Isolation and Characterization of Retrotransposons in Wild and Cultivated Peanut Species

S Nielen[1,*], F Campos Fonseca[1], P Guimarães[1], S C Leal-Bertioli[1] & D Bertioli[2]

Abstract

Retrotransposons are considered a possible source for mutations due to their potential of spreading in the genome using a "copy and paste"-like mechanism. Here we report on the isolation and characterization of a new Ty3-*gypsy* retrotransposon from allotetraploid peanut (*Arachis hypogaea,* 2n=4x=40) and its diploid ancestors *A. duranensis* (AA-genome, 2n=20) and *A. ipaënsis* (BB-genome, 2n=20). We have identified two repetitive sequences, one showing high similarity at the amino acid level to the reverse transcriptase of *Athila*-type retrotransposons, the other being AT-rich with no similarities to gene bank sequences. Results from genome walking experiments gave first evidence that both sequences represented parts of the same Ty3-*gypsy* retrotransposon, the 5'-LTR (long terminal repeat)- and the *pol* (polyprotein)-region respectively. Fluorescent *in situ* hybridization (FISH) experiments showed that the element is dispersedly distributed on the chromosomes, absent from centromeres and telomeric regions, and more prominent in chromosomes of the A-genome. The element appeared to be moderately repetitive with copy numbers of about 830 (*A. ipaënsis),* 2,600 *(A. duranensis),* and 3,000 (*Arachis hypogaea)* per haploid genome. Phylogenetic analysis of the deduced amino acid sequences of 80 isolated reverse transcriptase clones from the three species shed light on its evolution within the peanut species. The isolated sequences contained multiple stop-codons and so far, no evidence has been found that the element is still active. An outlook is given regarding finding new tools for the advancement of *Arachis* breeding programs aimed at the transfer of resistance to biotic and abiotic stresses to peanut.

Introduction

Cultivated peanut *(Arachis hypogaea)* is an allotetraploid species (AABB genome, 2n = 4x = 40), which derived from a single hybridization event between two diploid *Arachis* species and subsequent spontaneous chromosome doubling. Recent studies including those based on fluorescent *in situ* hybridization using rDNA sequences as probes suggest that *A. duranensis* (AA, 2n = 20) and *A. ipaënsis* (BB, 2n = 20) are the most probable ancestors [1, 2]. Its low genetic diversity is the major bottleneck that hampers crop improvement programs and genetic studies in peanut. In contrast, most wild *Arachis* species are diploid with high genetic diversity, and moreover, some are a source for resistance to important biotic and abiotic stresses. In order to elucidate the relationships of wild and cultivated peanut genomes, thereby improving our ability to efficiently introgress wild genes into the peanut crop, repetitive elements, especially retrotransposons, are an interesting research object.

Retrotransposons are mobile genetic elements present in many different organisms. Active elements replicate through a mechanism of reverse transcription and insertion of a new copy into new chromosomal sites, where they can cause mutations (for a recent review on their life cycle see [3]). Retroelements, particularly the LTR (long terminal repeat) retrotransposons, constitute the major part of repetitive DNA and contribute substantially to genome size [4, 5]. Based on phylogenetic analysis of their reverse transcriptase (*rt*) sequences and on structural differences LTR-, retrotransposons can be divided into two major lineages, one consisting of the Ty1-*copia* retrotransposons (*pseudiviridae*), and the other of the Ty3-*gypsy* retrotranspsons (*metaviridae*) [6, 7]. Retrovirus-like retrotransposons are characterized through an additional open reading frame (ORF) encoding transmembrane domains, which are characteristic for envelope (*env*) genes. Examples of retrovirus-like transposon families are the Ty3-*gypsy* retrotransposons *Athila* [8], *Cyclops* [9], and *Calypso* [10]. Several LTR-retrotransposons have been reported to be present in legumes. The first indication for presence of an LTR retrotransposon in peanut was given by Chavanne, *et al.* [9], who characterized the *gypsy*-like retrotransposon *Cyclops* in pea and detected hybridization of a fragment of its reverse transcriptase to genomic DNA of various legumes, including *A. hypogaea*. Yüksel, *et al.* [11] screened their BAC library of *A. hypogaea* and found a sequence with similarity to an *Arabidopsis copia* element.

