Perspectives in Plant Ecology, Evolution and Systematics 38 (2019) 13–23

Contents lists available at ScienceDirect

Perspectives in Plant Ecology, Evolution and Systematics

journal homepage: www.elsevier.com/locate/ppees

Do tropical plants have smaller genomes? Correlation between genome size and climatic variables in the Caesalpinia Group (Caesalpinioideae, Leguminosae)

Gustavo Souza[a,*], Lucas Costa[a], Maïté S. Guignard[b,c], Brena Van-Lume[a], Jaume Pellicer[b], Edeline Gagnon[d,e], Ilia J. Leitch[b], Gwilym P. Lewis[b]

[a] Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, Recife 50670-420, Brazil
[b] Comparative Plant and Fungal Biology Department, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK
[c] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK
[d] Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5NZ, UK
[e] Département de Biologie, Université de Moncton, E1A 3E9, Moncton, New Brunswick, Canada

ARTICLE INFO

Keywords:
Climate variables
Genome evolution
Genome size
Succulent Biome

ABSTRACT

While a number of studies have suggested that temperate plants typically have larger genome sizes (GS) than tropical ones, recent analyses have not convincingly confirmed this. We have selected the widely distributed Caesalpinia Group (Leguminosae) to test this hypothesis. We used flow cytometry to estimate the amount of DNA in the haploid genome (C-values) of 40 species of the Caesalpinia Group, aiming to explore the relationship between GS and latitude as well as a range of climatic variables. These comparisons were made using Phylogenetic Comparative Methods (PCM) in a spatio-temporal context. 2C-values varied 7.73-fold, ranging from 0.92 pg (in *Cenostigma bracteosum*) to 7.11 pg (in *Pomaria lactea*). By analyzing the data within a statistical framework that took into account phylogenetic relationships, we observed a positive correlation between GS and latitude as well as additional correlations with various temperature variables. While the genetic mechanism(s) underpinning the increase or decrease of genome size in a latitudinal gradient is unclear, we hypothesize that after the origin of the group (c. 55.9 Mya), fragmentation of the Succulent Biome led to the formation of small populations (island-like) which were subject to genetic drift. This in turn led to drastic and rapid changes in the repetitive DNA fractions of the genome correlated with temperature variables. Thus, we suggest that the environment has played a role in contributing to the diversity of genome sizes reported here for the Caesalpinia Group.

1. Introduction

Over the years, as data for plant C-values (or genome size (GS); i.e. the amount of DNA in the unreplicated gametic nucleus of a cell) have accumulated, it has become clear that GS can be correlated with a wide range of genomic and cellular traits (e.g. chromosome size, heterochromatin amount and transposable element composition), as well as environmental and ecological variables such as latitude, altitude, temperature, and precipitation (e.g. Rayburn et al., 1985; Beaulieu et al., 2007a, 2007b, 2008; Weng et al., 2012; Díez et al., 2013; Kang et al., 2014; Jordan et al., 2015; Du et al., 2017; Lyu et al., 2018; Qiu et al., 2018). Among these, latitude is the major factor investigated to understand how GS varies along latitudinal gradients, as well as related cytological traits such as chromosome size, heterochromatin content and composition vary across latitudinal gradients. Avdulov (1931) was the first to observe that tropical grasses typically have smaller chromosomes compared with grasses from the cooler temperate regions, suggesting a correlation between chromosome size (and hence GS when chromosome number is constant) and latitude. This was followed by the research of Stebbins (1966) who also noted that the tropical genera of Leguminosae and some primarily tropical and subtropical monocot genera (e.g. *Sansevieria* Thunb., *Asparagus* L. and *Smilax* L.) had smaller chromosomes compared with those in most of the temperate genera (e.g. as seen in the strictly temperate tribe Fabeae of Leguminoae, and *Lilium* L. and *Trillium* L. within Liliales). To test this relationship more extensively, Levin and Funderberg (1979) compared 892 angiosperm species from 218 genera and 38 families and showed that the mean and total chromosome lengths per diploid genome (which is correlated with

* Corresponding author.
E-mail address: lgrsouza@hotmail.com (G. Souza).

https://doi.org/10.1016/j.ppees.2019.03.002
Received 12 November 2018; Received in revised form 7 March 2019; Accepted 13 March 2019
Available online 15 March 2019
1433-8319/

GS) were significantly lower in tropical than in temperate species.

More recently, the increase in available GS data due to advances in flow cytometry (Bennett and Leitch, 2011), has led to a significant number of studies in which the relationship between GS and latitude has been investigated. Although some have supported the results of Levin and Funderberg (e.g., Kenton et al., 1986; Freshwater, 1988; Kang et al., 2014), others have reported a negative, or even a lack of, such a correlation between GS and latitudinal gradient depending on the genus (Rayburn et al., 1985; Ceccarelli et al., 1992; Creber et al., 1994; Reeves et al., 1998; Lysák et al., 2000; Grotkopp et al., 2004; Schmuths et al., 2004; Jakob et al., 2004; Mráz et al., 2009; Wang et al., 2011). Such observations therefore raise the question as to whether the reported relationships between GS and latitude are specific to particular plant lineages or arise because many of the individual studies are rather limited in the range of latitudes examined, as noted by Knight and Ackerly (2002). It is also possible that the statistical approaches and assumptions used to explore relationships between GS and latitude (as well as other environmental factors) are inappropriate. For example, many studies assume that the relationship will be linear (as in the studies listed above), whereas it may be unimodal, as discussed by Knight and Ackerley (2002). Indeed, it is still unclear as to what exactly is the impact of factors such as different levels of sampling, evolutionary relationships between the species studied, and the use of different statistical methods in these different studies.

Across the plant kingdom, a ~2,400-fold diversity in GS is observed (Pellicer et al., 2010, 2018; Hidalgo et al., 2017), and even within a single genus GS may range ~40-fold between species of the same ploidy level (Pellicer et al., 2010; Kelly et al., 2015). Nevertheless, analyses exploring the distribution of GS across different taxonomic levels show that there is considerable phylogenetic signal for this character (reviewed in Du et al., 2017; Costa et al., 2017). Therefore, to fully explore the relationship between environmental factors and GS, taxonomic sampling needs to be supported by a robust phylogenetic hypothesis. It is also recognized that analyses among closely related species within a single genus (or monophyletic group) are likely to provide greater interpretive power than analyses which compare more distantly related lineages at higher taxonomic levels (Hawkins et al., 2006; Du et al., 2017). To address such limitations, the study presented here is focused on the Caesalpinia Group (Leguminosae subfamily Caesalpinioideae) which includes species/genera that occupy diverse habitats along a broad environmental gradient that covers a wide latitudinal range (Gagnon et al., 2016).

The Caesalpinia Group has a complex taxonomic history and recently underwent significant nomenclatural changes based mainly on molecular phylogenetic data (Lewis, 1998; Manzanilla and Bruneau, 2012; Gagnon et al., 2013, 2015, 2016). The group has now been reorganized into 27 monophyletic genera comprising a total of 225 species of trees, shrubs and lianas that grow mostly in seasonally dry and semi-arid habitats [especially in the Succulent Biome], with the centre of diversity in the Neotropics (Lewis, 1998; Gagnon et al., 2016, 2019). Some species of the Caesalpinia Group are ornamental and hence are widespread beyond their natural habitat (e.g., *Caesalpinia pulcherrima* L.). Other species are widespread across coastal habitats (e.g., *Guilandina bonduc* L.), over large geographical distances. Biogeographical analysis clearly shows that 14 of the genera in the Caesalpinia Group have a disjunct distribution across several continents (Gagnon et al., 2016, 2019). For example, the genera *Hoffmannseggia* Cav. (Simpson et al., 2005), *Libidibia* (DC.) Schltdl. emend. Gagnon & G.P. Lewis, *Cenostigma* Tul. emend. Gagnon & G.P. Lewis, *Tara* Molina emend. Gagnon & G.P. Lewis, *Caesalpinia* L. emend. Gagnon & G.P.Lewis and *Erythrostemon* Klotzsch emend. Gagnon & G.P.Lewis have an amphitropical distribution across the Neotropics. Furthermore the genera *Pomaria* Cav. (Simpson et al., 2006), *Denisophytum* R.Vig emend. Gagnon & G.P.Lewis and *Haematoxylum* L. (Gagnon et al., 2013, 2016) have non-overlapping sister species occurring in highly localized endemic areas across Central America, the Caribbean, and South America, as well as having relatives in various dry regions of Africa and Madagascar.

The species of the Caesalpinia Group have been well studied cytologically and show that the group is karyotypically characterized by numerical stability. Nearly all species are reported to be diploid with $2n = 24$, comprising small meta/submetacentric and acrocentric chromosomes (~2 µm) (Beltrão and Guerra, 1990; Rodrigues et al., 2012; Van-Lume et al., 2017). Polyploids are very rare, being so far reported in only a few species (Alves and Custódio, 1989; Beltrão and Guerra, 1990; Caponio et al., 2012). Nevertheless, the group is highly variable in terms of the number and positions of cytological markers, such as chromomycin A3 (CMA^+)/ 4',6-Diamidine-2'-phenylindole (DAPI-) bands, on the chromosomes which have been shown to be correlated with geographical distribution and/or ecological niche preference (Van-Lume et al., 2017). The conserved ploidy level, chromosome number and karyotype structure, coupled with the wide geographical distribution (from the tropics to high latitudes) makes the Caesalpinia Group an excellent model for exploring the relationship between GS and environment.

