Genetica (2015) 143:113–125
DOI 10.1007/s10709-015-9820-y

First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species

Sergio Sebastián Samoluk · Germán Robledo · Maricel Podio · Laura Chalup · Juan Pablo A. Ortiz · Silvina Claudia Pessino · José Guillermo Seijo

Received: 2 May 2014 / Accepted: 23 January 2015 / Published online: 30 January 2015

Abstract Peanut is an allotetraploid ($2n = 2x = 40$, AABB) of recent origin. *Arachis duranensis* and *A. ipaënsis,* the most probable diploid ancestors of the cultigen, and several other wild diploid species with different genomes (A, B, D, F and K) are used in peanut breeding programs. However, the genomic relationships and the evolutionary pathways of genome differentiation of these species are poorly understood. We performed a sequence-based phylogenetic analysis of the L1 reverse transcriptase and estimated its representation and chromosome distribution in species of five genomes and three karyotype groups with the aim of contributing to the knowledge of the genomic structure and evolution of peanut and wild diploid relatives. All the isolated *rt* fragments were found to belong to plant L1 lineage and were named ALI. The best supported phylogenetic groups were not concordant with the genomes or karyotype groups. The copy number of ALI sequences was higher than the expected one for plants and directly related to genome size. FISH experiments revealed that ALI is mainly located on the euchromatin of interstitial and distal regions of most chromosome arms. Divergence of ALI sequences would have occurred before the differentiation of the genomes and karyotype groups of *Arachis*. The representation and chromosome distribution of ALI in peanut was almost additive of those of the parental species suggesting that the spontaneous hybridization of the two parental species of peanut followed by chromosome doubling would not have induced a significant burst of ALI transposition.

Keywords *Arachis* genomes · Peanut · LINEs · Reverse transcriptase · Fluorescent in situ hybridization

Electronic supplementary material The online version of this article (doi:10.1007/s10709-015-9820-y) contains supplementary material, which is available to authorized users.

S. S. Samoluk (✉) · G. Robledo · L. Chalup · J. G. Seijo
Instituto de Botánica del Nordeste (Facultad de Ciencias Agrarias, UNNE-CONICET), Casilla de Correo 209, 3400 Corrientes, Argentina
e-mail: samocarp@hotmail.com; ssamoluk@agr.unne.edu.ar

G. Robledo
e-mail: grobledo@agr.unne.edu.ar

L. Chalup
e-mail: lchalup@agr.unne.edu.ar

J. G. Seijo
e-mail: seijo@agr.unne.edu.ar

G. Robledo · J. G. Seijo
Facultad de Ciencias Exactas y Naturales y Agrimensura Universidad Nacional del Nordeste, Av. Libertad 5460, 3400 Corrientes, Argentina

M. Podio · J. P. A. Ortiz · S. C. Pessino
Laboratorio de Biología Molecular, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Campo Experimental J. Villarino CC 14, S2125ZAA Zavalla, Santa Fe, Argentina
e-mail: maricelpodio@yahoo.com.ar

J. P. A. Ortiz
e-mail: jortiz@unr.edu.ar

S. C. Pessino
e-mail: pessino@arnet.com.ar

Abbreviations

LINEs	Long interspersed nuclear elements
rt	Reverse transcriptase
GSS	Genomic survey sequences
ESTs	Expressed sequence tags
FISH	Fluorescent in situ hybridization
LTR	Long terminal repeats
SINEs	Short interspersed nuclear elements
NJ	Neighbor joining
GISH	Genome in situ hybridization
ORFs	Open reading frames
PCR	Polymerase chain reaction
dNTPs	Deoxynucleotide triphosphates
CSPD	Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2′-(5′-chloro)tricyclo [3.3.1.1^{3,7}]decan}-4-yl)phenyl phosphate
DAPI	4,6-Diamidino-2-phenylindole
TE	Transposable element

Introduction

Retrotransposons are classified into LTR and non-LTR groups, depending on whether they are flanked or not by long terminal repeats (Kumar and Bennetzen 1999). Non-LTR retrotransposons are subclassified into long or short interspersed nuclear elements (LINEs or SINEs, respectively), mainly based on their size and gene components (Ohshima et al. 1996; Schmidt 1999; Han 2010). Non-LTR retrotransposons are the most abundant transposable elements in mammalian genomes, and account for one-third of the human genome (Lander et al. 2001). In plant genomes, these retrotransposons are less abundant than in mammals (Wessler 2006), but they are ubiquitous (Noma et al. 1999) with great variation in size and sequence (Schmidt 1999). The first plant LINE retrotransposon to be described was Cin4 in *Zea mays* (Schwarz-Sommer et al. 1987); and since that pioneer study, numerous other LINEs have been identified in several monocot (Leeton and Smyth 1993; Vershinin et al. 2002) and dicot (Kubis et al. 1998; Sakamoto et al. 2000; Heitkam and Schmidt 2009) species. Phylogenetically, most plant LINEs are included in the L1 retrotransposon clade (Heitkam and Schmidt 2009; Smyshlyaev et al. 2013), although some others were identified as members of the RTE type (Zupunski et al. 2001; Wenke et al. 2011).

Comparative Southern blot experiments conducted in the genus *Beta* showed that LINEs are differentially represented in species belonging to different sections (Kubis et al. 1998). Moreover, FISH experiments in *Cannabis sativa* showed that differential accumulation of those elements onto the Y chromosome leads to sex chromosome heteromorphism (Sakamoto et al. 2000). Accordingly, it was postulated that LINEs may have played a crucial role in the genome differentiation among species of particular plant groups (Kubis et al. 1998) and even in the emergence of sequence variation among chromosomes of a given complement (Sakamoto et al. 2000).

Peanut (*Arachis hypogaea*) is one of the most important cultivated legumes in the world. It is an allotetraploid ($2n = 4x = 40$) with an AABB genome constitution (Smartt et al. 1978), and originated in South America approximately 3,500–5,000 years BP (Hammons 1994; Simpson et al. 2001). Cytogenetic and molecular data suggest that *A. ipaënsis* (BB) and *A. duranensis* (AA) are the wild diploid species most probably involved in the origin of the cultigen (Fernández and Krapovickas 1994; Kochert et al. 1996; Seijo et al. 2004; Grabiele et al. 2012). Both ancestors belong to the section Arachis, which is also composed of other 27 diploid species ($2n = 2x = 20$ or $2n = 2x = 18$) and *A. monticola,* a wild allotetraploid ($2n = 4x = 40$) (Krapovickas and Gregory 1994; Valls and Simpson 2005).

Wild species within the section Arachis were assigned to six different genomes: A, B, D, F, G and K (Gregory and Gregory 1979; Stalker 1991; Robledo and Seijo 2008, 2010; Silvestri et al. 2014). A-genome species have been further arranged into three different karyotype groups: Pantanal, Chiquitano and La Plata River Basin (Robledo et al. 2009). Chromosome pairing during meiosis in cultivated *A. hypogaea* (AABB) mostly occurs as bivalents (Husted 1936), which indicates the existence of considerable genetic divergence between A and B genomes, and/or potential genetic control of chromosome pairing. Accordingly, significant divergence of the repetitive DNA fraction of the A and B genomes was demonstrated by in situ hybridization analyses using genome-wide probes (Seijo et al. 2007). By contrast, the low-copy fraction of the genomes so far analyzed (including genes and expressed sequence tags) showed a high degree of intergenomic homology (Burow et al. 2001). Moreover, the comparison of genetic maps derived from the A and B genomes evidenced a high degree of gene order conservation, and indicated that only minor structural rearrangements occurred since their divergence (Moretzsohn et al. 2009; Bertioli et al. 2009; Shirasawa et al. 2013), estimated to have taken place 3–3.5 million years BP (Moretzhon et al. 2013). The available data support the hypothesis that the repetitive fractions of these genomes would have been the main driver to genome differentiation in *Arachis*, rather than the low-copy one (Seijo et al. 2007).

