GMR
The Original by FUNPEC-RP

Genome size and chromosome number of *Dipteryx alata* (Leguminosae): a model candidate for comparative genomics in Papilionoideae

Short communication

A.M. Antunes [1], C.P. Targueta [1], A.A. Castro [1], G. Souza [2], T.N. Soares [1] and M.P.C. Telles [1,3]

[1] Laboratório de Genética e Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia.,GO, Brasil
[2] Laboratório de Citogenética e Evolução Vegetal, Universidade Federal de Pernambuco, Recife, PE, Brasil
[3] Escola de Ciências Agrárias e Biológicas da Pontifícia Universidade Católica de Goiás, Goiânia, GO, Brasil

Corresponding author: A.M. Antunes
E-mail: adrianaantunesbio@gmail.com

Genet. Mol. Res. 19 (3): gmr18640
Received May 24, 2020
Accepted August 31, 2020
Published September 30, 2020
DOI http://dx.doi.org/10.4238/gmr18640

ABSTRACT. Genome size and chromosome number are basic biological characteristics that reveal a wide diversity in plants. The most widely used method to estimate genome size is flow cytometry; however, genomic estimation is only available for a small number of species. In this context, the DNA content of plants distributed in areas rich in species and endemism, such as the Brazilian Cerrado, remains little known. We examined genome size and number of chromosomes for the legume *Dipteryx alata* (popularly known as "baru"), a tree used for food, medicine, forage, recovery of degraded areas, landscaping and wood extraction. *Dipteryx alata* showed $2n$ = 16, with small chromosomes. Genome size or C-value was estimated at 1C = 0.825 pg which corresponds to 807.2 Megabases. This species is an important genetic resource, though it has a very small genome. In addition, this species is phylogenetically positioned in the first diverging lineage of Papilionoideae. Therefore, *D. alata* is a strong candidate among the tree species of the Brazilian Cerrado to

be a model species in studies of comparative genomics of Leguminosae.

Key words: Baru; DNA content; Flow cytometry; Genetic resource

INTRODUCTION

The genome size, commonly referred to as C-value, and the chromosome number are parameters that shows a wide diversity in plants (Doležel and Bartoš 2005; Hidalgo et al., 2015; Veleba et al., 2017). Among angiosperms, the variation in genome size is about 2,400-fold, from 1C = 0.06 pg (*Genlisea margaretae*) to 1C = 152.23 pg (*Paris japonica*) (Greilhuber et al., 2006; Pellicer et al., 2010). The variation in chromosome number is also large, the smallest number 2*n* = 4 found in species such as *Brachyscome dichromosomatica* and the largest number in fern species *Ophioglossum reticulatum* 2n = 1,400 (Castiglione and Cremonini, 2012; Dyer et al., 2013). The investigation of genome size and 2*n* of a species is relevant for many areas of plant biology, especially in plant genomics (Tyagi et al., 2019). In genomics, knowledge of genome sizes has practical implications because it allows us to estimate the coverage and cost of whole genome sequencing projects (Garcia et al., 2014; Wang et al., 2015; Pati et al., 2019). Despite this, genome size estimates are available for only ca. 7500 angiosperms, i.e. 2% of known species (Wang et al., 2015).

The Cerrado is a savannah-type biome that occurs in Central Brazil and is considered a biodiversity hotspot (Mendonça et al., 2008). Among the endemic species of the Cerrado is *Dipteryx alata* (Leguminosae), a tree popularly known as "baru". *Dipteryx alata* has potential for economic use due to its pulp and seeds that can be used fresh or in the manufacture of ice cream, cream, liqueur, roasted nuts, medicinal, forage, the recovery of degraded areas, landscaping and wood extraction (Sano et al., 2004). This species belongs to the earliest-branching papilionoids within clade ADA, which includes the reorganized monophyletic tribes Angylocalyceae, Dipterygeae, and Amburanae (Cardoso et al., 2013). Papilionoids include most model species of Leguminosae (Schmutz et al., 2010; Varshney et al., 2013; Schmutz et al., 2014; Tang et al., 2014). In this sense, the characterization of genomes of species of its first lineages (ADA clade) may help in future studies of comparative genomics. We examined genome size and number of chromosomes for *D. alata*, a genetic resource of the Brazilian Cerrado and a candidate for genomic sequencing. This work is the first report of genome size and a confirmation of chromosome number for *D. alata*.