In this study, we evaluated the GS diversity among species belonging to the Caesalpinia Group using flow cytometry. Our sampling covered 13 of the 27 major phylogenetic lineages, especially those distributed in the Neotropics (Gagnon et al., 2016) and included species with a broad latitudinal range that allowed us to test the Stebbins (1966) hypothesis that tropical genera of Leguminosae have smaller chromosomes (and hence smaller GS) compared with temperate genera. Therefore, the aims of this study were (1) to examine the distribution of GS across taxa of the Caesalpinia Group and the relationship with latitude and other environmental factors; (2) To assess the phylogenetic signal of genome size on Caesalpinia and 3) to investigate if the distribution of genome size diversity correlates with latitudinal gradients and other environmental factors.

2. Material and methods

2.1. Plant material

Table 1 lists the species belonging to the genera *Arquita* Gagnon, G.P.Lewis & C.E.Hughes, *Biancaea* Tod. emend. Gagnon & G.P.Lewis, *Caesalpinia* L. emend. Gagnon & G.P.Lewis, *Cenostigma* Tul. emend. Gagnon & G.P.Lewis, *Coulteria* Kunth emend. Gagnon, Sotuyo & G.P.Lewis, *Erythrostemon* Klotzsch emend. Gagnon & G.P.Lewis, *Guilandina* L., *Hoffmannseggia* Cav., *Libidibia* Schltdl. emend. Gagnon & G.P.Lewis, *Mezoneuron* Desf., *Paubrasilia* Gagnon, H.C.Lima & G.P.Lewis, *Pomaria* Cav. and *Tara* Molina emend. Gagnon & G.P.Lewis that were analysed here. The samples were obtained either from plants growing in the living collections held at the Royal Botanic Gardens, Kew (K), seed/seedling tissue from RBG Kew's Millennium Seed Bank (MSB), or from recently field-collected specimens. Herbarium vouchers are deposited in the herbaria listed in Table S1.

2.2. Genome size estimation using flow cytometry

Nuclear DNA contents were estimated using the one-step flow cytometry procedure described by Doležel et al. (2007). Briefly, c. 1 cm^2 of leaf material from the plant being studied (target sample) was placed in a petri dish with the same amount of tissue of the plant calibration standard. For the majority of species *Petroselinum crispum* 'Champion Moss Curled' (2C = 4.50 pg) was used. However, for *Arquita trichocarpa, Zea mays* 'CE-777' (2C = 5.43 pg) was used, while for *Cenostigma bracteosum*, *C. microphyllum*, *C. pluviosum C. pyramidale*, *Libidibia ferrea* and *Paubrasilia echinata, Glycine max* 'Polanka' (2C = 2.50 pg) was used (Doležel et al., 1998; Obermayer et al., 2002). The sample and calibration standard were co-chopped using a new razor blade and mixed together in a petri dish containing 1 mL of 'Ebihara buffer' or 'LB01 buffer' (Doležel et al., 1989; Ebihara et al., 2005). A further 1 mL buffer

Table 1
Samples analyzed of Caesalpinia Group species with provenance, genome size (2C), the mean coefficient of variation (CV%) of the flow cytometry peak, diploid number (2*n*) [when available], and references.

Genus/ species	Provenance of material	2C-value (pg)	CV (target, %)	2*n*	Reference (2C/ 2*n*)
Arquita Gagnon, G.P.Lewis & C.E.Hughes					
A. trichocarpa (Griseb.) Gagnon, G.P.Lewis & C.E.Hughes	Argentina	3.57	4.45	24	This study/ Van-Lume et al. (2017)
Biancaea Tod.					
B. decapetala (Roth) O.Deg.	Zimbabwe	1.93	2.88	24	This study/ Van-Lume et al. (2017)
Caesalpinia L.					
C. pulcherrima Sw.	Guatemala	1.61	3.03	24	This study/ Van-Lume et al. (2017)
–		3.60	–		Ohri et al. 2004
Cenostigma Tul.					
C. bracteosum (Tul.) Gagnon & G.P.Lewis	Brazil	0.92	1.08	24, 48	This study/ Alves and Custódio (1989)
C. eriostachys (Benth.) Gagnon & G.P.Lewis	Mexico	1.76	12.98	–	This study
C. microphyllum (Mart. ex G.Don) Gagnon & G.P.Lewis	Brazil	1.88	4.30	24	This study/ Van-Lume et al. (2017)
C. pluviosum (DC.) Gagnon & G.P.Lewis	Brazil	1.88	1.60	24	This study/ Van-Lume et al. (2017)
C. pyramidale (Tul.) Gagnon & G.P.Lewis	Brazil	1.80	4.07	24	This study/ Van-Lume et al. (2017)
Coulteria Kunth					
C. mollis Kunth	Guatemala	1.72	3.50	24	This study/ Mata-Sucre et al. (unpublished)
C. pumila (Britton & Rose) Sotuyo & G.P.Lewis	Mexico	1.92	3.30	24	This study/ Van-Lume et al. (2017)
Erythrostemon Klotzsch					
E. acapulcensis (Standl.) Gagnon & G.P.Lewis	Mexico	2.27	3.57	–	This study
E. angulatus (Hook. & Arn.) Gagnon & G.P.Lewis	Chile	4.03	2.57	24	This study/ Van-Lume et al. (2017)
E. calycinus (Benth.) L.P.Queiroz	Brazil	1.54	–	24	Rodrigues et al. (2018)/ Mata-Sucre et al. (unpublished)
E. coccineus (G.P.Lewis & J.L.Contr.) Gagnon & G.P.Lewis	Mexico	2.34	5.40	–	This study
E. coluteifolius (Griseb.) Gagnon & G.P.Lewis	Argentina	5.69	2.00	24	This study/ Van-Lume et al. (2017)
E. coulterioides (Griseb.) Gagnon & G.P.Lewis	Argentina	6.18	1.78	24	This study/ Van-Lume et al. (2017)
E. exilifolius (Griseb.) Gagnon & G.P.Lewis	Argentina	5.62	2.5	24	This study/ Van-Lume et al. (2017)
E. exostemma (Sessé & Moc. ex DC.) Gagnon & G.P.Lewis	Mexico	2.29	3.51	–	This study
E. gilliesii (Hook.) Klotzsch	Argentina	4.94	2.13	24	This study/ Van-Lume et al. (2017)
E. hintonii (Sandwith) Gagnon & G.P.Lewis	Mexico	2.29	3.48	–	This study
E. hughesii (G.P.Lewis) Gagnon & G.P.Lewis	Mexico	2.18	4.31	24	This study/ Van-Lume et al. (2017)
E. melanadenius (Rose) Gagnon & G.P.Lewis	Mexico	2.24	2.99	–	This study
E. mexicanus (A.Gray) Gagnon & G.P.Lewis	Mexico	4.01	–	24	Ohri et al. 2004/ Fedorov (1974)
E. nelsonii (Britton & Rose) Gagnon & G.P.Lewis	Mexico	2.27	4.29	–	This study
E. pannosus (Brandegee) Gagnon & G.P.Lewis	USA	2.47	3.51	24	This study/ Mata-Sucre et al. (unpublished)
E. placidus (Brandegee) Gagnon & G.P.Lewis	Mexico	2.59	2.87	24	This study/ Mata et al. (unpublished)
E. sousanus J.L.Contr., Sotuyo & G.P.Lewis	Mexico	3.41	7.06	–	This study
E. yucatanensis (Greenm.) Gagnon & G.P.Lewis	Mexico	2.15	4.02	–	This study
Guilandina L.					
G. bonduc L.	Brazil	1.34	4.78	24	This study/ Van-Lume et al. (2017)
Hoffmannseggia Cav.					
H. doelli Phil.	Chile	2.40	3.47	–	This study
Libidibia Schltdl.					
L. coriaria Schltdl.	USA	1.70	–	24	Ohri, 1998/ Kumari and Bir (1989)
L. ferrea (Mart. ex Tul.) L.P.Queiroz	Brazil	1.83	4.19	24, 48, 72	This study/ Van-Lume et al. (2017)
L. paraguariensis (D.Parodi) G.P.Lewis	Argentina	1.81	5.13	24	This study/ Cangiano and Bernardello (2005)
L. punctata (Willd.) Britton	Venezuela	1.70	1.76	24	This study/Mata et al. (unpublished)
Mezoneuron Desf.					
M. hildebrandtii Vatke	Madagascar	2.55	3.10	24	This study/ Mata et al. (unpublished)
Paubrasilia Gagnon, H.C.Lima & G.P.Lewis					
P. echinata (Lam.) Gagnon, H.C.Lima & G.P.Lewis	Brazil	2.89	4.28	24	Rodrigues et al. (2018)/ Van-Lume et al. (2017)
Pomaria Cav.					
P. lactea (Schinz) B.B.Simpson & G.P.Lewis	South Africa	7.11	2.87	24	This study/ Mata-Sucre et al. (unpublished)
Tara Molina					
T. cacalaco (Bonpl.) Molinari & Sanchez Och.	Mexico	2.98	3.28	24	This study/ Sarkar et al. (1982)
T. spinosa (Molina) Britton & Rose	Peru	2.65	3.06	24	This study/ Bandel (1974)
T. vesicaria (L.) Molinari, Sanchez Och. & Mayta	Mexico	2.78	3.01	24	This study

was then added and the resulting suspension was filtered through a 30 µm nylon mesh and the nuclei stained with 100 µL propidium iodide (1 mg/mL). Samples were incubated with RNAse at 37 °C for 20 min and the relative fluorescence of 5000 particles was then recorded using a Partec Cyflow SL3 flow cytometer (Partec GmbH, Münster, Germany) fitted with a 100 mW green solid-state laser (532 nm, Cobolt Samba, Solna, Sweden). Three leaves from different individuals were measured separately for each species and three replicates of each leaf were processed (although in some cases only one or two individuals were available). The output histograms were analysed with FlowMax

software v. 2.4 (Partec GmbH).

