Planta (2022) 256:50
https://doi.org/10.1007/s00425-022-03961-9

ORIGINAL ARTICLE

Comparative repeatome analysis reveals new evidence on genome evolution in wild diploid *Arachis* (Fabaceae) species

Sergio S. Samoluk[1] · Magdalena Vaio[2] · Alejandra M. Ortíz[1] · Laura M. I. Chalup[1] · Germán Robledo[1,3] · David J. Bertioli[4] · Guillermo Seijo[1,3]

Received: 21 March 2022 / Accepted: 12 July 2022

Abstract

Main conclusion **Opposing changes in the abundance of satellite DNA and long terminal repeat (LTR) retroelements are the main contributors to the variation in genome size and heterochromatin amount in *Arachis* diploids.**

The South American genus *Arachis* (Fabaceae) comprises 83 species organized in nine taxonomic sections. Among them, section *Arachis* is characterized by species with a wide genome and karyotype diversity. Such diversity is determined mainly by the amount and composition of repetitive DNA. Here we performed computational analysis on low coverage genome sequencing to infer the dynamics of changes in major repeat families that led to the differentiation of genomes in diploid species ($x = 10$) of genus *Arachis*, focusing on section *Arachis*. Estimated repeat content ranged from 62.50 to 71.68% of the genomes. Species with different genome composition tended to have different landscapes of repeated sequences. Athila family retrotransposons were the most abundant and variable lineage among *Arachis* repeatomes, with peaks of transpositional activity inferred at different times in the evolution of the species. Satellite DNAs (satDNAs) were less abundant, but differentially represented among species. High rates of evolution of an AT-rich superfamily of satDNAs led to the differential accumulation of heterochromatin in *Arachis* genomes. The relationship between genome size variation and the repetitive content is complex. However, largest genomes presented a higher accumulation of LTR elements and lower contents of satDNAs. In contrast, species with lowest genome sizes tended to accumulate satDNAs in detriment of LTR elements. Phylogenetic analysis based on repetitive DNA supported the genome arrangement of section *Arachis*. Altogether, our results provide the most comprehensive picture on the repeatome dynamics that led to the genome differentiation of *Arachis* species.

Keywords Genome differentiation · *Arachis* species · Athila retroelements · Satellite DNA · Genome size variation

Abbreviations

DAPI	4′,6-Diamidino-2-phenylindole
FISH	Fluorescent in situ hybridization
GISH	Genome in situ hybridization

Communicated by Anastasios Melis.

✉ Sergio S. Samoluk
samocarp31@gmail.com; ssamoluk@agr.unne.edu.ar

Magdalena Vaio
mvaio@fagro.edu.uy

Alejandra M. Ortíz
ortizalejandr@gmail.com

Laura M. I. Chalup
laurachalup@gmail.com

Germán Robledo
germanrobledo71@gmail.com

David J. Bertioli
bertioli@uga.edu

Guillermo Seijo
jgseijo@yahoo.com

[1] Instituto de Botánica del Nordeste (UNNE-CONICET), Facultad de Ciencias Agrarias, Corrientes, Argentina

[2] Laboratory of Plant Genome Evolution and Domestication, Department of Plant Biology, Faculty of Agronomy, University of the Republic, Montevideo, Uruguay

[3] Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

[4] Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA

Published online: 27 July 2022

LTR	Long terminal repeat
MYA	Million years ago
rDNA	Ribosomal DNA
satDNA	Satellite DNA
RepeatExplorer2	RE2
MULEs	Mutator-like elements

Introduction

Eukaryotic genomes contain large quantities of different classes of repetitive DNA sequences, which contribute largely to genome differentiation (Li et al. 2004; Mehotra and Goyal 2014). The repetitive sequences can be classified in two main categories, tandem (commonly known as satDNAs) and interspersed (transposable elements) repeats (Kumar and Bennetzen 1999; López-Flores and Garrido-Ramos 2012; Mehotra and Goyal 2014; Biscotti et al. 2015). Satellite DNA is formed by tandem arrays of monomeric units mainly localized in heterochromatic blocks of pericentromeric and subtelomeric regions (López-Flores and Garrido-Ramos 2012; Plohl et al. 2012; Biscotti et al. 2015; Garrido-Ramos 2015), but they also span into the functional centromere and telomere (Plohl et al. 2014; Robledillo et al. 2018). A library of satDNAs is commonly shared by closely related species, however, some members of the library can diverge and be differentially amplified leading to specific patterns (Plohl et al. 2008). Transposons are considered as genetic units that can move to new genome sites through a "cut and paste" mechanism (DNA transposons) or via reverse transcription of their mRNA (retrotransposons). Both tandem and dispersed repeats are differentially represented, even among closely related species, giving rise to species-specific landscapes called "repeatomes" (Woo et al. 2007). As a consequence of their high evolutionary rates, the different components of the repeatome can produce short- to long-term evolutionary changes in the size, structure* and function of host genomes leading, in this way, to species divergence (Biemont and Viera 2006; Ferree and Barbash 2009; Chénais et al. 2012; Biscotti et al. 2015; Alioto et al. 2020). Therefore, the comprehensive analysis of repeatomes provides invaluable information for understanding the genome evolution of closely related species.

The appearance of high-throughput sequencing technologies and development of new bioinformatics approaches allowed the characterization of repetitive DNA in many closely related groups of organisms (Ewing 2015; Staton and Burke 2015; Nelson et al. 2017; Negm et al. 2021). One of the most widely used approaches is RepeatExplorer, which performs a simultaneous and reliable characterization of diversity and abundance of repetitive DNA from low-coverage genome sequencing, without the need of previously assembled genomes (Novak et al. 2020). This approach is useful not only for the repeatome characterization of single species but also for the comparative analysis of multiple species, allowing the study of the evolutionary dynamics of homologous repeat families in the different genomes (Macas et al. 2015; Usai et al. 2017; Bolsheva et al. 2019; Gaiero et al. 2019; McCann et al. 2020; Mascagni et al. 2020). In addition, and considering that both relative abundance and sequence diversity of homologous repetitive elements evolve primarily by random genetic drift (Jurka et al. 2011, 2012), new methods have been developed to infer the phylogenetic relationships from comparative runs of RepeatExplorer (Dosworth et al. 2014; Vitales et al. 2020). These methods have proven to reconstruct reliable phylogenetic relationships in different genera such as *Solanum* (Dodsworth et al. 2016), *Nicotiana* (Dodsworth et al. 2017; Vitales et al. 2020), *Linum* (Bolsheva et al. 2019), *Melanopodium* (McCann et al. 2020) and *Heliophila* (Dogan et al. 2021).

A good system for investigating repetitive DNA evolution is the South American genus *Arachis* (Fabaceae) which diverged around 13.8 ± 1.7 MYA (Lavin et al. 2004). It is a natural group (Moretzsohn et al. 2004, 2013; Bechara et al. 2010; Wang et al. 2019) of 83 annual and perennial species organized in nine taxonomic sections (Krapovickas and Gregory 1994; Valls and Simpson 2005, 2017; Valls et al. 2013; Santana and Valls 2015; Seijo et al. 2021). The largest and most diverse is the section *Arachis* comprising 30 wild diploid species, 27 of them with $x = 10$ and three with $x = 9$ (Robledo and Seijo 2008, 2010; Robledo et al. 2009; Silvestri et al. 2014). It also includes two allotetraploids ($2n = 4x = 40$), the peanut (*A. hypogaea*) and a wild species of common origin (*A. monticola*). These two tetraploids arose from hybridization and spontaneous chromosome duplication of two wild species, *A. duranensis* and *A. ipaënsis* (Seijo et al. 2004, 2007; Bertioli et al. 2016, 2019). Section *Arachis* is a relatively young species complex with estimated divergence time of 2.33–4.99 MYA (Moretzsohn et al. 2013; Bertioli et al. 2016). Diploid species with $x = 10$ are highly diverse and they have been assigned to five different groups of genome, A, B, D, F and K (Smartt et al. 1978; Stalker 1991; Robledo and Seijo 2008, 2010; Robledo et al. 2009). Additionally, three karyotype groups (Chiquitano, Pantanal and La Plata River Basin) were identified among the A genome species (Robledo et al. 2009). This genome assignment was based on cross-compatible assays (Gregory and Gregory 1979; Stalker et al. 1991), but mainly on chromosome features (Smartt et al. 1978; Stalker et al. 1991; Fernández and Krapovickas 1994; Robledo et al. 2009; Robledo and Seijo 2010).

All species of section *Arachis* with $x = 10$ have symmetric karyotypes composed mostly of metacentric chromosomes, except *A. glandulifera*, that has a derived asymmetric karyotype probably through the occurrence of pericentric inversions and/or translocations (Stalker 1991). Even though they share

the same basic chromosome number, they have approximately 1.26-fold variation in genome size (Samoluk et al. 2015a). The annual species belonging to different genomes tend to have different genome sizes. However, the 2C values of the perennial species (all of them with A genome) were distributed almost along the whole range of genome sizes here observed. One of the most outstanding cytogenetic differences among these species is the proportion of $DAPI^+$ heterochromatin and the size and number of rDNA loci (Seijo et al. 2004; Robledo and Seijo 2008, 2010; Robledo et al. 2009). Also, GISH experiments performed both on natural and artificial amphidiploids demonstrated a high efficiency in discriminating genomes without the need of species-specific blocking DNA (Seijo et al. 2007, 2018). On these bases, it was suggested that the repetitive components would have played a major role in the karyotype evolution within section *Arachis*.

Initial analyses of few tandem repeats (Zhang et al. 2016; Samoluk et al. 2017, 2019) and transposable elements (Patel et al. 2004; Nielen et al. 2010, 2012; Gowda et al. 2011; Shirasawa et al. 2012; Bertioli et al. 2013, 2016, 2019; Samoluk et al. 2015b) supported the hypothesis that repetitive elements had played a major role in the genome divergence of *Arachis* species. However, most of them were mainly focused on the A and B genomes and on the effects of the allopolyploidization process that gave rise to the cultivated peanut. Therefore, a comprehensive analysis of the repeatome composition in representative species of different genomes of *Arachis* is still lacking for a better understanding of the evolution and the diversity of genomes at diploid level.

Here we make a comprehensive characterization of the repeatomes to gain insights, at genome-wide scale, on the evolutionary pathways behind the observed genome and karyotype diversity in wild diploid ($x = 10$) *Arachis* species. For this purpose, we performed a comparative similarity-based clustering (Novak et al. 2020) of low coverage read data in eight species belonging to different genomes and karyotype groups of section *Arachis*, using two species of sections *Procumbentes* and *Erectoides* as outgroups. Our results highlight the important role of Athila family retrotransposons in the genome diversification of section *Arachis* and probably in the diversification of the genus. In addition, they evidenced that the differential and parallel amplification of closely related satDNA sequences modeled the global patterns of AT-rich heterochromatin characteristic to different genomes.