MATERIAL AND METHODS

Samples of leaves of *D. alata* were collected from three specimens of the germplasm collection maintained by the Agronomy School of the Federal University of Goiás. This germplasm collection was planted in 2011 (latitude 16°35'58.96" S, longitude 49°16'49.55" W, altitude 736 m) and has 600 individual accesses from 25 populations of *D. alata* sampled in Cerrado regions in the Brazilian states of Goiás, Mato Grosso, Tocantins, Bahia and Minas Gerais (Guimarães et al., 2019a). For an internal standard, leaves of *Glycine max* (2C-Value = 2.5 pg, Graham et al., 1994) from germinated seeds supplied by

Dr. Jaroslav Doležel (Experimental Institute of Botany, Czech Republic) were used to obtain a genome size estimation.

The estimation of nuclear DNA content (C-value) was performed according to the protocol described by Dolezel and Gohde (1995). Small fragments of leaf tissue (40 mg) from the sample and the standard were co-minced with a razor blade in a petri dish containing 0.75 mL of WPB buffer (Tris-HCL 0.02 M, $MgCl_2 \cdot 6H_2O$ 4 mM, Triton X 100 1%, EDTA $Na_2 \cdot 2H_2O$ 2 mM, NaCl 86 mM, sodium metabisulfite 10 mM 2 PVP-10 1%). The samples were fragmented quickly to minimize the release of cytosolic compounds. Then, the contents of the petri dish were filtered using a mesh with 30 μm pores in a flow cytometry tube to remove cell debris and large debris. Cell nuclei were stained with propidium iodide 50 μL (1 mg/mL). This preparation was repeated in triplicate for each sample of *D. alata* and three measurements were carried out on consecutive days. Samples were analyzed on a Partec CyFlow Space (Partec, Muenster, Germany) equipped with an argon laser operating at 488 nm. The fluorescence intensity of 10,000 particles was recorded. The results, number of cores and coefficient of variation (CV) were obtained for *D. alata* and internal standard (*G. max*) samples using a Partec FloMax software (Version 2.4). For this, the sample and standard peaks were manually defined in the histograms. The nuclear DNA content (2C value) of *Dipteryx alata* was determined according to the following formula: (average peak sample / average standard x standard peak 2C-value in pg).

The determination of the chromosome number was performed from roots of germinated seeds. The seeds were obtained from the plants of the aforementioned germplasm collection. Root tips were pretreated with 0.002 M 8-hydroxyquinoline at room temperature for 20h, fixed overnight in ethanol / glacial acetic acid 3: 1 (v / v) and stored at -6°C. Root tips were digested using 18 U/mL cellulase and 18 U/mL pectinase for 120 min at 37°C. The meristem was dissected in acetic acid (45%) and crushed under a coverslip. Conventional staining was performed with Giemsa 10%. Capture was done under a Leica microscope and with the LAS EZ software (Leica). Diploid information was determined by counting chromosomes from four individuals. The proposed chromosome order was done using Photoshop CS5.

RESULTS AND DISCUSSION

The C-value of the genome size of *D. alata* was 1C = 0.825 pg corresponding to 807.2 Megabases (Table 1, Figure 1). The coefficient of variation for samples and standard were less than 5%, indicating that the results are consistent. In all prometaphases and metaphases analyzed, *D. alata* showed $2n = 16$ chromosomes (Figure 2).

Table 1. Genome size and chromosome number of *Dipteryx alata* and other species of Dypterygeae.