Some elements seem to be constitutively expressed, for instance the *Ogre* element of pea [12]. Others are silent and can be activated upon certain stress signals such as tissue culture (*Tos17* retrotransposon in rice [13]), ionizing irradiation (Ty1 in *Saccharomyces cerivisiae* [14]), wounding (*Tto1* in tobacco [15]), or allopolyploidization (*Wis2-1A* in synthetic wheat allotetraploids [16]). The ability of some retrotransposons to become active again after stimulation makes them an ideal tool for inducing desired genetic variability, and also for reverse genetics approaches, as it was shown by the generation of about 50,000 rice insertion lines with *Tos 17* [17]. Furthermore, their abundance and ability to transpose make them good potential markers in form of the PCR-based techniques IRAP (inter-retrotransposon amplified polymorphism) and REMAP (retrotransposon-microsatellite amplified polymorphism) that detect retrotransposon integration events in the genome [18].

So far, with the exception of rDNA sequences [1] relatively little is known about the features and localization of repetitive elements in *Arachis* species. This applies especially to retrotransposons. Here we report on the isolation and characterization of a new *Athila*-like retrotransposon in *Arachis* species.

Materials and Methods

Plant materials and DNA extraction: Leaf and root tissue were obtained from *A. hypogaea* cv. Tatu, *A. duranensis* (accession V14167c), *A. stenosperma* (accession V10309), and *A. ipaënsis* (accession KG30076c). Plants were grown from seeds under greenhouse conditions. All plants were obtained from the Brazilian *Arachis* germplasm collection, maintained at Embrapa Genetic Resources and Biotechnology — CENARGEN (Brasília-DF, Brazil).

Cloning of reverse transcription sequences and phylogenetic analysis: To enable phylogenetic analysis of the retroelement in *A. hypogaea*, *A.*

[1] Embrapa Recursos Genéticos and Biotechnologia, Brasília – Distrito Federal, Brazil
[2] Universidade Católica de Brasília, Brasília – Distrito Federal, Brazil
* Corresponding author. E-mail: stephan@cenargen.embrapa.br

duranensis, and *A. ipaënsis* a set of three forward (F) and three reverse (F) PCR primers have been designed to amplify the total sequence of the *Athila*-like reverse transcriptase (*rt*):
Rep-RT-F1 5'-AAGGACACACAAGACAGCTC-3';
Rep-RT-F2 5'-GTACGCACAAGATCCTATTG-3';
Rep-RT-F3 5'-CTAAATCCAGCCATGAAGG-3';
Rep-RT-R1 5'-GTCAGCTACAAGGAGATTGC-3';
Rep-RT-R2 5'-GGAGATGATAGGTGCAGAAG-3';
Rep-RT-R3 5'-TCACACATCAGTTCAAATGG-3'.

Sequence analysis and construction of contigs has been done using the Staden Package software [19]. For multiple alignment of amino acid- and DNA sequences the program ClustalX [20] has been used. The results of alignments have been utilized for construction of phylogenetic trees using the program Mega 3.1 [21].