Materials and methods

Plant material, DNA extraction and sequencing data

The study was conducted in 10 diploid species of the genus *Arachis* with $x = 10$ (sections *Arachis*, *Procumbentes* and *Erectoides*). Plants of *A. trinitensis*, *A. helodes*, *A. glandulifera* and *A. paraguariensis* were grown in greenhouses at the Center for Applied Genetic Technologies at the University of Georgia (Athens, GA, USA). Total DNA from young fresh leaves of each plant was extracted using a DNA extraction kit (Qiagen, Hilden, Germany) and libraries were prepared using the Illumina TruSeq DNA Library Prep kit-600 bp insert size (Illumina Inc.). These libraries were sequenced on Illumina MiSeq platform using the reagent kit v3-600 cycle (cat. MS-102-3003) to obtain 2×300 bp paired-end reads. In addition, Illumina DNA sequences of *A. duranensis*, *A. cardenasii*, *A. batizocoi*, *A. rigonii*, *A. magna* and *A. ipaënsis* were retrieved from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra). The details of taxa included in this study are described in Table 1.

Repeat identification

Adapter removing and read quality analysis were performed with Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). All sequence reads were then trimmed to 100 bp and filtered by quality with 95% of bases equal to or above the quality cut-off value of 10. Two different strategies were used for the genome-wide analysis of repetitive sequences. The first consisted of independent clustering analyses using a set of 500,000 reads for each species, while the second comprised a concatenated dataset of randomly selected reads covering the $0.01 \times$ of the genomes. The latter strategy allowed the identification of shared and specific repeats across all *Arachis* genomes. These datasets were analyzed using a similarity-based read clustering method implemented in the RepeatExplorer2 (RE2) pipeline (https://repeatexplorer-elixir.cerit-sc.cz; Novák et al. 2020). The clustering was performed using the default settings of 90% similarity over 55% of the read length. After removing clusters classified as plastid and contamination, the remaining top clusters (i.e. clusters with at least 0.01% of the total reads used as input) were annotated by similarity searches against different sources. These included a reference database of transposable element protein domains (Viridiplantae, version 3.0), a custom library of repetitive sequences of *Arachis* species (https://www.peanutbase.org/data/v2/Arachis/hypogaea/repeats/) and a database of *Arachis* satDNAs. The latter was built with previously characterized satDNAs (Zhang et al. 2012, 2016; Samoluk et al. 2017, 2019) and those satDNA repeats characterized here through the independent analysis of 500,000 reads from each *Arachis* species with TAREAN tool (Novák et al. 2017). The relationships among genome proportion of repeats, heterochromatin content and genome size were analyzed by linear regression in scatter plots and Pearson's correlation tests. For the aforementioned analysis, the previously published genome sizes (Samoluk et al. 2015a) and heterochromatin

Table 1 General information of the wild diploid *Arachis* species used in the clustering analysis including provenance and collection number, botanical section, four-letter code, genome assignment, genome size, heterochromatin content, and life cycle

Species	Provenance, collectors and collection number	Letter code	Section	Genome (karyotype group)	GS (Mb)	HC (%)[a]	LC
A. cardenasii Krapov. and W. C. Gregory	Bolivia, Dept. Santa Cruz, Prov. Chiquitos, Roboré. G, K, P 10017	Acar	*Arachis*	A (Chiquitano)	1.374	11.90	p
A. helodes Martius ex Krapov. and Rigoni	Brazil, St. Mato Grosso, Cuiabá. K, G 30029	Ahel	*Arachis*	A (Pantanal)	1.472	12.42	p
A. duranensis Krapov. and W. C. Gregory	Argentina, Salta, San Martín, Campo Durán. K 7988	Adur	*Arachis*	A (La Plata River Basin)	1.247	14.59	a
A. ipaënsis Krapov. and W. C. Gregory	Bolivia, Dept. Tarija, Prov. Gran Chaco, Ipa. K, G, B, P, Sc, S 30076	Aipa	*Arachis*	B	1.560	0	a
A. magna Krapov., W.C. Gregory and C. E. Simpson	Bolivia, Dept. Santa Cruz, Prov. Velasco, San Ignacio. K, G, Sc, S 30097	Amag	*Arachis*	B	1.575	0	a
A. trinitensis Krapov. and W. C. Gregory	Bolivia, Dept. Beni, Prov. Cercado, Trinidad. Wi 1117	Atri	*Arachis*	F	1.389	5.89	a
A. batizocoi Krapov. and W. C. Gregory	Bolivia, Dept. Santa Cruz, Prov. Cordillera, Parapetí. G, K, B, S, P, Sc 30082	Abat	*Arachis*	K	1.384	11.36	a
A. glandulifera Stalker	Bolivia, Dept. Santa Cruz, Prov. Velasco, San Ignacio. G, K, S, Sc 30099	Agla	*Arachis*	D	1.315	13.09	a
A. rigonii Krapov. and W. C. Gregory	K47[b]	Arig	*Procumbentes*	E	1.560	12.85	p
A. paraguariensis Chodat and Hassl	Brasil. BRA 017-591[c]	Apar	*Erectoides*	E	1.511	8.83	p

GS 1C genome size, *Mb* megabases, *HC* heterochromatin content, *LC* life cycle, *p* perennial, *a* annual

[a]Expressed as percentage of the total chromosome length

[b]As reported by Yin et al. (2017)

[c]As detailed in GRIN-GLOBAL passport

content (Robledo and Seijo 2008, 2010; Robledo et al. 2010; Ortiz et al. 2017) of *Arachis* species were considered. The holoploid genome size of *A. paraguariensis* subsp. *paraguariensis* (KGP 30109) and *A. rigonii* (GKP 10034) were estimated by flow cytometry according to the procedure described by Samoluk et al. (2015a), and using *Paspalum intermedium* Munro ex Morong. accession Sch 28857 (diploid, $2C = 1.42$ pg; Vaio et al. 2007) as internal standard. The *Cx* values, representing the DNA content of one non-replicated monoploid genome with the chromosome number *x* (Greilhuber et al. 2005), were calculated as the 2*C* nuclear DNA content divided by ploidy level and expressed in pg (1 pg of DNA = 978 Mbp; Doležel and Bartos 2005).

The sequence similarities among the characterized satDNA monomers were performed by BLASTn searches (Altschul et al. 1990) with *e*-value $< 10^{-10}$, and visualized with Circoletto (http://tools.bat.infspire.org/circoletto/; Darzentas 2010). To test the tandem organization of satDNAs, sequence comparisons against the chromosomal pseudomolecules of *A. hypogaea* cv. Tifrunner (Bertioli et al. 2019) were performed with BLAT (Kent 2002), using a match score ≥ 90%. The selected genome regions were analyzed for tandem repeat organizations using the YASS genomic similarity tool (http://bioinfo.lifl.fr/yass/yass.php; Noé and Kucherov 2005).

Estimation of temporal dynamics of Athila family retrotransposons

Dating the insertion of retrotransposons in *Arachis* species was inferred by calculating pairwise genetic divergences from alignments of encoding RT sequences (Piegu et al. 2006; Ammiraju et al. 2007; Mascagni et al. 2018) of *Athila*, the most abundant and variable monophyletic lineage of retrotransposons. For this purpose, those contigs that encoded for RT domains of *Athila* were retrieved by DANTE search tool (implemented in the RepeatExplorer Galaxy web-interface) and separately aligned by MAFFT

7.017 (Katoh et al. 2002) in order to obtain a consensus sequence. Then, Illumina reads of 100 nt-long of each species (used as input of RE2) were mapped to the consensus sequence using Geneious Read Mapper implemented in Geneious Prime 2021 (https://www.geneious.com). The intraspecific pairwise divergence values between mapped reads under the Kimura two-parameter model of sequence evolution (Kimura 1980) were determined using MEGA version 7.0.18 (Kumar et al. 2016). Kimura distances were converted to MYA using the equation $t = K/2r$, where t is the age (MYA), K is the pairwise divergence between reads and r is the nucleotide substitution rate ($r = 1.3 \times 10^{-8}$ substitution/site/year; Ma and Bennetzen 2004). Finally, we constructed a histogram representing the frequency distribution of the divergence values where the peaks are interpreted as transposition bursts.

Phylogenetic analysis

The phylogenetic inference based on genome proportions of repeats was performed according to Dodsworth et al. (2016), using the top clusters obtained from the clustering analysis of a concatenated dataset of *Arachis* species. All read numbers in the comparative data matrix were divided by a correction factor in order to convert all of them ≤ 65. Then, this matrix was used as input in the software TNT (Goloboff et al. 2008). The consensus tree was inferred using maximum parsimony and 10,000 bootstrap replicates with symmetrical resampling, using *A. paraguariensis* as outgroup. In addition, we performed a phylogenetic inference based on the pairwise sequence similarities among the reads of the species (matrices of observed/expected number of edges) of the 100 most-abundant clusters identified by RE2, as described by Vitales et al. (2020). Then, distance matrices were obtained by calculating the inverse of the values of the similarity matrices. Those matrices lacking pairwise similarity values were excluded from this analysis. The neighbor-joining trees were constructed with the package *ape* (Paradis and Schliep 2018) in R (R Core team 2018) and then exported to nexus format. Finally, the networks (Holland and Moulton 2003) were constructed using the program SplitsTree4 (Huson and Bryant 2006).

Statistical analysis

The identification of repeats in *Arachis* genomes using RE2 was done through three replicated runs to confirm the consistency of the results. The genome sizes of *A. paraguariensis* and *A. rigonii* estimated here are expressed as the average of three biological and three technical replicates. The software Infostat (Di Rienzo et al. 2013) was used to perform regression analysis and Pearson's correlation tests.

Data availability and GenBank accession numbers

The original reads of the species sequenced here (*A. trinitensis*, *A. paraguariensis*, *A. helodes* and *A. glandulifera*) were deposited in the NCBI short-read archive (SRA) and are available under accession numbers SRR19084327–SRR19084330. In addition, Illumina DNA sequences of *A. duranensis*, *A. cardenasii*, *A. batizocoi*, *A. rigonii*, *A. magna* and *A. ipaënsis* were retrieved from the NCBI Sequence Read Archive (Accession Nos. SRR8481368, SRR8481366, SRR8481357, SRR13188218, SRR8784101, and SRR8784095).

Results

Overall genome composition of repetitive sequences in diploid species of *Arachis*

The global composition of repetitive DNA sequences in the genomes of *Arachis* species was investigated following two strategies. Firstly, a clustering analyses using individual set of reads for each species (Fig. S1); and secondly, an analysis with a concatenated dataset of paired-reads from all diploid species covering 0.01 × of each genome. Since results from both strategies were comparable, only the results obtained from the concatenated dataset are presented in detail (Fig. 1 and Table 2). A total of 231 clusters with at least 0.01% of the total number of reads analyzed (top clusters) were obtained. Most of them could be assigned to different classes of known repetitive DNA (i.e. retrotransposons, DNA transposons, satDNA and rDNA); however, a small fraction (0.74–2.15%) remained as unclassified. The genome fraction of repetitive DNA among species of section *Arachis* ranged from 66.21% in *A. trinitensis* to 71.68% in *A. helodes*. The repetitive fraction in the outgroup species were 62.50% in *A. paraguariensis* (section *Erectoides*) and 71.11% in *A. rigonii* (section *Procumbentes*).