Specie	2n	1C-value (pg)	1C-value (Mpb)	Reference
Dipteryx alata	16	0.82	807.2	This study
Dipteryx magnifica	-	0.98	958.4	Madrigal (2018)
Dipteryx odorata	32	1.63	1594.1	Madrigal (2018)
Dipteryx oleifera	-	1.93	1887.5	Madrigal (2018)
Dipteryx rosea	-	1.26	1.232.28	Madrigal (2018)

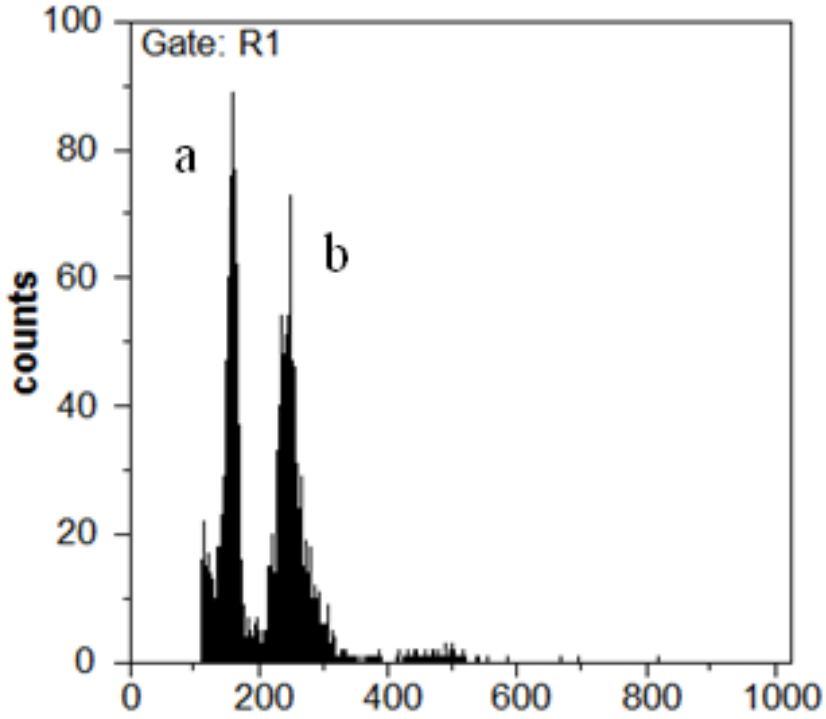

Figure 1. Cytogenetic characterization of *Dipteryx alata*. Relative fluorescence histogram obtained by flow cytometry. a = *D. alata* peak (2C = 1.64 pg) and b = internal standard peak (2C = 2.5 pg).

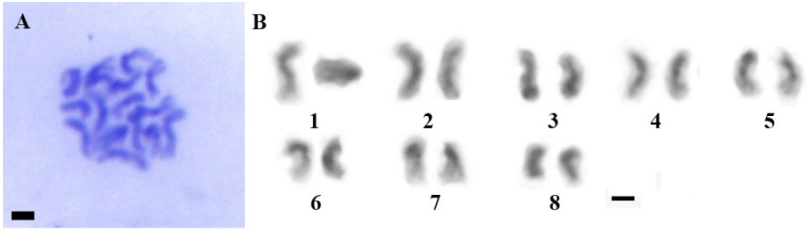

Figure 2. Cytogenetic characterization of *Dipteryx alata*. Metaphase stained with Giemsa 10% (A) and pairing showing 2n = 16 chromosomes (B). Bar = 2 µm.

The chromosome number reported here for *D. alata* (2*n* = 16) confirmed the result obtained by Torres (2001). The only other species of *Dipteryx* with a known chromosome number is *D. odorata* with 2n = 32 (Federov, 1969). The genome size estimate for *D. alata* (0.825 pg) can be considered very small, both considering flowering plants as a whole (Leitch et al., 2005) and other Leguminosae of the Dipterygeae clade (Madrigal, 2018) (Table 1). Because of its small and diploid genome, in addition to its phylogenetic position in Papilionoideae (Cardoso et al., 2013) the species *D. alata* becomes a strong candidate among the tree species of the Brazilian Cerrado to start sequencing projects of its nuclear genome. Currently, studies of discovery and characterization of microsatellite markers and studies of population genetics are found in the literature for *D. alata* (Soares et al., 2012, 2015; Collevatti et al., 2013; Guimarães et al., 2017, 2019b). Genomic studies for *D. alata*