Genome walking: In order to extend sequence information into the flanking genomic regions of the two pseudo-contigs a modified Genome Walker™ strategy has been applied. Genomic DNA of *A. hypogaea* (1.5 µg) was digested with 15 units of the restriction enzyme *Pvu*II. After complete digestion the samples were purified with Phenol/Chloroform, precipitated and re-suspended in 20 µl $TE_{(0.1).}$ 4 µl of the digested was used for ligation of a *Pvu*II adaptor, consisting of 5'-ACTCGATTCTCAACCCGAAAGTATAGATCCCA-3' (long arm) and 5'-Phosphate-TGGGATCTATACTT-H_2N-3'. PCR reactions have been performed with the adapter primer AP1 (5'-ACTCGATTCTCAACCCGAAAG-3') and genome specific primers directed outwards the known sequences. For Rep-1 primer Rep-1-Out-Right-2 (5'- GTTGCCGGGGATTGTTC-3') was used to amplify sequences downstream, for Rep-2 it were the primers Rep-2-Out-Left (5'-AATCATGTCCTCAATTACGC-3') and Rep-2-Out-Right (5'-TGGTGACTGAAGGAATTGTC-3') for upstream and downstream amplification. PCR products have been cloned and sequenced using Sp6 and T7 sequencing primers.

Chromosome preparation and Fluorescent in situ hybridization: Root tips were collected from young plants in the greenhouse or from rooting leaf petioles that have been cultivated in humid petri dishes in an incubator with average temperature of 25° C [22]. Chromosome preparations have been made according to Maluszynska and Heslop-Harrison [23]. Pretreatment, hybridization, washing and detection procedures essentially followed protocols published by Schwarzacher and Heslop-Harrison [24] with post-hybridization washes at 83% stringency.

Results

Isolation and structure of the Arachis retroelement: Based on a dot blot survey of short insert libraries of *A. duranensis* and *A. ipaënsis* that were probed with genomic DNA of both species, several clones were identified resembling repetitive sequences. Using the sequences of these clones and additional gene bank sequences, two pseudo-contigs, Rep-1 and Rep-2, were established using the Staden-Package software [19]. A BLASTx search [25] revealed high similarities between one of those contigs (Rep-2) and the reverse transcriptase region of the Ty3-*gypsy* retrotransposon *Athila*. On the other hand, the Rep-1 contig was AT-rich with no similarities to any genic sequences in gene bank. Extended sequence information from both repeats was gained through a modified genome walker strategy, involving the ligation of specific adaptors to genomic restriction fragments and subsequent amplification of sequences adjacent to the repeat clones using adaptor- and repeat-specific primers. From 24 clones isolated from the right site of Rep-1, three had inserts of 1200 to 2000 bps (**Fig. 1**). Those three inserts showed significant similarity ($2e^{-32}$) to the *gag-pol* region of the Ty3-*gypsy* retrotransposons *Ogre* (pea) and *Athila* (*Arabidopsis*). Additionally, the new sequences resembled at their 5'-end the primer-binding site (PBS) of *Calypso*- and *Athila* like retrotransposons, which is complementary to the 3'-end of the Asp tRNA. Translated sequences received from Rep-2 genome walking experiments revealed presence of conserved amino acid motifs that are characteristic for the retrovirus-like Ty3-*gypsy* elements. These results gave first evidence that Rep-1 and Rep-2 are parts of the LTR- and the *pol* region of the same Ty3-*gypsy* retrotransposon. We aimed to obtain the full sequence of this element using the recently generated BAC libraries of the *A. duranensis* and *A. ipaënsis* genomes [26]. Using PCR screening and Southern hybridization, four BAC clones have been identified that contain at least one *rt*- and two or more LTR-fragments, which implies the possibility of isolation of a complete element. According to preliminary results from PCR cloning the total length of the element is bigger than 11,000 bps and does not contain an *env*-typical ORF.

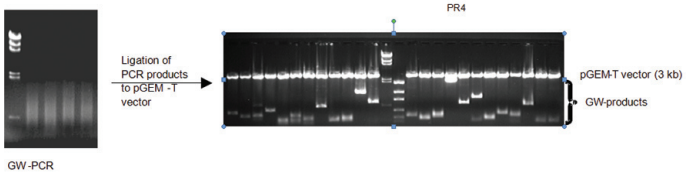

Figure 1 Cloning of sequences downstream of the isolated AT-rich repeat Rep-1. The translated sequence of clone PR4 revealed significant similarity ($2e^{-32}$) to the gag-pol region of the Ty3-gypsy retrotransposons *Athila* in *A. thaliana* and *Ogre* in pea. (GW: genome walking).

