Planta
https://doi.org/10.1007/s00425-019-03096-4

ORIGINAL ARTICLE

Heterochromatin evolution in *Arachis* investigated through genome-wide analysis of repetitive DNA

Sergio S. Samoluk[1] · Laura M. I. Chalup[1] · Carolina Chavarro[2] · Germán Robledo[1,3] · David J. Bertioli[2] · Scott A. Jackson[2] · Guillermo Seijo[1,3]

Received: 30 October 2018 / Accepted: 18 January 2019

Abstract

Main conclusion **The most conspicuous difference among chromosomes and genomes in *Arachis* species, the patterns of heterochromatin, was mainly modeled by differential amplification of different members of one superfamily of satellite DNAs.**

Divergence in repetitive DNA is a primary driving force for genome and chromosome evolution. Section Arachis is karyotypically diverse and has six different genomes. *Arachis glandulifera* (D genome) has the most asymmetric karyotype and the highest reproductive isolation compared to the well-known A and B genome species. These features make *A. glandulifera* an interesting model species for studying the main repetitive components that accompanied the genome and chromosome diversification in the section. Here, we performed a genome-wide analysis of repetitive sequences in *A. glandulifera* and investigated the chromosome distribution of the identified satellite DNA sequences (satDNAs). LTR retroelements, mainly the Ty3-gypsy families "Fidel/Feral" and "Pipoka/Pipa", were the most represented. Comparative analyses with the A and B genomes showed that many of the previously described transposable elements (TEs) were differently represented in the D genome, and that this variation accompanied changes in DNA content. In addition, four major satDNAs were characterized. Agla_CL8sat was the major component of pericentromeric heterochromatin, while Agla_CL39sat, Agla_CL69sat, and Agla_CL122sat were found in heterochromatic and/or euchromatic regions. Even though Agla_CL8sat belong to a different family than that of the major satDNA (ATR-2) found in the heterochromatin of the A, K, and F genomes, both satDNAs are members of the same superfamily. This finding suggests that closely related satDNAs of an ancestral library were differentially amplified leading to the major changes in the heterochromatin patterns that accompanied the karyotype and genome differentiation in *Arachis*.

Keywords Repetitive sequences · Chromosome structure · Genome evolution · Satellite DNA · *Arachis* species

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00425-019-03096-4) contains supplementary material, which is available to authorized users.

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

Laura M. I. Chalup
laura_cien@hotmail.com

Carolina Chavarro
mcch@uga.edu

Germán Robledo
germanrobledo71@gmail.com

David J. Bertioli
bertioli@uga.edu

Scott A. Jackson
sjackson@uga.edu

Guillermo Seijo
jgseijo@yahoo.com

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

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

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

Published online: 24 January 2019

Abbreviations

satDNA	Satellite DNA
FISH	Fluorescent in situ hybridization
rDNA	Ribosomal DNA
LTR	Long terminal repeat
NGS	Next-generation sequencing
TEs	Transposable elements

Introduction

Plant genomes are composed of a high proportion of repetitive DNA, which are largely diverse even among closely related species (Hemleben et al. 2007; Heslop-Harrison and Schwarzacher 2011; Mehrotra and Goyal 2014; Biscotti et al. 2015). Although the role of the repetitive DNAs is not fully understood, it has been shown that speciation events are consistently accompanied by salient fluctuations of these components (Charlesworth et al. 1994; Kidwell and Lisch 2000; Hemleben et al. 2007). The diverse functional impact of repeated sequences and their intrinsic contribution to genomic plasticity suggests that this fraction plays an important role in the genome diversification and, ultimately, can contribute to the divergence of species (Presgraves 2010; Ferree and Prasad 2012).

Among repetitive elements, long terminal repeat (LTR) retrotransposons constitute the major part of repetitive DNA of plant genomes and contribute substantially to the genome size variation among species (SanMiguel and Bennetzen 1998; Schnable et al. 2009). Moreover, mobile elements are one of the key factors accelerating eukaryotic evolution (Naito et al. 2009), since they can transpose into genes or their flanking regions, resulting in the disruption or promotion of gene expression (Shirasawa et al. 2012), induction of chromosome re-arrangements (Raskina et al. 2008), and promoting the loss of colinearity of genomes (Buchmann et al. 2012), among other effects. Tandem repeats are also important components of the repetitive fraction of genomes and several reports have documented that they have a central role in gene expression and heterochromatin assembly (Martienssen 2003; Sharma and Raina 2005; Kloc and Martienssen 2008; Pezer et al. 2012). In general, tandem repeats evolve quickly and, hence, can vary in nucleotide composition, genome representation and chromosomal distribution, affecting genome structure and/or genome sizes (Plohl et al. 2008).

The genus *Arachis* includes more than 80 species native to South America (Krapovickas and Gregory 1994; Valls and Simpson 2005, 2017; Valls et al. 2013; Santana and Valls 2015). The 31 species accepted within section Arachis are the most diverse and economically important, since they constitute the secondary gene pool of peanut (Burow et al. 2001; Simpson 2001; Mallikarjuna 2002; Mallikarjuna et al. 2004). Twenty-nine of these species are wild diploids ($2n=2x=20$, 18) and were assigned to the A, B, D, F, G, and K genomes (Robledo and Seijo 2008, 2010; Robledo et al. 2009, Silvestri et al. 2014). The other two, *A. monticola* and the cultivated peanut *A. hypogaea* are allopolyploids ($2n=4x=40$) with an AABB genome constitution (Seijo et al. 2004).

There are only a few molecular and cytogenetic information on the repetitive fraction of *Arachis*, mostly on LTR and non-LTR retrotransposons (Nielen et al. 2010, 2012; Samoluk et al. 2015b), DNA transposons (Gowda et al. 2010, 2011; Patel et al. 2004; Shirasawa et al. 2012), and satDNA sequences (Zhang et al. 2012, 2016; Samoluk et al. 2017), but they were mostly done in the cultivated peanut and its wild diploid progenitors, *A. duranensis* (A genome) and *A. ipaënsis* (B genome). These reports suggested that repetitive sequences have been a driving force for A and B genome differentiation. The recent description of the genomes of the diploid progenitors of the cultivated peanut provided a detailed characterization of the TEs that composed their repetitive fractions (Bertioli et al. 2016; Chen et al. 2016; Lu et al. 2018), but a little information was delivered about tandem repeat elements. In general, the data generated so far have not been suitable for analysis of the evolutionary dynamics of the repetitive fraction as a whole, and especially for the satellite fraction, both at chromosome and genomic levels in diploid *Arachis* species.

