Isolation, characterization and phylogenetic analysis of *Bagy*-2 retrotransposon *envelope*-domain in the Egyptian cotton *G. barbadense*

(Received: 09.12.2004; Accepted: 26.12.2004)

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

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

[**]Genetic Engineering & Biotechnology Research Institute, GEBRI, Research Area, Borg El Arab, Post Code 21934, Alexandria, Egypt.

* Corresponding author: Essam A. Zaki, Current Address: Department of Biological Sciences, 1392 Lilly Hall of Life Sciences, West Lafayette, IN 47907-1392, USA.

E-mail: ezaki@purdue.edu.

ABSTRACT

In this study, the presence of Bagy-2 env domains in the Egyptian cotton, G. barbadense, using modified PCR program was investegated Comparative DNA sequence of cotton env clones revealed the presence of sequence length polymorphisms and deletions. Nevertheless, the coding information seems to be preserved. The ratio of synonymous to nonsynonymous changes indicates that the env domain in cotton is evolving under purifying selection. Moreover, env sequences in cotton have evolved under functional constraints and likely to play a role in the life cycle of these elements. Our phylogenetic analysis illustrates that cotton env sequences closest homologue is that of barley Bagy-2 retroelement.

Key words: *Drosophila, Envelope, Gossypium, Metavirus, Retroelements, Retrotransposons, Retroviruses.*

Abbreviations

Env: envelope gene. LTR: long terminal repeat. ORF: open reading frame. PCR: polymerase chain reaction.

INTRODUCTION

Retrotransposons have been found in the genomes of most eukaryotes (for review, 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 *Ty1/copia* and the T*y*3/*gypsy* families *Pseudoviridae* and *Metaviridae* respectively (Boeke *et al*., 2000). The *Metaviridae* is further classified according to the presence of the *env* gene (genus *Errantivirus*) or its absence (genus *Metavirus*) (Hull, 2001).

The plant retrotransposons T*y*3/*gypsy* group with *env*-like genes have been previously 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 (Peterson-Burch *et al*., 2000), the *Bagy*-2 elements in barley (Vicient *et al.*, 2001), and the *GM*5 and *GM*6 elements in *Gossypium* (Abdel Ghany and Zaki, 2002). Interestingly, a unique *Ty1/copia* group *env*-containing element, *SIRE*-1 has also been described for soybean (Laten *et al*., 1998).

The *Bagy*-2 element from barley was recently shown to be widely spread in the grasses (Vicient *et al*., 2001). However, the authors reported the inability to amplify *Bagy*-2 *env* domains from plant species outside the grasses. The current work reports the isolation, characterization and phylogenetic analysis of *Bagy*-2 *env* domains in the Egyptian cotton *G. barbadense*.

MATERIALS AND METHODS

DNA extraction

Total DNA was extracted from *Gossypium barbadense* cultivar S14 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`-TCAGTTGCAAGAAA-GTCGCCG-3`) and (5`-CCTCTATCAGTG-TTTCGGGGC-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 1 min and extension step at 72°C for 2 min), followed by a final extension step at 72°C for 10 min. Extension temperature was modified by a ramping of lower 5% of the default value.

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). Phylogenetic and molecular evolutionary analyses were conducted using MEGA 2.1 (Kumar *et al*., 2001) from CLUSTALW alignments.

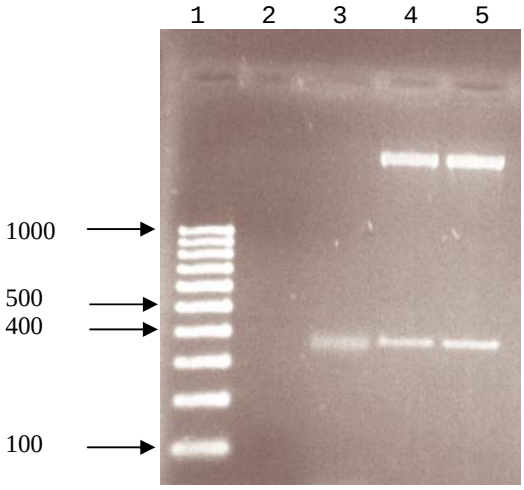
Env protein motif analysis

The presence of transmembrane domains was predicted using SOSUI (http://sosui.proteome.bio.tuat.ac.jp;

(Hirokawa *et al*., 1998), PHDhtm and TMpred (Hofmann and Stoffel, 1993; Rost *et al*., 1995).