The LTR retrotransposons were the most abundant repetitive sequences (from 50.86% in *A. duranensis* to 62.83% in *A. rigonii*) in all species analyzed, with a large predominance of the Ty3-Gypsy superfamily (31.73% in *A. glandulifera* to 49.40% in *A. rigonii*) over Ty1-Copia elements (0.41% in *A. rigonii* to 6.38% in *A. paraguariensis*). Pararetrovirus, Trim and LINE elements were also identified in all species; however they were scarce (less than 1% of the genomes). A large proportion of the elements identified as LTR remained as unclassified, mainly in species with B and F genomes in which the proportion accounted approximately 18% of the genome. Proportions of DNA transposons ranged from 2.98% (*A. rigonii*) to 8.07% (*A. glandulifera*), and they included CACTA, PIF/Harbinger, Hat, Helitron and Mule lineages. The highest representation of DNA transposons

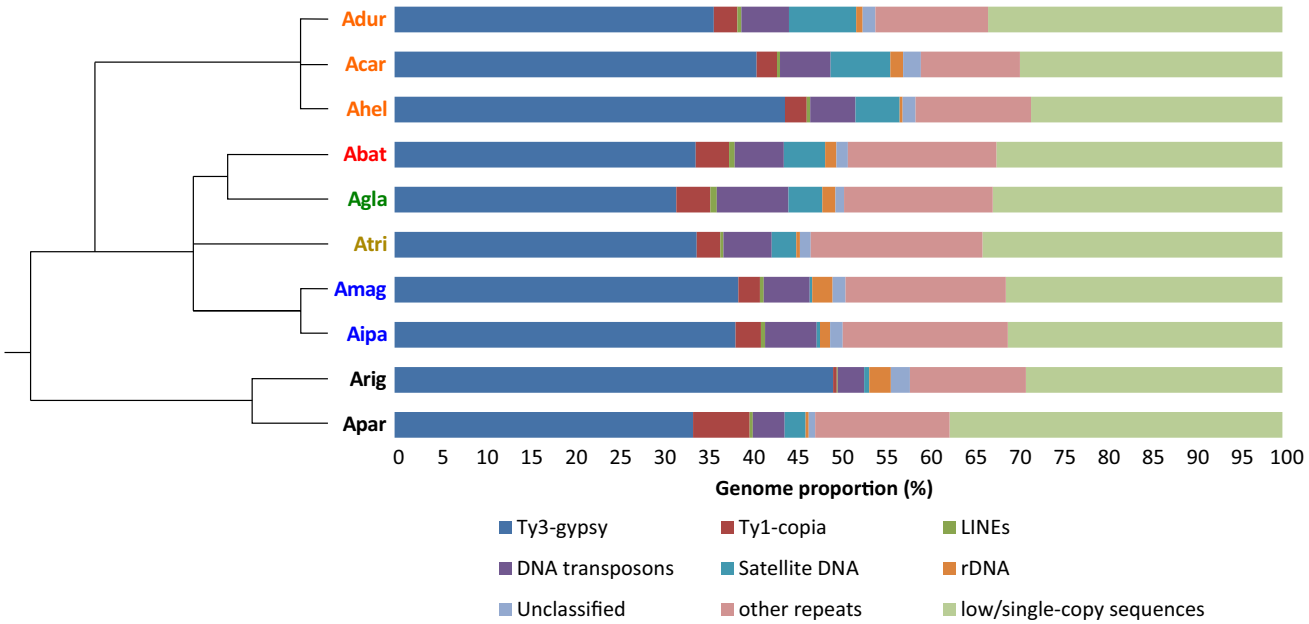


Fig. 1 Genome proportions of different classes of repetitive DNA and low/single-copy sequences identified in 10 *Arachis* species. Trim, Pararetrovirus and unclassified LTR elements are denoted as "other repeats". Species are identified with the letter codes mentioned in Table 1. Species of section *Arachis* belonging to the same genome group are highlighted with the same color (A genome: orange; B genome: blue; D genome: green; F genome: brown; and K genome: red). Species of sections *Procumbentes* and *Erectoides* are indicated in black. The schematic tree based on the ITS sequences (adapted from Fig. 1 of Friend et al. 2010) was used to display relationships among the analyzed *Arachis* species

observed in *A. glandulifera* was mainly accounted by the increase of Mule elements (5.21%), while other species of *Arachis* showed values between 1.34 and 2.40% for this element.

Satellite DNA and ribosomal DNA (rDNA) were identified as other classes of abundant sequences. The fraction of rDNA varied widely among *Arachis* species, from 0.34% (*A. helodes*) to 2.44% (*A. rigonii*). Genome proportions of satDNA repeats within section *Arachis* ranged from 2.77 to 7.56% in the species having $DAPI^+$ heterochromatic bands (*A. helodes*, *A. cardenasii*, *A. duranensis*, *A. trinitensis*, *A. glandulifera* and *A. batizocoi*). In contrast, the content of satDNA was extremely low in species with B genome, which are deprived of conspicuous heterochromatic bands (0.37% in *A. ipaënsis* and 0.29% in *A. magna*). Among the outgroups, this fraction was very low in *A. rigonii* (0.57% of the genome), but moderate in *A. paraguariensis* (2.36%).

Lineage diversity and genome proportions within Ty1-copia and Ty3-gypsy superfamilies

As outlined in Table 2, the Ty1-Copia superfamily was the most diverse among LTR retroelements, despite its lower genome proportions compared to the Ty3-Gypsy elements. Seven lineages (Ale, Bianca, SIRE, Ikeros, Ivana, Tar and Tork) were found among Ty1-Copia elements while five lineages within Ty3-Gypsy ones (Chromoviruses CRM, Tekay and Galadriel, and non-Chromoviruses OTA/Tat/Retand and OTA/Tat/Athila). Genome fractions of 3.04–6.21 and 0.02–0.21% identified as Ty3-gypsy and Ty1-copia, respectively, could not be assigned to any known lineage and were named as unclassified Ty3 and Ty1 retroelements.

Among Ty1-Copia lineages, SIRE retroelements were the most represented in all species analyzed (0.18% in *A. rigoni* to 5.06% in *A. paraguariensis*), with similar values among species of section *Arachis* (from 0.96% in *A. cardenasii* to 1.97% in *A. batizocoi*). Athila lineage was by far the most abundant Ty3-Gypsy lineage, and showed the largest variation in genome representation among the species analyzed (from 20.21% in *A. paraguariensis* to 35.44% in *A. rigoni*). Within section *Arachis*, species having A and B genomes showed the highest proportions of Athila elements (from 26.57%, in *A. ipaënsis*, to 33.39%, in *A. helodes*), compared with those observed in the F, K and D genome species (23.90%, 22% and 19.43%, respectively).

Diversity of satDNA repeats

The analysis of individual datasets from different *Arachis* species with TAREAN allowed the characterization of 44 satDNAs sequences, each one with at least 0.01% of genome proportion (Doc. S1). The number of detected

Table 2 Genome proportions of repetitive sequences identified in the genome of analyzed *Arachis* species

Repetitive DNA	*Lineage*	Ahel (A)	Acar (A)	Adur (A)	Aipa (B)	Amag (B)	Atri (F)	Abat (K)	Agla (D)	Arig	Apar
Total LTR		58.89	53.78	50.86	59.51	58.90	55.57	53.85	51.54	62.83	54.80
Ty3-gypsy		43.97	40.77	35.95	38.40	38.73	34.04	33.92	31.73	49.40	33.62
	Tekay	3.37	3.08	3.14	4.57	3.76	3.07	4.27	4.13	3.33	5.54
	Galadriel	0.05	0.05	0.10	0.07	0.09	0.11	0.21	0.13	0.02	0.07
	CRM	0.03	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.00	0.03
	OTA/Tat/Retand	3.09	3.39	2.71	2.63	2.82	2.30	3.43	3.45	4.72	1.56
	OTA/Athila	33.39	30.56	26.94	26.57	27.34	23.90	22.00	19.43	35.44	20.21
	Unclassified	4.04	3.67	3.04	4.52	4.69	4.62	3.97	4.55	5.89	6.21
Ty1-copia		2.45	2.30	2.66	2.87	2.43	2.64	3.75	3.83	0.41	6.38
	Bianca	0.52	0.42	0.50	0.52	0.48	0.50	0.70	0.66	0.09	0.45
	SIRE	1.17	0.96	1.09	1.18	1.13	1.01	1.97	1.89	0.18	5.06
	Ale	0.27	0.38	0.45	0.52	0.35	0.44	0.35	0.44	0.09	0.39
	Ikeros	0.14	0.20	0.20	0.21	0.16	0.25	0.20	0.22	0.01	0.17
	Ivana	0.02	0.03	0.04	0.02	0.02	0.04	0.03	0.02	0.00	0.04
	TAR	0.02	0.02	0.03	0.02	0.02	0.04	0.05	0.09	0.00	0.04
	Tork	0.15	0.14	0.17	0.21	0.13	0.21	0.25	0.26	0.02	0.16
	Unclassified	0.16	0.14	0.18	0.17	0.15	0.15	0.21	0.26	0.02	0.08
LTR unclassified		12.48	10.71	12.24	18.24	17.75	18.89	16.18	15.98	13.02	14.80
LINEs		0.39	0.35	0.45	0.46	0.42	0.36	0.65	0.73	0.10	0.34
Trim		0.11	0.10	0.10	0.10	0.10	0.07	0.12	0.14	0.02	0.05
Pararetrovirus		0.42	0.35	0.35	0.24	0.21	0.39	0.43	0.62	0.02	0.27
Total DNA transposons		5.11	5.70	5.37	5.82	5.17	5.44	5.53	8.07	2.98	3.59
	CACTA	3.17	3.01	2.95	2.90	2.73	2.85	2.67	2.33	0.91	1.95
	PIF/Harbinger	0.02	0.04	0.03	0.03	0.02	0.02	0.03	0.03	0.00	0.03
	hAT	0.49	0.49	0.46	0.50	0.48	0.47	0.35	0.36	0.54	0.15
	Mule	1.34	2.08	1.80	2.28	1.87	1.97	2.40	5.21	1.51	1.35
	Helitron	0.09	0.09	0.13	0.11	0.06	0.12	0.07	0.13	0.01	0.11
Satellite DNA		4.96	6.71	7.56	0.37	0.29	2.77	4.65	3.82	0.57	2.36
rDNA		0.34	1.44	0.73	1.13	2.29	0.41	1.27	1.47	2.44	0.35
Unclassified		1.46	2.01	1.43	1.43	1.46	1.20	1.28	0.98	2.15	0.74
Total repeats in top clusters		**71.68**	**70.45**	**66.85**	**69.05**	**68.83**	**66.21**	**67.78**	**67.37**	**71.11**	**62.50**

Diploid species are identified with the four-letter codes mentioned in Table 1. Letters in parentheses indicate the genome assignment of the species of section *Arachis*

satDNAs per species ranged between eight (*A. batizocoi*) and two (*A. helodes*). Sequence comparisons revealed that almost all monomeric sequences can be grouped into four main closely related groups of satDNA sequences, named as AraSat1–AraSat4 (Table S1 and Fig. S2). The Group AraSat1 is composed of monomers of *ca.* 300 bp and near 80% of AT nucleotides, while the Group AraSat2, has monomer sizes of approximately 100 bp with 50% of AT. Conspicuous differences in both monomeric length and nucleotide composition were observed in sequences of groups AraSat3 (approximately 20 bp and 60–75% AT) and AraSat4 (190 bp and 61–62% AT). Other monomeric sequences showed a wide variation in length (31–445 bp) and nucleotide composition (42–76% AT), but they could not be assigned to any of the above-mentioned groups. All satDNA groups identified in this survey were found as tandem arrays in both scaffolds and pseudomolecules of *A. hypogaea* cv. Tifrunner (Fig. S3).

