmidt-Lebuhn[1] and David J. Cantrill[2]

[1] CSIRO, Centre for Australian National Biodiversity Research, Clunies Ross Street, Canberra, Australian Capital Territory 2601, Australia; e-mail: alexander.s-l@csiro.au

[2] Royal Botanic Gardens Victoria, Birdwood Ave, Melbourne, Victoria 3004, Australia; e-mail: david.cantrill@rbg.vic.gov.au

Abstract

Genome sizes were estimated with flow cytometry for candidate species for genome sequencing by the Genomics for Australian Plants consortium. Sampled species represent taxonomic groups where no or few genome sizes have previously been published. None of the three analysed Cunoniaceae produced clear peaks; it is possible that their genomes are very small. *Roepera similis* showed strong endopolyploidy, and because of the lack of a suitable standard and time constraints, no formal measurement was made. The first genome size estimates are presented for *Apodasmia brownii*, *Flindersia xanthoxyla*, *Gastrolobium racemosum*, *Hedycarya angustifolia*, *Hypolaena fastigiata*, *Pittosporum angustifolium*, *Phyllanthus gunnii*, *Quintinia fawkneri*, and *Ripogonum discolor*.

***Keywords*:** endopolyploidy, flow cytometry, genome sequencing, genome size, polyploidy

Introduction

The Genomics for Australian Plants (GAP) consortium supports the sequencing and assembly of reference genomes for selected species as resources for the study of the evolution and conservation of the Australian flora (https://www.genomicsforaustralianplants.com/reference-genomes/). As of the time of writing in April 2023, 23 reference genome projects have commenced, with some already having proceeded to publication (Chen *et al.* 2022; McLay *et al.* 2022).

When considering the selection of species for sequencing, genome size is an important factor, because larger genomes require more investment in sequencing coverage and may be bioinformatically more challenging to assemble than species with small genomes (Hamilton & Robin Buell 2012). When choosing a species of a given genus for sequencing, a diploid with low DNA content per cell would therefore be preferred over a hexaploid relative, for example.

Currently, the most efficient approach for the determination of genome sizes is flow cytometry (Doležel *et al.* 2007). Living tissue, generally leaves, from a sample plant and a standard plant of known genome size, is finely chopped together in an extraction buffer to release intact cell nuclei. DNA in these nuclei is then stained with a fluorescent dye and observed in a flow cell (McKinnon 2018). The ratio of the fluorescence intensity peaks of the sample and of the standard allows the calculation of the absolute size of the sample genome in picograms.

ISSN: 0077-1813 (print) · ISSN: 2204-2032 (online)

The laboratory of the Australian National Herbarium in Canberra has a flow cytometer that was set up to study ploidy levels in plants and has previously been used to estimate genome sizes, mostly for taxonomic or conservation genetic studies (Castelli *et al.* 2017; Chen *et al.* 2019).

In April 2023, the first author analysed a set of samples provided by Royal Botanic Gardens Victoria of candidate species for genome sequencing by the GAP consortium. The species were selected for flow cytometric analysis to represent so far under-studied groups with few or no published genome size measurements (Fig. 1). Here, the results of these analyses are reported because both the successful measurements and the taxonomically biased failures may be of interest to colleagues studying genome size evolution or planning future measurements in the same plant groups.

Methods

Leaf or (in the case of Restionaceae) stem samples of 13 species (Table 1) were collected at Royal Botanic Gardens Victoria's Melbourne and Cranbourne sites on 03 April 2023 and maintained in zip lock bags on ice. They were transported to Canberra on 04 April and analysed on 05, 06, and 12 April.