include the assembly and characterization of the chloroplast genome (Antunes et al., 2020b). Despite advances in sequencing technology, smaller genomes have been prioritized due to lower costs for obtaining the total sequences set, which in plant includes sequences from the nuclear and organellar genomes. In addition, small nuclear genomes facilitate assembly of the sequences, a complex and difficult computational job for large genomes (Kelly et al., 2012). *Dipteryx alata* genome size is close to other legume species sequenced as *Medicago sativa* (1C = 0.86 pg - Blondon et al., 1994) and *Cajanus cajan* (1C = 0.90 pg - Bennett and Smith, 1976) and it is the first species of the ADA clade and endemic of the Cerrado that to be sequenced. From the results of the present study, the species *D. alata* was selected for a genomic sequencing project. A draft of the species genome has already been sequenced and annotated (Antunes et al., 2020a) and there is a prospect of generating more sequences to increase the coverage and depth of the *D. alata* genomic assembly.

ACKNOWLEGDMENTS

This work was supported by "Geographic Genetics and Regional Planning for natural resources in the Brazilian Cerrado" (GENPAC) from CNPq/MCT/CAPES/FAPEG (projects no. 564717/2010-0, 563727/2010-1 and 563624/2010-8), CNPq Universal (475182/2009-0) and by by the CERGEN (Núcleo de Excelência em Recursos Genéticos Vegetais do Cerrado -PRONEX/FAPEG/CNPq 07-2012). The authors also are thankful for the PhD and postdoctoral grant from “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (Capes) and productivity grants from CNPq. Our study has also been developed in the context of the National Institutes for Science and Technology in Ecology, Evolution and Biodiversity Conservation (INCT_EECBio), supported by MCTIC/CNPq (process #465610/2014-5) and FAPEG.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Antunes AM, Nunes R, Novaes E, et al. (2020a). Large number of repetitive elements in the draft genome assembly of *Dipteryx alata* (Fabaceae). *Genet. Mol. Res.* 19: GMR18463.

Antunes AM, Soares TN, Targueta CP, et al. (2020b). The chloroplast genome sequence of *Dipteryx alata* Vog. (Fabaceae: Papilionoideae): genomic features and comparative analysis with other legume genomes. *Braz. J. Bot.* 43: 271-282.

Bennett MD and Smith JB (1976). Nuclear DNA amounts in angiosperms. *Philos. T. R. Soc. B.* 274: 227-274.

Blondon F, Marie D, Brown S and Kondorosi A (1994). Genome size and base composition in *Medicago sativa* and *Medicago truncatula* species. *Genome.* 37: 264-270.

Castiglione MR and Cremonini R (2012). A fascinating island: 2n = 4. *Plant Biosyst.* 146: 711-726.

Collevatti RG, Telles MPC, Nabout JC, et al. (2013). Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. *Heredity.* 111: 97-105.

Doležel J and Bartoš J (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* 95: 99-110.

Doležel J and Godhde W (1995). Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high-resolution flow cytometry. *Cytometry.* 19: 103-106.

Dyer RJ, Pellicer J, Savolainen V, Leitch IJ, et al. (2013). Genome size expansion and the relationship between nuclear DNA content and spore size in the *Asplenium monanthes* fern complex (Aspleniaceae). *BMC Plant Biology.* 13: 219.

Federov AA (1969). Chromosome numbers of flowering plants. Academy of Sciences of the USSR. Komarov Botanical Institute, Leningrad.

Garcia S, Leitch JL, Anadon-Rosell A, Canela ML, et al. (2014). Recent updates and developments to plant genome size databases. *Nucleic Acids Res*. 42: D1159-D1166.

Graham M, Nickell C and Rayburn A. (1994). Relationship between genome size and maturity group in soybean. *Theor. Appl. Genet.* 88: 429-432.

Greilhuber J, Borsch T, MAller K, Worberg A, et al. (2006). Smallest angiosperm genomes found in Lentibulariaceae with chromosomes of bacterial size. *Plant Biol.* 8: 770-777.

Guimarães RA, Marques K, Miranda C, et al. (2019a). Assessing genetic diversity and population structure in a *Dipteryx alata* germplasm collection utilizing microsatellite markers. *Crop Breed Appl. Biotechnol.* 19: 329-336.