The comparative analysis of genome proportions of satDNAs (Fig. 2) revealed that seven clusters (Clusters 4, 36, 48, 96, 114, 118 and 141) were shared by at least four *Arachis* species and that five (Clusters 4, 36, 48, 96 and 141) were found in all species analyzed. Hereafter, the clusters are named with the letter "C" and the respective number retrieved from RE2 (eg. C4). Among the clusters shared by all species, clusters "C4, C36 and C48" (Group AraSat-1) accounted for more than 90% of the total reads classified as satDNAs. These clusters were especially abundant in species having conspicuous bands of DAPI$^+$ heterochromatin, particularly in species with A, D, F and K genomes (section *Arachis*), and *A. paraguariensis* (section *Erectoides*). The cluster C4 was the most abundant in species with A genome (3.22–5.57%), had intermediate values in F and K genome species (1.42% and 1.83%), and it was scarce in *A. glandulifera* (0.12%). In the E genome species, *A. paraguariensis*, this cluster represented 0.39% of the genome. The clusters C36 and C48 were the most abundant in *A. glandulifera* (each one representing 1.80% of the genome), they showed intermediate values in *A. batizocoi* (approximately 1.28%), and less than 1% in species of the A and F genomes. *A. paraguariensis* showed similar values of C36 and C48 to those observed for *A. batizocoi*. The other two clusters shared by all species, C96 (AraSat3) and C141 (AraSat4), were among the less represented satDNA clusters, with genome proportions between 0.01 and 0.82%. Among the clusters of satDNAs shared by only four species, C114 (AraSat2) was detected in *A. cardenassi* (0.15%), *A. duranensis* (0.24%), *A. trinitensis* (0.34%) and *A. batizocoi* (0.05%); while the cluster C118 (AraSat4) was only found in all species of the A genome (0.04–0.15%) and in *A. rigonii* (0.34%).

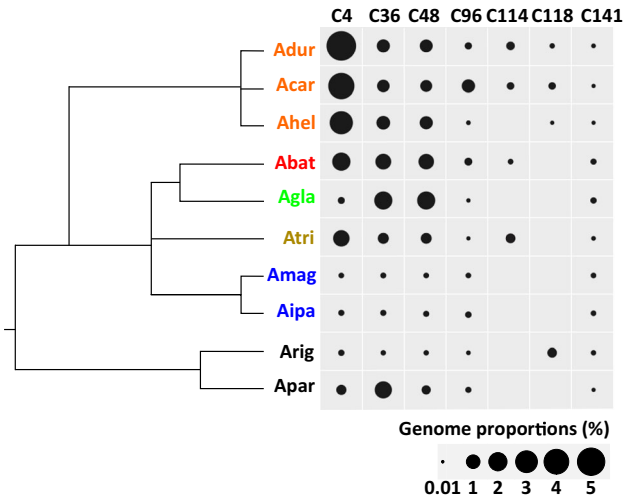


Fig. 2 Quantification of satDNA repeats obtained from the clustering analysis on a concatenated dataset of randomly selected reads of the *Arachis* genomes. The genome proportions are expressed as percentage per haploid genome. Species are identified with the letter codes mentioned in Table 1. Species of section *Arachis* belonging to the same genome group are highlighted with the same color (A genome: orange; B genome: blue; D genome: green; F genome: brown; and K genome: red). Species of sections *Procumbentes* and *Erectoides* are indicated in black. The schematic tree based on the ITS sequences (adapted from Fig. 1 of Friend et al. 2010) was used to display relationships among the analyzed *Arachis* species

Repetitive DNA and genome size

Genome sizes of diploid *Arachis* species are detailed in Table 1. Additionally, results of genome size estimations of *A. paraguariensis* and *A. rigonii* are shown in Table S2. The relationships between genome size (C*x*), heterocromatin content and proportions of different types of repetitive DNA were analyzed by regression analysis and Pearson's correlation tests (Fig. 3 and Table S3). Even though the total repetitive DNA fraction tends to be positively related with the genome sizes when all species were included in the analysis, this relationship was not statistically supported ($r = 0.167$, $p = 0.644$). However, this relationship strengthened when only annual species were considered ($r = 0.807$, $p = 0.053$).

The contribution of different classes of DNA repeats to genome size variation across *Arachis* was further investigated. A significant positive relationship was observed between the genome proportion of total LTR with genome size ($r = 0.757$, $p = 0.011$), however, this tendency was not statistically supported when any of the mayor superfamilies Ty3-gypsy, Ty1-copia or LINEs were considered separately. A negative relationship between genome size and the genome proportion of Pararetrovirus ($r = -0.752$, $p = 0.012$) was observed. Regarding to the relationship between the satDNA fraction and genome size, a negative and statistically supported relationship was observed

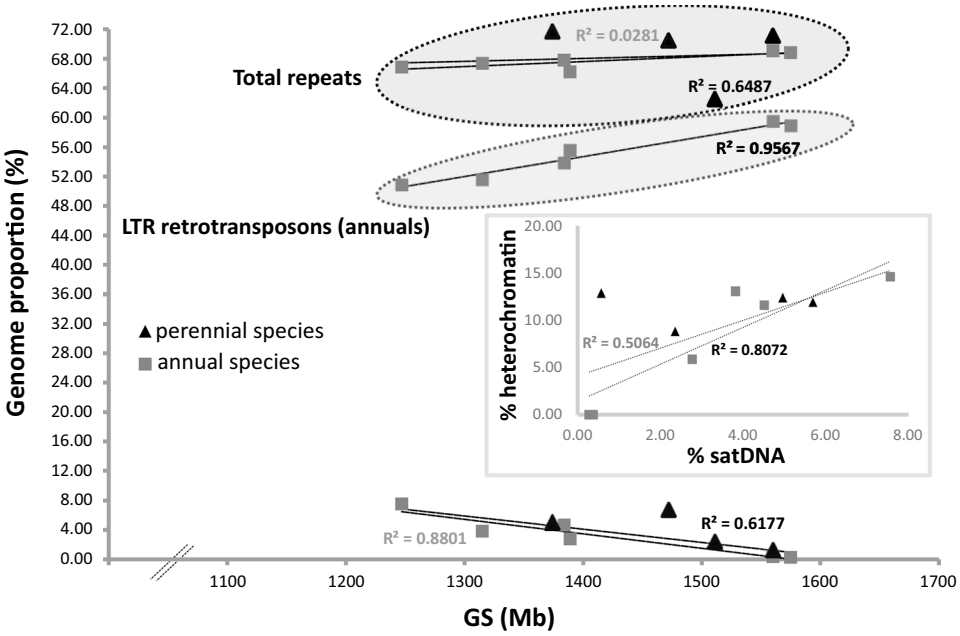


Fig. 3 Scatter plots showing the relationships among genome proportions of the total repetitive DNA fraction, LTR-retrotransposons, satDNAs, heterochromatin (expressed as percentages) and genome size (Mb/C*x*) in *Arachis* genomes. The R^2 values for annual and perennial species are indicated beside regression lines

($r = -0.786$, $p = 0.007$), which became more significant when only annual species were considered ($r = -0.938$, $p = 0.006$). Similarly, the percentage of heterochromatin, expressed as a proportion of the total chromosome length, was inversely related with genome sizes considering all ($r = -0.646$, $p = 0.044$) and only annual ($r = -0.96$, $p = 0.002$) species. The satDNA fraction showed a strong relationship with the percentage of heterochromatin, especially in the annual species of section *Arachis* ($r = 0.932$, $p = 0.007$).

Temporal dynamics of LTR-REs

The timing of retrotranspositional events was estimated for the most abundant LTR retroelements (Athila) in species representatives of different genomes (Fig. 4). The patterns of retrotranspositional waves of the Athila retroelement suggest differential times and magnitudes of activity in the evolutionary history of genomes/species within section *Arachis* and outgroups used here. The A genome species, together with *A. ipaënsis* and *A. rigonii*, showed a single peak of retrotranscriptional activity that increased from around 10 MYA to 2 MYA, when it entered on a plateau. However, in *A. cardenasii* and *A. duranensis* this activity

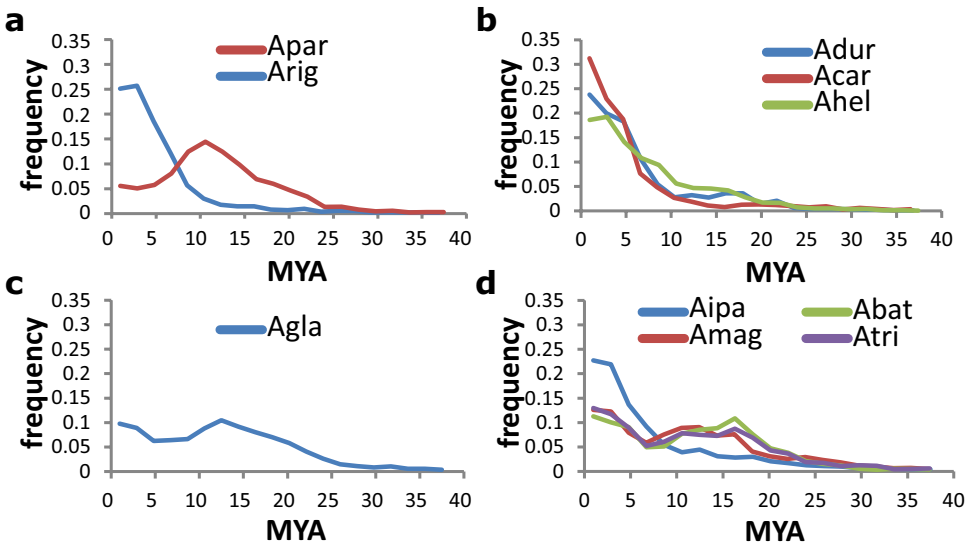


Fig. 4 Timing of the Athila retrotranspositional activity in *Arachis* species based on the divergence estimated from pairwise comparisons of Illumina reads that match RT-encoding sequences. Species are identified with the letter codes mentioned in Table 1. **a** Species of the sections *Erectoides* (*A. paraguariensis*) and *Procumbentes* (*A. rigonii*). **b** Species of the section *Arachis* with A genome. **c** Species of the section *Arachis* with D genome. **d** Species of the section *Arachis* with B, F and K genomes

seems to continue even in the last 2 MYA. The other species showed an ancient proliferation peak between 20 MYA and 7–8 MYA. Among this group of species, *A. trinitensis*, *A. batizocoi*, *A. glandulifera* and *A. magna* also had a burst of retrotranspositional activity that begun around 5 MYA and lasted around 2–3 MYA. This latter burst peak occurred in a similar time frame to that observed in the A genome species.