2.3. Phylogenetic relationships in the Caesalpinia Group

A phylogenetic hypothesis of the evolutionary relationships in the Caesalpinia Group was built based on plastid [maturase K gene (*mat*K), ribosomal protein S16 (*rps*16), intergenic spacer *trn*D-T, intergenic spacer *trn*L, and intergenic spacer *ycf*6] and nuclear ribosomal ITS [internal transcribed spacer ITS1 – 5.8S – ITS2] DNA sequences. All plastid and nuclear data were downloaded from Genbank (Table S1). Each set of sequences was aligned separately using Geneious 7.1.9 (Kearse et al., 2012) and subsequently concatenated into a single matrix. A Bayesian phylogenetic inference (BI) was conducted on this combined matrix. To estimate the appropriate evolutionary model for all sequences, the Akaike Information Criterion (AIC) was used as implemented in jModelTest 0.1.1 (Posada, 2008). The best estimated model for our dataset was GTR. The Bayesian analysis was subsequently performed using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003). The analysis was conducted for two independent runs and for 10×10^6 generations, sampling every 1000 trees. To verify the effective sampling of all parameters and assess convergence of independent chains, we examined their posterior distributions in Tracer 1.6. (Rambaut et al., 2014). The first 25% of the sampled trees were discarded as the burn-in period. The subsequent trees were retained and posterior probabilities (PP) were estimated by constructing a 50% majority-rule consensus tree. The trees were edited in FigTree 1.3.1 (Rambaut, 2009).

2.4. Divergence time estimates

To explore the evolution of GS in the Caesalpinia Group a molecular clock analysis was performed. Time-calibrated phylogenies were constructed using BEAST 1.8.0 (Drummond et al., 2012). The dataset was analysed using two partitions, the first consisted of the five plastid markers and the second contained ITS. We used the AIC values in jModelTest 2 (Darriba et al., 2012) to select the best evolutionary model, which identified the General Time Reversible + Gamma + Invariant sites (GTR + G + I) model for both partitions. Analyses were run using an uncorrelated log normal relaxed clock and a Yule Process speciation model. Two independent runs of 20×10^6 generations each were performed, sampling every 1000 generations. To verify the effective sampling of all parameters and assess convergence of independent chains, we examined their posterior distributions in Tracer 1.6 and the MCMC sampling was considered sufficient at effective sampling sizes (ESS) higher than 200. After removing 25% of samples as burn-in, the independent runs were combined and a maximum subgenus-credibility (MCC) tree was constructed using TreeAnnotator 1.8.2. (Drummond et al., 2012).

The phylogeny was calibrated using three calibration priors: two fossils previously used in time-calibrated phylogenetic analyses of Leguminosae and one secondary calibration prior derived from previous phylogenetic analyses. The first fossil calibration point is based on fossil fruits attributed to *Senna* (included here as the outgroup), found in South East USA and Mexico, estimated at 45 Mya (Herendeen, 1992; Calvillo-Canadell and Cevallos-Ferriz, 2005). The second fossil calibration point is based on fossil winged fruits attributed to *Mezoneuron*, found at several sites across the South East and Western USA (the Middle Eocene Claiborne Formations in western Tennessee, the Eocene Green River Formation in Idaho, and the Miocene Clarkia, Whitebird, and Oviatt Creek sediments in Idaho), with the oldest of these fossils estimated to be 45 Mya (Herendeen and Dilcher, 1991). These two fossil calibrations were placed on the stem node of the genus *Senna* (included here as outgroup but not shown in the phylogenetic tree of Fig. 5) and *Mezoneuron* as a minimum age of diversification.

In addition to these two fossil calibrations, we also constrained the root height of the tree, corresponding to the crown node of the MCC tree (oldest node of the Caesalpinia Group). To do this, we took into account the results by Bruneau et al. (2008) and Simon et al. (2009). The former estimated the age of the MCC tree to be 58.6 Ma (58.8 − 58.5, 95% confidence interval of penalised likelihood analysis repeated on 100 trees from the posterior distribution of a Bayesian analysis). The latter, based on a *mat*K phylogeny of the legume family, estimated the crown to be 60.8 Ma (height range: 66.4 − 56.5, BEAST analysis). Based on these Bayesian inferences (BI), we specified a normal prior for the root height, with a minimum age corresponding to the lowest value of the height range of Simon et al. (2009) and a maximum age corresponding to the age of the MCC clade estimated by Simon et al. (2009).

2.5. Geographical distribution and ecological traits of the Caesalpinia Group

To investigate biogeographical patterns within the Caesalpinia Group, species occurrence data were downloaded from the Global Biodiversity Information Facility (GBIF) website (https://www.gbif.org) and a distribution map was plotted using the software QGIS v. 2.18 (QGIS Development Team, 2014). Two criteria were used to minimize the effect of erroneous GBIF distribution data: i) we carried out a taxonomic survey to filter only valid names based on the most recent Caesalpinia Group classification (Gagnon et al., 2016); ii) only species with vouchers deposited in herbaria were recorded. The data were cleaned to exclude e.g. oceanic points, and locations that were unlikely to be natural occurrence (e.g., occurrences in botanical gardens at more northern latitudes). From the collection sites of each species [see Fig. 2], we extracted climatic variables from the WorldClim 1.4 (5 min) generic grid format (Hijmans et al., 2005), utilizing the package "raster" 2.6–7 (Hijmans, 2017) implemented in the R software 3.3.3 (R Core Team, 2017). We calculated species means for absolute latitude, altitude, and the 19 climatic variables. Correlations between these environmental variables were assessed with Pearson's correlation coefficient (Table S2) and principal components analysis (Fig. S1), to avoid the selection of highly correlated variables ($r > 0.75$). We also viewed scatter plots between GS and each climatic variables (Fig. S2), checked model assumptions with simple linear regressions, ran diagnostics to test that variables met evolutionary assumptions (see below), and linearity between variables (Quader et al., 2004). Temperature variables were highly correlated with each other and/or with latitude (e.g. latitude and temperature annual range: $r = 0.86$). Therefore we selected just one temperature variable: minimum temperature of coldest month (BIO_6) which is highly correlated with annual mean temperature ($r = 0.96$) but may be a more concise predictor. Maximum temperature of the warmest month showed no impact on species distributions in terms of their GS with phylogenetic regressions (Fig. S3). Annual precipitation and precipitation seasonality did not meet evolutionary assumptions, thus we selected precipitation of the wettest month (BIO_{13}, $r = 0.93$ with annual precipitation). Precipitation of the driest month was not correlated with species GS, hence we included only one precipitation variable (Fig. S2, S3).

2.6. Statistical analyses

New GS estimates for 36 species together with previously available published data (Table 1) were compiled with environmental variables and species locations. The original BIO_6 and BIO_{13} variables were divided by ten, to obtain minimum temperature in C°, and precipitation of wettest month in cm. To account for non-independence in evolution (i.e. traits may be more similar between closely related species (Garland et al., 1992)), and non-independence in geographical space (traits may be more similar in species which are nearer to each other (Legendre, 1993)), we applied methods to account for 1) phylogenetic correlation, and 2) spatial auto-correlation. For each method we fitted: species means ($n = 38$) with 1) univariate regressions; 2) a multiple regression; and 3) the complete data ($n = 4368$) with a polynomial model (details

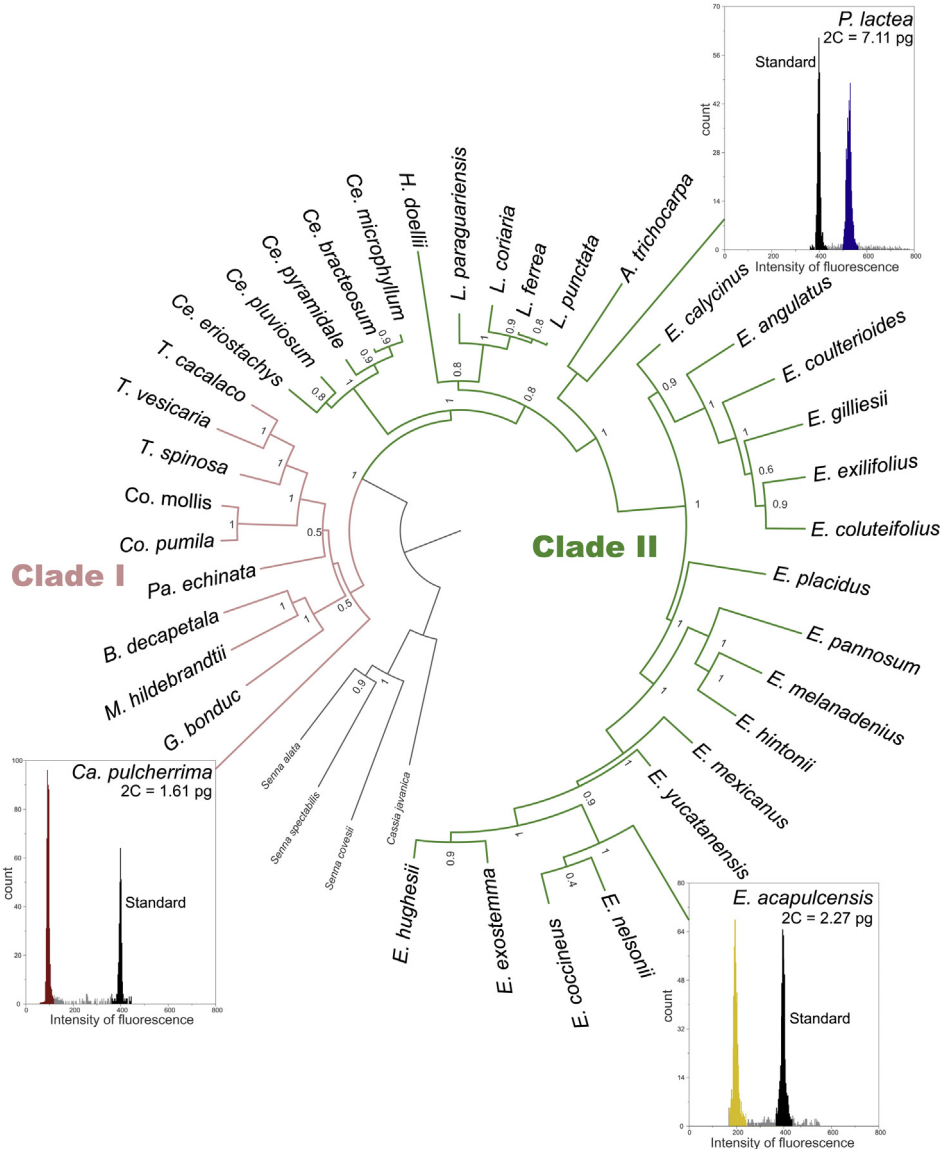

Fig. 1. Bayesian inference (BI) majority-rule consensus topology of combined nuclear internal transcribed spacer (ITS) and plastid (*mat*K, *rps*16, *trn*D-T, *trn*L and *ycf*6) data for the 39 species in the Caesalpinia Group with genome size data estimated in this study.