CyStain PI Absolute P (Sysmex Partex GmbH, Görlitz, Germany) was used for sample preparation following the manufacturer's instructions, except for halved reaction volumes. A tissue sample of approximately 0.5–1.0 cm^2 from both sample and standard was placed in 300 µL of extraction buffer in a Petri dish and chopped manually with a razor blade. The liquid was filtered through a 40 µm cell strainer and then transferred to a sample tube. The sample was mixed with 1,000 µL staining solution (staining buffer, propidium iodide, and RNAse). It was loaded into the flow cytometer (BD Accuri C6 Plus equipped with a 488 nm laser and a BD CSampler Plus, BD Biosciences, San Jose, CA, USA) and run at a flow rate of 14 mm/min per min. Histogram data were collected using the FL2 detector while eliminating events with a value of less than 80,000 on FL2-H. Analysis was performed with the BD Accuri C6 Software version 1.0.23.1. Soy (*Glycine max* (L.) Merr. 'Polanka', 2C = 2.50 pg) (Doležel *et al.* 1994) and pea (*Pisum sativum* L.

Table 1. Sample collection information, genome size standard used, and (where successful) genome size estimates for 13 Australian angiosperms. MEL accession numbers are those of voucher specimens deposited at the National Herbarium of Victoria, Melbourne. RBG accession numbers are those of the plants in the living collections of Royal Botanic Gardens Victoria (RBGM = Melbourne; RBGC = Cranbourne). *Hypolaena fastigiata* was a wild collection at RBGV Cranbourne.

Species	Family	Voucher	Standard	Measurements	Genome size (2C)
Callicoma serratifolia Andrews	Cunoniaceae	*D.J. Cantrill 2359* (MEL 2529498A; RBGM 501996)	NA	0	NA
Eucryphia lucida (Labill.) Baill.	Cunoniaceae	*G.D. Holmes 174* (MEL 2529901A; RBGC 194355)	NA	0	NA
Pullea stutzeri (F.Muell.) Gibbs	Cunoniaceae	*D.J. Cantrill 2364* (MEL 2529503A; RBGM 951570)	NA	0	NA
Gastrolobium racemosum (Turcz.) Crisp	Fabaceae	*G.D. Holmes 157* (MEL 2529504A; RBGC 114382)	pea	1	c. 1.69 pg
Hedycarya angustifolia A.Cunn.	Monimiaceae	*G.D. Holmes 175* (MEL 2529902A; RBGC 104027)	pea	3	1.38 (±0.02) pg
Quintinia fawkneri F.Muell.	Paracryphiaceae	*G.D. Holmes 155* (MEL 2529504A; RBGC 194330)	pea	3	2.35 (±0.06) pg
Phyllanthus gunnii Hook.f.	Phyllanthaceae	*D.J. Cantrill 2363* (MEL 2529502A; RBGM 170598)	pea	3	1.99 (±0.03) pg
Pittosporum angustifolium Lodd., G.Lodd. & W.Lodd.	Pittosporaceae	*G.D. Holmes 176* (MEL 2529903A; RBGC 014073.1)	soy	3	8.75 (±0.25) pg
Apodasmia brownii (Hook.f.) B.G.Briggs & L.A.S.Johnson	Restionaceae	*G.D. Holmes 177* (MEL 2529904A; RBGC 114125)	pea	3	1.55 (±0.04) pg
Hypolaena fastigiata R.Br.	Restionaceae	*G.D. Holmes 173* (MEL 2529905A)	2x pea, 1x soy	3	1.23 (±0.02) pg
Ripogonum discolor F.Muell.	Ripogonaceae	*D.J. Cantrill 2360* (MEL 2529499A; RBGM 941157)	pea	3	20.18 (±0.18) pg
Flindersia xanthoxyla (A.Cunn. ex Hook.) Domin	Rutaceae	*V. Stajsic 8999* (MEL 2457406A; RBGM 534828)	pea	3	0.86 (±0.03) pg
Roepera similis (H.Eichler) Beier & Thulin	Zygophyllaceae	*B.A. Swartz 15* (MEL 2478506A; RBGC 194032)	NA	0	c. 2.0–2.5 pg?

Figure 1. Four of the species whose genome sizes were estimated: a. *Gastrolobium racemosum* (Fabaceae). b. *Hedycarya angustifolia* (Monimiaceae). c. *Flindersia xanthoxyla* (Rutaceae). d. *Hypolaena fastigiata* (Restionaceae). Image sources: Australian Plant Image Index, photo numbers a.30121, a.27788, a.27376, and dig.35369, © Murray Fagg.