Studies based on retrotransposons at a genomic scale are scarce in *Arachis*, and have been focused on peanut and its proposed wild diploid progenitors (*A. ipaënsis* and *A. duranensis*). To date, Ty3-*gypsy* (Nielen et al. 2009) and Ty1-*copia* (Nielen et al. 2011) LTR-retrotransposons have been characterized and quantified in these three species. A significant differential representation of Ty3-*gypsy*

retrotransposons, but not of Ty1-*copia* retrotransposons, was described in the two analyzed diploid species. Notably, the representation of these elements per monoploid genome (C*x*) in cultivated peanut was almost the mean of the values observed in its wild progenitors. A more recent study comparing 1.26 Mb of homeologous A and B genomes BAC clones evidenced the existence of a diverse group of complete and truncated copies of the LTR retrotransposons fraction that covered more than 40 % of the sequences analyzed (Bertioli et al. 2013).

The representation and diversity of non-LTR retrotransposons in the genomes of *Arachis* species is completely unknown. Most of the studies that explored the importance of LINEs and other non-LTR retrotransposons in the evolution of eukaryote genomes used phylogenetic analysis of conserved-domain sequence of the reverse transcriptase (*rt*) as a feasible approach to evaluate their diversity (Malik et al. 1999; Kubo et al. 2001; Schön and Arkhipova 2006) and dot blot assays to evaluate the genome wide copy number of these elements (Ichimura et al. 1997). Here, we analyzed the role of L1 retrotransposons in the evolution of the genomes of peanut and its wild relative species. For this purpose, we performed an L1 *rt* sequence-based neighbor joining (NJ) analysis and estimated the representation and the chromosome distribution of the cloned *rt* fragment in *A. hypogaea* (AABB) and in several wild relative species ($x = 10$), representative of five different genomes (A, B, F, K and D) and of the three karyotype groups of the A genome (Pantanal, Chiquitano and La Plata River Basin).

Materials and methods

Plant material and DNA extraction

Leaf and root samples were obtained from cultivated peanut (*A. hypogaea*) and from representative diploid ($2n = 20$) species of the different genomes and karyotype groups of section Arachis. The eight species used in this work, their provenance, genome constitution and karyotype group are listed in Table 1. All the plant materials were obtained from the *Arachis* germplasm collection maintained at the Instituto de Botánica del Nordeste, Corrientes, Argentina. Genomic DNA was extracted from young leaves using double precipitation with CTAB (Grattapaglia and Sederoff 1994). DNA concentrations were determined by spectrophotometry.

Amplification of the reverse transcriptase gene

Generally, full length L1 retrotransposons contain two open reading frames (ORFs) and a poly A tail in the 3′ end that defines the extreme of the element (Fig. 1). The ORF1 contains the RNA recognition motif (RRM) (Martin and Bushman 2001; Smyshlyaev et al. 2013) and the ORF2 includes the genes for the apurinic/apyrimidinic endonuclease (APE), the *rt* and the RNAase H (Flavell 1995; Kumar and Bennetzen 1999; Han 2010; Smyshlyaev et al. 2013). An initial alignment was constructed by recovering sequences with homology to the *rt* fragment of L1 retrotransposons from GSS of *Arachis*, deposited in public databases, using published LINE *rt* sequences (domains II to IV) of other species (Kubis et al. 1998). Degenerate primers were designed on conserved amino acid residues of domain II [forward (*ALIfwd*): 5′-GGCCDATTAGYATG GTGDGG-3′] and IV [reverse (*ALIrev*): 5′-GGATCWCCT TGTCTNARNCC-3′] (Fig. 1). PCR reactions were performed in a final volume of 10 µL in the presence of 20 ng of total genomic DNA, 2 µM of each primer, 0.2 mM of dNTPs (Roche), 2.5 mM of $MgCl_2$ and 0.5 U of Taq polymerase (Promega) in the corresponding PCR buffer (Promega). After an initial denaturation step at 94 °C for 5 min, amplifications were conducted for 35 cycles (94 °C for 30 s, 54 °C for 1 min, 72 °C for 1 min) with a final elongation step at 72 °C for 5 min. PCR products were analyzed by electrophoresis on 1.4 % agarose gel.

Cloning and sequence data analysis

PCR fragments of about 400 bp were gel-purified using *Wizard® SV Cleanup System* (Promega), cloned into *pGEM®-T easy Vector Systems* (Promega) and the inserts from five colonies of each species (four of *A. glandulifera*) were sequenced. The clones were denominated ALI, with the letter "A" standing for the *Arachis* genus, followed by "LI", in reference to LINEs. For specific identification of the clones, the first three letters (of the specific epithet of the *Arachis* species they were isolated from), and a distinctive number for each clone were used (e.g. ALIdur 2 refers to the L1 *rt* sequence isolated from *A. duranensis*, clone 2). Verified *rt* sequences were deposited in GenBank under accession numbers KP255357-KP255395.

Nucleotide sequences of the isolated LINEs were aligned and manually edited using ClustalW software (Thompson et al. 1994). Deduced amino acid sequences were manually edited, and gaps were introduced where necessary to correct frameshifts due to insertions or deletions. Genetic distances were computed using the number of differences in pairwise comparison of nucleotide and amino acid sequences without considering the positions containing gaps and missing data. A NJ tree (Saitou and Nei 1987) was constructed using the amino acid alignment. Both alignment and NJ tree with bootstrap analysis of 1,000 replicates were performed using MEGA software version 5 (Tamura et al. 2011).

A new alignment including all amino acid sequences from *Arachis* and 10 sequences belonging to the four major clades (L1, I, Jockey and CR1) was performed (Online

Table 1 *Arachis* species analyzed, provenances, genomes and karyotype groups, DNA content and copy number of ALI sequences per monoploid chromosome complement (C*x*)

Species	Provenance and collection number	Genome/Karyotype group	DNA content 2C (pg)	Copy number (copies/1C*x*)
A. helodes Martius ex Krapov. and Rigoni	Brasil, Mato Grosso, Cuiabá. K, G 30029 (CTES)	AA/Pantanal	2.81 (±0.017)	129 ± 20
A. duranensis Krapov. and W. C. Gregory	Argentina, Salta, San Martín, Campo Durán. K 7988 (CTES)	AA/La Plata River Basin	2.55 (±0.071)	121 ± 2
A. cardenasii Krapov. and W.C. Gregory	Bolivia, Santa Cruz, Chiquitos, Roboré. K, S, Sc 36015 (CTES)	AA/Chiquitano	3.01 (±0.070)	151 ± 6
A. ipaënsis Krapov. and W. C. Gregory	Bolivia, Tarija, Gran Chaco, Ipa. K, G, B, P, Sc, S 30076 (CTES)	BB	3.19 (±0.075)	218 ± 27
A. trinitensis Krapov. and W.C. Gregory	Bolivia, Beni, Cercado, Trinidad. Wi 1117 (CTES)	FF	2.84 (±0.030)	145 ± 24
A. batizocoi Krapov. and W. C. Gregory	Bolivia, Santa Cruz, Cordillera, Parapeti. K 9484 (CTES)	KK	2.83 (±0.084)	129 ± 31
A. hypogaea L. subsp. *hypogaea* var. *hypogaea* (race Guaycurú)	Argentina, Corrientes. FCA 27	AABB	5.70 (±0.150)	148 ± 31
A. glandulifera Stalker	Bolivia, Santa Cruz, Velasco, San Ignacio. Se, Sn 3263 (CTES, LPB)	DD	2.69 (±0.182)	103 ± 3