Copy Number Estimation: Copy number estimations have been made on the basis of non-radioactive dot blot hybridizations of a retrotransposon subclone against dilution series of itself and of genomic DNA from the three *Arachis* species. Subclone dilutions represented one to 100,000 copies of the retrotransposon sequence in 500 ng of genomic DNA. The strength of hybridization signals in the chemilumigraphs was quantified using the MultiGauge software (Fucifilm) and the slope of the regression curves of the resulting plots have been used for calculating the copy numbers. The representative *rt*-subclone Ah-9 was estimated to be present in the *A. ipaënsis* genome (2C = 2.8 pg) with about 830 copies (equivalent to 0.7% of the genome), in *A. duraensis* (2C = 2.61 pg) with about 2,600 copies (equivalent to 2.3% of the genome), and in *A. hypogaea* (2C = 5.93 pg) with 3,000 copies (equivalent to 1.2% of the genome) per haploid genome. (Genome sizes according to Temsch and Greilhuber [27, 28].)

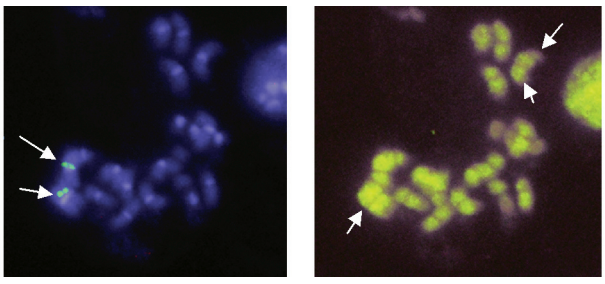

Figure 2 FISH of chromosome spread of A. stenosperma (A-genome). Left: hybridization with the wheat 5S-rDNA probe [29] reveals one sub-centromeric 5S-rDNA site (arrows). Right: re-hybridization with Dig-labeled LTR-probe. Note the lack of hybridization to centromeres and telomeric regions (arrows).

Chromosomal Distribution of retroelement: To determine the chromosomal distribution of the retrotransposon metaphase spreads of *A. duranensis*, *A. stenosperma* (A-genome), and *A. hypogaea* have been hybridized with Dig-labeled LTR- and reverse transcriptase probes.

DAPI staining of the chromosomes revealed centromeric bands, which are typical for the *Arachis* A-genome, whereas B-genome chromosomes do not show such bands [1]. Hybridization of A-genome chromosome spreads with the LTR probe showed dispersed distribution of the element in euchromatic regions of chromosomes and absence from heterochromatic regions, such as the centromer and telomeric regions (**Fig. 2**). When hybridized to metaphase spreads of *A. hypogaea*, the same probe appeared to be localized preferably in the A-genome of *Arachis,* which became obvious by comparison with the DAPI stained spreads allowing differentiation between A- and B- genome chromosomes. This kind of preferential hybridization to the A-genome was also detected when using a clone representing the reverse transcriptase as a probe (data not shown).

Phylogenetic Analysis: From 87 reverse transcriptase sequences isolated and cloned from *A. hypogaea*, *A. duranensis* , and *A. ipaënsis*, 80 have been selected for analysis of their evolution. Using Mega 3.1 the genetic distances between the sequences were calculated (**Table 1**). Three groups are shown, each of them contained all sequences of one species. The distances within each group and between the groups have been analyzed with regard to the nucleotide and as well the amino acid sequences. The data reveal that the DNA sequences isolated from *A. duranensis* suffered less modifications as compared to the ones from *A. hypogaea* and *A. ipaënsis*, which revealed the largest genetic distances. This trend is even more obvious when looking at the amino acid sequences, probably a result of changes in the reading frame that could have generated different amino acids.