below). Genome size (2C-value in picograms) and precipitation of wettest month were log-transformed to meet model assumptions of linearity and evolution. Analyses were carried out in R 3.3.3 (R Core Team, 2017).

2.7. Phylogenetic comparative methods (PCM)

As most of the methods employed in this study do not work with phylogenetic trees containing polytomies, we used the algorithm [multi2di] in 'phytools' 0.6–44 (Revell, 2012) to resolve polytomies. Phylogenetic signal in log GS, a measure of how related species resemble each other more than species drawn at random from the same tree, was assessed using Pagel's lambda (Pagel, 1999) with phytools 0.6–44 (Münkemüller et al., 2012; Revell, 2012).

We first fitted phylogenetic independent contrasts (PICs) with caper 1.0.1 (Orme et al., 2018) between GS and mean species values for the environmental variables. Diagnostics revealed the node for *Cenostigma microphyllum* and *C. bracteosum* to be a substantial outlier. We removed *C. microphyllum*, which resolved the issue of an influential outlier. Subsequently, this taxon was removed from the data for all analyses. We ran diagnostics to test whether GS, absolute latitude and environmental variables met the assumptions of a Brownian motion model using the caic.diagnostics function; these include assumptions that the amount of variation is proportional to branch length, and that the magnitude of the trait values does not influence nodal values (Cooper et al., 2016; Orme et al., 2018) (Table S2a). After the selection of environmental variables (i.e. BIO_6 - minimum temperature of coldest month, and BIO_{13} - precipitation of wettest month), univariate phylogenetic GLS were fitted with the function pgls. The lambda parameter was optimised to find the maximum likelihood value. A multiple PGLS was then fitted with GS regressed on absolute latitude and the two environmental variables. Model assumptions were checked with standard diagnostic plots, and possible multicollinearity assessed with variance inflation factors (VIF) with the package "car" 3.0 (Fox et al., 2012).

To further explore the relationships between GS, latitude, and minimum temperature, we fitted a phylogenetic generalized mixed model. Instead of means, all locations for the 38 species were fitted (n = 4368) within a Bayesian framework with MCMCglmm (Hadfield, 2010). We tested whether GS was significant in predicting a taxon's value for minimum temperature, when latitude is accounted for. We used actual latitude (rather than absolute value of latitude, as above),

which was fitted with a second-order polynomial term. Species (to account for multiple measurements within a taxon) and phylogeny were treated as random effects. Models were run with 4×10^6 iterations, a burn in of 50000, and sampling of the chain at intervals of 500, with an inverse Wishart prior (V = 1, nu = 0.002). An extended prior was also tested (V = 1, nu = 1, alpha.mu = 0, alpha. V = 1000), which did not change regression results. Model diagnostics included checking autocorrelation within the chains of fixed and random effects, viewing trace plots, checking chain convergence with Heidelberger and Welch's diagnostic, and convergence of separate chains (n = 3) with Gelman and Rubin's convergence diagnostic; the latter two were implemented in package coda 0.19–1 (Plummer et al., 2010).

2.8. Spatial autocorrelation methods

The degree of spatial autocorrelation was estimated with Morans's I, as a function of Euclidean distance, using ape 5.0 (Paradis et al., 2004). We first fitted generalized least squares (GLS) to investigate the individual effects of mean absolute latitude and mean values of each of the two climatic variables on species GS using nlme 3.1 (Pinheiro et al., 2012). Four correlation structures were tested: exponential, spherical, linear, and Gaussian spatial correlation structures; a fifth model was fitted that did not account for spatial auto-correlation. Fitted models were compared with second-order AIC (AICc), which is more suitable for small sample sizes (MuMIn 1.40.4 (Bartón, 2009)). The model with the lowest AICc was then compared to a model with no correlation structure using a Chi-square test, and the more complex model was retained if the reduction in residual sum of squares was significantly improved (alpha = 0.05). Finally, we fitted a polynomial regression using all data points (n = 4368), to test whether the link between GS and minimum temperature remained significant, when latitude was included as an effect on temperature. Residual plots, semivariograms, and VIFs were used to assess model fits.

3. Results

3.1. Phylogenetic reconstruction of the Caesalpinia Group

As noted above, phylogenetic relationships within the Caesalpinia Group taxa were based on plastid (*mat*K, *rps*16, *trn*D-T, *trn*L and *ycf*6) and nuclear ribosomal DNA (ITS1-5.8S-ITS2) sequences available from previous studies (Table S1). The topology generated here using Bayesian Inference (BI) (Fig. 1) comprises a sub-sample of species from the most comprehensive phylogenetic study currently available which contains 172 species (76.4% of the extant diversity), including representatives of all genera recognized in the Caesalpinia Group (Gagnon et al., 2016). The phylogenetic tree presented in Fig. 1 only contains the 39 species analysed in our study for which GS data are available (N.B. it was not possible to include *Erythrostemon sousanus* J.L.Contr., Sotuyo & G.P.Lewis in the phylogenetic analysis although its GS was determined for this study). Nevertheless, the relationships recovered support the division of the Caesalpinia Group into two major clades in agreement with the more comprehensive phylogenetic study by Gagnon et al. (2016). Clade I comprises species belonging to *Biancaea* (represented here by 1 sp.), *Caesalpinia* (1 sp.), *Coulteria* (2 spp.), *Guilandina* (1 sp.), *Mezoneuron* (1 sp.), *Paubrasilia* (1 sp.), and *Tara* (3 spp.); while clade II includes representatives of *Arquita* (1 sp.), *Cenostigma* (5 spp.), *Erythrostemon* (18 spp.), *Hoffmannseggia* (1 sp.), *Libidibia* (4 spp.), and *Pomaria* (1 sp.) (Fig. 1). Overall, we obtained strong phylogenetic support for most of the major clades resolved in Gagnon et al. (2016).

3.2. Nuclear DNA content in the Caesalpinia Group

Table 1 lists the 2C-values for 40 species of the Caesalpinia Group, of which 36 are reported for the first time. Flow cytometric profiles revealed high resolution histograms (Fig. 1) with the CVs of the 2C peaks for the target taxa and the reference below 5% for most (32 out of 36 taxa) species. However, for *Cenostigma eriostachys*, *Erythrostemon coccineus*, *E. sousanus*, and *Libidibia paraguariensis* only seed material (non-viable seeds with a long storage capacity) was available for GS analyses and these gave lower quality histograms (CV% 5.13–12.98).

Overall, 2C-values in the Caesalpinia Group ranged 7.73-fold, from 0.92 pg/2C in *Cenostigma bracteosum* to 7.11 pg/2C in *Pomaria lactea* (Schinz) B.B.Simpson & G.P.Lewis (Table 1, Fig. 2). In general, species belonging to Clade II had larger genomes (2C-values range 1.70–7.11 pg), with a mean 2C-value of 3.26 pg compared with Clade I where the 2C-values ranged from 0.92 to 2.98 pg, with a mean 2C-value of 2.03 pg. Although the mean values in Clades I and II were different, our ANOVA analysis did not find this difference to be statistically significant ($p = 0.07$).