'UTAS Line 107', 2C = 9.09 pg) (Price 2010) were used as internal standards, depending on the size of the sample peak as determined by preliminary analyses without using standards.

Results and discussion

The estimated genome sizes, standards used, and sample voucher information are summarised in Table 1.

All three Cunoniaceae (*Callicoma serratifolia*, *Eucryphia lucida*, *Pullea stutzeri*) failed to produce distinct fluorescence peaks. It is possible that their genomes are very small, and peaks were obscured by the debris field (stained material other than intact nuclei; Fig. 2A). Unfortunately, the Plant DNA C-values database (https://cvalues.science.kew.org/search/angiosperm, accessed 07 April 2023) does not contain records for any other Cunoniaceae, so it remains unclear whether our interpretation is correct.

Roepera similis (Zygophyllaceae) showed strong endopolyploidy (Fig. 2B), *i.e.*, duplication of genomes without mitosis (Leitch & Dodsworth 2017). At the time of analysis, only two standards were at hand; the smaller one (soy) would have been too close to the *Roepera* 2C peak, and the larger one (pea) too close to the *Roepera* 8C peak. Therefore, no formal measurement was produced. The approximate position of the 2C peak just slightly left of where soy would have been suggests that the genome size of *Roepera similis* would be around 2C = c. 2.0–2.5 pg.

The analysis of *Gastrolobium racemosum* (Fabaceae) showed a large debris field that the 2C peak of the sample barely exceeded, leading to a poor measurement. Its genome size is estimated at 2C = c. 1.70 pg.

The remaining species were analysed using triplicate measurements, each time from a separate extraction: *Apodasmia brownii* (Restionaceae), 2C = 1.55 (±0.04) pg; *Flindersia xanthoxyla* (Rutaceae), 2C = 0.86 (±0.03) pg; *Hedycarya angustifolia* (Monimiaceae), 2C = 1.38 (±0.02) pg (Fig. 2C); *Hypolaena fastigiata* (Restionaceae), 2C = 1.23 (±0.02) pg; *Pittosporum angustifolium* (Pittosporaceae), 2C = 8.75 pg (±0.25) pg (Fig. 2D); *Phyllanthus gunnii* (Phyllanthaceae), 2C = 1.99 (±0.03) pg; *Quintiania fawkneri* (Paracryphiaceae), 2C = 2.35 (±0.06) pg; *Ripogonum discolor* (Ripogonaceae), 2C = 20.18 (±0.18) pg.

The genome size of *Pittosporum angustifolium* was unexpectedly large, as the only two measurements of the same genus in the C-values database were 2C = 0.92 pg and 2C = 1.22 pg, respectively (Hanson *et al.* 2001; Horjales *et al.* 2003). What is more, chromosome counts are available for numerous species of *Pittosporum* Banks ex Sol. (http://ccdb.tau.ac.il, accessed 12 April 2023), all consistently 2n = 24, including one of *P. angustifolium* (as *P. phillyreoides* DC.). Further research will be required to establish if this species is unexpectedly a polyploid complex or if genome size variation at the same ploidy level is strong in the genus.

A genome size of one picogram of DNA is equivalent to 0.978 gigabases (Doležel *et al.* 2003). However, for the purposes of sequencing and bioinformatic genome assembly, the relevant size is that of the haploid genome, so that the 2C values discussed here can be halved to obtain an approximate haploid size in gigabases.

Conversely, flow cytometry cannot provide information on ploidy level except through direct comparison of the genome sizes of close relatives, e.g., when finding that some samples have approximately twice or four times the genome size as others from the same genus. Because sampling was not designed in this way, the ploidy level of samples analysed in this study remains unknown.