Arachis glandulifera is the only species with 'D genome'. It has the most distinct karyotype (Stalker 1991; Fernandez and Krapovickas 1994) and the largest proportion of heterochromatin of section Arachis (Robledo and Seijo 2008). Moreover, it is reproductively isolated from the other diploid species of the section (Stalker 1991; Robledo and Seijo 2008). For these reasons, the analysis of its genome composition and chromosome structure is essential for understanding the genome evolution in the section Arachis.

Here, we analyze the repeatome of *A. glandulifera* to provide insights into the repetitive DNA of a distant species of those so far analyzed. We characterize the major DNA repeats of the *A. glandulifera* genome using a low-depth Illumina sequencing and computational analyses. Moreover, the chromosome distribution of rDNAs and satDNAs is investigated by FISH. This work delivers the first comprehensive repeatome of a wild diploid *Arachis* species, and it is compared with those available for representative species of the A and B genomes.

Materials and methods

Plant material

For genome sequencing, the genomic DNA was isolated from young fresh leaves of *A. glandulifera* GKSSc30099

of plants growing in greenhouses at the Center for Applied Genetic Technologies at the University of Georgia (Athens, GA, USA). Hybridization experiments were made using chromosome preparations obtained from root tips of *A. glandulifera* SeSN2990 and SeSN3263, from the *Arachis* germplasm collection at the Instituto de Botánica del Nordeste, Corrientes, Argentina.

DNA extraction and whole-genome sequencing

Total DNA was extracted using a DNA extraction kit (Qiagen, Hilden, Germany) and a library of *A. glandulifera* was prepared using the Illumina TruSeq DNA Library Prep kit- 600 bp insert size (Illumina Inc.). This library was sequenced on Illumina MiSeq platform using the reagent kit v3-600 cycle (cat. MS-102-3003) to obtain 2 × 300 bp paired-end reads.

Identification of repetitive sequences

The genome-wide analysis of repetitive sequences was conducted using the first version of the software RepeatExplorer (Novak et al. 2010, 2013), implemented within the Galaxy environment (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/). From low coverage sequencing data, repeat sequences were identified and quantified using a graph-based clustering analysis. The resulting clusters were classified by similarity searches against the Conserved Domain Database for the functional annotation of proteins (Marchler-Bauer et al. 2011), Repbase Update (Jurka et al. 2005), and a custom library containing repetitive sequences already annotated for *Arachis* species (PeanutBase, https://www.peanutbase.org/). The proportion of the repetitive fraction in the *A. glandulifera* genome was calculated by the number of clustered reads versus the total amount of reads used for the analysis. Similarly, the contribution of each repeated sequence to the total genome size was estimated as the proportion of reads in each cluster compared to the total amount of reads used.

For the analysis of satDNAs, those clusters with star-like and circular graphical representation were selected. Derived contigs were aligned and putative monomers were detected using Tandem Repeats Finder (Benson 1999) and Dot-plots to characterize the putative monomeric sequences. The logos of different monomeric sequences were created from multiple sequence alignments using WebLogo (Crooks et al. 2004). To investigate possible homology among the characterized satDNAs, we used the criterion proposed by Ruiz-Ruano et al. (2016), which consider that monomeric sequences showing from 50 to 80% of identity belong to different families of the same superfamily of satDNAs, sequences with more than 80% of identity are variants of the same family, and those showing identity higher than 95% are considered the same monomeric variant.

BLAST was used to make sequence comparisons against the public database NCBI and the reference genomes of *Arachis* species deposited in the PeanutBase (https://www.peanutbase.org/).

PCR amplifications

Amplifications of satDNA sequences were performed using specific primers (Table 1) in a final volume of 10 µl in the presence of 50 ng of total genomic DNA, 1 µl 10 × Taq buffer, 2 mM each primer, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, and 0.5 U of Taq polymerase (Promega). After an initial denaturation step at 94 °C for 5 min, amplifications were conducted for 35 cycles at 30 s denaturation at 94 °C, 1 min annealing at 40–54 °C (depending on the primer pairs), and 1 min elongation at 72 °C, with a final elongation step at 72 °C for 5 min.

Synthesis of labeled probes

Clusters identified as satDNA sequences (Agla_CL8sat, Agla_CL39sat, Agla_CL69sat, and Agla_CL122sat) were considered for developing FISH probes. Probe labeling of monomeric sequences was made by PCR (using the same amplification conditions described above) using biotin-11-dUTP (Roche, Mannheim, Germany) or digoxigenin-11-dUTP (Roche, Mannheim, Germany). Due to the high percentage of similarity (89% of identity over 100% of the query length) between Agla_CL39sat and ATR-2 satDNA, a probe derived from the AhyTR-2c6 clone isolated from *A. hypogaea* (Samoluk et al. 2017) was used for FISH experiments. The chromosome distributions of the 5S and 18S-26S rDNAs were also analyzed. For this purpose, the probes of

Table 1 List of primer pairs used to amplify the different tandem repeats in *A. glandulifera*

Primer name	Nucleotide sequence	Tm (°C)
Agla_CL8-fwd	GATCTGTTGTTTACTAGATAMTGC TGG	52.1
Agla_CL8-rev	ATTCTTAATTTGTGAACAAGA GCAGC	54.5
Agla_CL39-fwd[a]	AAAATAGGATTTCAAAAGACAAG	46.7
Agla_CL39-rev[a]	TTTTTATTCATCGGATAGTGC	46.8
Agla_CL69-fwd	TGAGANATGAYATACTTGGAC	41.6
Agla_CL69-rev	CTTTRAGATCATAAAARCTATC	40.4
Agla_CL122-fwd	GTCCGCAGTAGTATAAAACTT TGG	53.8
Agla_CL122-rev	ATTAACTTAGTTGTTGTGTGA TGCC	54.0

fwd forward primer, *rev* reverse primer, *Tm* melting temperature

[a]Primers used to amplify AhyTR-2c6 clone isolated from *A. hypogaea* (Samoluk et al. 2017)

the rRNA genes (5S and 18-26S) of *A. hypogaea* (Robledo and Seijo 2008) were used to detect the ribosomal loci as comparative chromosome landmarks.

Preparations of mitotic chromosomes

Root tips were pretreated with 2 mM 8-hydroxyquinoline for 3 h at room temperature (24 °C) and then fixed in 3:1 absolute ethanol:glacial acetic acid. Somatic chromosome spreads were prepared using the SteamDrop method (Kirov et al. 2014).

