International Scholars Journals

African Journal of Virology Research ISSN 2756-3413 Vol. 14 (7), pp. 001-005, July, 2020. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Molecular distribution of *gypsy*-like retrotransposons in cotton *Gossypium* Spp.

Essam A. Zaki[1, *] and Abdel Ghany A. Abdel Ghany[2]

[1]Genetic Engineering and Biotechnology Research Institute, GEBRI, Research Area, Borg El Arab, Post Code 21934, Alexandria

[2]Institute of Efficient Productivity, Zagazig University, Zagazig, Egypt.

Accepted 18 April, 2020

PCR primers specific for conserved domains of the reverse transcriptase (RT) genes of *gypsy*-like retrotransposons amplified their corresponding gene in two *Gossypium barbadense* cultivars. Analysis with the FASTA software showed a high DNA sequence homology to pine, *gypsy* LTR-retrotransposon. Using the PCR product as a hybridization probe, *gypsy*-like retrotransposons were detected in wild type species of *Gossypium*, suggesting that *gypsy*-like retrotransposons are present in the *Gossypium* genome. This supports the view that *gypsy*-like retrotransposons are major components of plant genomes. Our results suggest *gypsy*-like retrotransposons have played a fundamental role in the shaping and evolution of the *Gossypium* genome.

Key words: *Gossypium*, *gypsy*, polyploidy, retroelements, retrotransposons, retroviruses, reverse transcriptase.

INTRODUCTION

Long terminal repeat, LTR-retrotransposons, are mobile genetics elements that encode their own reverse transcriptase (RT) for transposition (Malik et al., 2002, for review). Typically, retrotransposons contain two characteristic open-reading frames (ORFs). The *gag* gene (ORF1) encodes structural proteins and *pol* gene (ORF2) encodes protease, RT, RNase H, and integrase domains (Wilhelm and Wilhelm, 2001). Based on the order of the RT, and integrase domains in their pol genes, LTR-retrotransposons are divided into two groups: T*y*1/*copia* and T*y*3/*gypsy* elements (Xiong and Eickbush, 1990). The *gypsy* group elements have a similar organization to retroviruses and some encode a third ORF in the position occupied by the envelope, *env*, gene in retroviruses (Song et al., 1994).

The reverse transcriptase (RT) genes have conserved amino acid domains, some of which are characteristic of each retroelement group (Xiong and Eickbush 1990). In plants, degenerate oligonucleotide primers have been designed to amplify these domains by PCR and used for detection and assessment of their distribution and evolution (Flavell et al., 1992; Voytas et al., 1992; Matsuoka 1999; Friesen et al., 2001). These results showed that retrotransposons are ubiquitous throughout the plant kingdom and constitute a major portion of the nuclear genomes (Kumar and Bennetzen, 1999). Therefore, the detailed characterization of different plant taxa with respect to the content, variability, and physical distribution of retrotransposons makes a major contribution to the understanding of host genome organization and evolution (Bennetzen, 2000).

The cotton genus, *Gossypium* L., comprise approximately 45 diploid and five allopolyploid species distributed throughout the arid and semi-arid regions of Africa, Central and South America, the Indian subcontinent, Arabia, the Galapagos, and Hawaii (reviewed by Wendel and Cronn, 2002). The diploid *Gossypium* species fall into eight cytological groups or genome designated A through G, and K. The five allotetraploid species, indigenous to the New World, derive from a single allopolyploidization event that united the Old World A genome with the New World D genome, in an A genome cytoplasm (Wendel, 1989). Here, we aimed to characterize *gypsy* -like retrotransposons in cotton. Using PCR primers specific for conserved domains of RT genes of *gypsy*-like retroelements, we studied their molecular distribution in representative *Gossypium* species. The analysis of the molecular existence and distribution of *gypsy*-like retrotransposons in cotton allows retrotransposon evolution to be studied in two levels: first,

*Corresponding author: Phone (765) 494-9837, Fax: (765) 496-1496, E-mail: ezaki@purdue.edu.

Current Address: Department of Biological Sciences, 1392 Lilly Hall of Life Sciences, West Lafayette, IN 47907-1392, USA.

Abbreviations: Env, envelope gene; LTR, long terminal repeat; ORF, open-reading frame; PCR, polymerase chain reaction; RT, reverse transcriptase gene.

in the diploids, subsequent to their divergence from a common ancestor, and second, in the polyploid, after retrotransposons have been reunited in a common genome (Abdel Ghany and Zaki, 2002).

MATERIALS AND METHODS

Plant materials

Gossypium wild type species and cultivars were kindly provided by Dr. A. Percival. Their genome descriptions are listed in Table 1.