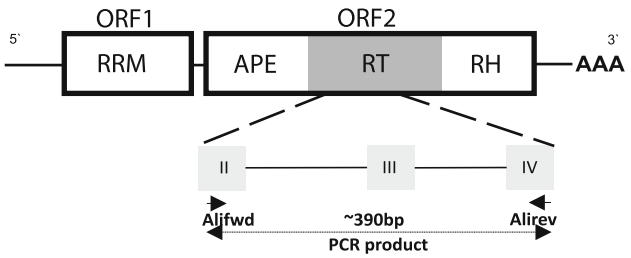

Fig. 1 Typical structure of a L1 retrotransposon of plant genomes. The figure shows the open reading frames 1 and 2 (ORF1 and ORF2) with the relative positions of the RNA recognition motif (RRM), the apurinic/apyrimidinic endonuclease (APE), the reverse transcriptase (RT) and the ribonuclease H (RH) genes, as well as the poly A tail at the 3′ end. The region of the RT gene where the primers *ALIfwd* (PISMVG) and *ALIrev* (RGLRQGD) were designed is shown by *arrows* below the conserved domains II and IV, respectively. The expected PCR product was about 390 bp

Resource 1) to classify the sequences isolated from *Arachis* within the previously described LINEs lineages (Malik et al. 1999). The selected LINEs sequences were retrieved from GenBank with accession numbers: L47193 (*Arabidopsis thaliana* Ta11-1), FM993987 (*Beta vulgaris* BvL2), Y00086 (*Zea mays* Cin4), AJ270056 (*Hordeum vulgare* BLIN), AB013908 (*Cannabis sativa* LINE-CS), U83119 (*Rattus norvegicus* L1 rat), U93574 (*Homo sapiens* L1 Hs), M14954 (*Drosophila melanogaster* I element), M22874 (*D. melanogaster* Jockey), and U88211 (*Gallus gallus* CR1). The aligned sequences were manually edited and further used for constructing a NJ tree.

Searches in ESTs databases

Expression of *rt* sequences of L1 retrotransposons was inferred through searches in the public databases of expressed sequence tags from different tissues and taxa of *Arachis*. These databases contain 254,541 ESTs. Searches in the EST database were performed with BLAST (Altschul et al. 1990) using all the *rt* sequences isolated from *Arachis* as query.

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 according to Schwarzacher et al. (1980). Root apices were treated in 1 % (m/v) cellulase plus 10 % (v/v) pectinase dissolved in 40 % glycerol, 0.01 M citric acid/sodium citrate buffer (pH 4.8) at 37 °C for 1 h. The meristems were squashed in 45 % acetic acid; the preparations were frozen with CO_2 and coverslips were removed. Finally, slides were air dried at room temperature for 1 day and then kept at −20 °C until use.

Synthesis of labeled probe and in situ hybridizations

To investigate the chromosome distribution of LINEs in *Arachis* species, a probe spanning domains II-IV of the *rt* sequence was labeled with 1 mM biotin-11-dUTP (Roche

Applied Science) by PCR using the ALIbat 4 clone as template and the same amplification conditions described for the isolation of the *rt* sequences from *Arachis*.

In situ hybridizations were performed according to Moscone et al. (1996) and Seijo et al. (2004). The hybridization mix used contained DNA probe at a concentration of 2.5–3.5 ng/μL, with a stringency to allow sequences with 70–75 % identity to remain hybridized. The first and second sets of antibodies consisted of mouse anti-biotin and rabbit anti-mouse conjugated to tetramethyl-rodamine isothiocyanate (TRITC), respectively (Sigma-Aldrich, St. Louis, Missouri, USA). Vectashield medium (Vector Laboratories) containing 2 mg/mL of 4′-6-diamidino-2-phenylindole (DAPI) was used for counterstaining and mounting the preparations. After this staining, a C-banding-like pattern was obtained with major heterochromatic bands fluorescing more intensely than euchromatin (Seijo et al. 2004).

Fluorescence microscopy and image acquisition

Preparations of 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 and blue images were captured in black and white using IM 1000 Leica software. The images were merged and adjusted (brightness and contrast) using Photoshop CS3 Extended version 10.0 (Adobe, San Jose, California, USA).

DNA content determinations

Nuclei of *Paspalum dilatatum* Chirú Q4081 (2C = 3.57 pg, Vaio et al. 2007) and *Paspalum intermedium* Munro ex Morong. accession Sch 28857 (2C = 1.42 pg, Vaio et al. 2007) were used as internal standards. Approximately 15 mg of young leaves and flowers of *Arachis* species were chopped together with 15 mg of leaf tissues of the internal standard in 0.5 mL buffer (High resolution DNA kit, Partec GmbH-Münster, Germany). The suspension was filtered through a 30-μm nylon mesh. Then, the nuclei suspension was incubated for 2–5 min at room temperature with 1.5 mL of staining solution containing 1 μg/μL propidium iodide and 5 U/mL of RNAase. The genomic DNA content of about 5,000 nuclei was estimated from measurements of three replicates for each *Arachis* species with a Partec CA-II flow cytometer. Data were analyzed using the Partec PA II FloMax software (Partec GmbH, Münster, Germany). The DNA content of each sample was calculated using the formula: DNA content of the sample = (peak of sample × G1 DNA content (2C) of the standard)/G1 peak of the standard (Doležel and Bartos 2005).

Dot blot assays and statistics

The ALIbat 4 clone, with >70 % similarity to the other *rt* sequences isolated in this work was used for calibration purposes and as probe. A dilution series of genomic DNA (ranging from 10 to 500 ng/μL) for the eight *Arachis* species analyzed here were dot-blotted onto Zeta-Probe® Blotting Membranes (Bio-Rad Laboratories) together with a dilution series, ranging from 10 to 1,000 pg, of the DNA plasmid from ALIbat 4 clone (equivalent from 5.36×10^6 to 5.36×10^8 copies of cloned *rt* sequence). Each point of the plasmid DNA dilution series was spotted by duplicate in each membrane. Dilution series of Lambda phage DNA was used as negative control. The membranes were incubated at 80 °C for 1 h and then maintained under UV light for cross-linking for 15 min. The *rt* probe was labeled with digoxigenin-dUTP (Roche) using PCR as described for FISH experiments, and hybridized onto the membranes at a concentration of 25 ng/mL at 65 °C (overnight). Hybridization and post-hybridization washes were adjusted to detect sequences with >70 % similarity. After incubation with CSPD substrate (Roche), the membrane was exposed to a Hyperfilm ECL (Amersham) between 15 min and 1 h and developed. Hybridization signal intensity was determined using the Quantity One program version 4.6.1 (Bio-Rad). Only non-saturated chemiluminographic dots were considered for the analysis. The linear or logarithmic relationship between the applied DNA amounts and the signal intensities was verified; then, the copy number of the L1 *rt* sequences in each *Arachis* species was estimated by comparing the signal intensities of the plant genomic DNA with those of the calibration curves of the ALIbat 4 clone. The average and standard deviation of the copy number per haploid genome (C*x* values) were calculated from three independent hybridization experiments.

Statistical significance of the variation in copy number of L1 retrotransposons observed among species as well as the existence of amplifications/reductions of these retrotransposons during the allopolyploidization process that originated the cultivated peanut were evaluated using a one-way analysis of variance (ANOVA) using the Tukey test at a significance level of 5 % ($\alpha = 0.05$). The copy numbers of ALI sequences for each diploid species here analyzed was correlated with their respective genome sizes using the Pearson coefficient. All the analyses were performed using Infostat software (Di Rienzo et al. 2013).

