International Scholars Journals

Frontiers of Agriculture and Food Technology ISSN 7295-2849 Vol. 10 (2), pp. 001-004, February, 2020. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Sequence heterogeneity of the *envelope*-like domain in the Egyptian cotton *Gossypium barbadense*

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

[1]Institute of Efficient Productivity, Zagazig University, El Zagazig, Egypt. [2]Genetic Engineering and Biotechnology Research Institute, GEBRI, Research Area, Borg El Arab, Post Code 21934, Alexandria, Egypt.

Accepted 15 September, 2019

The current study aimed to investigate the evolution of *env*-like sequences in the Egyptian cotton *Gossypium barbadense*. DNA sequence determination and analysis of *env*-like sequences revealed that these sequences are heterogeneous in *G. barbadense*. The observed sequence diversity, however, seems to preserve the coding information. Phylogenetic analysis demonstrated that plant *env* -like sequences group together, suggesting their monophyletic origin. *Gossypium env*-like sequences are, however, more closely related to elements present in other plant species. Our result suggests that *env*-like sequences in cotton have evolved under functional constraint and likely to play a role in the life cycle of these elements.

Key words: Envelope, *Gossypium*, *gypsy*, plant retroviruses, retrotransposons, retroviruses, reverse transcriptase.

INTRODUCTION

Retrotransposons have been found in the genomes of most eukaryotes (for review see Eickbush and Malik, 2002). Their integrated proviral forms consist of two long open reading repeats (LTRs) flanking an internal region which contains one to three open reading frames (ORFs) coding for structural and enzymatic functions for their replication cycle (Wilhelm and Wilhelm, 2001). Based on their reverse transcriptase (RT) domains, retrotransposons were divided into two major groups: the T*y*1/*Copia* and the T*y*3/*Gypsy* families (Xiong and Eickbush, 1990). They differ by the order of enzymatic domains in the *pol* gene. Moreover, the T*y*3/*gypsy* family is more closely related to vertebrate retroviruses. The viral envelope (*env*) gene of the retroviruses distinguishes them from retrotransposons. Structural and functional data converged when it was shown that the *gypsy* element of *D. melanogaster* was able to function as a retrovirus (Kim et al., 1994, Song et al., 1994). Recently, the International Committee on Taxonomy of Viruses (ICTV) has proposed to term the T*y*1/*Copia* and the T*y*3/*Gypsy* families *Pseudoviridae* and *Metaviridae*, respectively (Boeke et al., 2000). The *Metaviridae* are further classified according to the presence of the *env* gene (genus *Errantivirus*) or its absence (genus *Metavirus*) (Hull, 2001).

Plant retrotransposons with *env*-like genes have been reported (Zaki 2003, for review). They include: the *Athalia*/*Tat*1 clade of *Arabidopsis thaliana* (Wright and Voytas, 1998), the related legume elements *Cyclops* of pea and *Calypso* of soybean (Chavanne et al., 1998;

*Corresponding author; Essam A. Zaki, Current Address: Department of Biological Sciences, 1392 Lilly Hall of Life Sciences, West Lafayette, IN 47907-1392, Phone (765) 494-9837 Fax (765) 496-1496, E-mail: ezaki@purdue.edu.

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

Peterson-Burch et al., 2000), the *Bagy*-2 elements of barely (Vicient et al., 2001), and the *GM*5 and *GM*6 elements of cotton (Abdel Ghany and Zaki, 2002). Interestingly, a unique T*y*1/*Copia* *env*-containing element, *SIRE*-1 has also been described for soybean (Laten et al., 1988).

Phylogenetic analyses based on reverse transcriptase amino acid sequences strongly suggest that the retroviral *env* gene transduced an *env* gene from a baculoviral source (Malik et al., 2000). In plants, a recent study has indicated that *gypsy*-like retrotransposon: *Bagy*-2 of barely defines a lineage of endogenous plant retroviruses (Vicient et al., 2001). In this regard, the fact that *gypsy*-like elements and *env*-like genes have been previously described in *Gossypium* (Abdel Ghany and Zaki, 2002, Zaki and Abdel Ghany, 2003), has promoted the initiative to search for *Bagy*- 2 *env* -domain in the cotton genome. In addition, this study also aims to investigate the evolution of *env*-like sequences in the Egyptian cotton *G. barbadense*.