Acknowledgements

We are grateful to Gareth Holmes (RBGV) and Todd McLay (RBGV) for collecting samples, to Ashley Jones (ANU) and Callum Lawler (CSIRO) for transferring the samples between institutions, to Francisco Encinas Viso (CSIRO) for assistance in the laboratory, and to Nigel England (CSIRO) for caring for the standard plants.

References

Castelli, M., Miller, C.H. and Schmidt-Lebuhn, A.N. (2017). Polyploidization and genome size evolution in Australian billy buttons (*Craspedia*, Asteraceae: Gnaphalieae). *International Journal of Plant Sciences* **178**, 352–361. doi:10.1086/691460

Chen, S.H., Guja, L.K. and Schmidt-Lebuhn, A.N. (2019). Conservation implications of widespread polyploidy and apomixis: a case study in the genus *Pomaderris* (Rhamnaceae). *Conservation Genetics* **20**, 917–926. doi:10.1007/s10592-019-01184-2

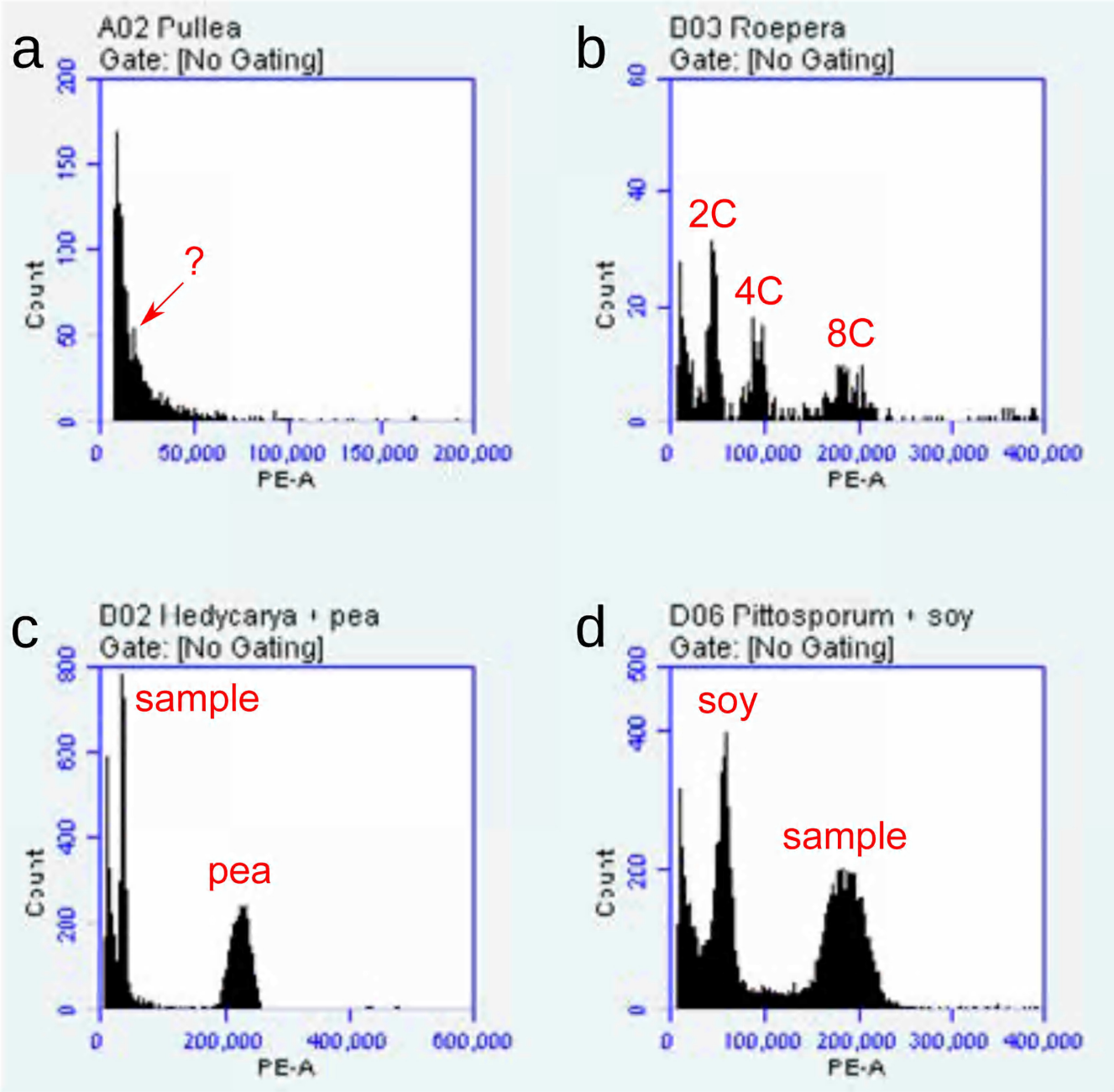

Figure 2. Flow cytometry histograms of selected analyses showing on the y-axis the frequency of nuclei with relative genome sizes indicated on the x-axis. a. *Pullea stutzeri* showing no peak that could with certainty be differentiated from the debris field on the left. The other two Cunoniaceae showed the same result. b. Endopolyploidy in *Roepera similis*, with strong 4C and 8C peaks. No standards were included in a. and b. c. *Hedycarya angustifolia* (2C = 1.38 pg) with pea (2C = 9.09 pg) as the standard. d. *Pittosporum angustifolium* (2C = 8.75 pg) showing an unexpectedly large with soy (2C = 2.5 pg) as the standard.

Chen, S.H., Rossetto, M., van der Merwe, M., Lu-Irving, P., Yap, J.-Y.S., Sauquet, H., Bourke, G., Amos, T.G., Bragg, J.G. and Edwards, R.J. (2022). Chromosome-level de novo genome assembly of *Telopea speciosissima* (New South Wales waratah) using long-reads, linked-reads and Hi-C. *Molecular Ecology Resources* **22**, 1836–1854. doi:10.1111/1755-0998.13574