FISH experiments

Hybridization experiments were performed according to Seijo et al. (2004) with hybridization mixes containing DNA probes with an approximate concentration of 2.5 ng/µl. The hybridization and post-hybridization washes were carried out with more than 80% stringency. The first set of antibodies consisted of goat anti-biotin (B3640, Sigma-Aldrich) and mouse anti-digoxigenin (F3523, Sigma-Aldrich). The second sets of antibodies included the anti-goat conjugated to TRITC produced in rabbit (T6028, Sigma-Aldrich) and the anti-mouse conjugated to FITC produced in sheep (2883, Sigma-Aldrich). Chromosomes were counterstained with 2 mg/ml of 4′-6-diamidino-2-phenylindole (DAPI) in Vectashield anti-fade (Vector Laboratories, USA).

Microscopy and image processing

Slides of hybridized mitotic chromosomes were photographed with a Leica DMRX epi-fluorescence microscope (Leica, Heerbrugg, Switzerland) coupled to a computer-assisted Leica DC 350 digital camera system. Red, green, and blue images were captured in black and white using IM 1000 Leica software. Finally, the images were merged and pseudocolored. The chromosome pairs were named according to Robledo and Seijo (2008).

Results

Global genome composition

A total of 1,311,216 paired-end reads (covering approximately 0.20× of the genome) were used as input for similarity-based clustering analysis, with a minimum overlap of 55% and a similarity of 90%. From this, 856,474 reads (65.32%) were grouped in clusters, while the remaining 454,742 reads were classified as singletons.

Based on the cluster shapes and sequence similarity searches, clusters were additionally classified in different kind and families of repetitive sequences (Fig. 1 and Suppl. Table S1). We found that the genome of *A. glandulifera* is comprised by 54.19% of TEs, 4.21% of rDNAs and 3.29% satDNAs. A small fraction (3.64%) of the clustered reads was not clearly associated with any kind of repetitive sequences and they were referred as unclassified repeats.

Characterization of TEs

The LTR retrotransposons were the most abundant repetitive sequences, accounting for 50.99% of the genome of *A. glandulifera*. Among them, the Ty3-gypsy superfamily was the main constituent with 46.04%, while the Ty1-copia superfamily was much less represented (3.98% of the genome). Twenty families of the Ty3-gypsy elements were identified (Suppl. Table S1) of which "Fidel/Feral" and "Pipoka/Pipa" were the most abundant, accounting for 15.36% and 14.29%, respectively. Twelve families of Ty1-copia group were identified (Suppl. Table S1), among which, the "RE128" was the most abundant one (2.20%). A small amount (0.95%) of the LTR group was assigned to other retroelements (unclassified, trim, and pararetrovirus).

Non-LTR retrotransposons were poorly represented in *A. glandulifera*, accounting for only 1.32% of the genome (Suppl. Table S1). LINEs accounted for 1.28% as compared to SINEs at 0.04%. DNA transposons (including TIRs and Helitrons) represented only 1.87% (Suppl. Table S1). This fraction included elements belonging to the CACTA, Harbinger, hAT, Helitron, and MULE superfamilies. The latter was the most represented among DNA transposons (0.86%).

The comparative analysis of the genomic proportions of the main classes of TEs among *A. glandulifera*, *A. duranensis,* and *A. ipaënsis* is shown in the Suppl. Table S2.

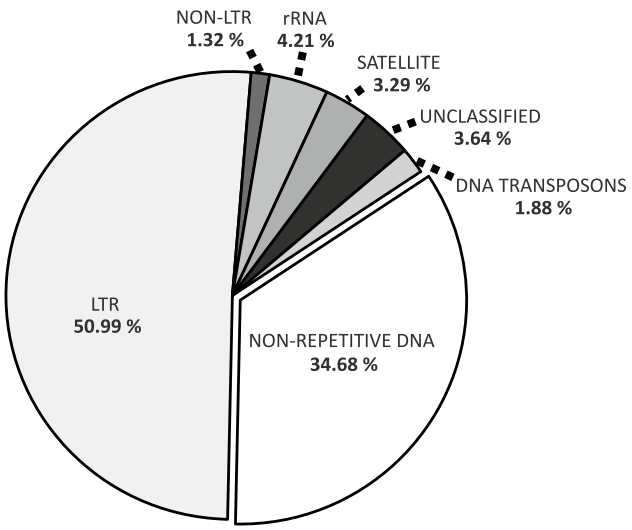

Fig. 1 Proportions (expressed as percentages) of non-repetitive and different kinds of repetitive DNA sequences in the genome of *A. glandulifera*, based on the graph-based clustering analysis

Characterization of rDNAs and satDNAs

The output of Repeat Explorer showed five clusters with the typical star-like or circular graph (Suppl. Fig. S1). Four of them (CL8, CL39, CL69, and CL122) corresponded to different satDNA families (hereafter, referred as Agla_CL8sat, Agla_CL39sat, Agla_CL69sat, and Agla_CL122sat). The remaining cluster, CL21, corresponded to the 5S rDNA (hereafter referred as Agla_CL21-5S). Sequence logos of monomeric units of the individual satDNAs and 5S rDNA are shown in the Suppl. Fig. S2. Representative repetitive units of the four satDNA families (Agla_CL8sat, Agla_CL39sat, Agla_CL69sat, and Agla_CL122sat) and the Agla_CL21-5S were deposited in NCBI GenBank (MK111638–MK111642).

The rDNA sequences represented 4.21% of the genome of *A. glandulifera*. The reads belonging 18–26S rDNA represented 3.85% of the genome, while those of the 5S rDNA constituted only 0.36%. The alignment with the other well-described plant 5S rRNA genes (Waminal et al. 2014; Cloix et al. 2000) helped to delimit the boundaries of the coding region and the intragenic promoter of Agla_CL21-5S (Suppl. Fig. S3). With a length of 305 bp, the Agla_CL21-5S repetitive unit contains an intergenic spacer of 216-bp.

Comprising approximately 3.11% of the genome, Agla_CL8sat was the most abundant satDNA of *A. glandulifera*. The monomeric sequence had 359 bp with 82% A + T content. The comparison of the consensus sequence against public databases revealed an identity of 70–75% (over 61–71% of the query length) with the previously published ATR-2 satDNA (GenBank KX685960–KX686015) and 72–75% of identity (over 58–62% of the query length) with the H-b-Ad and H-b-Ai satDNAs (GenBank KF957859 and KF957860, respectively).