Table 1. Description of *Gossypium* species studied.

Taxon	Genome*
G. herbaceum	A_1
G. arboreum	A_2
G. raimondii	D_5
G. gosspioides	D_6
G. barbadense	$(AD)_2$
G. darwinii	$(AD)_5$
G. mustelinum	$(AD)_4$
G. hirsutum	$(AD)_1$

*Genomic designations taken from Wendel (1989).

DNA extraction

Total DNA was extracted using Qiagen DNeasy kit (Qiagen, Hilden, Germany).

Isolation of T*y3*/*gypsy*-like reverse transcriptase gene in *Gossypium*

Total DNA was subjected to PCR with primers specific to the T*y3*/*gypsy* reverse transcriptase gene. The forward and reverse primer sequences are (5`-CTAAGCTTTAYCAYCARHTNMGNAT-3`) and (5` - TCGAATTCTGNCCNARNMDYTYNAC-3`), where N= A+C+G+T, R= A+G, and Y= T+C (Xiong and Eickbush 1990). DNA amplifications were carried in an ABI GeneAmp PCR system 9700 cycler with a denaturing step at 95°C for 5 min and the step cycle program set for 45 cycles (with a cycle consisting of denaturing 94°C for 30s, annealing at 50°C for 30s and extension step at 72°C for 30s), followed by a final extension step at 72°C for 10 min. PCR fragments were then cloned in pCR 4-TOPO vector with TOPO TA cloning kit (Invitrogen, USA) in the competent *E. coli* strain TOPO 10. Plasmid DNA was isolated using QIA Spin mini-prep kit (Qiagen, Hilden, Germany). Plasmid DNA was sequenced in both directions using BigDye Sequencing Kit and ABI 3700 DNA sequencer (ABI, USA). Pairwise DNA sequence alignments were carried out using CLUSTALW (http://www2.ebi.ac.uk/clustalw; Thompson et al. 1994).

DNA HYBRIDIZATION

PCR amplified probes were labelled with [α-^{32}P] dCTP using the random primer method (Feinberg and Vogelstein, 1983), and used for DNA/RNA, and slot -blot hybridization as described (Sambrook et al., 1989). Filters were hybridized overnight at 42°C in a solution containing (50% foramide, 5 x SSC, 10 x Denhardt's, and 0.5% SDS). Hybridization wash was carried out at 50°C in 0.1 x SCC containing 0.5% SDS for 1 h.

RESULTS

The search for T*y3*/*gypsy*-like reverse transcriptase gene in two cultivated *G. barbandense* cultivars, Giza 45 and 84, was carried out using specific oligonucleotide for the T*y3*/*gypsy* reverse transcriptase gene (Xiong and Eickbush, 1990). This led in each case to the detection of 310 bp fragment. These fragments were designated G45 and G84 respectively. These fragments were cloned in pCR 4-TOPO vector, and plasmid DNAs were sequenced in both directions. G45 and G84 complete DNA sequence is 310 bp (Figures 1 and 2). G45 and G84 DNA sequences were deposited in the NCBI nucleotide sequence database, GenBank; with the accession numbers U75247, U75248, respectively.

```
TCGAGATTCGGACGTACCCAAAACTGCTTTCAGAACGAGGTACGGTCACTACGAGTTCTTAGTGA
TGCCGTTCGGGCTCACTAATGCCCCTGCGGTATTTATGGATTTGATGAATCGGATCTTCAGACAG
TATTTGGACCGGTTCGTAGTTGTGTTCATTGATGACATCTTGGTCTATTCAGGAGATGAGACCGA
ACATGCTGAGCACCTGAGATTAGTGTTGCAAATTTTGCGGGATAAGCAGTTATATGCTAAGTTCA
GTAAGTGTGAGTTCTGGTTAAGAGAGGTTAGCTTCTTGGGTCATGTGGTA
```

Figure 1. DNA nucleotide sequence of G45. DNA sequence is given 5`-3`.

```
TGCGCGAGGGAGATGAGTGGAAAATAGCATTTAAAACTAAACATAGTTTGTACGAGTGGTTGGTA
ATGCCATTCGGTCTTACTAACACTTCTAGTACCTTTATGCGTTTAATGAACCATGTTCTGCGAGC
TTTCATTGGTAAATTTTGTGTGGTTTACTTTGATGATATTCTTGTTTATAGTAGATCCTTAGATG
ATCACTTAAAGCATTTACGAGCTGTTCTTGATGTGTTGCGAAAGGAAAATTTATATGCCAATCTT
AAGAAATGCACATTTTGTTCTAACCAAGTTGTTTTCCTAGGATTTGTTGT
```