Guimarães RA, Marques K, Miranda C, et al. (2019b). Mating system and pollen dispersal in *Dipteryx alata* Vogel (Leguminosae): comparing in situ and ex situ conditions. *Tree Genet. Genomes*. 15: 1-10.

Guimarães RA, Telles MPC, Antunes AM, et al. (2017). Discovery and characterization of new microsatellite loci in *Dipteryx alata vogel* (Fabaceae) using next-generation sequencing data. *Genet. Mol. Res*. 16: gmr16029639.

Hidalgo O, Garcia S, Garnatje T, Mumbru M, et al. (2015). Genome size in aquatic and wetland plants: fitting with the large genome constraint hypothesis with a few relevant exceptions. *Plant Syst. Evol.* 31: 1927-1936.

Kelly LJ, Leitch AR, Fay MF, Renny-Byfield S, et al. (2012). Why size really matters when sequencing plant genomes. *Plant Ecol. Divers*. 5(4): 415-425.

Leitch IJ, Soltis DE, Soltis OS and Bennett MD (2005). Evolution of DNA amounts across land plants (embryophyta). *Ann. Bot.* 95: 207-217.

Madrigal JPJ (2018). Next-Generation Sequencing Technologies in Tree Improvement and Conservation Genetics of *Dipteryx oleifera*. Doctoral Thesis. North Carolina State University, United States. 128p. Available in: https://repository.lib.ncsu.edu/bitstream/handle/1840.20/35585/etd.pdf?sequence=1&isAllowed=y

Mendonça RC, Felfili JM, Walter BMT, Silva JMC, et al. (2008). Flora vascular do Bioma Cerrado: checklist com 12.356 especies. In: Sano SM, Almeida SP, Ribeiro JF (eds) Cerrado: ecologia e flora. *Embrapa Cerrados*. 423-1279.

Pati K, Zhang F and Batley J (2019). First report of genome size and ploidy of the underutilized leguminous tuber crop Yam Bean (*Pachyrhizus erosus* and *P. tuberosus*) by flow cytometry. *Plant Genet. Res*. 17(5): 456-459.

Pellicer J, Fay MF and Leitch IJ (2010). The largest eukaryotic genome of them all? *Bot. J. Linn. Soc*. 164: 10-15.

Sano SM, Ribeiro JF and Brito MA (2004). Baru: biologia e uso. Planaltina, DF: Embrapa Cerrados, 52p.

Soares TN, Diniz-Filho JAF, Nabout JC, et al. (2015). Patterns of genetic variability in central and peripheral populations of *Dipteryx alata* (Fabaceae) in the Brazilian Cerrado. *Plant Syst. Evol.* 301: 1315-1324.

Soares TNA, Melo DBO, Resende LVI, et al. (2012). Development of microsatellite markers for the neotropical tree species *Dipteryx alata* (Fabaceae). *Am. J. Bot.* 99(2): e72-e73.

Torres GA (2001). Morfologia e aspectos evolutivos dos cromossomos mitóticos de baru (*Dipteryx alata* Vog.). Doctoral Thesis. Universidade Federal de Lavras, Brasil. 97p. Available at: http://repositorio.ufla.br/jspui/bitstream/1/32102/1/TESE_Morfologia%20e%20aspectos%20evolutivos%20dos%20cromossomos%20mit%c3%b3ticos%20de%20baru%20%28Dipteryx%20alata%20Vog.%29.pdf.

Tyagia A, Sandhyaa, Sharmaa P, Saxenaa S, et al. (2019). The genome size of clusterbean (*Cyamopsis tetragonoloba*) is significantly smaller compared to its wild relatives as estimated by flow cytometry. *Gene*. 707: 205-211.

Veleba A, Smarda P, Zedek FS, Horova L, et al. (2017). Evolution of genome size and genomic GC content in carnivorous holokinetics (Droseraceae). *Ann. Bot.* 119: 409-416.

Wang J, Liu J and Kang M (2015). Quantitative testing of the methodology for genome size estimation in plants using flow cytometry: a case study of the *Primulina* genus. *Front Plant Sci.* 6: 354.