MATERIALS AND METHODS

Plant materials and genomic DNA extraction

Total genomic DNA was extracted from the *Gossypium barbadense* cultivar S14, young seedlings, using Qiagen DNeasy kit (Qiagen, Germany).

Isolation of *Bagy*-2 *env* domains in *Gossypium*

Total DNA was subject to PCR with primers specific to the *env*-domain of *Bagy*-2 retrotransposon, (5`-TCAGTTGCAAGAAAGTCG CCG-3`) and (5`-CCTCTATCAGTGTTTCGGGGC-3`) (Vicient et al., 2001). 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 55°C for 30s and extension step at 72°C for 30s), followed by a final extension step at 72°C for 10 min.

Cloning and sequencing of PCR-amplified fragments

Expected PCR-amplified fragments were excised from the agarose gel and purified using Qiagen Gel Extraction kit (Qiagen, Germany). Purified DNA 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, Germany). Plasmid DNA was sequenced in both directions using BigDye Sequencing Kit and ABI 377 DNA sequencer (ABI, USA).

Alignments and phylogenetic analysis

Pairwise and multiple DNA sequence alignment were carried out using CLUSTALW (1.82) (http://www2.ebi.ac.uk/clustalw; Thompson et al. 1994). Bootstrap neighbor-joining tree (Saitou and Nei, 1987) was generated using MEGA 2.1 (Kumar et al., 2001) from CLUSTALW alignments.

RESULTS AND DISCUSSION

PCR amplification with primers specific for the *env* domain of the barely *Bagy* -2 retrotransposon (Vicient et al., 2001) was employed to search for the *env*-like domains in *G. barbadense*. Expected amplicons were cloned in pCR 4-TOPO vector. Two *G. barbadense* recombinant clones were randomly selected and further studied by DNA sequence analysis. These clones were designated GB and GB1, respectively. GB and GB1 sequences have been deposited in the NCBI nucleotide sequence database, GenBank; with the accession numbers AY257162 and AY257163 respectively. Blast search confirmed the *env* nature of the cloned products. Furthermore, GB and GB1 derived amino acid sequences are compared to the *Bagy*-2 *env* domain (Vicient et al., 2001) in Figure 1, with amino acid similarities of 67% and 80%, respectively. The high amino acid similarities observed supports the interpretation that GB and GB1 represent portions of the *env* gene of *Bagy*-2 retrotransposon.

```
1 QETRRDKQGLRLLPLVREALLELHMSASRLRWRSLLFIGTRLFLPLGIIVLFLVNGPAIWFQ
2 QEARRDTQGLRLLPMVREALLELHMSASRLRWRILLFIGTRSFLPLGLIVLFDVSGPAIWFQ
3 QEARRDKQGLRLLPMVREALLQLHMSVSRLRWRILLFIGTRSSLPPWLILLFLIRPPTIWFP
  ** *** ******************** ****** *******  **  *  **    * ***
```

Figure 1. Alignment of the putative *envelope*-like domain of *G. barbadense*: GB, GB1, and barley *Bagy*-2 retrotransposon (Vicient et al., 2001). Identical amino acids are identical by asterisks. Key; 1: Bagy-2, 2: GB1, 3: GB.

Comparative nucleotide and amino acid sequences analysis of GB and GB1 using ClustalW program revealed identities of 75% and 72%, respectively (Figure 2). The level of nucleotide and amino acid identities observed for GB and GB1 is comparable to that reported for the *Bagy*-2 element, where 86% identity between the genomic copies was observed (Vicient et al., 2001). Despite the fact that multiple gaps were introduced for GB and GB1 at the nucleotide sequence analysis to compensate for the sequence length polymorphism and deletions, yet it seems to preserve the coding information evident to the overall high amino acids homology. A similar pattern of length variation, deletions and coding information conservation was recently reported in the *SIRE*-1 elements of soybean (Laten et al., 2003).