3.3. Phylogenetic regressions

Pagel's λ was found to be 1.001 ($p < 0.0001$), indicating a strong significant phylogenetic signal for GS in the Caesalpinia Group. Phylogenetic signal was also strong for absolute latitude (0.882, $p = 0.0055$, Table S3a), minimum temperature (0.942, $p < 0.0001$), and

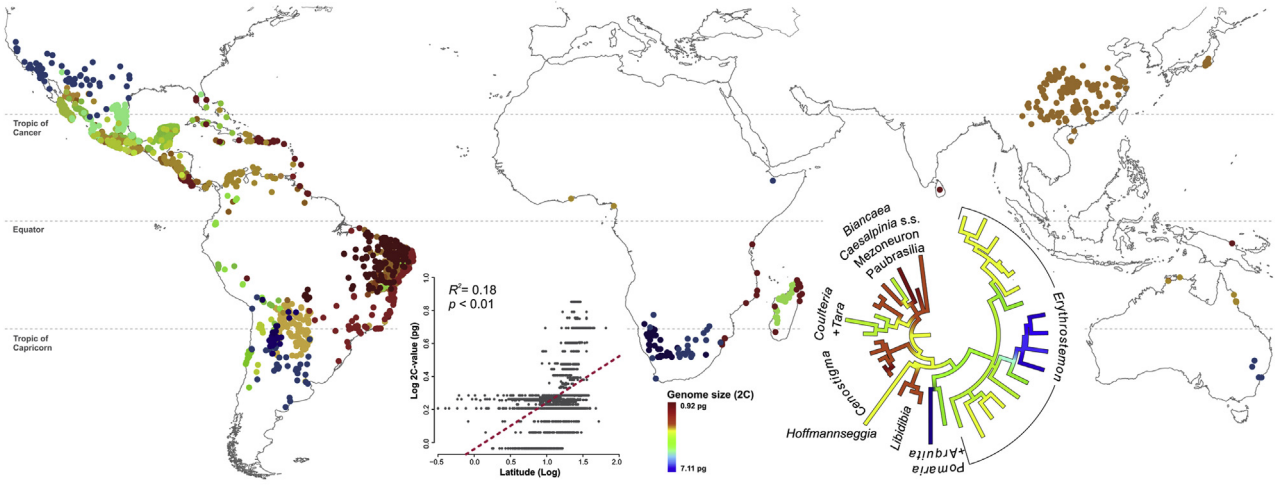

Fig. 2. Geographical distribution of genome size (GS) variation for species in the Caesalpinia Group included in this study with reconstruction of ancestral 2C-values (pg) across the phylogeny. Plot of latitude versus genome size graph showing the correlation between these variables.

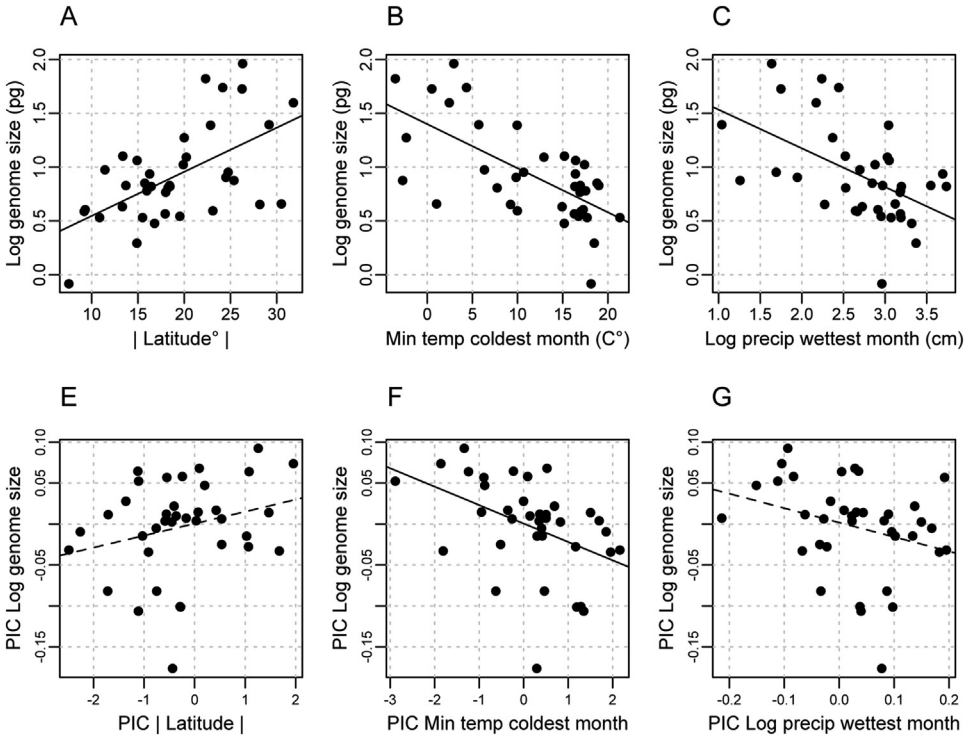

Fig. 3. Relationships between genome size (2C-values) and environmental traits using simple linear regression analyses showing correlations between log-transformed GS and (A) absolute latitude ($F = 17.52$, $R^2 = 0.309$, $p = 0.0002$); (B) minimum temperature of the coldest month ($F = 28.04$, $R^2 = 0.422$, $p < 0.0001$); and (C) log-transformed precipitation of the wettest month ($F = 13.55$, $R^2 = 0.253$, $p = 0.0008$). Phylogenetic independent contrasts (PIC) analyses between log GS and (E) absolute latitude ($F = 2.80$, $r^2 = 0.046$, $p = 0.1032$); (F) minimum temperature ($F = 8.81$, $r^2 = 0.174$, $p = 0.0053$); and (G) log-transformed precipitation of the wettest month ($F = 3.592$, $r^2 = 0.065$, $p = 0.0661$).

maximum precipitation (0.958, $p = 0.0002$). A univariate PGLS showed that GS increased significantly with lower minimum temperature of the coldest month (BIO_6), or put another way, GS decreased by 2.21% for each 1 °C increase in the minimum temperature of the coldest month y ($p = 0.0053$). For latitude and precipitation, GS increased by 1.59% for each degree increase in absolute latitude, and decreased by 1.61% for each 10% increase in precipitation; these were marginally significant ($p = 0.0758$ and 0.0661 respectively) (Fig. 3, Table S3b).

A multiple PGLS model showed that GS in the Caesalpinia Group is most associated with minimum temperature, while absolute latitude and maximum precipitation do not have a significant association with GS. With latitude and precipitation accounted for, GS increased by 1.86% with each 1 °C decrease in minimum temperature of the coldest month ($p = 0.0568$) (Table S3c).

The MCMCglmm showed, as expected, lower minimum temperatures at latitudes further from the equator. It also supports the significant contribution of GS in influencing species distributions relative to the minimum temperature of the coldest month (Fig. 3). A ten percent increase in GS is associated with a 0.6 °C decrease in minimum temperature ($p = 0.0111$) (Table S3d).

3.4. Spatial autocorrelation regressions

Moran's I was used to measure whether species closer to each other were more similar than would be expected if the species distributions were random. It ranges between [-1] (dispersed), 0 (fully random), and 1 (clustered). Species GS was weakly clustered, Moran's $I = 0.141$ ($p = 0.0251$), and minimum temperature of the coldest month was the most spatially correlated of the environmental variables, with $I = 0.366$ ($p < 0.0001$) (Table S4a).

In univariate models accounting for spatial autocorrelation, absolute latitude, minimum temperature of the coldest month, and maximum precipitation of the wettest month were all significantly correlated with GS. The best-fit spatial correlation was a Gaussian correlation structure for all three regressions. These show (i) GS increasing by 4.23% per latitudinal degree increase ($p = 0.0009$); (ii) GS decreasing by 3.82% with a 1 °C increase in minimum temperature ($p = 0.0001$), and (iii) GS decreasing by 2.31% with a 10% increase in maximum precipitation ($p = 0.0099$) (Table S4b).

A multiple regression GLS fitting GS on absolute latitude, minimum temperature, and maximum precipitation identifies minimum temperature as the most significant effect. Accounting for spatial autocorrelation, GS increases by 2.16% as minimum temperatures decrease by 1 °C ($p = 0.0985$) (Fig. 4a, Table S4c). A polynomial regression including all 4374 specimen locations, fitting minimum temperature onto latitude and GS, confirms the significant link between GS and temperature. An exponential spatial correlation structure rendered the best-fit model (χ^2 test < 0.0001). In this regression, a 10% increase in GS is associated with a 0.39 °C decrease in minimum temperature ($p < 0.0001$) (Fig. 4, Table S3d).

4. Discussion

4.1. Distribution of GS within the Caesalpinia Group phylogeny

Here we present the first comprehensive study exploring the diversity and evolution of GS and its correlation with climate variables within the Caesalpinia Group *sensu* Gagnon et al. (2016). Our results considerably extend the variation of DNA content reported for the group from 1.83-fold in a previous analysis of 10 species belonging to *Cenostigma*, *Erythrostemon*, *Libidibia* and *Paubrasilia* (Rodrigues et al., 2018) to 7.73-fold reported here. The remarkably high diversity of GS observed here is not associated with polyploid (whole genome duplication) events, which are reported to be rare in the Caesalpinia Group (Van-Lume et al., 2017). In the absence of polyploidy, previous genome-wide analyses of other genera showing large variations in GS have commonly been attributed to the differential amplification of repetitive elements, especially Ty3/gypsy LTR-retrotransposons (e.g. Kelly et al., 2015; Macas et al., 2015; Vu et al., 2015; Mascagni et al., 2017). Indeed, a preliminary genomic analysis for some species of the Caesalpinia Group has revealed a high proportion of Ty3/gypsy elements, with the Chromovirus lineage being the most abundant repeat and accounting for up to 38% of the genome in some species (Souza, personal communication). Alternative sources of GS variation in the

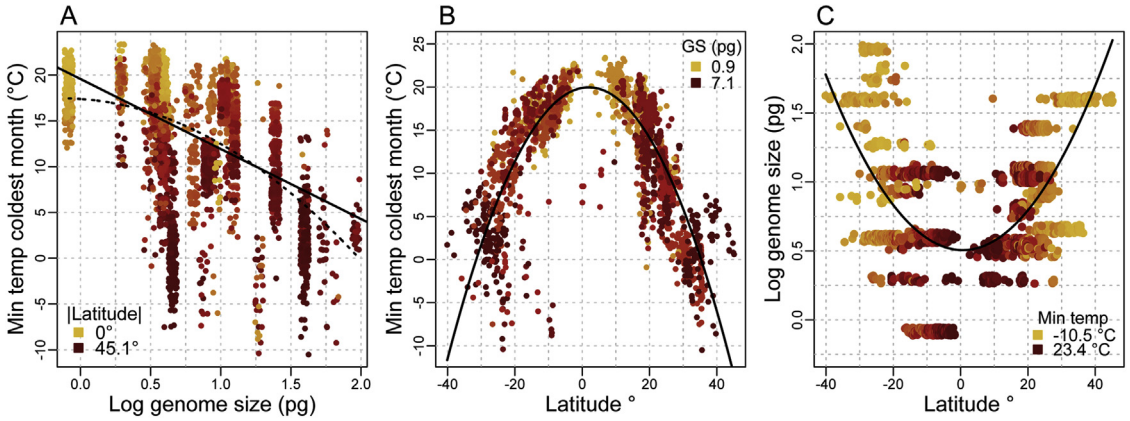

Fig. 4. Relationships between log-transformed genome size, latitude, and minimum temperature of wettest month. Minimum temperature was regressed on latitude and genome size, accounting for phylogenetic non-independence (Table S3d) and spatial auto-correlation (Table S4d), which showed significant relationships between minimum temperature of the coldest month (BIO_6) and GS (A), and the polynomial correlation between latitude and minimum temperature (B). Mapped onto each plot by colour (darker = higher values) is the third variable. Also shown in (C) is the relationship between GS and latitude with increasing GS associated with increasing distances from the equator (simple polynomial regression: $F(2, 4372) = 894.8$, $R^2 = 0.29$, $p < 0.0001$), it also shows the decreasing minimum temperatures of the coldest month at higher latitudes.