Results

Isolation and characterization of L1 *rt* sequences

We obtained amplifications from eight *Arachis* species, representative of six different genomes and three karyotype

	A. trinitensis	*A. batizocoi*	*A. ipaënsis*	*A. cardenasii*	*A. glandulifera*	*A. duranensis*	*A. helodes*	*A. hypogaea*
A. trinitensis	91 **88** 86 / 87 **78** 74							
A. batizocoi	92 **87** 82 / 85 **75** 71	91 **85** 82 / 84 **76** 71						
A. ipaënsis	91 **87** 83 / 85 **77** 72	89 **85** 80 / 84 **75** 70	99 **87** 83 / 98 **77** 71					
A. cardenasii	93 **89** 86 / 86 **77** 74	92 **87** 82 / 86 **78** 74	92 **86** 83 / 77 **75** 73	94 **89** 86 / 85 **80** 76				
A. glandulifera	93 **88** 82 / 85 **77** 72	92 **87** 81 / 87 **78** 72	92 **86** 80 / 82 **75** 72	93 **89** 84 / 84 **79** 74	92 **88** 83 / 80 **77** 73			
A. duranensis	92 **87** 85 / 87 **77** 72	92 **86** 81 / 86 **76** 71	92 **85** 82 / 84 **76** 71	92 **88** 85 / 93 **80** 73	91 **87** 81 / 82 **77** 72	100 **87** 83 / 99 **78** 74		
A. helodes	97 **88** 81 / 91 **76** 84	92 **86** 80 / 85 **76** 71	91 **86** 80 / 90 **75** 69	95 **88** 81 / 91 **78** 73	92 **87** 81 / 82 **76** 71	94 **86** 81 / 89 **77** 72	91 **86** 82 / 79 **75** 71	
A. hypogaea	93 **88** 85 / 87 **77** 73	95 **87** 82 / 84 **77** 72	94 **86** 82 / 82 **75** 71	97 **89** 85 / 92 **79** 74	92 **88** 83 / 83 **77** 73	100 **89** 83 / 99 **80** 73	94 **88** 81 / 82 **77** 71	94 **89** 94 / 90 **78** 75

Fig. 2 Matrix of identity of ALI-like reverse transcriptase sequences. Amino acid (*white boxes*) and nucleotide (*gray boxes*) identities among ALI *rt* sequences isolated from *Arachis* species ($x = 10$) representative of different genomes and karyotype groups of section *Arachis*. The arithmetic mean, maximum and minimum values of identities between the compared nucleotide and amino acid sequences among the eight species analyzed are expressed in percentages. The *largest number* in each *box* indicates the mean identity value of the compared sequences, and the *smallest numbers* represent the ranges of sequence identity. All the values are expressed in percentages and were calculated based on an amino acid consensus sequence of 132 amino acids and a nucleotide consensus sequence of 396 nucleotides

groups within the section Arachis (Table 1). Four *rt* clones originated from *A. glandulifera* and five from the remaining seven species were isolated and sequence-analyzed (39 sequences in total). They showed strong homology to LINE retrotransposons and were named ALI. The sequences varied between 378 and 387 bp in length. They putatively encoded 126–129 amino acids, spanning the conserved *rt* domains II - IV (Online Resource 1). Twenty-two of these sequences showed premature stop codons and in many of the sequences the introduction of reading frame shifts was necessary to allow informative protein alignments. Pairwise comparisons showed moderate nucleotide divergence, with average similarity values ranging from 75 to 80 % (Fig. 2). The nucleotide sequence divergence observed at the intraspecific level was similar to that detected among species. The comparison of the derived amino acid sequences showed high average similarity values ranging between 85 and 89 %.

BLAST search within ESTs

The search in *Arachis* ESTs database for sequences with homology to the *rt* fragments here isolated revealed very few low scored hits in the *A. hypogaea* database and no hits were found in the databases of any of the other species of the genus.

NJ tree analysis

The NJ analysis of the ALI sequences showed that most branches of the tree were so deep and bootstrap values so low that only a few groups of closely related sequences emerged (Fig. 3). The best supported groups were not related to the species, genomes, or karyotype groups from which they were isolated.

The NJ tree (Fig. 4) derived from the analysis of amino acid sequences (with about 130 amino acids each) representative of the major LINEs lineages together with the ALI sequences isolated in this work (49 sequences in total), revealed that all the plant sequences grouped into the L1 clade in a different subclade from that of the mammalian L1 LINEs. Moreover, all the ALI *rt* fragments clustered together and separately from all the other L1 retroelements here included with a 100 % bootstrap.

Genome size of *Arachis* species and copy number estimation of ALI sequences

The flow cytometer measurements carried out to determine the genome size (2C) of the species here analyzed are presented in Table 1. Dot blot hybridization experiments were used to estimate the copy number of L1 sequences in the *Arachis* genomes (Fig. 5). The calibration curve was constructed from dots made of serial dilutions of a recombinant plasmid containing the *A. batizocoi* sequence ALIbat 4 clone. Only values that led to regression coefficients $R^2 > 0.95$ were considered. The calibration dots hybridized with the ALIbat 4 clone digoxigenin-labeled probe produced signals of variable intensity, whereas no signal was obtained onto lambda phage dots (negative control) (Fig. 5). Among the diploid species, the estimated

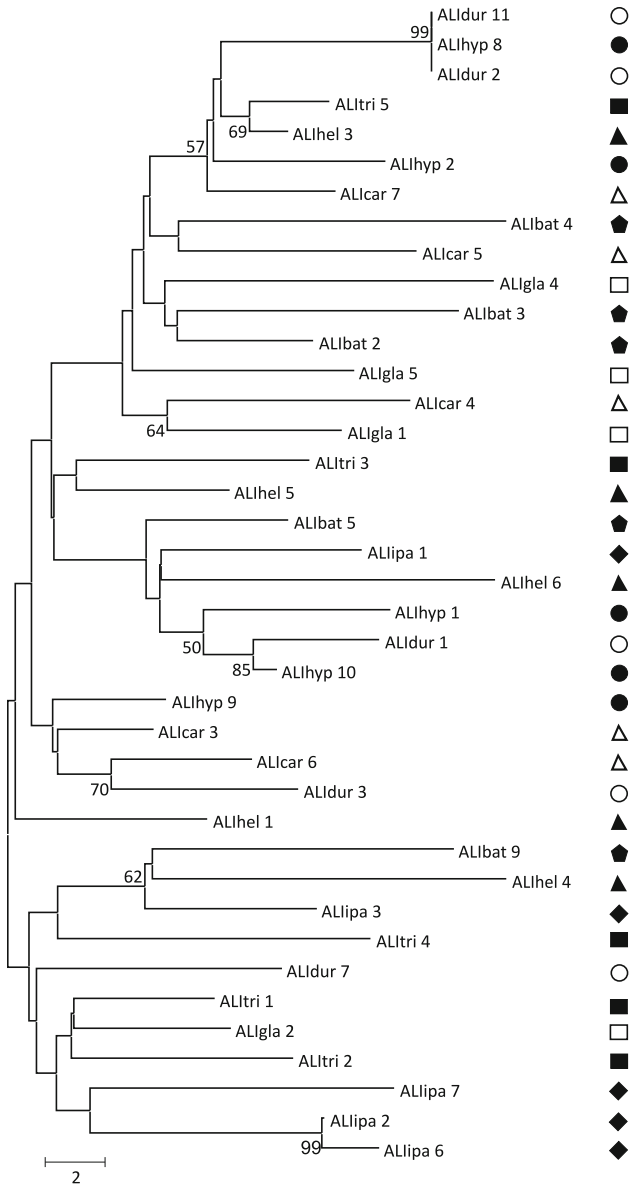

Fig. 3 Dendrogram of ALI sequences isolated from *Arachis* species. The NJ tree was constructed based on deduced amino acid sequences of reverse transcriptase domains of ALI retroelements isolated from *Arachis* species belonging to different genomes and karyotype groups. The *rt* sequences isolated from the same species are indicated with the same symbol in the *right-hand column*: *A. cardenasii* (*open triangle*), *A. duranensis* (*open circle*), *A. helodes* (*closed triangle*), *A. batizocoi* (*closed pentagon*), *A. ipaënsis* (*closed diamond*), *A. trinitensis* (*closed square*), *A. hypogaea* (*closed circle*) and *A. glandulifera* (*open cicle*). The *numbers* adjacent to the branches indicate the bootstraps supporting a particular cluster

ALI copy number per monoploid genome (C*x*) ranged from 103 in *A. glandulifera* (D genome) to 218 in *A. ipaënsis* (B genome) (Table 1). The only species that presented statistically significant differences in copy number was *A. ipaënsis* ($p = 0.0002$). Within the A genome, the annual *A. duranensis* showed a significantly lower ($p = 0.0492$) copy number (121 copies) than the perennial *A. helodes* (129 copies) and *A. cardenasii* (151 copies). In all cases, the copy number detected in each diploid species was directly correlated with its C*x* genome size ($r = 0.86$; $p = 0.01$) (Fig. 6).