Table 1. Genetic distances between DNA sequences and between amino acid sequences of isolated *rt*-genes from *A. hypogaea* (Ah), *A. duranensis* (Ad), and *A. ipaënsis* (Ai). The mean value corresponds to all 80 sequences of DNA and/or amino acids.

Groups	Distance between the sequences (%)	
	Seq. DNA	Seq. amino acids
Ah x Ah	8.68	18.83
Ad x Ad	8.20	17.44
Ai x Ai	8.95	19.87
Mean value	8,61	18.71

Discussion

The present work evolved from our efforts in isolating repetitive elements from the A- and B- genome of *Arachis* aimed at finding out more about the genomic relationships between the ancestors of cultivated peanut. The described element is the first characterized retrotransposon in *Arachis*. Preliminary sequencing results revealed that it belongs to the Ty3-*gypsy* retrotransposon family with high similarities to retrovirus-like elements, particularly to the *Athila* element from *Arabidopsis*. However, it also became clear that this new element has its own characteristics, which make it different from the related elements. FISH analysis has shown its localization in the euchromatic region of the chromosomes, and absence from centromeric and telomeric regions. This kind of distribution pattern was shown to be typical for Ty1-*copia* elements, such as BARE [30]. In contrast, the *Athila* element in *Arabidopsis* is associated with centromeric and pericentromeric regions [31]. We have detected preferential hybridization of the *Arachis* element to A-genome chromosomes. These results were substantiated by copy number estimations, where more than three times less copies were estimated for *A. ipaënsis* than for *A. duranensis*. Since multiple insertions of the element were found in *A. hypogaea*, *A. ipaënsis*, *A. duranensis*, and *A. stenosperma*, it can be assumed that it is an ancient component of the genus. In a phylogenetic analysis of 80 *rt*-sequences isolated from all three species, fewer mutations were found between the sequences derived from *A. duranensis* as compared to the ones from *A. ipaënsis*. We conclude that the element was present in the common ancestor of the *Arachis* wild species at low copy number. After differentiation it amplified in *A. duranensis*, whereas in *A. ipaënsis* the amplification was low, the copy number remained stable or was even reduced through elimination. We could not find evidence for activity of the element by screening all relevant EST databases. It appeared that all sequences isolated included stop codons. Using BAC sequences we are now on the way to a more detailed analysis of this element. Further efforts are directed towards identifying elements, which can be activated thus enabling their use for induced mutations approaches.

ACKNOWLEDGEMENTS

The presented work was funded by EU INCO-DEV (ARAMAP: ICA4-2001-10072), and the Generation Challenge Program (Project # 31, and TLI).