Phylogenetic analysis of *Arachis* species based on the genome proportions and similarities of repeats

The phylogenetic tree obtained from the genome proportions of repetitive DNA (Fig. 5) showed two well supported groups: (1) comprised the F (*A. trinitensis*) and B (*A. ipaënsis* and *A. magna*) genome species; (2) included all A genome species (*A. duranensis*, *A. helodes* and *A. cardenasii*) together with *A. rigonii* (section *Procumbentes*). The species *A. glandulifera* (D genome) and *A. batizocoi* (K genome) were not grouped with other *Arachis* species and they diverged independently as independent and more distant linages.

After the exclusion of matrices with missing similarity values, the consensus network based on the pairwise sequence similarities was done using 94 top clusters identified in RE2 (Fig. 6a). The topology was almost consistent with that obtained from the analysis of the genome proportions of the repetitive elements. *Arachis paraguariensis* diverged firstly while the remaining species were included in two main groups. One of them composed of A genome species and *A. rigonii* and another formed by species belonging to the B, D, F and K genomes.

To investigate the patterns of divergence of different repetitive linages in the *Arachis* genomes, phylogenetic analyses based on pairwise sequence similarities were separately done for Athila retroelements and the three major clusters of satDNAs (C4, C36 and C48). The consensus network of Athila elements (Fig. 6b) showed a similar topology to the consensus network obtained from the analysis of top clusters. The networks derived from the three most abundant satDNAs are shown in Fig. 6c–e. These networks showed larger differences in branch lengths and topologies among each other, and also with those derived from Athila and from the 94 top clusters. As expected, *A. paraguariensis* appeared as the most distant species in all satDNA networks. The C4 network (Fig. 6c) did not show any species/genome specific divergent cluster, instead a single lineage of sequences was observed, with *A. paraguariensis* in one of the extremes. The first species connected to *A. paraguariensis* were those of the B genome, then those of the A genome, and at the other end of the network, the species of the D, F and K genomes split in short branches. The topologies found for C36 (Fig. 6d) and C48 (Fig. 6e) networks were very similar with a clear segregation of the A genome species and *A. rigonii* from the clade of *A. magna*, *A. batizocoi* and *A. glandulifera*. However, *Arachis trinitensis* (F genome) showed an intermediate position in the two trees and *A. ipaënsis* (B genome) in the C36 one (Fig. 6d, e).

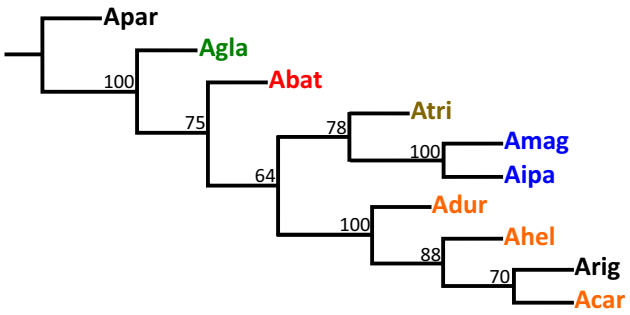


Fig. 5 Phylogenetic relationships of *Arachis* species based on the relative abundance of 231 top clusters. Species are identified with the letter codes mentioned in Table 1. Species of section *Arachis* belonging to the same genome group are highlighted with the same color (A genome: orange; B genome: blue; D genome: green; F genome: brown; and K genome: red). Species of sections *Procumbentes* and *Erectoides* are indicated in black. Numbers in the tree represent bootstrap values

Discussion

Repetitive landscape in *Arachis* species

The medium/high content of repetitive sequences here observed in *Arachis* genomes (63.44–69.80%) is consistent with the 64% previously obtained by re-association kinetics studies in the genome of the tetraploid *A. hypogaea* (Dhillon et al. 1980). It is also in the range (60% and 69%) obtained from the genome assemblies of the wild diploids *A. duranensis* (A genome) and *A. ipaënsis* (B genome) (Bertioli et al. 2016). A detailed analysis of the clusters retrieved by RE2 revealed that the largest fraction of the repetitive elements of *Arachis* accounted for dispersed repeats, mainly represented by LTR retrotransposons, as reported for other groups of plant species (Macas et al. 2015; Mascagni et al. 2018; McCann et al. 2020; Ribeiro et al. 2020). The genome proportions of these retroelements were nearly similar in all species analyzed, ranging from 52.88% (*A. helodes*) to 55.22% (*A. ipaënsis*). In spite of the fraction Ty3-gypsy was less diverse in terms of number of lineages, its genome proportion was in average 13.8 times higher than Ty1-copia. The higher representation of Ty3-gypsy compared to Ty1-copia elements is in accordance with previous records in legumes with large genomes like *Vicia*, *Lens*, *Pisum* and *Lathyrus* (Macas et al. 2015; Kreplak et al. 2019) of the Fabeae tribe.

The prevalence of particular lineages of retroelements in a genome is usually related to amplification events and activity of these elements (Feschotte et al. 2002). The highest, and at the same time variable, genomic representation of the Athila lineage suggests its preponderant role during the divergence of the genomes of section *Arachis* and probably of the whole genus, as deduced from the outgroups used here. Our results are consistent with previous analyses of Fidel, a highly abundant Athila family of retrotransposons in *Arachis*, and its non-autonomous parasite Feral (Nielen et al. 2010; Bertioli et al. 2013). Similar results were also observed in *Sthylosanthes* (Silva Oliveira et al. 2021), highlighting the preponderance of Athila elements in medium sized genome species of the early-branching Dalbergieae clade within the subfamily Faboideae. The highest and variable representation of SIRE lineage among the Ty1-copia retroelements suggests its differential rerotranspositional activity among *Arachis* species. In spite of the amount of this element remained with similar amount within section *Arachis*, its representation was five times higher in *A. paraguariensis* but very low in *A. rigonii*. Therefore, this Ty1-copia linage may have also be important in the diversification of the genomes within the genus.

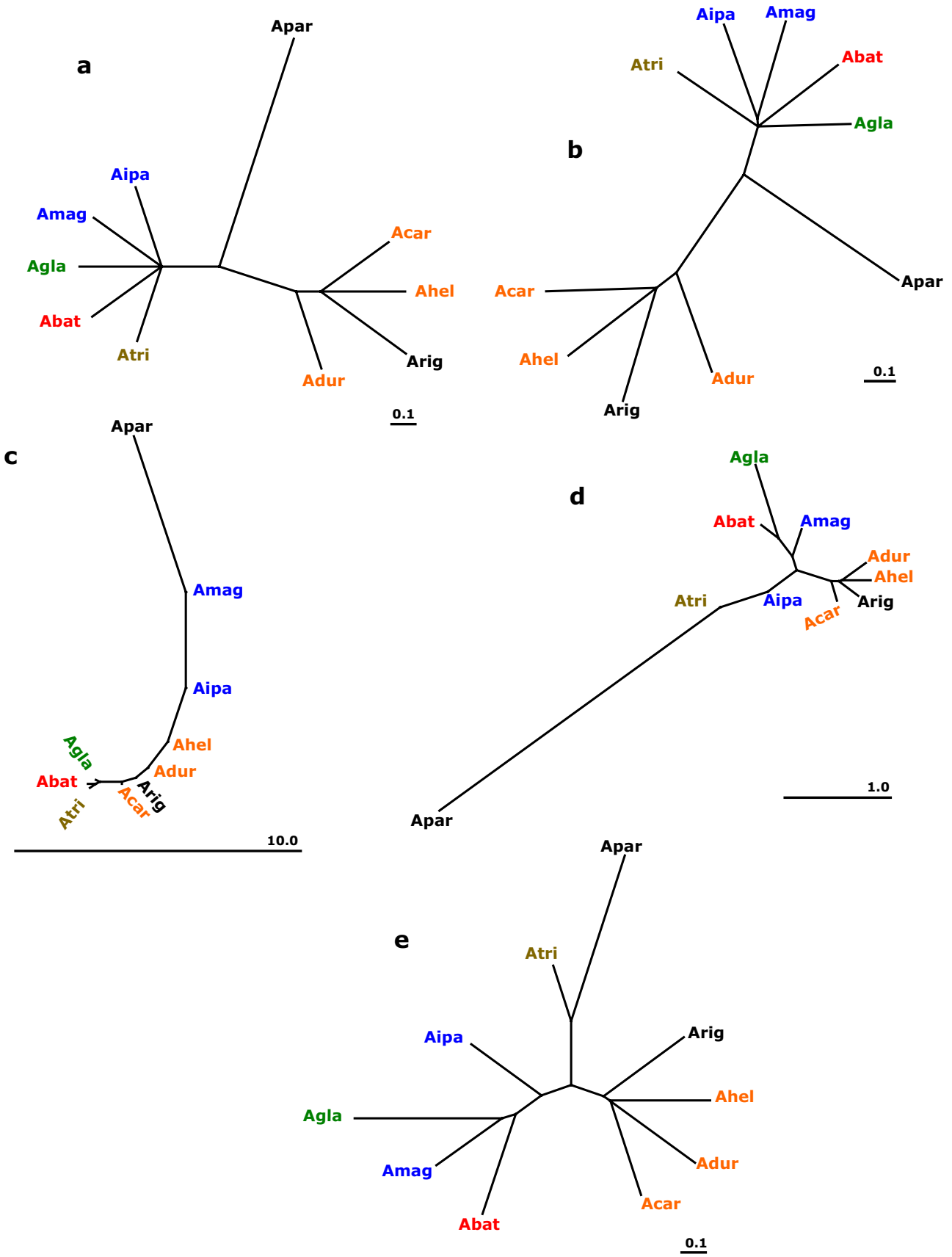
The variation in the genome proportion of the Athila lineage in the genomes of the species here analyzed suggests that this element would have contributed significantly to the genomic diversification of *Arachis*. To estimate the proliferation times of Athila retroelements in the genomes of different *Arachis* species we translated the pairwise genetic distances between RT-encoding sequences into insertion dates (Piegu et al. 2006; Ammiraju et al. 2007; Mascagni et al. 2018). Although we are aware that the translation of nucleotide distances into insertion times could be subject to errors, this method has been a useful approach to make a rough estimation of the transposition waves in different lineages of retroelements among related species (Mascagni et al. 2018). Keeping this consideration in mind, the period with the highest activity of Athila in almost all *Arachis* species (except *A. paraguariensis*) is observed from 10 to 2 MYA. The initial increase of activity was detected around the initial divergence of the genus *Arachis* estimated in 13.8 ± 1.7 MYA (Lavin et al. 2004); and the peak of activity was coincident with the beginning of genome divergence within section *Arachis*, estimated in 2.33–4.99 MYA (Moretzsohn et al. 2013; Nielen et al. 2010). The continuous increase in the retrotranspositional activity in *A. duranensis* and *A. cardenasii*, even in the last 2 MYA, suggests that this element is still modeling the repetitive landscapes of their genomes.