The representation of ALI sequences per monoploid genome in *A. hypogaea* (148 copies) was lower (although not statistically different; $p = 0.3155$) than the expected mean value (170 copies) calculated from the copy number observed in the diploid parental species, *A. duranensis* (121 copies) and *A. ipaënsis* (218 copies). Accordingly, the genome size per monoploid genome in *A. hypogaea* (C*x* = 1.425 pg) was lower (but not statistically different; $p = 0.7765$) than the expected mean value (1.435 pg) inferred from those of the ancestor species.

Chromosomal distribution of ALI sequences

In FISH experiments, the ALI sequences showed a similar pattern of distribution in all the genomes, with widespread hybridization signals in almost all the chromosomes of the analyzed species (Fig. 7a–g). The strongest signals were observed at the interstitial and distal chromosome regions, whereas at proximal ones the signals were sparser or even absent.

In the three A genome diploid species, two chromosome pairs were unequivocally identified: the smallest chromosome of the complement (A9), which had a strong pericentromeric band of $DAPI^+$ heterochromatin, and the SAT chromosome (A10), in which the satellite segment was usually located far from the proximal segments of the long arm, even in advanced metaphases. Notably, A9 and A10 chromosomes were poorly hybridized, except for the satellite portion of the A10 chromosome, which showed a similar hybridization pattern to that of the other chromosomes of the complement. The hybridization signals in the B10 chromosome of *A. ipaënsis* were similar to those observed in the A10 chromosome. Hybridization patterns in the species having other genomes (D, F and K) were similar in all the chromosomes.

In somatic peanut metaphases, chromosomes corresponding to the A and B genomes were unmistakably detected after DAPI counterstaining. The 20 chromosomes of the former genome featured strong centromeric DAPI bands, whereas those of the latter were devoid of centromeric heterochromatin (Fig. 7h). The hybridization signals of ALI sequences were observed in almost all the chromosomes of the A and B genomes. Similarly to diploid species, the intensity of the hybridization signals was stronger in the interstitial and distal regions of the chromosome arms than in the proximal ones. Particularly, the A9 pair and the SAT chromosomes of both genomes (pairs

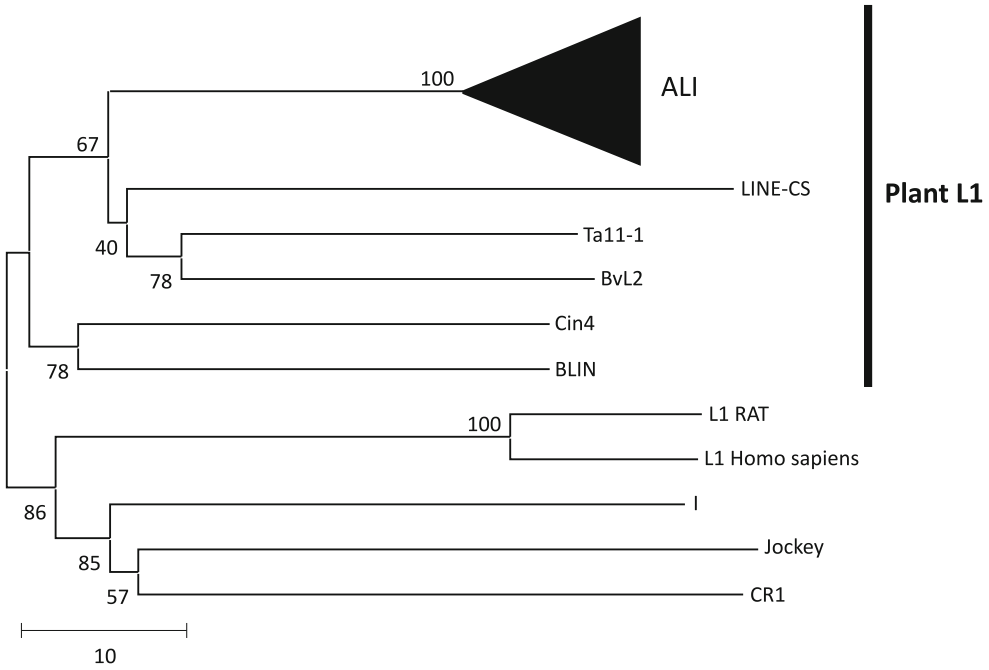

Fig. 4 Dendrogram showing the relatedness of the L1 elements with other LINE elements. L1 *rt* sequences isolated from *Arachis* species (ALI) are indicated in *bold*. Sequences belonging to four different clades of LINEs were included in the analysis: L1 (LINE-CS from *Cannabis sativa*, Ta*11-1* from *Arabidopsis thaliana*, BvL2 from *Beta vulgaris*, BLIN from *Hordeum vulgare*, Cin4 from *Zea mays,* L1 rat from *Rattus norvegicus* and L1 *Homo sapiens*), CR1 (CRE1 from *Gallus gallus*), Jockey (Jockey from *Drosophila melanogaster*) and I (I from *Drosophila melanogaster*). The dendrogram was produced using the NJ algorithm. Only bootstrap values higher than 50 % are shown next to the branches

A10 and B10) showed the lowest hybridization signals, as observed in the respective diploid ancestors.

Discussion

The sequence analysis indicated that all of the *rt* fragments isolated from *Arachis* species correspond to L1 retrotransposons, which is in agreement with the primers used for the assay. The presence of stop codons and the need to introduce frameshifts to maintain the alignment of amino acid sequences demonstrates the presence of several defective LINEs in the genomes of the analyzed species. The few hits with homology to the *rt* fragments of L1 retrotransposons recovered from the EST databases available for *Arachis* species suggest that most ALI elements may be inactive and only few elements may have very low levels of activity. These findings are consistent with previous studies on plant LINEs, in which these retroelements were found to be mostly inactive or with very low levels of transcription (Meyers et al. 2001; Rossi et al. 2001).

The fact that all ALI sequences formed a separate group within the plant cluster of the L1 clade evidences that all LINEs here isolated from the *Arachis* species belong to a phylogenetically related group of elements, and different from the groups characterized to date. The absence of species, karyotype group, or genome specific clusters of ALI sequences in the NJ tree suggests that these sequences diverged before the differentiation of the genomes analyzed here. Considering the date estimated for the divergence of *Arachis* A and B genomes (Moretzhon et al. 2013), the divergence of LINEs should have occurred before 3 Mya. The polyphyly observed within the genomes and karyotype groups for these LINEs might be explained by retaining part of the ancestral polymorphism.