The second most abundant satDNA was Agla_CL39sat, with 0.11% of the genome. As observed for Agla_CL8sat, its 330 bp monomeric consensus sequence had high percentage of A + T nucleotides, 88%. Searches in public databases revealed an identity of 81–94% (over 72% of the query length) with ATR-2 satDNA and of 90–93% (over 100% of the query length) with H-b satDNA. The comparison between Agla_CL8sat and Agla_CL39sat showed 75% of similarity.

The other two satDNA sequences, Agla_CL69sat and Agla_CL122sat, were less represented in the genome. The former had monomeric units of 155 bp and with 63% of A + T content and it accounted for 0.05% of the genome. The latter had a consensus monomeric unit of 209 bp, 70% of A + T nucleotides, and it constituted 0.02% of the genome. The comparisons of these two satDNAs against the NCBI database did not show any similarity with the other previously characterized satDNAs; however, both monomeric sequences had similarity hits to genomic sequences of *A. duranensis*, *A. ipaënsis,* and *A. hypogaea* in the PeanutBase (https://www.peanutbase.org/).

FISH experiments

To gain more insights into chromosome organization and sequence composition of the heterochromatin in this D genome species, mitotic chromosomes of *A. glandulifera* were in situ hybridized using probes of the four identified satDNA families (Agla_CL8sat, Agla_CL39sat, Agla_CL69sat, and Agla_CL122sat), together with the ribosomal genes (5S and 18-26S) as chromosomal landmarks. Results of hybridization experiments and the relative positions of hybridization loci on the karyotype are shown in Figs. 2 and 3, respectively.

In situ hybridization using rDNA probes revealed one 5S rDNA loci in the subterminal position on the long arms of the D5 chromosomes (Fig. 2b) and three conspicuous 18-26S rDNA loci: one located along the secondary constrictions of the D2 chromosomes and the remaining two on the long arms of the subtelocentric D9 and D10 chromosomes (Fig. 2e). The 18–26S rDNA sites on chromosomes D1 and D6 detected by Robledo and Seijo (2008) in the accession Se3263 were very tiny and rarely observed here. Therefore, only the three constant hybridization signals were represented in Fig. 3.

Seven chromosome pairs (D1, D2, D4, and D7–D10) showed fluorescent signals with the Agla_CL8sat, Agla_CL39sat, Agla_CL69sat, and Agla_CL122sat probes, while D3, D5, and D6 did not evidence any detectable hybridization signals of satDNAs.

Agla_CL8sat probe revealed fluorescent signals on most of the $DAPI^+$ heterochromatic bands (Fig. 2c). The hybridization signals were detected on the pericentromeric bands of D1, D2, and D7–D10 chromosomes, and on the interstitial bands of long arms of D7 chromosomes. The size and strength of the fluorescent signals corresponded to the size of the heterochromatic bands.

The probe of Agla_CL39sat (sequence AhyTR-2c6) showed a total of three pairs of hybridization signals. Two of them on the D9 chromosomes, one pair on $DAPI^+$ pericentromeric bands and the other on distal regions of the long arms; the third in the interstitial location on the long arms of D7 chromosomes (Fig. 2d). All signals were of similar size and strength. The non-pericentromeric signals of D7 chromosomes overlapped with the interstitial $DAPI^+$ heterochromatic bands, while those of D9 chromosomes mapped on the euchromatin region.

Hybridization signals of the Agla_CL69sat probe were observed in the four subtelocentric (D7, D8, D9, and D10) chromosome pairs (Fig. 2e). All the signals were small and of similar in size and strength, and showed different locations. Distal hybridization signals were observed on the

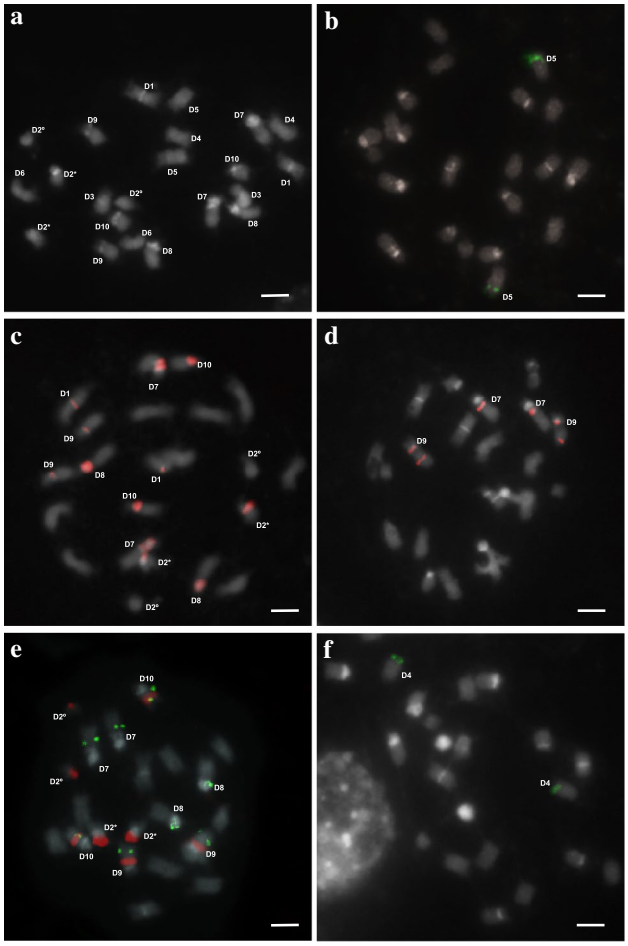

Fig. 2 Fluorescent in situ hybridizations of tandem repeats on *A. glandulifera* chromosomes. Metaphase chromosomes stained with DAPI are pseudocolored in gray, whereas green fluorescein isothiocyanate (FITC) and red tetramethyl-rhodamine isothiocyanate (TRITC) signals indicate hybridization signals of tandem repeats. **a** Metaphase stained with DAPI. **b** 5S rDNA loci (green). **c** Agla_CL8sat (red). **d** Agla_CL39sat (red). **e** Agla_CL69sat (green) and 18–26S rDNA (red). **f** Agla_CL122sat (green). The chromosomes are identified according to the nomenclature of Robledo and Seijo (2008). The short arm and the proximal segment of the long arm of pair D2 are indicated by an asterisk, and the separated satellite is marked by a degree sign. Bars represent 3 µm

euchromatin of the long arms of the D9 chromosomes, adjacent to signals for Agla_CL39sat. Interstitial signals were detected on the long arms of D7 and D10 pairs. On pair D7, the hybridization signals of this satDNA mapped to euchromatic regions of the long chromosome arms. The signals on the D10 pair co-localized with the hybridization signals of one inactive 18-26S rDNA loci. The pericentromeric signals mapped within the hybridization sites of Agla_CL8sat, on the DAPI$^+$ heterochromatin of D8 pair.