Doležel, J., Bartoš, J., Voglmayr, H. and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry Part A* **51A**, 127–128. doi:10.1002/cyto.a.10013

Doležel, J., Doleželová, M. and Novák, F.J. (1994). Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biologia Plantarum* **36**, 351. doi:10.1007/BF02920930

Doležel, J., Greilhuber, J. and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* **2**, 2233–2244. doi:10.1038/nprot.2007.310

Hamilton, J.P. and Robin Buell, C. (2012). Advances in plant genome sequencing. *The Plant Journal* **70**, 177–190. doi:10.1111/j.1365-313X.2012.04894.x

Hanson, L., McMahon, K.A., Johnson, M.A.T. and Bennett, M.D. (2001). First Nuclear DNA C-values for 25 Angiosperm Families. *Annals of Botany* **87**, 251–258. doi:10.1006/anbo.2000.1325

Horjales, M., Redondo, N., Blanco, A. and Rodriguez, M.A. (2003). Cantidades de DNA nuclear en arboles y arbustos. *Nova Acta Cientifica Compostelana (Bioloxia)* **13**, 23–33

Leitch, I.J. and Dodsworth, S. (2017). Endopolyploidy in plants. 'Encycl. Life Sci.' pp. 1–10. (John Wiley & Sons, Ltd) doi:10.1002/9780470015902.a0020097.pub2

McKinnon, K.M. (2018). Flow cytometry: an overview. *Current Protocols in Immunology* **120**, 5.1.1-5.1.11. doi:10.1002/cpim.40

McLay, T.G.B., Murphy, D.J., Holmes, G.D., Mathews, S., Brown, G.K., Cantrill, D.J., Udovicic, F., Allnutt, T.R. and Jackson, C.J. (2022) A genome resource for *Acacia*, Australia's largest plant genus. *PLOS ONE* **17**, e0274267. doi:10.1371/journal.pone.0274267

Price, A. (2010). *Advanced breeding and deployment methods for tropical Acacias*. University of Tasmania, Internal Report FST/2008/007.