Sequence comparisons among the conserved domains of the *rt* from different retrotransposons within the same genome provide some hints about the divergence and evolution of each element type (Xiong and Eickbush 1988, 1990; Doolittle et al. 1989). In this sense, the fact that the genetic distances of the ALI *rt* nucleotide sequences were higher than those of the Ty3-*gypsy* (FIDEL) retrotransposons of cultivated peanut (Nielen et al. 2009) is consistent with the hypothesis that proposed the non-LTR retrotransposons to be more ancient than the LTR-retrotransposons (Xiong and Eickbush 1990).

Copy number estimation demonstrated that the representation of ALI sequences in the eight *Arachis* species analyzed is relatively high when compared to the LINEs

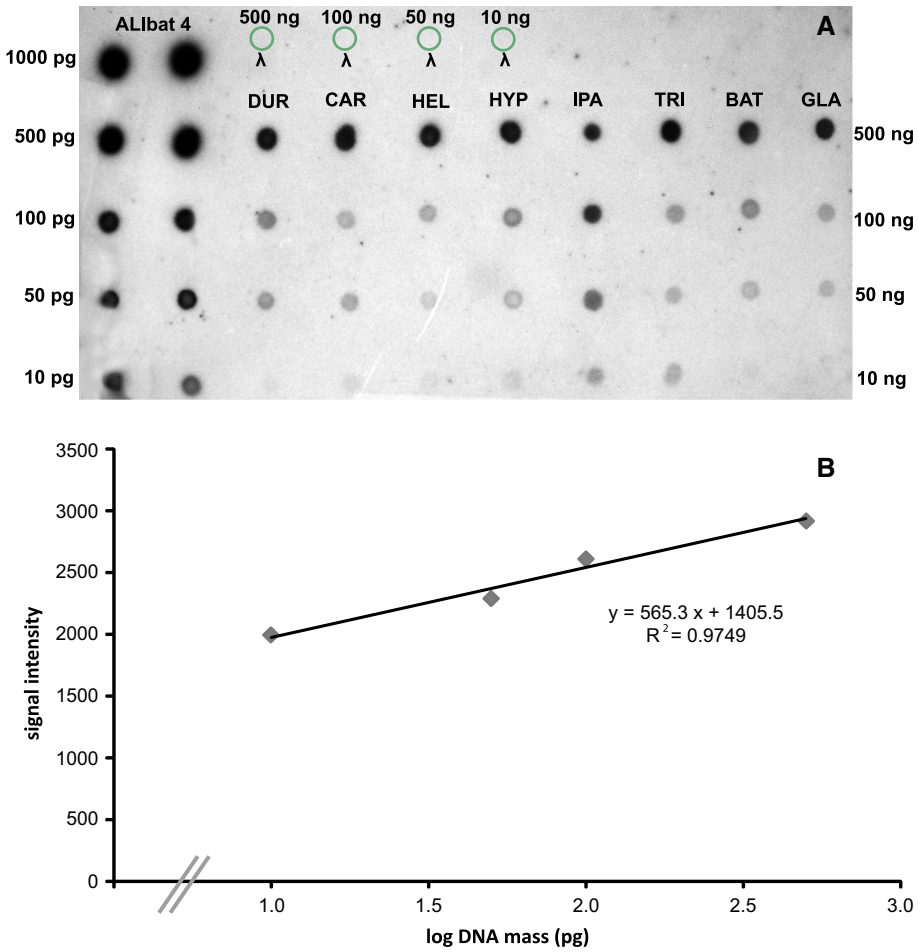

Fig. 5 Copy number estimation of L1 elements in different species of section Arachis. **a** Chemiluminograph of a *dot blot* assay with dilution series of a plasmid containing the ALIbat 4 clone for calibration, of a lambda phage (λ) as negative control and of genomic DNA from eight *Arachis* species (DUR: *A. duranensis*, CAR: *A. cardenasii*, HEL: *A. helodes*, HYP: *A. hypogaea*, IPA: *A. ipaënsis*, TRI: *A. trinitensis*, BAT: *A. batizocoi*, GLA: *A. glandulifera*) hybridized with a digoxigenin-labeled probe (ALIbat 4). The amount of genomic DNA is indicated in the *right-hand* of the graph and the mass of ALIbat 4 clone in the *left-hand*. **b** Calibration *curve* of the ALIbat 4 clone. The calibration *curve* was constructed with the average values of signal intensities from the individual dots of two dilution series. Data from dots oversaturated were not included in the analysis

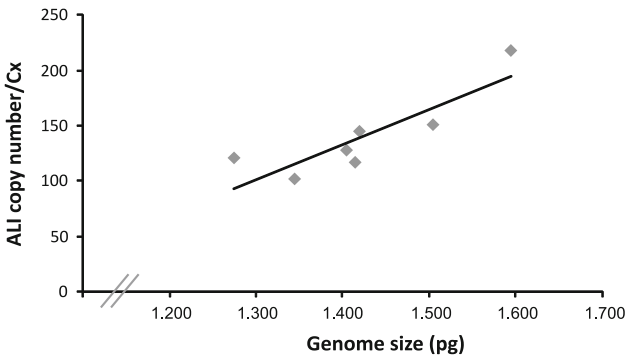

Fig. 6 Relationship between the copy number of ALI sequences and genome size (monoploid genome, C*x*). Copy numbers were estimated by *dot blot* and DNA contents by flow cytometry

present in other plant species. Although our estimation might underestimate the real representation of L1 LINEs, the detected copy number is still higher than the estimations published for other plant groups, e.g. 25–50 copies of the Cin4 LINE per haploid genome in *Z. mays* (Schwarz-Sommer et al. 1987). The correlation observed between the genome sizes of the species and the ALI copy numbers may suggest that these elements have been involved in or accompanied the evolution of genome size in *Arachis*. However, the lack of significant differences observed in most of the *Arachis* species evidences that this LINE family would have had a low impact in the recent genome evolution. Therefore, it seems that the variation in copy number of ALI occurred passively during genome size changes; i.e., changes in representation of ALI sequences occurred as a consequence of any intrinsic mechanism of genome size increase or decrease rather than by self-amplification or selective deletion.

LINEs are generally less abundant than LTR-retroelements in the plant genomes (Noma et al. 1999; Alix and Heslop-Harrison 2004; Hawkins et al. 2006). Within the section Arachis, the available data (Nielen et al. 2009, 2011; Bertioli et al. 2013) allowed us to compare the genome abundance of repetitive elements only in *A. duranensis* (A genome) and *A. ipaënsis* (B genome). Considering these data, the copy number of ALI sequences is three (B genome) to 20 (A genome) times smaller than that observed for Ty3-*gypsy* sequences (about 820 and 3,000 copies per haploid genome for *A. ipaënsis* and *A. duranensis*, respectively).