The similar genome proportion of DNA transposons in *Arachis* species (except in *A. glandulifera*) suggests a minor role of these elements in genome differentiation. The exception is *A. glandulifera* in which the approximately double proportion of DNA transposons was explained mainly by Mutator-like elements (MULEs). The high proportion of these elements deserves special consideration since it is well documented that DNA transposons have the ability to mediate chromosomal rearrangements by at least two possible types of mechanisms: (1) indirect by homologous recombination or (2) direct by an alternative transposition process (Gray 2000). As *A. glandulifera* is the only *Arachis* species with a highly asymmetric karyotype (Robledo and Seijo 2008) evolved through chromosome pericentric inversions and/or translocations (Stalker 1991), the accumulation of MULE transposons in this genome provides an initial framework for future investigations on the karyotype evolution of this species.

Concerning the satDNA fraction, it is widely known that numerous families of these sequences can coexist in eukaryote genomes as part of a library shared by closely related species (Plohl et al. 2008; Ugarkovic and Plohl 2002). Consistently, we found that the satDNA fraction of *Arachis* species is mainly constituted of four groups of closely related sequences of a same superfamily, which are characterized by differences in monomeric length and nucleotide composition. However, these satDNA repeats were differentially amplified in the species here analyzed leading to specific profiles for each taxa.

The karyotypes of *Arachis* species belonging to different genomes are characterized by variable patterns of constitutive $DAPI^+$ heterochromatin. The karyotypes of A, K and D genomes presents bands of variable size with a preferential localization in the pericentromeric region of most chromosomes, and the F genome has tiny bands in six-to seven chromosome pairs (Seijo et al. 2004; Robledo and Seijo 2008, 2010; Robledo et al. 2009). In contrast, B genome species lack detectable $DAPI^+$ heterochromatin bands in their karyotypes (Seijo et al. 2004; Robledo and Seijo 2010), although small dots of more condensed chromatin may appear in the centromeres after strong treatments with proteases for FISH procedures (Seijo pers. obs.). The variation in the proportions of the satDNA fraction in *Arachis* genomes from 0.29% (*A. magna*) to 7.56% (*A. duranensis*) is directly proportional to the $DAPI^+$ heterochromatin content, evidencing that the satDNAs here identified are the major components of this heterochromatin. Fluorescent in situ hybridization with some representatives of these satDNAs on chromosome complements of few diploid *Arachis* species supports this hypothesis (Samoluk et al. 2017, 2019; Zhang et al. 2012, 2016).

Among the seven shared, but differentially represented clusters (C4, C36, C48, C96, C114, C118 and C141) in *Arachis* species, the most abundant ones (C4, C36 and C48) belong to the group AraSat1 and accounted for more than 90% of the fraction. These clusters includes sequences with homology to the previously characterized sequences ATR-2 (Samoluk et al. 2017), H-b (Zhang et al. 2016) and Agla_CL8sat (Samoluk et al. 2019). These satDNAs are the main


a
Apar
Aipa
Amag
Agla
Abat
Atri
Acar
Ahel
Arig
Adur
0.1
b
Aipa
Amag
Atri
Abat
Agla
Apar
Acar
Ahel
Adur
Arig
0.1
c
Apar
Amag
Aipa
Ahel
Agla
Adur
Abat
Arig
Atri
Acar
10.0
d
Agla
Abat
Amag
Adur
Ahel
Atri
Aipa
Arig
Acar
Apar
1.0
e
Apar
Atri
Aipa
Arig
Agla
Ahel
Amag
Adur
Acar
Abat
0.1

◂**Fig. 6** Phylogenetic relationships of *Arachis* species based on repeat sequence similarities. **a** 94 top clusters identified with RE2. **b** Clusters classified as Athila retrotransposons. **c** satDNA from cluster 4 (satDNA C4). **d** satDNA from cluster 36 (satDNA C36). **e** satDNA from cluster 48 (satDNA C48). Species are identified with the letter codes mentioned in Table 1. Species of section *Arachis* belonging to the same genome group are highlighted with the same color (A genome: orange; B genome: blue; D genome: green; F genome: brown; and K genome: red). Species of sections *Procumbentes* and *Erectoides* are indicated in black

component of the DAPI$^+$ heterochromatin in the A, F, K and D genomes. However, they are barely represented in the species of the B genome, which are deprived of heterochromatic bands. Among the less represented satDNAs (C96, C114, C118 and C141), C96 (group AraSat2) showed homology with the previously characterized satDNA "clone 119" of *A. hypogaea* (Zhang et al. 2012), while the remaining ones would constitute new satDNA families. The results suggest that the high rates of turnover occurred in this superfamily of satDNAs led to the different amounts of DAPI$^+$ heterochromatin, one of the main features that modeled the genomes in section *Arachis*.

Repetitive DNA and genome size evolution in *Arachis*

The genome size (*C*-value) variation among species of a particular plant genus and ploidy level (Bennett 1987) is mainly explained by different mechanisms of expansion and contraction of repetitive DNA (Bennetzen et al. 2005). Here we investigated the quantitative changes in different fractions of repetitive DNA that may explain the *ca.* 1.26-fold genome size variation among wild *Arachis* diploids with $x = 10$ (Samoluk et al. 2015a). Considering both annual and perennial species, we found that the genome size variation tend to be determined by the differential amplification of the repetitive DNA fraction as a whole, as observed in other plant species (SanMiguel and Bennetzen 1998; Macas et al. 2015). A deeper analysis showed that genome sizes are more strongly correlated with the contents of LTR-retrotransposons, particularly when only annual species of section *Arachis* are considered.

It has been documented that recent retrotransposition bursts of LTR retrotransposons is the main mechanism of genome size increase in plant genomes (Bennetzen and Wang 2014; Zhang and Gao 2017). Accordingly, our results evidenced that the last burst of activity of Athila elements (between 5 and 2 MYA) may have been one of the main processes involved in the genome amplification of the species studied here. However, the tendency toward the obesity of the genomes is not the only one described in eukaryotes (Bennetzen and Kellog 1997). Indeed, reduction in genome size has been documented via LTR retroelements removal from host genomes through deletions and recombinations (Devos et al. 2002; Baidouri and Panaud 2013). In this context, the lowest content of LTRs found in the *Arachis* species with the smallest genome size can be interpreted as a result of an active erosion of these elements from an ancestral genome with higher content, like in the annual *A. duranensis*. From comparisons of genome assemblies it was proposed that *A. duranensis* (A genome), after its divergence from a common ancestor with *A. ipaënsis* (B genome), experienced several inversions followed by a genome size reduction through a dramatic loss of LTR retroelements in the inverted regions, likely through a mechanism of illegitimate recombination (Ren et al. 2018). This hypothesis is supported by the high proportion of solo-LTRs found in *A. duranensis* (Bertioli et al. 2016) and suggest a complex relationship between the karyotype structure and the proportion of LTRs that are maintained in each chromosomal region, which ultimately affect the genome size of the species.

The contribution of the satDNA content to the genome size variation is variable in different plant groups. The proportion of satDNAs may be directly related (Emadzade et al. 2014), inversely related (Sader et al. 2021) or unrelated (Ambrožová et al. 2011) with genome size variations. The inverse relationship observed between the proportion of satDNAs and the genome size in *Arachis* species, and the fact that the largest genomes are deprived of conspicuous heterochromatic bands, clearly denotes that this fraction cannot explain the observed genome size variation. Although satDNAs showed a wide range of variation among species, it is masked by the huge variations in the LTRs elements.

The contrasting relationships of genome sizes with the proportions of LTR elements and satDNAs suggest different ways of genome size evolution at diploid level within section *Arachis*. One of them shows a tendency towards the accumulation of satDNA sequences in genomes with medium to low genome size, while another towards the accumulation/deletion of LTRs. Our results suggest that the mechanisms involved in the accumulation/deletion of both types of sequences (LTRs and satDNA) in *Arachis* may be constrained by the genome size and its nucleotype effects (Bennet et al. 1987) in a dynamic balance. That is, satDNA amplification should have been accompanied or tolerated as long as there was a reduction in the LTRs proportions. On the contrary, an increase in the size of the genome by amplification of LTRs would imply a restriction for a concomitant amplification of satDNAs or would induce a reduction in the proportion of these tandem repeats. We propose that the evolutionary dynamics of these two repetitive fractions under genome size and chromosomal constraints may have been the principal pathways that modeled the modern genomes and main karyotype features in *Arachis* species.

Repetitive DNA reflects phylogenetic relationships in *Arachis*

The phylogenetic network based on comparative analysis of the repeatomes showed that the A genome species are well differentiated from those with other genome types within section *Arachis*. These results are in agreement with the proposed genome classification (Robledo et al. 2010; Robledo and Seijo 2010) and previous phylogenetic analysis (Bechara et al. 2010; Moretzsohn et al. 2004, 2013). However, the relative positions of the D, F and K genomes reflects the controversial relationships found in the literature. While the repeat proportion tree suggests that the F genome is closely related to the B genome, being the D and K genomes in an intermediate position between A and B-F clusters, the network based on the sequence diversity of 94 top clusters showed that B, D, F, and K genome species clustered together, separated from the A genome species. The latter result is in agreement with the previous species relationship analysis based on nuclear ribosomal DNA (ITS regions and 5.8 S; Bechara et al. 2010), Tentative Orthologous Gene (TOGs; Leal-Bertioli et al. 2015) and plastid trnT–S, trnY–T regions (Grabiele et al. 2012).

Regarding to the diploid species belonging to other sections, the repeatome data confirm the distant relationship between *A. paraguariensis* (section *Erectoides*) and the species of section *Arachis*, as it was previously observed (Bechara et al. 2010). However, the phylogenetic position of the accession annotated in the GenBank under the name of *A. rigonii* Krapov. and W.C. Gregory do not agree with the accepted taxonomic position for this species. The only one known accession of this species (GKP 10034) belongs to section *Procumbentes*, due to the horizontal growth of the peg and the highest percentage of fertility in hybrids with species of that section (Krapovickas and Gregory 1994). Most likely, this incongruence can be attributed to a mismatch between the source of the DNA sequences and the species name registered in the GenBank. The extremely low proportion of satDNAs observed here for this species is weird, and not expected for a species with medium size heterochromatic DAPI$^+$ bands in all or most of their chromosomes, such as *A. rigonii* (Ortiz et al. 2017). Since the original passport data and/or an international voucher specimen were not provided in Yin et al. (2017), who generated the Illumina sequences (SRR13188218), further analysis with a properly documented material of this species are necessary to shed light into its relationships with species of section *Arachis*.