The less represented family Agla_CL122sat was exclusive of the metacentric pair D4. This satDNA produced weak hybridization signals revealed in subterminal position of the long arms of this chromosome pair (Fig. 2f).

Discussion

The characterization of the repetitive DNA has become key for the understanding, at a broad scale, the organization, and the structure of plant genomes (Biemont and Vieira 2006). However, since repeat sequences are extremely variable both in abundance and sequence length (Hemleben et al. 2007; Heslop-Harrison and Schwarzacher 2011), the analysis of this fraction is still a great challenge given technical constraints of the traditional methods. By contrast, high-throughput in silico analysis of NGS data has transformed the study of repetitive DNA. The impact of these technologies has been particularly important for species still lacking a reference genome, opening new opportunities for studying the organization and evolution of plant genomes (Barghini et al. 2014; Heitkam et al. 2015; Iwata-Otsubo et al. 2016; Sveinsson et al. 2013; Sousa et al. 2017; Kirov et al. 2017). Here, we provide a detailed composition of the repeatome of *A. glandulifera* through the analysis of Illumina paired-end reads to identify repetitive sequences with those reported previously for the peanut diploid progenitors.

Composition of the *A. glandulifera* repeatome

Using a graph-based clustering approach in *A. glandulifera,* we found that 65.32% of its 1315 Mbp (1C value) genome (Samoluk et al. 2015a) was composed of medium/highly repetitive sequences. This proportion was similar to that reported for *A. hypogaea* (60%) in the pioneering study of Dhillon et al. (1980) and to those reported for two wild diploid *Arachis* species through the analysis of whole-genome assemblies, in which the repetitive fraction was 60–69% and 74–75% of *A. duranensis* (1C = 1247 Mbp) and *A. ipaënsis* (1C = 1560 Mbp), respectively (Bertioli et al. 2016; Chen et al. 2016; Lu et al. 2018). Considering that the non-repetitive fraction is highly conserved among diploid species (Bertioli et al. 2016), our results provide additional evidences that genome size variation among *Arachis* genomes is the outcome of the differential accumulation and elimination of repetitive DNA sequences.

Like in many studied plant genomes, LTR retroelements are the dominant repetitive fraction in *A. glandulifera*, comprising 50.99% of its genome. This proportion is higher than those estimated for *A. duranensis* (44.74%) and *A. ipaënsis* (45.72%) by whole-genome assemblies (Bertioli et al. 2016). However, these estimations are significantly different compared to those found in the draft genomes of *A. duranensis* (40.07%; Chen et al. 2016) and *A. ipaënsis* (64.15%; Lu et al. 2018). Despite the absolute values published for the different

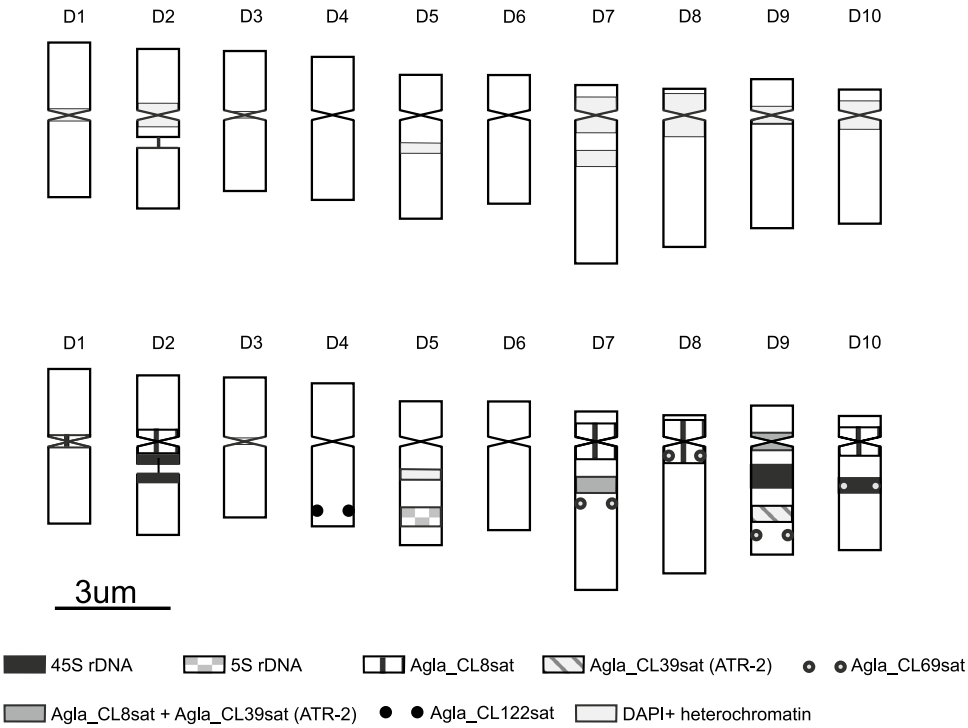

Fig. 3 Schematic representation of *A. glandulifera* chromosomes according to the distribution of the DAPI$^+$ heterochromatic bands, the rDNA loci (5S and 18–26S) and the satDNA sequences (Agla_CL8sat, Agla_CL39sat, Agla_CL69sat, and Agla_CL122sat). Somatic chromosomes were ordered by morphology and size

species, it is clear that the proportion of LTR fraction is directly related to the genome size and that those elements have been major players in evolutionary dynamics of the genome size among *Arachis* species.

Among the LTR retroelements, the superfamilies Ty3-gypsy and Ty1-copia were differentially represented in the D genome. Ty3-gypsy elements (46.04%) were more than 11 times higher than the estimated for Ty1-copia (4.01%). This clear prevalence of Ty3-gypsy over the Ty1-copia group was also observed in A and B genome *Arachis* species (Bertioli et al. 2016; Chen et al. 2016; Lu et al. 2018). Large-scale comparative analysis of repeats in different species from four genera (*Vicia, Lathyrus, Lens,* and *Pisum*) of the legume tribe Fabeae revealed that the Ty3-gypsy group is the main contributor to the genome size, with genome representations much higher than that observed for Ty1-copia retroelements (Macas et al. 2015). Thus, the high prevalence of Ty3-gypsy over Ty1-copia seems to be a general characteristic for legume species.