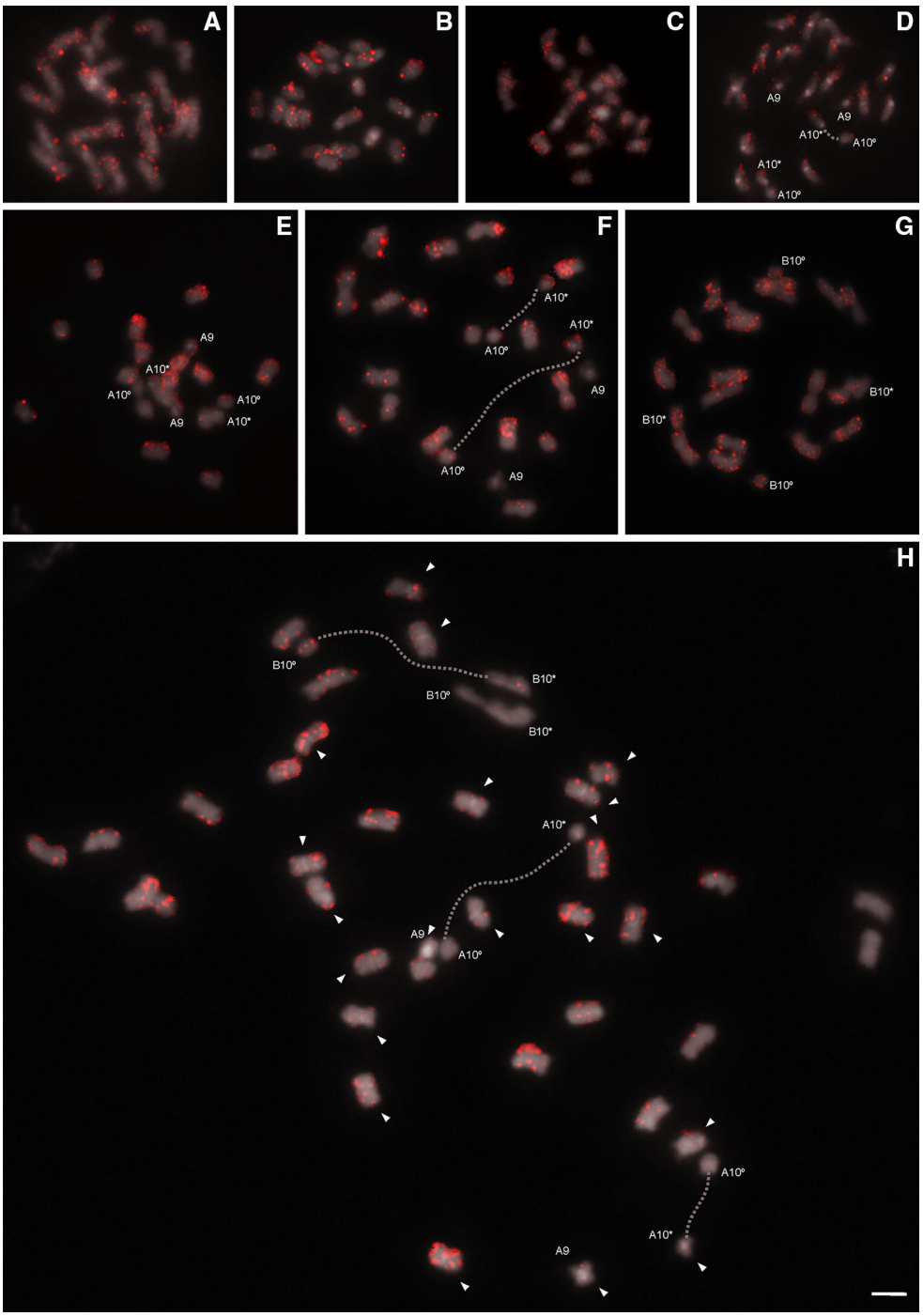

Fig. 7 Chromosome distribution of ALI sequences on mitotic chromosomes of different *Arachis* species. Hibridization signals of the ALIbat 4 clone (in *red*) appear with a dispersed pattern in the insterstitial and distal portion of most chromosomes in all the species analyzed. Chromosomes counterstained with DAPI subsequent to the in situ hybridization were pseudocolored in grey. **a** *A. batizocoi*, **b** *A. glandulifera*, **c** *A. trinitensis*, **d** *A. cardenasii*, **e** *A. duranensis*, **f** *A. helodes*, **g** *A. ipaënsis*, **h** *A. hypogaea*. Extended secondary constrictions of the A10 and B10 chromosomes were connected with a *grey dotted line*. The short arm and the proximal segment of the long arm are indicated by an *asterisk* (A10*, B10*) while the separated satellites are marked by a *degree sign* (A10°, B10°). In the metaphase of *A. hypogaea* (H), the chromosomes of the A genome complement are indicated by *arrowheads*. *Scale bar* 3 μm. (Color figure online)

Conversely, the representation of ALI sequences is similar to that reported for Ty1-*copia* sequences (about 135 per haploid genome, either A or B). Therefore, the present results evidence that the L1 LINEs are well represented in *Arachis* genomes, and that their abundance is comparable to that of some families of the LTR retrotransposons.

FISH results evidenced that, despite the dispersed pattern, the ALI sequences are non-randomly distributed in the genomes of diploid *Arachis* species. They tended to be preferentially distributed in the euchromatin of the interstitial and distal regions of the chromosomes and are almost completely absent in the proximal and centromeric regions of all the analyzed species. This preferential distribution of ALI sequences is in agreement with that reported for LINEs in other plant species, in which these elements are completely absent from the heterochromatic regions (Schmidt 1999; Kubis et al. 1998). The similar distribution and hybridization intensities observed in all the analyzed *Arachis* species indicated a high conservation of the copy number and organization of ALI sequences among the genomes and karyotype groups of section Arachis. These results are in complete agreement with the results obtained using dot blot assays. However, the fact that some chromosome regions or even some entire chromosomes showed very few hybridization signals, while others had strong localized signals might reflect the existence of specific insertional patterns of ALI sequences, which may be different in the genomes analyzed. In particular, the lower hybridization observed in the pairs A9 and A10 than in the other chromosomes of the A genome is consistent with the results obtained using GISH analyses (Seijo et al. 2007) and FISH using probes specific for LTR retroelements (Nielen et al. 2009, 2011; Bertioli et al. 2013). These differential hybridization patterns suggest that these pairs may have a particular sequence composition, with a lower representation of mobile elements than the remaining chromosomes.

It has been shown that the genomic shock occurring during allopolyploidization events produces changes in the retrotransposon activity that may lead to significant changes in the representation of these elements in the duplicated genomes (Liu and Wendel 2000; Kashkush et al. 2002). Moreover, the available evidence indicates that TE proliferation in the short- or long-term after allopolyploidization may be restricted to a few TEs in specific polyploid systems (Madlung et al. 2005; Lim et al. 2007). The results obtained here indicate that ALI sequences would have not undergone important changes in their genomic representation during the hybridization/genome doubling events that gave rise to *A. hypogaea.* Similar findings were reported for the LTR-retrotransposons (Ty1-*copia* and Ty3-*gypsy*), since their copy numbers did not vary between *A. hypogaea* and its diploid parentals, *A. duranensis* and *A. ipaënsis* (Nielen et al. 2009, 2011). The bulk of evidence suggests that during the allopolyploidization process that originated peanut, the interaction of the A and B genomes within a single nucleus did not induce significant modifications in the retrotransposon component of their repetitive fractions. Similar results were reported for *Brassica* species (Alix and Heslop-Harrison 2004), in which three diploid species (*B. rapa, B. oleracea and B. nigra*) and their corresponding allopolyploids (*B. napus, B. carinata and B. juncea*) did not show any retrotransposon amplification.

The lower (non-statistically significant) copy number of ALI concomitantly with a lower (non-statistically significant) genome size observed in *A. hypogaea* than expected suggests that, if any reduction in copy number of ALI sequences occurred, it might have occurred passively, as a consequence of genome size reduction that usually occurs in polyploids. In this sense, extant data indicate that illegitimate recombination in the TE genome fraction immediately after allopolyploidization may lead to losses of TE sequences. But more important, the recombination among distant TE may lead to losses of large portions of other sequences that lay in between those recombinant elements, with the concomitant genome size shrinkage (Petrov 2002; Devos et al. 2002; Wicker et al. 2003; Ma et al. 2004; Bennetzen et al. 2005).

Concerning the genomic impact of ALI sequences during allopolyploidization, the conservation of the copy number as well as the distribution and intensities of the FISH signals in the complements of the diploid parental species (*A. duranensis* and *A. ipaënsis*) compared to those in the A and B genomes of *A. hypogaea* revealed that the pattern of the tetraploid is approximately the sum of those patterns observed in the diploids. The conservation of the chromosome complement in the tetraploid was also revealed by the high homology detected by FISH mapping of ribosomal genes (Seijo et al. 2004) and GISH experiments (Seijo et al. 2007). All these findings suggest a strong genomic stability during polyploid hybridization of *A. duranensis* and *A. ipaënsis*.