Different repetitive sequences within a genome can provide variable levels of phylogenetic signal, therefore affecting tree topology and resolution (Dodsworth et al. 2016; Weiss-Schneeweiss et al. 2015). While the phylogenetic networks obtained from the sequence diversity of top clusters and from Athila retroelements have the same topology, the networks obtained from satDNAs (C4, C36 and C48) show some discrepancies. In the satDNA C4 tree, the sequences of the F, K and D genomes are closer to those of the A genome than to those of the B genome. The sequences of *A. ipaënsis* and *A. magna* were the closest to *A. paraguariensis* (in one edge) and showed larger relative distances from each other than the distances observed among the other species. The sequences of the D, F and K genomes showed the shortest distances at the other edge of the network, close to those of the A genome, suggesting a recent amplification of a few elements of the C4 lineage in these genomes. The trees of satDNAs C36 and C48 were similar and revealed a higher similarity among the sequences of the K, D and B genomes compared to those of the A genome. This finding is in agreement with the topologies published based on both nuclear (Leal-Bertioli et al. 2015) and plastid (Grabiele et al. 2012) DNA sequences. The intermediate position of *A. trinitensis* suggests that these sequences would have had a less evolutionary dynamic than those present in the other species of section *Arachis*. This fact is probably related to the small sizes of heterochromatin blocks observed in the species with F genome (Robledo and Seijo 2010). If the topology revealed by the similarity of sequences is accepted, the coincidence in the amounts of heterochromatin among K, D and A genomes would suggest a parallel amplification of the different versions of the same family of satDNAs in these genomes. Furthermore, considering a similar mutation rate for the three elements, the distances suggest that the amplifications of the different elements would have occurred at different evolutionary times. For satDNAs C36 and C48, the divergence would have occurred at an initial moment of genomic divergence within section *Arachis* (probably even earlier in C48). By contrast, the satDNA C4 network suggests a recent divergence of motifs within the K, D, F genomes with higher similarity to those in the A genome. With these results, we propose that the parallel amplification of different satDNAs in the K, D, F and A genome species may explain the high similarity in the heterochromatic pattern of these species, especially those of the A and K genomes.

Author contribution statement Project planning and coordination: SSS, GS and GR. DNA library and whole genome sequencing: SSS and DJB. Genome size estimations: AMO. Bioinformatic analysis of repetitive DNA and temporal dynamics of retrotransposons: SSS and MV. Phylogenetic analysis of repeats: SSS and LMIC. Manuscript writing: SSS and GS. All authors read and approved the final version of the manuscript.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00425-022-03961-9.

Acknowledgements This work was supported by Grants (PICT 2017-3220; PICT 2018-3664) of Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. We thank United States Department of Agriculture (USDA) for providing seeds of some accessions and Center for Applied Genetic Technologies, University of Georgia (Athens, GA, USA), for supporting the genome sequencing.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Alioto T, Alexiou KG, Bardil A et al (2020) Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. Plant J 101:455–472

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Ambrožová K, Mandáková T, Bureš P, Neumann P, Leitch IJ, Koblížková A, Macas J, Lysak MA (2011) Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. Ann Bot 107:255–268

Ammiraju JS, Zuccolo A, Yu Y, Song X, Piegu B, Chevalier F, Walling JG, Ma J, Talag J, Brar DS, SanMiguel PJ, Jiang N, Jackson SA, Panaud O, Wing RA (2007) Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. Plant J 52:342–351

Bechara MD, Moretzsohn MC, Palmieri DA, Monteiro JP, Bacci M, Martins J, Valls JFM, Lopes CR, Gimenes MA (2010) Phylogenetic relationships in genus Arachis based on ITS and 5.8 S rDNA sequences. BMC Plant Biol 10:255

Bennett MD (1987) Variation in genomic form in plants and its ecological implications. New Phytol 106:177–200

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genomic obesity? Plant Cell 9:1509–1514

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530

Bertioli DJ, Vidigal B, Nielen S, Ratnaparkhe MB, Lee TH, Leal-Bertioli SCM, Kim C, Guimarães PM, Seijo G, Schwarzacher T, Paterson AH, Heslop-Harrison P, Araujo ACG (2013) The repetitive component of the A genome of peanut (*Arachis hypogaea*) and its role in remodelling intergenic sequence space since its evolutionary divergence from the B genome. Ann Bot 112:545–559

Bertioli DJ, Cannon SB, Froenicke L et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 8:438–446

Bertioli DJ, Jenkins J, Clevenger J et al (2019) The genome of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet 51:877–884

Biémont C, Vieira C (2006) Junk DNA as an evolutionary force. Nature 443:521–524

Biscotti MA, Olmo E, Heslop-Harrison JP (2015) Repetitive DNA in eukaryotic genomes. Chromosome Res 23:415–420

Bolsheva NL, Melnikova NV, Kirov IV, Dmitriev AA, Krasnov GS, Amosova AV, Samatadze TE, Yurkevich OY, Zoshchuk SA, Kudryavtseva AV, Muravenko OV (2019) Characterization of repeated DNA sequences in genomes of blue-flowered flax. BMC Evol Biol 19:49

Chenais B, Caruso A, Hiard S, Casse N (2012) The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene 509:7–15

Darzentas N (2010) Circoletto: visualizing sequence similarity with Circos. Bioinformatics 26:2620–2621

Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079

Dhillon SS, Rake AV, Miksche JP (1980) Reassociation kinetics and cytophotometric characterisation of peanut (*Arachis hypogaea* L.) DNA. Plant Physiol 65:1121–1127

Di Rienzo JA, Casanoves F, Balzarini MG, González L, Tablada M, Robledo CW (2013) InfoStat version 2013. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. http://www.infostat.com.ar

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-Schneeweiss H, Leitch AR (2014) Genomic repeat abundances contain phylogenetic signal. Syst Biol 64:112–126

Dodsworth S, Chase MW, Särkinen T, Knapp S, Leitch AR (2016) Using genomic repeats for phylogenomics: a case study in wild tomatoes (*Solanum* section Lycopersicon: Solanaceae). Biol J Linn Soc 117:96–105

Dodsworth S, Jang TS, Struebig M, Chase MW, Weiss-Schneeweiss H, Leitch AR (2017) Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Plant Syst Evol 303:1013–1020

Dogan M, Pouch M, Mandáková T, Hloušková P, Guo X, Winter P, Chumová Z, Van Niekerk A, Mummenhoff K, Al-Shehbaz IA, Mucina L, Lysak MA (2021) Evolution of tandem repeats is mirroring post-polyploid cladogenesis in *Heliophila* (Brassicaceae). Front Plant Sci 11:1944

Doležel J, Bartos J (2005) Plant DNA flow cytometry and estimation of nuclear genome size. Ann Bot 95:99–110

El Baidouri M, Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol Evol 5:954–965

Emadzade K, Jang TS, Macas J, Kovařík A, Novák P, Parker J, Weiss-Schneeweiss H (2014) Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). Ann Bot 114:1597–1608

Ewing AD (2015) Transposable element detection from whole genome sequence data. Mob DNA 6:24

Fernandez A, Krapovickas A (1994) Cromosomas y evolucion en *Arachis* (Leguminosae). Bonplandia 8:187–220

Ferree PM, Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. PLoS Biol 7:e1000234

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329–341

Fransz P, Linc G, Lee CR, Aflitos SA et al (2016) Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. Plant J 88:159–178

Friend SA, Quandt D, Tallury SP, Stalker HT, HiluK W (2010) Species, genomes, and section relationships in the genus *Arachis* (Fabaceae): a molecular phylogeny. Plant Syst Evol 290:185–199

Gaiero P, Vaio M, Peters SA, Schranz ME, de Jong H, Speranza PR (2019) Comparative analysis of repetitive sequences among species from the potato and the tomato clades. Ann Bot 123:521–532

Garrido-Ramos MA (2015) Satellite DNA in plants: more than just rubbish. Cytogenet Genome Res 146:153–170

Goloboff PA, Farris JS, Nixon K (2008) TNT, a free program for phylogenetic analysis. Cladistics 24:774–786

Gowda MVC, Bhat RS, Sujay V, Kusuma P, Bhat S, Varshney RK (2011) Characterization of AhMITE1 transposition and its

association with the mutational and evolutionary origin of botanical types in peanut (*Arachis* spp.). Plant Syst Evol 291:153–158
Grabiele M, Chalup L, Robledo G, Seijo G (2012) Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences. Plant Syst Evol 298:1151–1165
Gray YH (2000) It takes two transposons to tango. Trends Genet 16:461–468
Gregory MP, Gregory WC (1979) Exotic germoplasm of *Arachis* L. interspecific hybrids. J Hered 70:185–193
Greilhuber J, Lysák M, Doležel J, Bennett MD (2005) The origin, evolution and proposed stabilization of the terms 'genome size', and 'C-value' to describe nuclear DNA contents. Ann Bot 95:255–260
Holland B, Moulton V (2003) Consensus networks: a method for visualising incompatibilities in collections of trees. In: Benson G, Page RDM (eds) International workshop on algorithms in bioinformatics. Springer, Berlin, pp 165–176
Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267
Jurka J, Bao W, Kojima KK (2011) Families of transposable elements, population structure and the origin of species. Biol Direct 6:44
Jurka J, Bao W, Kojima KK, Kohany O, Yurka MG (2012) Distinct groups of repetitive families preserved in mammals correspond to different periods of regulatory innovations in vertebrates. Biol Direct 7:36
Katoh K, Misasa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acids Res 30:3059–3066
Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12:656–664
Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120
Krapovickas A, Gregory WC (1994) Taxonomy of the genus *Arachis* (Leguminosae). Bonplandia 8:1–186
Kreplak J, Madoui MA, Cápal P et al (2019) A reference genome for pea provides insight into legume genome evolution. Nat Genet 51:1411–1422
Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532
Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
Lavin M, Schrire BD, Lewis GP, Pennington RT, Delgado-Salinas A, Thulin M, Hughes CE, Beyra-Matos A, Wojciechowski MF (2004) Metacommunity processes rather than continental tectonic history better explain geographically structured phylogenies in legumes. Philos Trans R Soc Lond Biol Sci 359:1509–1522
Leal-Bertioli SCM, Santos SP, Dantas KM, Inglis PW, Nielen S, Araujo ACG, Silva JP, Cavalcante U, Guimarães PM, Brasileiro ACM, Carrasquilla-Garcia N, Penmetsa RV, Cook D, Moretzsohn MC, Bertioli DJ (2015) *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea*) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Ann Bot 115:237–249
Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core *Triticeae* genome. Plant J 40:500–511
López-Flores I, Garrido-Ramos MA (2012) The repetitive DNA content of eukaryotic genomes. Genome Dyn 7:1–28
Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410
Macas J, Novák P, Pellicer J, CíŽková J, KoblíŽková A, Neumann P, Fuková I, Doležel J, Kelly LJ, Leitch IJ (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
Mascagni F, Usai G, Natali L, Cavallini A, Giordani T (2018) A comparison of methods for LTR-retrotransposon insertion time profiling in the *Populus trichocarpa* genome. Caryologia 71:85–92
Mascagni F, Vangelisti A, Giordani T, Cavallini A, Natali L (2020) A computational comparative study of the repetitive DNA in the genus *Quercus* L. Tree Genet Genomes 16:1–11
McCann J, Macas J, Novák P, Stuessy TF, Villaseñor JL, Weiss-Schneeweiss H (2020) Differential genome size and repetitive DNA evolution in diploid species of *Melampodium* sect. Melampodium (Asteraceae). Front Plant Sci 11:362
Mehrotra S, Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genom Proteom Bioinform 12:164–171
Moretzsohn MC, Hopkins MS, Mitchell SE, Kresovich S, Valls JFM, Ferreira ME (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biol 4:11
Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ (2013) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann Bot 111:113–126
Negm S, Greenberg A, Larracuente AM, Sproul JS (2021) RepeatProfiler: a pipeline for visualization and comparative analysis of repetitive DNA profiles. Mol Ecol Resour 21:969–981
Nelson MG, Linheiro RS, Bergman CM (2017) McClintock: an integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. G3: Genes Genom Genet 7:2763–2778
Nielen S, Campos-Fonseca F, Leal-Bertioli SCM, Guimaraes PM, Seijo G, Town C, Arrial R, Bertioli D (2010) FIDEL—a retrovirus-like retrotransposon and its distinct evolutionary histories in the A and B-genome components of cultivated peanut. Chromosome Res 18:227–246
Nielen S, Vidigal B, Leal-Bertioli SCM, Ratnaparkhe M, Paterson A, Garsmeur O, D'Hont A, Guimaraes P, Bertioli D (2012) Matita, new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis* A–B genome divergence. Mol Genet Genomics 287:21–38
Noé L, Kucherov G (2005) YASS: enhancing the sensitivity of DNA similarity search. Nucl AcIds Res 33:540–543
Novák P, Robledillo LA, Koblızková A, Vrbová I, Neumann P, Macas J (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucl Acids Res 45:e111
Novák P, Neumann P, Macas J (2020) Global analysis of repetitive DNA from unassembled sequence reads using repeatexplorer2. Nat Protoc 15:3745–3776
Ortiz AM, Robledo G, Seijo G, Valls JFM, Lavia GI (2017) Cytogenetic evidences on the evolutionary relationships between the tetraploids of the section Rhizomatosae and related diploid species (*Arachis*, Leguminosae). J Plant Res 130:791–807
Paradis E, Schliep K (2018) Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528
Patel M, Jung S, Moore K, Powell G, Ainsworth C, Abbott A (2004) High-oleate peanut mutants result from a MITE insertion into the FAD2 gene. Theor Appl Genet 108:1492–1502
Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: dynamics of retrotransposition driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res 16:1262–1269