Caesalpinia Group include differences in the amount of satellite DNA sequences present in the highly variable heterochromatin CMA^+ bands, which have been observed to vary in number, size and distribution across the genome (Van-Lume et al., 2017). Indeed, by comparing the GS data obtained in this work with previously published reports comparing the distribution of CMA/DAPI bands between species (Van-Lume et al., 2017), there is a clear trend of increasing GS being associated with increasing amounts (%) of GC-rich heterochromatin.

4.2. Potential evolutionary and ecological drivers of GS variation

The significant associations between GS and climate variables observed in the Caesalpinia Group (Fig. 3) are consistent with broad patterns that have previously been reported across other plant groups (Beaulieu et al., 2008; Veselý et al., 2012; Díez et al., 2013; Kang et al., 2014; Jordan et al., 2015). Our findings on latitude contrast with some but not all previous studies that explored correlations between latitude and GS. For example, previous work on *Pinus* (Grotkopp et al., 2004), *Zea* (Rayburn et al., 1985), *Festuca* (Ceccarelli et al., 1992), and *Arabidopsis* (Schmuths et al., 2004) showed a negative correlation between genome size and latitude, in striking contrast to the hypotheses of Avdulov (1931) and Stebbins (1966). On the other hand, other studies corroborate the Avdulov hypothesis (e.g., Freshwater, 1988; Kenton et al., 1986; Kang et al., 2014). Nevertheless, as noted by Knight and Ackerley (2002), most of these studies were based on a geographically restricted sampling (many of them comparing just a few localities in Europe) and did not take phylogenetic relationships into account. Indeed, it is notable that those analyses which did include a wide geographical sample and did use phylogenetically-informed analytical methods showed a significant positive correlation between GS and latitudinal gradient (e.g., Kang et al., 2014; Du et al., 2017), and such findings are confirmed here. Indeed, the most impactful ecological predictors in the Caesalpinia Group are directly related to latitudinal variation (i.e. annual temperature and precipitation).

In tropical latitudes, both total annual rainfall and rainy season length have large spatial variations, but generally decrease from equatorial to subtropical and temperate regions (Allen et al., 2017). The most characteristic feature of the Succulent Biome, where the majority of the Caesalpinia Group is distributed, is an extended dry season with the majority of the precipitation (˜80%) occurring during a short wet season (Maas and Burgos, 2011). Despite rainfall seasonality and variation, there is no significant association between GS and precipitation in our analysis. This ecologically restrictive environment may be related to the impact of seasonality variables (temperature) on GS variation. Several studies have demonstrated that genomic signatures of adaptation can be strongly linked to environmental traits, even with respect to subtle environmental differences, with lineages continuing to change and adapt to the shared features of their environments (e.g., Deatherage et al., 2017; Bilinski et al., 2018; Qiu et al., 2018). In the Caesalpinia Group studied here, the contrasting impact of high temperature stress on species occupying the Succulent Biome (including the Seasonally Dry Tropical Forest [SDTF] in South America) compared with species in the Temperate zones less pronounced water or temperature stress, may contribute to explaining the contrasting trends in genome size evolution between these regions. Our results represent examples of how the environment may play a role in influencing nuclear DNA amounts (Knight and Ackerley, 2002). However, the exact mechanisms by which contrasting environmental conditions impact and influence the genome sizes of species occupying different climatic regions are still poorly understood.

In addition to the correlation with climate variables, GS variation can also be driven by plant life cycle/morphological traits. Although based on a relatively small sampling, our data also suggests a relationship between GS and plant habit as postulated by Stebbins (1971). Woody angiosperms, including long-lived trees, have been observed to be characterised by possessing small GSs in part because of the supposed restriction on the maximum nuclear size and hence cell size needed to ensure the efficient movement of water and nutrients through the longer xylem pathways (Stebbins, 1971; Ohri, 1998; Cavalier-Smith, 2005). In our study, the smallest nuclear DNA contents were found in tree species such as *Cenostigma bracteosum* (2C = 0.92 pg) and *Paubrasilia echinata* (2C = 1.15 pg), while the largest genomes were found in shrubby species, e.g. *Pomaria lactea* (2C = 7.11 pg). Thus, the differences in DNA amounts observed in taxa of the Caesalpinia Group may be driven in part by the impact of genome size on various adaptive traits operating at the nuclear, cellular and organismal levels (Cavalier-Smith, 2005).

4.3. GS evolution and the "Evolutionary islands" hypothesis

While adaptive processes may indeed play some role in contributing to the diversity and geographical distribution of GSs encountered in the Caesalpinia group, as noted above, the observation of high phylogenetic signal in the GS data suggests there may also be a role for genetic drift and hence neutral evolutionary processes. Kamilar and Cooper (2013) It is recognized that strong phylogenetic signal does not necessarily

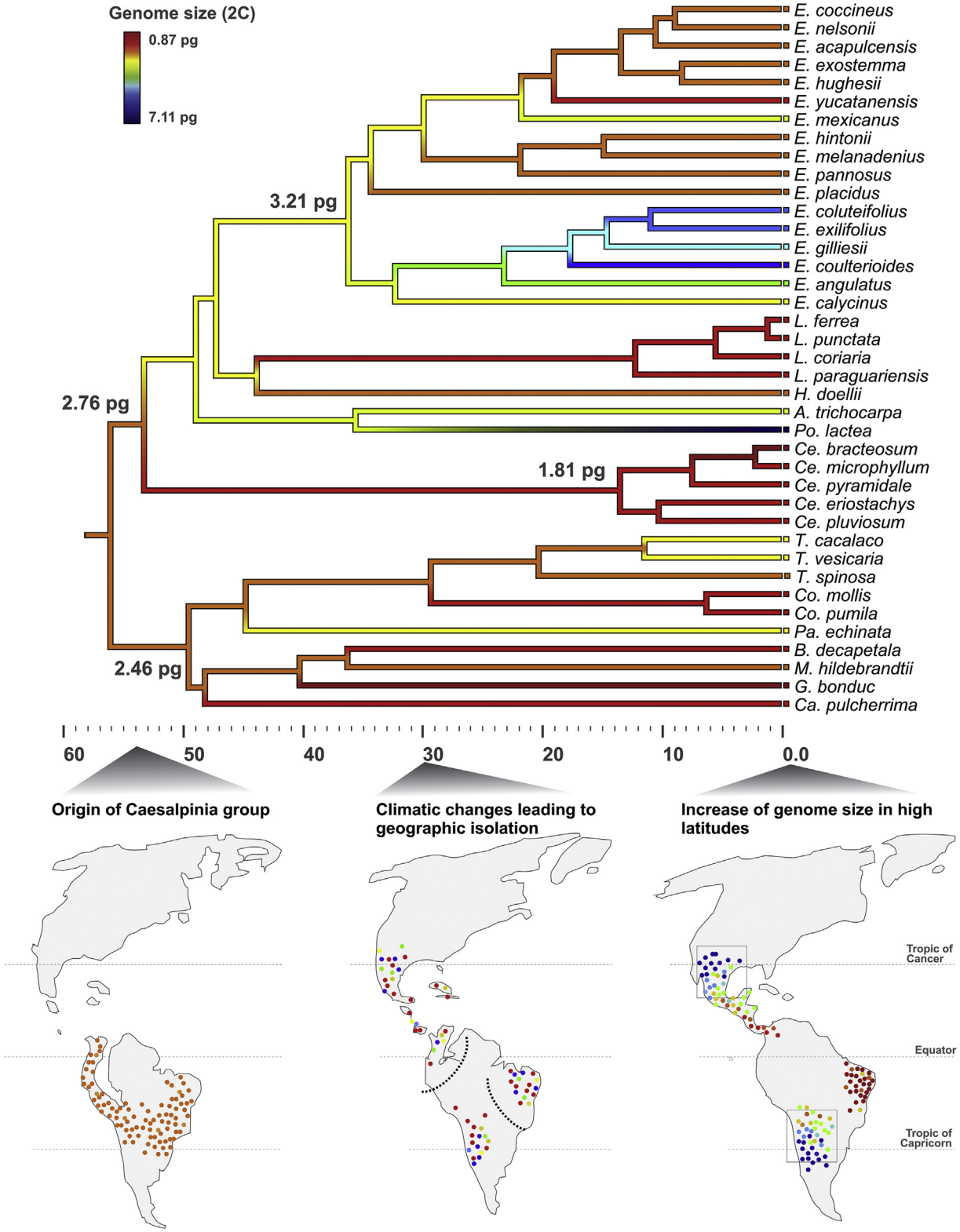

Fig. 5. Dated Bayesian phylogeny of the Caesalpinia Group radiation and palaeo-tectonic evolution of the American continents. A time scale is shown spanning the full evolutionary history of the group. Upper part of the figure, ancestral character state reconstruction of 2C-values (pg) along branches and nodes: hot colors (i.e. red/yellow) correspond to relatively low 2C-values, whereas cold colors (blue) represent high values. Lower part of the figure, maps of the hypothetical distribution of Caesalpinia Group lineages at 55 Mya (group origin), 30 Mya (Andean uplift) and present, from left to right respectively (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

confirm that neutral evolutionary processes are operating (e.g. see Hansen et al. (2008); Revell, 2012). The proposal by Särkinen et al. (2012) that the SDTF (the South American component of the succulent biome where many Neotropical species of the Caesalpinia group occur), formed stable and strongly isolated "evolutionary islands" as a consequence of the Andean uplift, likely led to small population sizes in which drift in genome size may well have occurred.