Considering the 20 families identified within the Ty3-gypsy superfamily, a few autonomous and their non-autonomous retrotransposon were very abundant, while the others were poorly represented in *A. glandulifera*. The high representativeness of "Fidel/Feral" (15.36%) was similar to that estimated for *A. duranensis* (16.8%) by Bertioli et al. (2013), but the estimated value of "Pipoka/Pipa" (14.29%), "Gordo" (6.09%), and "Curu/Bravo/Golf/Hemera/Hera" (3.34%) were almost twice than those published for A genome species. Inversely, "Grilo/Gilo" and "Mico" families were poorly represented in both species (less than 7.3%). All the Ty1-copia families were scarcely represented in the D genome, excepting "RE128" family that represented almost half (2.20%) of the amount of these elements. The low representation of Ty1-copia elements was similar to that reported for *A. duranensis* (Bertioli et al. 2013).

Among the non-LTR retrotransposons of *A. glandulifera*, the representation of LINEs (1.28%) was significantly lower than that reported by Bertioli et al. (2016) for *A. duranensis* and *A. ipaënsis* (7.77% and 11.73%, respectively), but it was between the values reported by Chen et al. (2016) and Lu et al. (2018) (1.26% and 2.97% for *A. duranensis* and *A. ipaënsis*, respectively). The representation of SINEs (0.04%) was similar to those reported by Chen et al. (2016) and Lu et al. (2018) (0.01% and 0.05% of the genomes in *A. duranensis* and *A. ipaënsis*, respectively), but slightly lower compared to those reported by Bertioli et al. (2016) in *A. duranensis* and *A. ipaënsis* (0.14 and 0.13% of the genome, respectively). Irrespective of the absolute values, the representation of LINEs is relatively high in *Arachis* compared to the other legume species (Macas et al. 2015). In spite of the representation of LINEs which was directly related to the genome size in *Arachis* species (Samoluk et al. 2015b), our results suggest that the impact of non-LTRs in the genome size changes would have been less important than that of LTR elements.

The representation of DNA transposons found in the genome of *A. glandulifera* was almost three-to-six times

lower than that reported for A and B species of *Arachis* (Bertioli et al. 2016; Chen et al. 2016; Lu et al. 2018). However, the composition of this fraction in the D genome at the superfamily level followed the overall pattern found in A and B genomes; with higher representation of TIRs over the Helitrons. Among the former, the high representation of MULEs and CACTAs is comparable to the pattern found in *A. duranensis* (Bertioli et al. 2016).

The impact of the activity of TEs, as a whole, on *Arachis* genomes was directly demonstrated by the comparison of homeologous regions in the A and B genomes (Bertioli et al. 2013), showing that the microsyntenic regions are flanked by completely different repetitive DNA components. At a different scale, the magnitude of the activity of these elements in the differentiation of the euchromatic regions was clearly evidenced in GISH experiments in the AABB allotetraploids, since the chromosomes of each chromosome complement showed very limited cross-hybridization (Seijo et al. 2007, 2018). Moreover, the analysis of a few TEs in FISH experiments showed preferential hybridization on either the A or B genome complements of the allopolyploids (Nielen et al. 2010; Bertioli et al. 2013; do Nascimento et al. 2018). All these comparative analyses, together with the results provided here, evidence that only a few families of LTR retrotransposons have been the major contributors to the differentiation of the euchromatin in *Arachis* species.

rDNAs, satDNAs, and heterochromatin in *A. glandulifera*

The rDNA fraction was composed of the 18–26S (3.85%) and 5S (0.36%) rRNA genes. The 18–26S rDNA mapped constantly in three loci (D2, D9, and D10) and in complete coincidence with the CMA^+ heterochromatic bands (data not shown). However, the loci located on D2 pair were the only ones observed to be active, with extended secondary constrictions as previously reported (Robledo and Seijo 2008). It is well known that all the transcriptionally inactive genes in the active NORs and those in the inactive loci remain as CMA^+ heterochromatin (Guerra 2000; Preuss et al. 2008). Accordingly, the arrangement of 18–26S rRNA genes in heterochromatin (partially in the active loci and totally the inactive ones) completely explains the pattern of CMA^+ heterochromatic bands observed in *A. glandulifera*. As expected from the previous FISH analysis (Robledo and Seijo 2008), the 5S rDNA sequences mapped in unique loci not related to heterochromatic bands. The genome proportions here estimated for both rDNAs genes through bioinformatic approaches are in accordance to the number and size of the loci observed for each rDNA family in the karyotype.

Arachis glandulifera is one of the species with highest percentage of heterochromatin of section *Arachis*, detected as $DAPI^+$ blocks mainly localized in the pericentromeric regions of subtelocentric chromosomes (Robledo and Seijo 2008). Considering that tandem repeats are usually localized in the heterochromatic regions of chromosomes (Hemleben et al. 2007), special attention was dedicated to this genomic fraction. In agreement with the amount of $DAPI^+$ heterochromatin detected in the karyotype of this species (13.09% of the total chromosome length), the graph-based clustering analysis indicated that tandem repeat sequences were, after LTR retrotransposons, one of the most abundant sequences in the genome.

The conspicuous $DAPI^+$ heterochromatic bands of *A. glandulifera* are located in the centromeres of six chromosome pairs and in the interstitial regions of two chromosome pairs. The hybridization of three of the four satDNAs here characterized on most of the heterochromatic bands suggests that these families are the major components of the $DAPI^+$ heterochromatin.

The satDNA fraction of *A. glandulifera* includes four major families, defined by their sequence and length. The families Agla_CL8sat and Agla_CL39sat had the longer monomeric units (330 bp and 359 bp, respectively), while the families Agla_CL69sat and Agla_CL122sat showed the shorter ones (130 bp and 209 bp, respectively). According to the criterion proposed by Ruiz-Ruano et al. (2016), which classify the satDNA families based on their sequence similarities, Agla_CL8sat and Agla_CL39sat are different families of the same superfamily of satDNAs. While Agla_CL8sat is a new described satDNA family for *Arachis* species, Agla_CL39sat belongs to the same family of ATR-2 (Samoluk et al. 2017) and H-b (Zhang et al. 2016), with around 90% of nucleotide similarity. The sequences Agla_CL69sat and Agla_CL122sat constitute new satDNAs which have not yet been described for the *Arachis* species.