Figure 2. DNA nucleotide sequence of G84. DNA sequence is given 5`-3`.

```
G45 -TCGAGATTCGGACGTACCCAAAACTGCTTTCAGAACGAGGTACGGTCACTACGAGTTCT 59
G84 TGCGCGAGGGAGATGAGTGGAAAATAGCATTTAAAACTAAACATAGTTTGTACGAGTGGT 60
      ** **    ** *     ****       *********   *  **   *******  *

G45 TAGTGATGCCGTTCGGGCTCACTAATGCCCCTGCGGTATTTATGGATTTGATGAATCG-G 118
G84 TGGTAATGCCATTCGGTCTTACTAACACTTCTAGTACCTTTATGCGTTTAATGAACCATG 120
    * ** ***** ***** ** *****  *   **      ****** *** ***** *   *

G45 ATCTTCAGACAGTATTTGGACCGGTTCGTAGTTGTGTTCATTGATGACATCTTGGTCTAT 178
G84 TTCTGCGAGCTTTCATTGGTAAATTTTGT-GTGGTTTACTTTGATGATATTCTTGTTTAT 179
     *** *    *  * ****       ** ** ** ** * * ******* **  * ** ***

G45 TCAGGAGATGAGACCGAACATGCTGAGCACCTGAGATTAGTGTTGCAAATTTTGCGGGAT 238
G84 AGTAGATCCTTAGATGATCACTTAAAGCATTTACGAGCTGTTCTTGATGTGTTGCGAAAG 239
       **          ** **      ****  *  **   **  * *   * *****  *

G45 AAGCAGTTATATGCTAAGTTCAGTAAGTGTGAGTTCTGGTTAAGAGAGGTTAGCTTCTTG 298
G84 GAAAATTTATATGCCAATCTTAAGAAATGCACATTTTGTTCTAACCAAGTTGTTTTCCTA 299
     *  * ******** **  * *   ** **       *****  *  * ***    *** *

G45 GGTCATGTGGTA 310
G84 GGATTTGTTGT- 310

    **   *** **
```

Figure 3. Comparative DNA sequence analysis of G45 and G84 using CLUSTALW.

Comparative nucleotide sequence analysis of G45 and G84 using the ClustalW program revealed that G45 and G84 are closely related, producing alignment score of 88.8% (Figure 3). Analysis with the FASTA software (Pearson and Lipman, 1988) using EMBL and EMBL

plant database information indicated high homology of the pine, *Ananas comosus* , *gypsy* LTR-retrotransposon (Thomson et al., 1998), 68.5% identity (72.619% ungapped) and 68.5% identity (72.615% ungapped) with G45 and G84 respectively. It is noteworthy that G45 and G84 did not show close correspondence to previously reported T*y*1/*copia* retrotransposons in *G. hirsutum* (Vanderwiel et al., 1993), nor to the major dispersed repeats in the cotton genome (Zhao et al., 1998), suggesting that a very complex population of such elements exists in cotton.

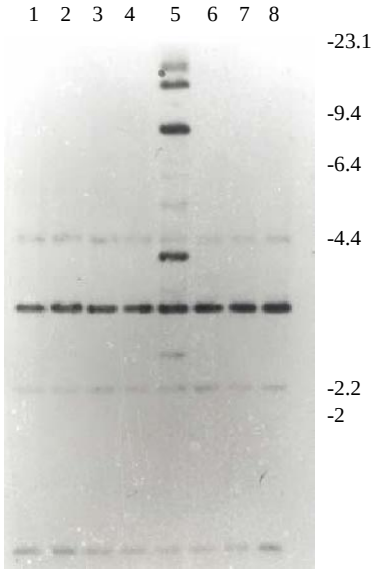

Figure 4. Molecular distribution of *gypsy*-like-related sequences in *Gossypium* species. *Gossypium* DNAs were digested with *Hin*dIII and hybridized with G45 ^{32}P-labeled specific probe. Lane 1: *G. herbaceum*; Lane 2: *G. arboreum*; Lane 3: *G. raimondii*; Lane 4: *G. gosspioides*; Lane 5: *G. barbadense*; Lane 6: *G. darwinii*; Lane 7: *G. mustelinum*; Lane 8: *G. hirsutum*. Sizes of molecular weight markers, in kb, are indicated on the right.