Combining these data together with our divergence time analysis and ancestral genome size reconstructions, we hypothesize that following the origin of the Caesalpinia group c. 55 Mya with a predicted ancestral genome size of 2.46–2.76 pg/2C, climatic and, more recently geological processes resulted in (i) the formation of isolated 'evolutionary islands' in the SDTF ecosystems with small population sizes in which genetic drift dominated the evolution of genome size, while (ii) the dispersal of species to higher latitudes was accompanied by selection for larger genomes in the more temperate regions (Fig. 5).

Currently there are insufficient genomic data to unpick the relative contributions of drift versus selection in the evolution of genome size diversity across the geographical distribution of the Caesalpinia group. Nevertheless, with the rapid growth in high throughput sequencing approaches and the increasingly powerful models available to obtain insights into population level processes from small numbers of samples (e.g. Pairwise sequentially Markovian coalescent models Li and Durbin, 2011), the ability to resolve such questions are fast become tractable and will likely lead to a more holistic understanding of the evolutionary and environmental processes underpinning genomic evolution in the Caesalpinia group.

Acknowledgements

The authors wish to thank the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE - APQ- 0970-2.03/15) for financial support and a post-doc grant to G.S. by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Estágio Pós-Doutoral 88881.119479/2016-01). G.S. receives a productivity fellowship from CNPq (process number PQ - 310693/2018-7).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ppees.2019.03.002.

References

Allen, K., Dupuy, J.M., Gei, M.G., Hulshof, C., Medvigy, D., Pizano, C., Salgado-Negret, B., Smith, C.M., Trierweiler, A., Van Bloem, S.J., Waring, B.G., Xu, X., Powers, J.S., 2017. Will seasonally dry tropical forests be sensitive or resistant to future changes in rainfall regimes? Environ. Res. Lett. 12, 023001.

Alves, M.A.O., Custódio, A.V.C., 1989. Citogenética de Leguminosas coletadas no estado do Ceará. Rev. Bras. Genet. 12, 81–92.

Avdulov, N.P., 1931. Karyo-systematische untersuchungen der Familie Gramineen. Bull. App. Bot. Genet. Plant. Breed. 4, 1–428.

Bartón, K., 2009. Package "MuMIn: Multi-model inference" for R. R Package.

Beaulieu, J.M., Leitch, I.J., Knight, C.A., 2007a. Genome size evolution in relation to leaf strategy and metabolic rates revisited. Ann. Bot. 99, 495–505.

Beaulieu, J.M., Moles, A.T., Leitch, I.J., Bennett, M.D., Dickie, J.B., Knight, C.A., 2007b. Correlated evolution of genome size and seed mass. New Phytol. 173, 422–437.

Beaulieu, J.M., Leitch, I.J., Patel, S., Pendharkar, A., Knight, C.A., 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytol. 179, 975–986.

Beltrão, G.T.A., Guerra, M., 1990. Citogenética de angiospermas coletadas em Pernambuco – III. Ciência e Cultura 42, 839–845.

Bennett, M.D., Leitch, I.J., 2011. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Ann. Bot. 107, 467–590.

Bilinski, P., Albert, P.S., Berg, J.J., Birchler, J., Grote, M., Lorant, A., Quezada, J., Swarts, K., Yang, J., Ross-Ibarra, J., 2018. Parallel altitudinal clines reveal adaptive evolution of genome size in *Zea mays*. PLoS Genet. 14, e1007162.

Bruneau, A., Mercure, M., Lewis, G.P., Herendeen, P.S., 2008. Phylogenetic patterns and diversification in the caesalpinioid legumes. Botany 86, 697–718.

Calvillo-Canadell, L., Cevallos-Ferriz, S.R.S., 2005. Diverse assemblage of Eocene and Oligocene leguminosae from Mexico. Int. J. Plant Sci. 166, 671–692.

Caponio, I., Anton, A.M., Fortunato, R.H., Norrmann, G.A., 2012. Ploidy dimorphism and reproductive biology in *Stenodrepanum bergii* (Leguminosae), a rare South American endemism. Genome 55, 1–7.

Cavalier-Smith, T., 2005. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann. Bot. 95, 147–175.

Ceccarelli, M., Falisfocco, E., Cionini, P.G., 1992. Variation in genome size and organisation within hexaploid Festuca arundinaceae. Theor. Appl. Genet. 83, 273–278.

Cooper, N., Thomas, G.H., FitzJohn, R.G., 2016. Shedding light on the 'dark side' of phylogenetic comparative methods. Methods Ecol. Evol. 7, 693–699.

Costa, L., Oliveira, Á., Carvalho-Sobrinho, J., Souza, G., 2017. Comparative cytomolecular analyses reveal karyotype variability related to biogeographic and species richness patterns in Bombacoideae (Malvaceae). Plant Syst. Evol. 303, 1131–1144.

Creber, H.M.C., Davies, M.S., Francis, D., Walker, H.D., 1994. Variation in DNA C value in natural populations of *Dactylis glomerata*. New Phytol. 128, 555–561.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

Deatherage, D.E., Kepner, J.L., Bennett, A.F., Lenski, R.E., Barrick, J.E., 2017. Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. Proc. Natl. Acad. Sci. U. S. A. 114, E1904–E1912.

Díez, C.M., Gaut, B.S., Meca, E., Scheinvar, E., Montes-Hernandez, S., Eguiarte, L.E., Tenaillon, M.I., 2013. Genome size variation in wild and cultivated maize along altitudinal gradients. New Phytol. 199, 264–276.

Doležel, J., Binarova, P., Lucretti, S., 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. Biol. Plantarum 31, 113–120.

Doležel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysák, M.A., Nardi, L., Obermayer, R., 1998. Plant genome size estimation by flow cytometry: inter-laboratory comparison. Ann. Bot. 82, 17–26.

Doležel, J., Greilhuber, J., Suda, J., 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nat. Protoc. 2, 2233–2244.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Du, Y.P., Bi, Y., Zhang, M.F., Yang, F.P., Jia, G.X., Zhang, X.H., 2017. Genome size diversity in *Lilium* (Liliaceae) is correlated with karyotype and environmental traits. Front. Plant Sci. 8, 1303.

Ebihara, A., Ishikawa, H., Matsumoto, S., Lin, S.J., Iwatsuki, K., Takamiya, M., Watano, Y., Ito, M., 2005. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. Am. J. Bot. 92, 1535–1547.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., et al., 2012. Package 'car'. R Foundation for Statistical Computing, Vienna.

Freshwater, D.W., 1988. Relative genome-size differences among populations of *Spartina alterniflora* Loisel (Poaceae) along east and gulf coasts of USA. J. Exp. Mar. Bio. Ecol. 120, 239–246.

Gagnon, E., Lewis, G.P., Sotuyo, J.S., Hughes, C.E., Bruneau, A., 2013. A molecular phylogeny of *Caesalpinia* sensu lato: increased sampling reveals new insights and more genera than expected. S. Afr. J. Bot. 89, 111–127.

Gagnon, E., Hughes, C.E., Lewis, G.P., Bruneau, A., 2015. A new cryptic species in a new cryptic genus in the *Caesalpinia* group (Leguminosae) from the seasonally dry inter-Andean valleys of South America. Taxon 64, 468–490.

Gagnon, E., Bruneau, A., Hughes, C.E., de Queiroz, L., Lewis, G.P., 2016. A new generic system for the pantropical *Caesalpinia* group (Leguminosae). PhytoKeys 71, 1.

Gagnon, E., Ringelberg, J.J., Bruneau, A., Lewis, G.P., Hughes, C.E., 2019. Global succulent biome phylogenetic conservatism across the pantropical Caesalpinia group. New Phytol. https://doi.org/10.1111/nph.15633.

Garland, T., Harvey, P.H., Ives, A.R., 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41, 18–32.

Grotkopp, E., Rejmánek, M., Sanderson, M.J., Rost, T.L., 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. Evolution 58, 1705–1729.

Hadfield, J.D., 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J. Stat. Softw. 33, 1–22.

Hansen, T.F., Pienaar, J., Orzack, S.H., 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62, 1965–1977.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., Wendel, J.F., 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res. 16, 1252–1261.

Herendeen, P.S., 1992. The fossil history of the Leguminosae from the Eocene of southeastern North America. In: Herendeen, P.S., Dilcher, D.L. (Eds.), Advances in Legume Systematics, Part 4: the Fossil Record. Royal Botanic Gardens, Kew, United Kingdom, pp. 85–160.