Plohl M, Luchetti A, Mestrovic N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. Gene 409:72–82

Plohl M, Meštrović N, Mravinac B (2012) Satellite DNA evolution. Genome Dyn 7:126–152

Plohl M, Meštrović N, Mravinac B (2014) Centromere identity from the DNA point of view. Chromosoma 123:313–325

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Ren L, Huang W, Cannon EKS, Bertioli DJ, Cannon SB (2018) A mechanism for genome size reduction following genomic rearrangements. Front Genet 9:454–454

Ribeiro T, Vasconcelos E, Dos Santos KGB, Vaio M, Brasileiro-Vidal AC, Pedrosa-Harand A (2020) Diversity of repetitive sequences within compact genomes of *Phaseolus* L. beans and allied genera *Cajanus* L. and *Vigna* Savi. Chromosome Res 28:139–153

Robledillo LA, Koblížkova A, Petr N, Böttinger K, Vrbová I, Neumann P, Schubert I, Macas J (2018) Satellite DNA in *Vicia faba* is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. Sci Rep 8:5838

Robledo G, Seijo JG (2008) Characterization of *Arachis* D genome using physical mapping of heterochromatic regions and rDNA loci by FISH. Genet Mol Biol 31:717–724

Robledo G, Seijo G (2010) Species relationships among the wild B genome of *Arachis* species (section *Arachis*) based on FISH mapping of rDNA loci and heterochromatin detection: a new proposal for genome arrangement. Theor App Genet 121:1033–1046

Robledo G, Lavia GI, Seijo G (2009) Species relations among wild *Arachis* species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. Theor App Genet 118:1295–1307

Sader M, Vaio M, Cauz-Santos LA, Dornelas MC, Vieira MLC, Melo N, Pedrosa-Harand A (2021) Large vs small genomes in *Passiflora*: the influence of the mobilome and the satellitome. Planta 253:1–18

Samoluk SS, Chalup LMI, Robledo G, Seijo JG (2015a) Genome sizes in diploid and allopolyploid *Arachis* L. species (section *Arachis*). Genet Res Crop Evol 62:747–763

Samoluk SS, Robledo G, Podio M, Chalup L, Ortiz JPA, Pessino SC, Seijo JG (2015b) First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. Genetica 143:113–125

Samoluk SS, Robledo G, Bertioli D, Seijo JG (2017) Evolutionary dynamics of an at-rich satellite DNA and its contribution to karyotype differentiation in wild diploid *Arachis* species. Mol Genet Genomics 292:283–296

Samoluk SS, Chalup LMI, Chavarro C, Robledo G, Bertioli DJ, Jackson SA, Seijo G (2019) Heterochromatin evolution in *Arachis* investigated through genome-wide analysis of repetitive DNA. Planta 249:1405–1415

San Miguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 82:37–44

Santana SH, Valls JF (2015) *Arachis veigae* (Fabaceae), the most dispersed wild species of the genus, and yet taxonomically overlooked. Bonplandia 24:139–150

Seijo G, Lavia GI, Fernández A, Krapovickas A, Ducasse D, Moscone EA (2004) Physical mapping of 5S and 18S–25S rRNA genes evidences that *Arachis duranensis* and *A. ipaensis* are the wild diploid species involved in the origin of *A. hypogaea* (Leguminosae). Am J Bot 91:2293–2303

Seijo G, Lavia GI, Fernández A, Krapovickas A, Ducasse D, Bertioli DJ, Moscone EA (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*-Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971

Seijo JG, Kovalsky IE, Chalup LMI, Samoluk SS, Fávero A, Robledo G (2018) Karyotype stability and genome specific nucleolar dominance in peanut, its wild 4× ancestor and in a synthetic AABB polyploidy. Crop Sci 58:1671–1683

Seijo GJ, Atahuachi M, Simpson CE, Krapovickas A (2021) *Arachis inflata*. Bonplandia 30:169–174

Shirasawa K, Koilkonda P, Aoki K, Hirakawa H, Tabata S, Watanabe M, Hasegawa M, Kiyoshima H, Suzuki S, Kuwata C, Naito Y, Kuboyama T, Nakaya A, Sasamoto S, Watanabe A, Kato M, Kawashima K, Kishida Y, Kohara M, Kurabayashi A, Takahashi C, Tsuruoka H, Wada T, Isobe S (2012) In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol 12:80

Silva Oliveira MA, Nunes T, Dos Santos MA, Ferreira Gomes D, Costa I, Van-Lume B, Marques da Silva SS, Oliveira RS, Simon M, Lima GS, Gissi DS, Almeida CC, Souza G, Marques A (2021) High-throughput genomic data reveal complex phylogenetic relationships in *Stylosanthes* Sw (Leguminosae). Front Genet 12:727314

Silvestri MC, Ortiz AM, Lavia GI (2014) rDNA loci and heterochromatin positions support a distinct genome type for '$x = 9$ species' of section *Arachis* (*Arachis*, Leguminosae). Plant Syst Evol 301:555–562

Smartt J, Gregory WC, Gregory MP (1978) The genomes of *Arachis hypogaea*. 1. Cytogenetic studies of putative genome donors. Euphytica 27:665–675

Stalker HT (1991) A new species-section *Arachis* of peanuts with D genome. Am J Bot 78:630–637

Stalker HT, Dhesi JS, Parry DC, Hahn JH (1991) Cytological and interfertility relationships of *Arachis* section *Arachis*. Am J Bot 78:238–246

Staton SE, Burke JM (2015) Evolutionary transitions in the Asteraceae coincide with marked shifts in transposable element abundance. BMC Genomics 16:623

Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles-causes and effects. EMBO J 2:5955–5959

Usai G, Mascagni F, Natali L, Giordani T, Cavallini A (2017) Comparative genome-wide analysis of repetitive DNA in the genus *Populus* L. Tree Genet Genomes 13:96

Vaio M, Mazzella C, Porro V, Speranza P, López-Carro B, Estramil E, Folle GA (2007) Nuclear DNA content in allopolyploid species and synthetic hybrids in the grass genus *Paspalum*. Plant Syst Evol 265:109–121

Valls JFM, Simpson CE (2005) New species of *Arachis* from Brazil, Paraguay, and Bolivia. Bonplandia 14:35–64

Valls JFM, Simpson CE (2017) A new species of *Arachis* (Fabaceae) from Mato Grosso, Brazil, related to *Arachis matiensis*. Bonplandia 26:143–149

Valls JFM, Da Costa LC, Custodio AR (2013) A novel trifoliolate species of *Arachis* (Fabaceae) and further comments on the taxonomic section Trierectoides. Bonplandia 22:91–97

Vitales D, Garcia S, Dodsworth S (2020) Reconstructing phylogenetic relationships based on repeat sequence similarities. Mol Phylogenet Evol 147:106766

Wang J, Li Y, Li C, Yan C, Zhao X, Yuan C, Sun Q, Shi C, Shan S (2019) Twelve complete chloroplast genomes of wild peanuts: great genetic resources and a better understanding of *Arachis* phylogeny. BMC Plant Biol 19:1–18

Weiss-Schneeweiss H, Leitch AR, McCann J, Jang TS, Macas J (2015) Employing next generation sequencing to explore the repeat landscape of the plant genome. Next generation sequencing in plant systematics. Regnum Veget 157:155–179

Woo TH, Hong TH, Kim SS, Chung WH, Kang HJ, Kim CB, Seo JM (2007) Repeatome: a database for repeat element comparative analysis in human and chimpanzee. Genomics Inform 5:179–187

Yin D, Wang Y, Zhang X, Ma X, He X, Zhang J (2017) Development of chloroplast genome resources for peanut (*Arachis hypogaea* L.) and other species of *Arachis*. Sci Rep 7:1–11

Zhang QJ, Gao LZ (2017) Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* species. G3: Genes Genom Genet 7:1875–1885

Zhang L, Xu C, Yu W (2012) Cloning and characterization of chromosomal markers from a Cot-1 library of peanut (*Arachis hypogaea* L.). Cytogenet Genome Res 137:31–41

Zhang L, Yang X, Tian L, Chen L, Yu W (2016) Identification of peanut (*Arachis hypogaea*) chromosomes using a fluorescence in situ hybridization system reveals multiple hybridization events during tetraploid peanut formation. New Phytol 211:1424–1439