BIBLIOGRAPHY

1. Seijo, J.G., Lavia, G.I., Fernández, A., Krapovickas, A., Ducasse, D., Moscone, E.A. Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaënsis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). *Amer J Bot* **9**, 1294–1303 (2004).
2. Krapovikas, A., Gregory, W.C. Taxonomía del género *Arachis* (Leguminosae). Bonplandia **8**, 1-186 (1994).
3. Sabot, F., Schulman, A.H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* **97**, 381-388 (2006).
4. San Miguel, P., Bennetzen, J.L. Evidence that a recent increase in maize genome size was caused by a massive amplification of intergene retrotransposons. *Ann. Bot.* **82**, 37-44 (1998).
5. Vicient, C. M., Suoniemi, A., Anamthawat-Jónsson, K., Tanskanen, J., Beharay, A., Nevo, E., Schulman, A.H. Retrotransposon BARE-1 and its role in genome evolution in the genus Hordeum. *Plant Cell* **11**, 1769-1784 (1999).
6. Xiong, Y., Eickbush, T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* **9**, 3353-3362 (1990).
7. Kumar, A., Bennetzen, J.L. Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479-532 (1999).
8. Pelissier, T., Tutois, S., Deragon, J.M., Tourment, S., Genestier, S., Picard, G. Athila, a new retroelement from *Arabidopsis thaliana*. *Plant Mol. Biol.* **29**, 441-452 (1995).
9. Chavanne, F., Zhang, D.X., Liaud, M.F., Cerf, R. Structure and evolution of Cyclops: a novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. *Plant Mol Biol* **37** 363-375 (1998).
10. Wright, D.A., Voytas, D.F. Athila 4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses. *Genome Res.* **12**, 122-131 (2002).
11. Yüksel, B., Bowers, J.E., Estill, J., Goff, L., Lemke, C., Paterson, A.H. Exploratory integration of peanut genetic and physical maps and possible contributions from Arabidopsis. *Theor Appl Genet* **111**, 87–94 (2005).
12. Neumann, P., Požárkovà,D., Macas, J. Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. *Plant Mol Biol* **53**, 399-410 (2003).
13. Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., Kanda, M. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA* **93**, 7783-7788 (1996).
14. Sacerdot, C., Mercier, G., Todeschini, A.L., Dutreix, M., Springer, M., Lesage, P. Impact of ionizing radiation on the life cycle of *Saccharomyces cerevisiae Ty1* retrotransposon. *Yeast* **22**, 441-55 (2005).
15. Takeda, S., Sugimoto, K., Otsuki, H., Hirochika, H. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. *Plant Mol Biol.* **36** 365-76 (1998).
16. Kashkush K., Feldman, M., Levy, A.A. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* **160**, 1651-1659 (2002).
17. Miyao, A., Iwasaki, Y., Kitano, H., Itoh, J.I., Maekawa, M., Murata, K., Yatou, O., Nagato, Y., Hiroshika, H. A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol Biol.* **63**, 625-635 (2007).

18. Kalendar, R., Grob, T., Regina, M., Suoniemi, A., Schulman, A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor Appl Genet* **98**, 704-711 (1999).

19. Staden, R., Beal, K.F., Bonfield, J.K. The Staden Package. In: Misener S., Krawetz. Computer Methods in Molecular Biology. *Bioinformatics Methods and Protocols* **132**, 115-130 (1998).

20. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25** 4876-4882 (1997).

21. Kumar, S., Tamura, K., Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163 (2004).

22. Fávero A.P., Nass, L.L., Cargnin, A., Neto, E.L. Rooting performance from leaf petioles of accessions and hybrids of wild *Arachis* species. *Sci. Agric. (Piracicaba, Braz.)* **62**, 62-68 (2005).

23. Maluszynska, J., Heslop-Harrison, J.S. Physical mapping of rDNA loci in *Brassica* species. *Genome* **36**, 774-781 (1993).

24. Schwarzacher, T., Heslop-Harrison, J.S. Practical *in situ* Hybridization. Springer-Verlag, New York (2000).

25. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

26. Guimarâes, P.M, Garsmeur, O., Proite K., Leal-Bertioli, S.C., Seijo, G., Chaine, C., Bertioli, D., D'hont, A. BAC libraries construction from the ancestral diploid genomes of the allotetraploid cultivated peanut. *BMC Plant Biol.* **29**, 8-14 (2008).

27. Temsch, E.M., Greilhuber, J. Genome size variation in *Arachis hypogaea* and *A. monticola* re-evaluated. *Genome* **43** 449-451 (2000).

28. Temsch, E.M., Greilhuber, J. Genome size in Arachis duranensis: a critical study. *Genome* **44**, 826-830 (2001).

29. Gerlach, W.L., Dyer, T.A. Sequence organization of the repeating units in the nucleus of wheat, which contains 5S rRNA genes. Nucleic Acids Res **8**, 4851-4865 (1980).

30. Heslop-Harrison, J.S. *et al.* The chromosomal distributions of *Ty1*-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* **100**, 197-204 (1997).

31. The Arabidospis Genome Initiative, Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. *Nature* **408**, 423-426 (2000).

32. Hirochika, H. Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* **4**, 118–122 (2001).