We have previously identified *env*-like genes in *Gossypium* using specific oligonucleotides for the *Drosophila gypsy* *env*-gene (Abdel Ghany and Zaki, 2002). Comparative amino acid sequences of *env*-like sequences in *Gossypium* were performed (Figure 3). Moreover, nucleotide pairwise comparisons revealed a diversity range of 7 to 81% among Gossypium *env*-like elements. GM5 and GM6 are clearly closely related to each with nucleotide and amino acid sequences identities of 81% and 79%, respectively.

A.

```
GB   --GATT---GCCCATACCGATGCTTAGTTGTCTTCCTTT-GGTCGGTGATGGCAGGTGGA  54
GB1 GGGACTTGGAACCATATGGCTGGACCACTGACGTCGAAGAGGACTATGAGCCCAAGGGGA 60
       ** *     ***** * **        *****       ** *  ***   ** * ***

GB   GTTGTTGAAGCAGTCTTGA---------GCTTGTCTAATTTCCACTTGAGCAAAAAGTTC  105
GB1 AGGAACGAACGAGTTCCGATGAAGAGGAGGATCCTCCATCTCAACCTGGACGCACTCATG 120
            ***   *** **           *  *         ******** *  *    *

GB   TGCTCCCTTAGATCGTCATTTTCCTCCTCAAGTTTCAACACTCTATGGCAAAGTTCCTGC  165
GB1 TGGAGTTCAAGAAGTCAAGCCTCCCTGACCATAGGAAGAAGCCTAAGACCTTGTGTATCC 180
    **       ***      *   ***    * *    *  *     *****  **    * *

GB   TTACTCTCCTGCAAGAAGCGAGAAGGGATAAACAAGGTCTTAGGCTTCTTCCTATGGTCA  225
GB1 CTTCTCGCTTCTTGCAGGAGAGTAAGCAGGAACTTTGCCATAGAGTGTTGAAACTTGAGG 240
     ******            ********  ***   *****     *  *    * *

GB   GGGAGGCTTGACTTCTTCAACTCCACATGAGTGTCTCCCGGTTGAGAT--GGCGGATCCT  283
GB1 AGGAAAATGACGATCTAAGGGAGCAGAACTTTTTGCTCAAGTGGAAATTAGACAAGCTCA 300
     ***   *      ***       ** *    * *      *   ** **    *      *

GB   CCTCTTCATCGG-AACTCGTTCGTCCCTTCCCCCTTGGCTCATACTCCTCTTCCTCATCC  342
GB1 AGACTGGTTCAACAACTCCACCG-CCATCCCCACCAAAGGAAGAGAACTAAGCCTTGGGT 359
       **   **   ****     ** **  *** *        * *   **   ***

GB   GTCCGCCCACCATATGGTTTCCAGGTCCCATATACC 378
GB1 ATGGGCAATCCCCTT---------------------- 374
     *   **  **    *
```

B.

```
GB   --LPIPMLSCLPLVGDGRWSCSSLELVFPLEQKVLLPIVIFLLKFQHSMAKFLLTLLQEA  58
GB2 KGIAHTQGLVLFLWWGWRWSCTSLELVFPLEQKVLLPIVIFLLKFQHSMAKFLLTLLQEA 60
       .       * * .  ****:**************************************

GB   RRDKQGLRLLPMVREA-LLQLHMSVSRLRWRILLFIGTRSSLPPWLILLFLIRPPTIWFP  117
GB2 RRDTQGLRLLPMVREA-LLELHMSASRLRWRILLFIGTRSFLPLGLIVLFDVSGPAIWFQ 119
    ***.************ **:****.*************** **  **:** :  *:***

GB   GPIY 121
GB2 --VH 121
      ::
```