Acknowledgments This work is part of S. Samoluk Doctoral Thesis that will be presented in the Facultad de Ciencias Agrarias, Universidad Nacional de Rosario (Santa Fe, Argentina). It was supported by the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), Argentina, Project PICT 2007-1356 and PICT 2012-1875, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina, Project PIP No 11220090100613 and Universidad Nacional del Nordeste, PI No 189. S. Samoluk, M. Podio and L. Chalup received fellowships from CONICET. G. Robledo, S. Pessino, J.P.A. Ortiz and J. G. Seijo are research staff members of CONICET. We would like to thank INTA Manfredi Station, Córdoba, Argentina and the Texas Agriculture Experimental Station, Stephenville, Texas, for providing seeds of some accessions.

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

References

Alix K, Heslop-Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid *Brassica* species. Plant Mol Biol 54:895–909

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

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95:127–132

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC, Guimarães PM, Hougaard BK, Fredslund J, Schauser L, Nielsen AM, Sato S, Tabata S, Cannon SB, Stougaard J (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genome 10:45

Bertioli DJ, Vidigal B, Nielen S, Ratnaparkhe MB, Lee T-H, 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

Burow MD, Simpson CE, Starr JL, Paterson A (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

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

Di Rienzo JA, Casanoves F, Balzarini MG, González L, Tablada M, Robledo CW (2013) InfoStat version 2014. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. http://www.infostat.com.ar/

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

Doolittle RF, Feng DF, Johnson MS, McClure MA (1989) Origins and evolutionary relationships of retroviruses. Quart Rev Biol 64:1–30

Fernández A, Krapovickas A (1994) Cromosomas y evolución en *Arachis* (Leguminosae). Bonplandia 8:188–220

Flavell AJ (1995) Retroelements, reverse transcriptase and evolution. Comp Biochem Physiol B Biochem Mol Biol 110:3–15

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

Grattapaglia D, Sederoff R (1994) Genetic linkage Maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Gregory MP, Gregory WC (1979) Exotic germoplasm of *Arachis* L. interspecific hybrids. J Hered 70:185–193

Hammons RO (1994) The origin and history of the groundnut. In: Smartt J (ed) The groundnut crop: a scientific basis for improvement. Chapman & Hall, London, pp 24–39

Han JS (2010) Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mobile DNA 1:15

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16:1252–1261

Heitkam T, Schmidt T (2009) BNR-a LINE family from *Beta vulgaris*-contains a RRM domain in open reading frame 1 and defines a L1 sub-clade present in diverse plant genomes. Plant J 59:872–882

Husted L (1936) Cytological studies on the peanut *Arachis*. II. Chromosome number, morphology and behavior, and their application to the problem of the origin of the cultivated forms. Cytologia 7:396–423

Ichimura S, Mita K, Sugaya K (1997) A major non-LTR retrotransposon of *Bombyx mori*, L1Bm. J Mol Evol 45:253–264

Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651–1659

Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am J Bot 83:1282–1291

Krapovickas A, Gregory W (1994) Taxonomía del género *Arachis* (Leguminosae). Bonplandia 8:1–186

Kubis SE, Heslop-Harrison JS, Desel C, Schmidt T (1998) The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. Plant Mol Biol 36:821–831

Kubo Y, Okazaki S, Anzai T, Fujiwara H (2001) Structural and phylogenetic analysis of TRAS, telomeric repeat-specific non-LTR retrotransposon families in Lepidopteran insects. Mol Biol Evol 18:848–857

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Leeton PR, Smyth DR (1993) An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol Gen Genet 237:97–104

Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytol 175:756–763

Liu B, Wendel JF (2000) Retroelement activation followed by rapid repression in interspecific hybrid plants. Genome 43:874–880

Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869

Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L (2005) Genomic changes in synthetic *Arabidopsis* polyploids. Plant J 41:221–230

Malik HS, Burke WD, Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805

Martin SL, Bushman FD (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol 2:467–475

Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 11:1660–1676

Moretzhon 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

Moretzsohn MC, Barbosa AV, Alves-Freitas DM, Teixeira C, Leal-Bertioli SC, Guimarães PM, Pereira RW, Lopes CR, Cavallari MM, Valls JF, Bertioli DJ, Gimenes MA (2009) A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. BMC Plant Biol 9:40

Moscone E, Matzke M, Matzke A (1996) The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. Chromosoma 105:231–236

Nielen S, Campos-Fonseca F, Leal-Bertioli S, Guimaraes P, Seijo G, Town C, Arrial R, Bertioli D (2009) 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 S, Ratnaparkhe M, Paterson A, Garsmeur O, D'Hont A, Guimarães P, Bertioli D (2011) Matita,

a new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis* A–B genome divergence. Mol Genet Genomics 287:21–38

Noma K, Ohtsubo E, Ohtsubo H (1999) Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. Mol Gen Genet 261:71–79

Ohshima K, Hamada M, Terai Y, Okada N (1996) The 3′-ends of tRNA-derived short interspersed repetitive elements are derived from the 3′-ends of long interspersed repetitive elements. Mol Cell Biol 16:3756–3764

Petrov DA (2002) Mutational equilibrium model of genome size evolution. Theor Popul Biol 61:531–543

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

Rossi M, Gonçalves Araujo P, Van Sluys MA (2001) Survey of transposable elements in sugarcane expressed sequence tags (ESTs). Genet Mol Biol 24:147–154

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sakamoto K, Ohmido N, Fukui K, Kamada H, Satoh S (2000) Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. Plant Mol Biol 44:723–732

Schmidt T (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol Biol 40:903–910

Schön I, Arkhipova IR (2006) Two families of non-LTR retrotransposons, *Syrinx* and *Daphne,* from the Darwinulid ostracod, *Darwinula stevensoni*. Gene 371:296–307

Schwarzacher T, Ambros P, Schweizer D (1980) Application of Giemsa banding to orchid karyotype analysis. Plant Syst Evol 134:293–297

Schwarz-Sommer Z, Leclercq L, Göbel E, Saedler H (1987) Cin4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J 6:3873–3880

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. ipaënsis* 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

Shirasawa K, Bertioli D, Varshney R, Moretzsohn M, Leal-Bertioli S, Thudi M, Pandey M, Rami J, Foncéka D, Gowda M, Qin H, Guo B, Hong Y, Liang X, Hirakawa H, Tabata S, Isobe S (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. DNA Res 20:173–184

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. doi:10.1007/s00606-014-1092-y

Simpson CE, Krapovickas A, Valls JFM (2001) History of *Arachis* including evidence of *A. hypogaea* L. progenitors. Peanut Sci 28:78–80

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

Smyshlyaev G, Voigt F, Blinov A, Barabas O, Novikova O (2013) Acquisition of an Archaea-like ribonuclease H domain by plant L1 retrotransposons supports modular evolution. Proc Natl Acad Sci USA 110:20140–20145

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

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

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* (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia 14:35–63

Vershinin AV, Druka A, Alkhimova AG, Kleinhofs A, Heslop-Harrison JS (2002) LINEs and gypsy-like retrotransposons in *Hordeum* species. Plant Mol Biol 49:1–14

Wenke T, Döbel T, Sörensen TR, Junghans H, Weisshaar B, Schmidt T (2011) Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. Plant Cell 23:3117–3128

Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. Proc Natl Acad Sci USA 103:17600–17601

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. Plant Cell 15:1186–1197

Xiong Y, Eickbush TH (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol 5:675–690

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362

Zupunski V, Gubensek F, Kordis D (2001) Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. Mol Biol Evol 18:1849–1863