The clustering results showed that Agla_CL8sat was by far the most abundant satDNA (3.11% of the genome size). It has been hypothesized that the most abundant satellite repeat of any particular genome, generally localizes at the centromere and constitutes the binding site for the centromeric histone variant CENH3 (Melters et al. 2013). Our FISH results are in accordance with this hypothesis, since Agla_CL8sat arrays hybridized to most of the $DAPI^+$ pericentromeric heterochromatin.

The Agla_CL39sat was the second most abundant satDNA identified in the genome of *A. glandulifera* with 0.11% of the genome fraction. Out of the three pairs of hybridization signals, two pairs located on the pericentromeric and interstitial $DAPI^+$ heterochromatic bands of D9 and D7 pairs, respectively, co-localizing with the hybridization sites of Agla_CL8sat. The first interpretation of this result may be attributed to cross-hybridization. However, because of the high stringency conditions used in these FISH experiments, the cross-hybridization would be insignificant.

Under this assumption, the co-localization of these probes may be indicating that at least two different families of satDNA compose the heterochromatin in those bands.

Despite their low genome representations, Agla_CL69sat and Agla_CL122sat constitute new markers for chromosome identification. Agla_CL122sat was the less represented satDNA family (0.02% of the genome) identified in *A. glandulifera*. The exclusive hybridization of Agla_CL122sat in the subterminal region of D4 pair allowed its unequivocal identification from the D3 pair. The family Agla_CL69sat (0.05% of the genome) was particular, since, aside of hybridizing on one DAPI$^+$ heterochromatic band (pair D8) and on one CMA$^+$ heterochromatic band (pair D10); it showed hybridization signals on the euchromatic regions of two chromosomes pairs (D7 and D9). The partial co-localization of Agla_CL69sat with the 18-26S rDNA hybridization sites in D10 pair needs a special consideration. The lack of significant similarity of Agla_CL69 against 18-26S rDNA reads and related clusters in *A. glandulifera* suggests that this satDNA family did not evolved from rDNA sequences, as it was reported (for other satDNAs) in *Solanum lycopersicum* (Jo et al. 2009) and *Vicia sativa* (Macas et al. 2003). Thus, the presence of Agla_CL69sat within the 18–26S rDNA loci in D10 pair should be interpreted as a satDNA cluster of independent origin inserted within those NORs. Hybridization experiments, either on extended DNA fibers or on pachytene chromosomes, are needed to confirm this hypothesis.

The most notable karyotype difference among the genomes of the section Arachis is the relative percentage and distribution of DAPI$^+$ heterochromatin (Seijo et al. 2004; Robledo and Seijo 2008, 2010; Robledo et al. 2009). The species of the A, K, and D genome species have conspicuous DAPI$^+$ pericentromeric bands in most chromosomes; those of the F genome have tiny pericentromeric bands in six-to-seven chromosome pairs, while those of the B genome are deprived of large blocks of pericentromeric heterochromatin. The ATR-2 satDNA family is composed most of the heterochromatin in the A, F, and K genomes, but it is very low represented in the B and D genomes (Samoluk et al. 2017). Here, we found that the heterochromatin in the D genome was mainly composed of Agla_CL8sat. Both satDNA families are members of the same superfamily of sequences. These results suggest that different members of one superfamily of satDNAs were differentially amplified from an ancestral set of sequences, leading to the heterochromatin diversification at chromosome and genome levels in species of the section Arachis. This finding is completely compatible with the library model, which establishes that a common set of satDNA families shared by closely related species may show a differential spreading across the genomes due expansions and/or contractions of different variants from the library (Fry and Salser 1977; Plohl et al. 2008). The results now available for *Arachis* species suggest that the replacing of the main satDNA families in the heterochromatic regions occurred at very high rates, since the genome divergence estimated in less than 3 million years ago (Moretzsohn et al. 2014, Bertioli et al. 2016).

Conclusion

This is the first survey of the repetitive fraction of a wild *Arachis* species not involved in the origin of peanut. All major types of repeats were characterized in the wild diploid *A. glandulifera* ($2n = 2x = 20$, D genome). The satDNA families were used to develop chromosome markers, which revealed new insights into the genome organization of *Arachis* species. Altogether, the results presented herein provide a fundamental understanding of the chromosome structure in *A. glandulifera* that can be helpful for further studies of genome and karyotype evolution in *Arachis* species.

Author contribution statement SSS design of experiments and bioinformatic analysis, PCR amplifications, probe labeling, and writing the manuscript; LC and GR FISH experiments; CC library preparation and whole-genome sequencing; DJB, SAJ, and GS data analysis and writing the manuscript. All authors read and approved the manuscript.

Acknowledgements The authors gratefully acknowledge the financial support from the Agencia Nacional de Promoción Científica y Tecnológica, Argentina (Projects PICT 2007-01875 and PICT 2015-2804); Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina (Project PIP 11220120100192) under the "Exploring the Biological and Genetic Diversity of Arachis Germplasm" program.

Compliance with ethical standards

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

References

Barghini E, Natali L, Cossu RM, Giordani T, Pindo M, Cattonaro F, Scalabrin S, Velasco R, Morgante M, Cavallini A (2014) The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome. Genome Biol Evol 6:776–791

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

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 ipaënsis*, the diploid ancestors of cultivated peanut. Nat Genet 47:438–446

Biemont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. Nature 443:521–524

Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M (2015) Transcription of tandemly repetitive DNA: functional roles. Chromosome Res 23:463–477

Buchmann JP, Matsumoto T, Stein N, Keller B, Wicker T (2012) Interspecies sequence comparison of *Brachypodium* reveals how transposon activity corrodes genome colinearity. Plant J 71:550–563

Burow MD, Simpson CE, Starr JL, Paterson AH (2001) Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. Genetics 159:823–837

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Chen X, Li H, Pandey MK et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proc Natl Acad Sci USA 113:6785–6790

Cloix C, Tutois S, Mathieu O, Cuvillier C, Espagnol MC, Picard G, Tourmente S (2000) Analysis of 5S rDNA arrays in *Arabidopsis thaliana*: physical mapping and chromosome-specific polymorphisms. Genome Res 10:679–690