Figure 2. Comparative DNA (A) and amino acid (B) sequences analysis of *G. barbadense* putative *envelope*-like elements: GB and GB1 using CLUSTALW.

```
GB    --LPIPMLSCLPLVGDGRWSCSSLELVFPLEQKVLLPIVIFLLKFQHSMAKFLLTLLQEA 58
GB1   KGIAHTQGLVLFLWWGWRWSCTSLELVFPLEQKVLLPIVIFLLKFQHSMAKFLLTLLQEA 60
GM5   ------------          HGG-FARPSFS---PWGAPVLF----------------------          V 19
GM6   ------- LSCKKVADRG-FARPSFS---PWGAPVLF----------------------          V 26
                        :: .*:.    *    **:                          .

GB    RRDKQGLRLLPMVREA-LLQLHMSVSRLRWRILLFIGTRSSLPPWLILLFLIRPPTIWFP 117
GB1   RRDTQGLRLLPMVREA-LLELHMSASRLRWRILLFIGTRSFLPLGLIVLFDVSGPAIWFQ 119
GM5   KKKDGTMRMC-----------------   IDYRQLNKVTVKNKYP--- LPRNTDRG------ 53
GM6   KKKDGTMRMC-----------------   IDYRQLNKVTXKNKYP--- L------------- 53
      ::.    :*:                    ::**       ::.*   :

GB    GPIY 121
GB1   --VH 121
GM5   ----
GM6   ----
```

Figure 3. Comparative amino acid sequence analysis of *G. barbadense* putative *envelope*-like elements: GB, GB1, GM5 and GM6 using CLUSTALW.

Relationships among *Gossypium env*-like genes and other organisms were assessed by constructing a neighbor-joining tree (Saitou and Nei, 1987), with accession numbers on the tree, and the *Drosophila gypsy* as the outgroup (Figure 4). The phylogenetic analysis revealed high level of amino acid sequences diversity as evident by the branch lengths which are proportional to the degree of divergence. In addition, plant *env*-like sequences group together, suggesting their monophyletic origin. *Gossypium env* -like sequences are, however, more closely related to elements present in other plant species. GM5 has the strongest affinity with soybean element *SIRE*-1. On the other hand, GB and GB1 closest homologue is that of barley *Bagy*-2 element.

In this study, we investigated the evolution of *env*-like sequences in the Egyptian cotton. Our analysis revealed that these sequences are comprised of a very heterogeneous collection of *env*-like sequences. The observed sequence diversity, however, seems to maintain the ORFs and thus preserve coding information. This suggests that the *env*-like sequences in cotton have evolved under functional constraint and likely to play a role in the life cycle of these elements. This suggestion is supported by the presence of conserved ORFs coding for *env*-like sequences that can be identified across diverse plant taxa (Zaki, 2003). It is noteworthy that such functional constraint contrasts with what has been found in mammalian retroviral *env* genes, where adaptive selection results in high levels of variation to avoid the immune response (Coffin et al., 1997).

Phylogenetic analysis revealed that *Gossypium env*-likes sequences are more closely related to elements present in other plant species. We previously reported that *gypsy* group retrotransposons is a standard component of the *Gossypium* genome (Zaki and Abdel Ghany, 2003). The detection of *env* -like sequences in the cultivated *G. barbadense* suggests that these sequences are probably very ancient sequences maintained in the genome because they are located in chromosomal locations where the recombination rate is low and selection therefore less efficient. It should be noted that the chromosomal locations for *gypsy* group elements are

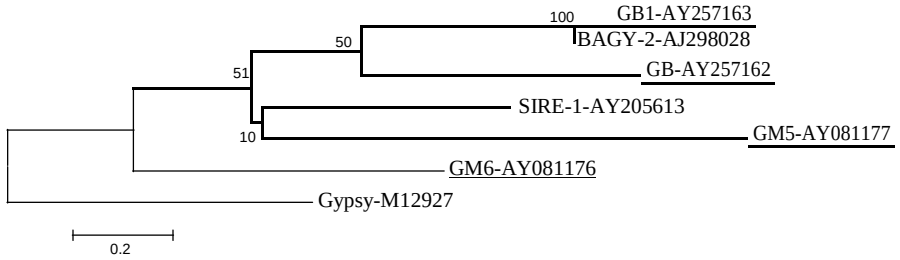

Figure 4. Phylogenetic tree showing relationship between *envelope* domain amino acid sequences of *G. barbadense* (underlined), plant and *Drosophila gypsy* group retrotransposons. The Neighbor-Joining method (Saito and Nei, 1987) was employed to construct the tree, with branch lengths propotional to the degree of divergence between the amino acid sequences. The numbers on the branches represent bootstrap value of 1,000 replicates. Names refer to the accession number of the nucleotide sequences that encode the corresponding *envelope* domain.