The molecular distribution of *gypsy*-like retrotransposons in wild type species of the genus *Gossypium* was investigated using Southern blot. Total genomic DNAs from different *Gossypium* species were digested with *Hin*dIII and hybridized with G45 reverse transcriptase domain as a probe as shown in Figure 4 and Table 2. The detection of multiple hybridizing bands suggests the repetitive nature of these retroelements in the *Gossypium* genome. In addition, the pattern of the distribution of *gypsy*-like retrotransposons in these genomes is similar, with four common hybridization bands of 6.8, 4, 2.5, and 0.3 kb respectively. The observation of different band intensities within a single track may be attributed to sequence variations between *gypsy*-like retrotransposon-related sequences in these genomes. The fact that these genomes display a similar set of *gypsy*-like retrotransposon-specific restriction

Table 2. Sizes of *gypsy* -like-related sequences in *Gossypium* species. A: *G. herbaceum*; B: *G. arboreum*; C: *G. raimondii*; D: *G. gosspioides*; E: *G. barbadense*; F: *G. darwinii*; G: *G. mustelinum*; H: *G. hirsutum*. Sizes are given in kb.

A	B	C	D	E	F	G	H
				25			
				23			
				17			
				10			
				9			
6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8
				6.4			
4	4	4	4	4	4	4	4
				3			
2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3

fragments suggest that the genomic locations of *gypsy*-like retrotransposon are similar within these different genomes. *G. Barbadense* (lane 5), however, showed a different pattern by possessing seven hybridization bands of 25, 23, 17, 10, 9, 6.4, and 3 kb respectively. The presence of sequences ranging from larger than 23 kb to smaller than 2 kb indicates the presence of a large family of related sequences.

DISCUSSION

Retrotransposons due to their abundance, diversity, and widespread distribution, make a major contribution to the shape as well as size of plant genomes (Bennetzen, 2002). PCR assays have been developed for isolation of amplified fragments of RT genes of retrotransposons using degenerate primers to conserved amino acid domains in many species (Flavell et al., 1992; Voytas et al., 1992; Matsuoka and Tsunewaki, 1999; Friesen et al., 2001). In the current study, degenerate primers designed for *gypsy* elements amplified reverse transcriptase gene in two cultivated *G. barbadense*. In addition, with the PCR product as a hybridization probe, it was revealed that *gypsy*-like retrotransposons can be detected in wild type species of *Gossypium*, suggesting that *gypsy*-like retrotransposons is a standard component of the *Gossypium* genome, and supporting the fact that *gypsy*-like retrotransposons represents a major component of the plant genome (Freschotte et al., 2002).

The distribution pattern of *gypsy*-like retrotransposons within the genus *Gossypium* is similar, with *G. barbadense* possessing additional hybridisation bands. The additional hybridization bands in *G. barbadense* can be explained by one or more of the following mechanisms: a) deletions or additions in or near the *gypsy*-like retrotransposon element, loss or acquisition of restriction sites by single base changes or by changes in DNA modification such as methylation; b) homologous recombination of two *gypsy*-like retrotransposon

elements resulting in deletion or inversion of the DNA between the *gypsy* -like retrotransposon elements on one chromosome, or in reciprocal translocation of chromosomal arms; c) transposition of a *gypsy*-like retrotransposon element including the reverse transcriptase domain to a new chromosomal location (Bennetzen, 2002). The cytosine residues in the CpG sequence are known to be frequently methylated in mammalian DNA (Barker et al., 1984). In plants, the cytosine residues in the CpNpG sequence are also methylated (Gardiner-Garden et al., 1992). *Hind*dIII, the restriction enzyme used in this study, is sensitive to cytosine methylation. Hence, a part of the additional hybridization bands detected with this enzyme may be explained by methylation. However, two lines of evidence strongly, suggest that *gypsy*-like retrotransposon is an active retrotransposon in *G. barbadense*. First, *gypsy*-like retrotransposon -encoded transcripts were detectable (Abdel Ghany and Zaki, 2002). Second, an envelope, *env*-like sequences were reported in cultivated *G. barbadense* (Abdel Ghany and Zaki, 2002). Detection of *gypsy*-like retrotransposon-encoded transcripts and *env*-like sequences, intermediates in the retrotransposition process (Kumar and Bennetzen, 1999), suggests that a subset of these molecules is competent for retrotransposition.