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190

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

do Nascimento EFMB, Vidigal dos Santos B, Marques LOC, Guimarães PM, Brasileiro ACM, Leal-Bertioli SCM, Bertioli DJ, Araujo ACG (2018) The genome structure of *Arachis hypogaea* (Linnaeus, 1753) and an induced *Arachis* allotetraploid revealed by molecular cytogenetics. Comparative Cyt 12:111

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

Ferree PM, Prasad S (2012) How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. Genet Res Int. https://doi.org/10.1155/2012/430136

Fry K, Salser W (1977) Nucleotide sequences of HS-A satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. Cell 12:1069–1084

Gowda MVC, Bhat RS, Motagi BN, Sujay V, Kumari V, Bhat S (2010) Association of high-frequency origin of late leaf spot resistant mutants with AhMITE1 transposition in peanut. Plant Breed 129:567–569

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

Guerra M (2000) Patterns of heterochromatin distribution in plant chromosomes. Genet Mol Biol 23:1029–1041

Heitkam T, Petrasch S, Zakrzewski F, Kögler A, Wenke T, Wanke S, Schmidt T (2015) Next-generation sequencing reveals differentially amplified tandem repeats as a major genome component of Northern Europe's oldest *Camellia japonica*. Chromosome Res 23:791–806

Hemleben V, Kovarik A, Torres-Ruiz RA, Volkov RA, Beridze T (2007) Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. Syst Biodivers 5:277–289

Heslop-Harrison JS, Schwarzacher T (2011) Organisation of the plant genome in chromosomes. Plant J 66:18–33

Iwata-Otsubo A, Radke B, Findley S, Abernathy B, Vallejos CE, Jackson SA (2016) Fluorescence in situ hybridization (FISH)-based karyotyping reveals rapid evolution of centromeric and subtelomeric repeats in common bean (*Phaseolus vulgaris*) and relatives. G3 Genes Genom Genet 6:1013–1022

Jo SH, Koo DH, Kim JF, Hur CG, Lee S, Yang TJ, Kwon SY, Choi D (2009) Evolution of ribosomal DNA-derived satellite repeat in tomato genome. BMC Plant Biol 9:42

Jurka J, Kohany O, Pavlicek A et al (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467

Kidwell MG, Lisch DR (2000) Transposable elements and host genome evolution. Trends Ecol Evol 15:95–99

Kirov I, Divashuk M, Van Laere K, Soloviev A, Khrustaleva L (2014) An easy "SteamDrop" method for high quality plant chromosome preparation. Mol Cytogenet 7:21

Kirov IV, Kiseleva AV, Van Laere K, Van Roy N, Khrustaleva LI (2017) Tandem repeats of *Allium fistulosum* associated with major chromosomal landmarks. Mol Genet Genomics 292:453–464

Kloc A, Martienssen R (2008) RNAi, heterochromatin and the cell cycle. Trends Genet 24:511–517

Krapovickas A, Gregory W (1994) Taxonomía del género *Arachis* (Leguminosae). Bonplandia 8:11–86

Lu Q, Li H, Hong Y et al (2018) Genome sequencing and analysis of the peanut B-genome progenitor (*Arachis ipaensis*). Front Plant Sci 9:604

Macas J, Navratilova A, Meszaros T (2003) Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. Chromosoma 112:152–158

Macas J, Novák P, Pellicer J et al (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

Mallikarjuna N (2002) Gene introgression from *Arachis glabrata* into *A. hypogaea, A. duranensis* and *A. diogoi*. Euphytica 124:99–105

Mallikarjuna N, Pande S, Jadhav, Sastri DC, Rao JN (2004) Introgression of disease resistance genes from *Arachis kempff-mercadoi* into cultivated groundnut. Plant Breed 123:573–576

Marchler-Bauer A, Lu S, Anderson J (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–D229

Martienssen RA (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. Nat Genet 35:213–214

Mehrotra S, Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genom Proteom Bioinform 12:164–171

Melters D, Bradnam K, Young H et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14(1):R10

Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ (2014) 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

Naito K, Zhang F, Tsukiyama T et al (2009) Unexpected consequences of sudden and massive transposon amplification on rice gene expression. Nature 461:1130–1134

Nielen S, Campos- Fonseca F, Leal- Bertioli S, Guimaraes P, 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. Chrom Res 18:227–246

Nielen S, Vidigal B, Leal-Bertioli S, Ratnaparkhe M, Paterson A, Garsmeur O, D'Hont A, Guimarães P, Bertioli D (2012) Matita, a new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis* A-B genome divergence. Mol Genet Genom 287:21–38

Novak P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinform 11:378

Novak P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793

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

Pezer Z, Brajkovic J, Feliciello I, Ugarkovic D (2012) Satellite DNA-mediated effects on genome regulation. Genome Dyn 7:153–169

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

Presgraves DC (2010) The molecular evolutionary basis of species formation. Nat Rev Genet 11:175

Preuss SB, Costa-Nunes P, Tucker S et al (2008) Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins. Mol Cell 32:673–684

Raskina O, Barber JC, Nevo E, Belyayev A (2008) Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res 120:351–357

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 Appl 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 Appl Genet 118:1295–1307

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM (2016) High-throughput analysis of the satellitome illuminates satellite DNA evolution. Sci Rep 6:28333

Samoluk SS, Chalup L, 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 Genom 292:283–296

SanMiguel 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

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

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. https://doi.org/10.2135/cropsci2018.02.0088

Sharma S, Raina SN (2005) Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. Cytogenet Genome Res 109:15–26

Shirasawa K, Koilkonda P, Aoki K et al (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

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

Simpson CE (2001) Use of wild *Arachis* species/introgression of genes into *A. hypogaea* L. Peanut Sci 28:114–116

Sousa A, Fuchs J, Renner SS (2017) Cytogenetic comparison of heteromorphic and homomorphic sex chromosomes in *Coccinia* (Cucurbitaceae) points to sex chromosome turnover. Chromosome Res 25:191–200

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

Sveinsson S, Gill N, Kane NC, Cronk Q (2013) Transposon fingerprinting using low coverage whole genome shotgun sequencing in Cacao (*Theobroma cacao* L.) and related species. BMC Genom 14:502

Valls JFM, Simpson CE (2005) New species of *Arachis* (Leguminosae) 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

Waminal NE, Ryu KB, Park BR, Kim HH (2014) Phylogeny of *Cucurbitaceae* species in Korea based on 5S rDNA non-transcribed spacer. Genes Genom 36:57–64

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