Herendeen, P.S., Dilcher, D.L., 1991. *Caesalpinia* subgenus *Mezoneuron* (Leguminosae, caesalpinioideae) from the tertiary of North America. Am. J. Bot. 78, 1–12.

Hidalgo, O., Pellicer, J., Christenhusz, M.J.M., Schneider, H., Leitch, A.R., Leitch, I.J., 2017. Is there an upper limit to genome size? Trends Plant Sci. 22, 567–573.

Hijmans, R.J., 2017. Raster: Geographic Data Analysis and Modeling. R Package Version 2.6-7. https://CRAN.R-project.org/package=raster.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Jakob, S.S., Meister, A., Blattner, F.R., 2004. The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. Mol. Biol. Evol. 21, 860–869.

Jordan, G.J., Carpenter, R.J., Koutoulis, A., Price, A., Brodribb, T.J., 2015. Environmental adaptation in stomatal size independent of the effects of genome size. New Phytol. 205, 608–617.

Kamilar, J.M., Cooper, N., 2013. Phylogenetic signal in primate behaviour, ecology and life history. Philos. Trans. Biol. Sci. 368, 20120341.
Kang, M., Tao, J., Wang, J., Ren, C., Qi, Q., Xiang, Q.Y., Huang, H., 2014. Adaptive and nonadaptive genome size evolution in Karst endemic flora of China. New Phytol. 202, 1371–1381.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Thierer, T., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kelly, L.J., Renny-Byfield, S., Pellicer, J., Macas, J., Novák, P., Neumann, P., Lysák, M.A., Day, P.D., Berger, M., Fay, M.F., et al., 2015. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol. 208, 596–607.
Kenton, A.Y., Rudall, P.J., Johnson, A.R., 1986. Genome size variation in *Sisyrinchium* L. (Iridaceae) and its relationship to phenotype and habitat. Bot. Gaz. 147, 342–354.
Knight, C.A., Ackerly, D.D., 2002. Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. Ecol. Lett. 5, 66–76.
Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74, 1659–1673.
Levin, D.A., Funderberg, S.W., 1979. Genome size in angiosperms: temperate *versus* tropical species. Am. Nat. 114, 784–795.
Lewis, G.P., 1998. Caesalpinia: a Revision of the Poincianella-erythrostemon Group. Royal Botanic Gardens (K-RBG). .
Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493.
Lysák, M.A., Rostková, A., Dixon, J.M., Rossi, G., Doležel, J., 2000. Limited genome size variation in *Sesleria albicans*. Ann. Bot. 86, 399–403.
Lyu, H., He, Z., Wu, C.I., Shi, S., 2018. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. New Phytol. 217, 428–438.
Maas, M., Burgos, A., 2011. In: Dirzo, R., Young, H.S., Mooney, H.A., Ceballos, G. (Eds.), Water Dynamics and the Ecosystem Level in Seasonally Dry Tropical Forests Seasonally Dry Tropical Forests: Ecology and Conservation. Island Press, Washington, DC, pp. 141–156.
Macas, J., Novak, P., Pellicer, J., Čížková, J., Koblížková, A., Neumann, P., Fuková, I., Doležel, J., Kelly, L.J., Leitch, I.J., 2015. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. PLoS One 10, e0143424.
Manzanilla, V., Bruneau, A., 2012. Phylogeny reconstruction in the Caesalpinieae grade (Leguminosae) based on duplicated copies of the sucrose synthase gene and plastid markers. Mol. Phylogenet. Evol. 65, 149–162.
Mascagni, F., Giordani, T., Ceccarelli, M., Cavallini, A., Natali, L., 2017. Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.). BMC Genomics 18, 634.
Mráz, P., Chrtek, J., Šingliarová, B., 2009. Geographical parthenogenesis, genome size variation and pollen production in the arctic-alpine species Hieracium alpinum. Bot. Helv. 119, 41–51.
Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., Thuiller, W., 2012. How to measure and test phylogenetic signal. Methods Ecol. Evol. 3, 743–756.
Obermayer, R., Leitch, I.J., Hanson, L., Bennett, M.D., 2002. Nuclear DNA C-valuesin 30 species double the familial representation in pteridophytes. Ann. Bot. 90, 209–217.
Ohri, D., 1998. Genome size variation in plant systematics. Ann. Bot. 82, 75–83.
Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., Pearse, W., 2018. Caper: Comparative Analyses of Phylogenetics and Evolution in R. R Package Version 1.0.1. https://CRAN.R-project.org/package=caper.
Pagel, M., 1999. Inferring the historical patterns of biological evolution. Nature 401, 877–884.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Pellicer, J., Fay, M.F., Leitch, I.J., 2010. The largest eukaryotic genome of them all? Bot. J. Linn. Soc. 164, 10–15.
Pellicer, J., Hidalgo, O., Dodsworth, S., Leitch, I., 2018. Genome size diversity and its impact on the evolution of land plants. Genes 9, 88.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2013. R Development Core Team (2012) Nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-103. R Foundation for Statistical Computing, Vienna.
Plummer, M., Best, N., Cowles, K., Vines, K., 2010. Coda: Output Analysis and Diagnostics for MCMC.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Qiu, F., Baack, E.J., Whitney, K.D., Bock, D.G., Tetreault, H.M., Rieseberg, L.H., Ungerer, M.C., 2018. Phylogenetic trends and environmental correlates of nuclear genome size variation in Helianthus sunflowers. New Phytol. https://doi.org/10.1111/nph.15465.
Quader, S., Isvaran, K., Hale, R.E., Miner, B.G., Seavy, N.E., 2004. Nonlinear relationships and phylogenetically independent contrasts. J. Evol. Biol. 17, 709–715.
Quantum Gis Development Team, 2014. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project.
R Core Team, 2017. R: a Language and Environment for Statistical Computing. R Version 3.3.3. URL. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rambaut, A., 2009. FigTree v1. 3.1: Tree Figure Drawing Tool. http://tree.bio.ed.ac.uk/software/figtree.
Rambaut, A., Drummond, A.J., Suchard, M., 2014. Tracer v1. pp. 6. http://beast.bio.ed.ac.uk.Tracer.
Rayburn, A.L., Price, H.J., Smith, J.D., Gold, J.R., 1985. C-band heterochromatin and DNA content in *Zea mays*. Am. J. Bot. 72, 1610–1617.
Reeves, G., Francis, D., Davies, M.S., Rogers, H.J., Hodkinson, T.R., 1998. Genome size is negatively correlated with altitude in natural populations of Dactylis glomerata. Ann. Bot. 82, 99–105.
Revell, L.J., 2012. Phytools: an R package for phylogenetic comparative biology (and other things). R version 0.6–44. Methods Ecol. Evol. 3, 217–223.
Rodrigues, P.S., Souza, M.M., Corrêa, R.X., 2012. Karyomorphology of *Caesalpinia* species (Caesalpinioideae: Fabaceae) from Caatinga and Mata Atlantica Biomes of Brazil. J. Plant Stud. 1, 82.
Rodrigues, P.S., Souza, M.M., Melo, C.A.F., Pereira, T.N.S., Corrêa, R.X., 2018. Karyotype diversity and 2C DNA content in species of the Caesalpinia group. BMC Genet. 19, 25.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Särkinen, T., Pennington, R.T., Lavin, M., Simon, M.F., Hughes, C.E., 2012. Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry tropical forests. J. Biogeog. 39, 884–900.
Schmuths, H., Meister, A., Horres, R., Bachmann, K., 2004. Genome size variation among accessions of *Arabidopsis thaliana*. Ann. Bot. 93, 317–321.
Simon, M.F., Grether, R., de Queiroz, L.P., Skema, C., Pennington, R.T., Hughes, C.E., 2009. Recent assembly of the Cerrado, a Neotropical plant diversity hotspot, by *in situ* evolution of adaptations to fire. Proc. Natl. Acad. Sci. U. S. A. 106, 20359–20364.
Simpson, B.B., Tate, J.A., Weeks, A., 2005. The biogeography of *Hoffmannseggia* (Leguminosae, Caesalpinioideae, Caesalpinieae): a tale of many travels. J. Biogeog. 32, 15–27.
Simpson, B., Larkin, L., Weeks, A., McDill, J., 2006. Phylogeny and biogeography of *Pomaria* (caesalpinioideae: leguminosae). Syst. Bot. 31, 792–804.
Stebbins, G.L., 1966. Chromosomal variation and evolution. Science 152, 1463–1469.
Stebbins, G.L., 1971. Chromosomal Evolution in Higher Plants. Chromosomal Evolution in Higher Plants.
Van-Lume, B., Esposito, T., Diniz-Filho, J.A.F., Gagnon, E., Lewis, G.P., Souza, G., 2017. Heterochromatic and cytomolecular diversification in the Caesalpinia group (Leguminosae): relationships between phylogenetic and cytogeographical data. Perspect. Plant Ecol. Evol. Syst. 29, 51–63.
Veselý, P., Bureš, P., Šmarda, P., Pavlícek, T., 2012. Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? Ann. Bot. 109, 65–75.
Vu, G.T., Schmutzer, T., Bull, F., Cao, H.X., Fuchs, J., Tran, T.D., Jovtchev, G., Pistrick, K., Stein, N., Pecinka, A., et al., 2015. Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. Plant Genome 8.
Wang, W., Kerstetter, R.A., Michael, T.P., 2011. Evolution of genome size in duckweeds (Lemnaceae). J. Bot. https://doi.org/10.1155/2011/570319.
Weng, M.L., Ruhlman, T.A., Gibby, M., Jansen, R.K., 2012. Phylogeny, rate variation, and genome size evolution of *Pelargonium* (Geraniaceae). Mol. Phylogenet. Evol. 64, 654–670.