Plant retrotransposons are largely inactive during normal development, but can be activated by stresses, such as wounding, pathogen attacks, and cell culture (Kumar and Bennetzen, 1999). Additionally, it has been documented that hybridization can activate retrotransposons (Liu and Wendel, 2000). Hybridization is a prominent feature of natural plant populations, and one that often leads to speciation either through polyploidization or at the diploid level (Wendel, 2000) . In this regard, the fact that *G. barbadense* represents an allopolyploid cotton that appears to have arisen as a consequence of trans-oceanic dispersal of an A-genome taxon to the New World followed by hybridization with an indigenous D-genome diploid (Wendel and Cronn, 2002) raises the question to the effect that hybridization may have had on the *gypsy*-like retrotransposons in *G. barbadense*. The detection of additional hybridization bands in *G. barbadense* could be the result of the hybridization process. The absence of these additional bands in the other allopolyploids, namely, *G. darwinii*, *G. hirsutum* and *G. mustelinum*, however, argues against this suggestion. Alternatively, these additional *gypsy*-elements could be the result of a massive amplification process of *gypsy*-like retrotransposons in *G. barbadense* since the divergence of *G. barbadense*, and the remaining allopolyploids from a common ancestor. This suggestion is supported by the massive amplification process observed in maize since its divergence from the common ancestor with sorghum (SanMiguel and Bennetzen, 1998). Further experimental data such as copy number determination, chromosomal distribution, and sequencing of large contiguous regions of the genome as it has been previously demonstrated in maize (SanMiguel and Bennetzen, 1998) will significantly add up fundamental knowledge about the role of *gypsy*-like retrotransposons in shaping and evolution of the *Gossypium* genome.

Finally, the unique properties of retrotransposons have been exploited as genetic tools for plant genome analysis (Kumar and Hirochika, 2001). In this regard, our results in this report and the detection of *env*-like sequences in *Gossypium* (Abdel Ghany and Zaki, 2002) promote the initiative to employ these retroelements in cotton genome analysis. *Gypsy*-like elements should provide the means to fulfil major applications such as determining phylogeny and genetic diversity and in functional analyses of genes in *Gossypium*.

ACKNOWLEDGEMENT

This work was supported by a grant from the US-Egypt Science and Technology Foundation to E.A. Zaki.

REFERENCES

Abdel Ghany AG, Zaki EA (2002). Cloning and sequencing of an *envelope*-like gene in *Gossypium*. Planta 216:351-353.

Barker D, Schafer M, White R (1984). DNA methylation in mammalian cells. Cell 36:131-138.

Bennetzen JL (2002). Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 115:29-36.

Bennetzen JL (2000). Transposable elements contributions to plant gene and genome evolution. Plant Mol. Biol. 42: 251-269.

Flavell AJ, Dunbar R, Anderson R, Pearce SR, Hartley R, Kumar A (1992). T*y*1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res. 20:3639-3644.

Feinberg AP, Vogelstein B (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6-13.

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

Friesen N, Brandes A, Heslop-Harrison J (2001). Diversity, origin and distribution of retrotransposons in conifers. Mol. Biol. Evol. 18:1176-1188.

Gardiner-Garden M, Sved JA, Frommer M (1992). DNA methylation in plant cells. J. Mol. Evol. 34:219-230.

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

Kumar A, Hirochika H (2001). Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci. 6:127-134.

Liu B, Wendel JF (2000). Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874-880.

Malik HS, Henikoff S, Eickbush TH (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. 10:1307-1318.

Matsuoka Y, Tsunewaki K (1999) Evolutionary dynamics of T*y*1-*copia* group retrotransposons in grass shown by reverse transcriptase domain analysis. Mol. Biol. Evol. 16:208-217.

Pearson WR, Lipman DJ (1988). Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA. 85:2444-2448.

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SanMiguel P, Bennetzen JL (1998). Evidence that a recent increase in

maize genome size was caused by a massive amplification of intergene retrotransposons Ann. Bot. 82:37-44.

Song SU, Gerasimova M, Kurkulos M, Boeke JD, Corces VC (1994). An *env*-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. Genes Dev. 8:2046-2057.

Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-6480.

Thompson KG, Thomas JE, Dietzgen RG (1998). Retrotransposon-like sequences integrated into the genome of pineapple, *Ananas comosus*. Plant Mol. Biol. 38:461-465.

Vanderwiel PS, Voytas DF, Wendel JF (1993). *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). J. Mol. Evol. 36:429-447.

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel S (1992). *Copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA. 89:7124-7128.

Wendel JF, Cronn R (2002). Polyploidy and the evolutionary history of cotton. Advances in Agronomy 78:139-186.

Wendel JF (2000). Genome evolution in polyploids. Plant. Mol. Biol. 42:225-249.

Wendel JF (1989). New world cottons contain old world cytoplasm. Proc. Natl. Acad. Sci. USA. 86:4132-4136.

Wilhelm M, Wilhelm FX (2001). Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. 58:1246-1262.

Zhao XP, Si Y, Hanson RE, Crane CF, Price JH, Stelly DM, Wendel JF, Paterson AH (1998). Dispersed repetitive DNA spread to new genomes since polyploid formation in cotton. Genome Res. 8:479-492.