yet to be determined in *Gossypium*. Alternatively, these sequences could be the result of a recent transposition burst event of *gypsy* group elements in *G. barbadense* or acquisition of new elements by horizontal transfer. The distribution pattern of *gypsy* group retrotransposons within the genus *Gossypium* is similar, with *G. barbadense* possessing additional hybridisation bands (Zaki and Abdel Ghany, 2003), supports the recent transposition suggestion. Similarly, a massive amplification process was observed in maize since its divergence of sorghum from a common ancestor (SanMiguel and Bennetzen, 1998). Currently, it is difficult to envisage the horizontal transfer of *gypsy* group elements in *G. barbadense*, especially that *gypsy* group retrotransposons have been detected in the genus *Gossypium*. Further experimental data such as copy number determination, chromosomal distribution, and sequencing of large contiguous regions of the *Gossypium* genome will significantly add up fundamental knowledge about the role of *gypsy* group retrotransposons in shaping and evolution of the *Gossypium* genome.

The mammalian retroviral *env* gene is a highly diverging sequence in relation to the highly diverse sequences of the receptor molecules with which *env* proteins interact for virus-cell interaction and entry (Coffin et al., 1997). Nevertheless, the elucidation of various retroviral *env* complexes show a highly conserved structural conservation possibly reflecting a common mechanism for mammalian retroviruses for triggering the fusion and entry process (Eickbush and Malik, 2002). In this regard, the functional role for plant retroviruses for viral propagation in the plant host is still unknown, and cell walls rule out membrane fusion as a suitable invasive strategy (Zaki, 2003). The identification of a replication-competent plant retrovirus is imperative to determine its functional significance. In addition, to test the hypothesis that plant retroviruses are infectious. Finally, to elucidate the unique biology of plants that has helped to restrict the pathogenicity of retroviruses within the animal kingdom.

ACKNOWLEDGEMENT

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

REFERENCES

Abdel Ghany AA, Zaki EA (2002) . Cloning and sequencing of an *envelope*-like gene in *Gossypium*. Planta 216:351-3.

Boeke JD, Eickbush TH, Sandmeyer SB, Voytas DF (2000). Metaviridea. In: Murphy FA (eds) Virus Taxonomy: ICTV VIIth Report. New York, USA.

Chavanne F, Zhang DX, Liaud MF, Cerff R (1998). Structure and evolution of T*y*3/*Gypsy* family highly amplified in pea and other legume species. Plant. Mol. Biol. 37:363-.375.

Coffin JM, Hughes SH, Varmus HE (1997). Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Eickbush TH, Malik HS (2002). Origins and evolution of retrotransposons. In: Craig et al. (eds) Mobile DNA II. ASM Press, USA, pp 1111-1144.

Hull R (2001). Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. Arch. Virol. 146:2255-2261.

Kim A, Terzian C, Santamaria P, Pelisson A, Purd'Homme N, Bucheton A (1994). Retroviruses in invertebrate: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA. 91:1285-1289.

Kumar SK, Tamura K, Jakobsen IB, Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244-1245.

Laten HM, Majumdar A, Gaucher EA (1998). *SIRE*-1, a *copia*/T*y*1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc. Natl. Acad. Sci. USA. 95:6897-6902.

Laten HM, Havecker ER, Farmer LM, Voytas DF (2003). SIRE-1, an endogenous family from Glycine max, is highly homogenous and evolutionary young. Mol. Biol. Evol. 20:1-13.

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

Peterson-Burch BD, Wright DA, Laten HM, Voytas DF (2000). Retroviruses in plants? TIG. 16:151-152.

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

SanMiguel P, Bennetzen JL (1998). Evidence that a recent increase in maize genome size was caused by a massive amplification of intergene retrotransposons. Ann. Botany 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.

Vicient CM, Kalendar R, Schulman AH (2001). *Envelope*-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res. 11:2041-2049.

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

Wright DA, Voytas DF (1998). Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. Genetics. 149:703-715.

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

Zaki, EA (2003). Plant retroviruses: structure, evolution and future applications. Afr. J. Biotechnol. 2:136-139.

Zaki EA, Abdel Ghany AA (2003). Molecular distribution of *gypsy*-like retrotransposons in cotton *Gossypium* Spp. Afr. J. Biotechnol. 2:124-128.