RESULTS AND DISCUSSION

Barely *Bagy*-2 element encodes a predicted *env* domain with conserved features (Vicient *et al*., 2001). Moreover, primers specific for its *env* domain amplified its corresponding domains only in other cereals. Vicient *et al.* (2001) suggested that this could be due to the extreme sequence heterogeneity of *env* domains in plants. Taking into consideration *Bagy*-2 widespread in grasses and the fact that Ty3/*gypsy* group plant retrotransposons represent a standard component of plant genomes (Feschotte *et al*., 2002), we favored an alternative suggestion to explain this result that is *env* primers were not provided with sufficient time to seek their corresponding sequences. To test this suggestion, a modified PCR program was employed to search for *Bagy*-2 *env* domains in *G. barbadense*. The modified program includes two alterations from the original program described by Vicient *et al*., (2001). First, annealing and extension temperatures were set for 1 and 2 min respectively. Secondly, a ramp with 5% slower of the default value was introduced in the extension temperature. The detection of an amplicon of approximately 400 bp in *G. barbadense* (Fig. 1, Lane 3), similar to the expected size previously detected in barley, suggests that this amplicon may represent *Bagy*-2 *env*-like domain. The fact that this amplicon was only detected in the modified program and not in the original (Fig. 1, Lane 2) suggests that the above mentioned alternations were effective not only to provide sufficient time for *env* primers to seek their corresponding sequences, but allowing the opportunity for the formation of a stable *env* complex substrate for DNA polymerase.



Lane 1: 100 bp ladder DNA marker.
Lane 2: Non-modified program.
Lane 3: Modified program.
Lane 4: Recombinant GB1 plasmid cut with *Eco*RI.
Lane 5: Recombinant GB2 plasmid cut with *Eco*RI.

Fig. (1): Amplification of Bagy-2 env domain in G. barbadense using modified PCR program.

Expected amplicons were excised, purified from agarose (Fig. 1, Lanes 4 and 5), and finally 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 GB1 and GB2, respectively. GB1 and GB2 DNA sequences were deposited in the NCBI nucleotide sequence database, GenBank; with the accession numbers AY257162, AY257163, respectively. GB1 and GB2 derived amino acid sequences are compared to the *Bagy*-2 *env* domain (Vicient *et al*., 2001) in (Fig. 2), with amino acid similarities of 56% to 70%, respectively. The high amino acid similarities observed support the interpretation that GB1 and GB2 sequences generated in this study represent portions of the *env* gene of *Bagy*-2 retrotransposon.

```
CLUSTALW (1.82) multiple sequence alignment

Sequence 1: GB1        121 aa
Sequence 2: GB2        121 aa
Sequence 3: HGV1       128 aa

Sequences (1:3) Aligned. Score:  67
Sequences (2:3) Aligned. Score:  80

GB2  KGIAHTQGLVLFLWWGWRWSCTSLELVFPLEQKVLLP-IVIFLLKFQHSMAKFLLTLLQE 59
HGV1 QGDCPYQCLVVFLWW-WWWSCSSLELVFPLEHKVLLLQIVIFLLKLQHSMAKFLLTLVQE 59
GB1  LPIPMLSCLPLVGDG--RWSCSSLELVFPLEQKVLLP-IVIFLLKFQHSMAKFLLTLLQE 57
           . * :.      ***:*********:****  *******:***********:**

GB2  ARRDTQGLRLLPMVREA-LLELHMSASRLRWRILLFIGTRSFLPLGLIVLFDVSGPAIWF 118
HGV1 TRRDKQGLRLLPLVREA-LLELHMSASRLR-RSLLFIGTRLFLPLGIIVLFLVNGPAIWF 117
GB1  ARRDKQGLRLLPMVREA-LLQLHMSVSRLRWRILLFIGTRSSLPPWLILLFLIRPPTIWF 116
     :***.*******:**** **:****.**** * *******  **  :*:** :  *:***
GB2  QVH------ 121
HGV1 QVPIDLYLS 126
GB1  PGPIY---- 121
```

Fig. (2):Comparative amino acids sequence analysis of G. barbadense GB1, GB2 and barley Bagy-2 env (HGV1) domain (Vicient et al., 2001) using CLUSTALW.

Key:

"*" means that the residues or nucleotides in that column are identical in all sequences in the alignment. ":" means that conserved substitutions have been observed. "." means that semi-conserved substitutions are observed.

Comparative nucleotide and amino acid sequences analysis of GB1 and GB2 using ClustalW program revealed homologies of 75% and 72%, respectively (Fig. 3). The level of nucleotide and amino acids identity observed for GB1 and GB2 is comparable to that reported for the *Bagy*-2 element, where 86% similarity between the genomic copies was observed (Vicient *et al*., 2001). Despite the presence of sequence length polymorphisms, deletions and several gaps at the nucleotide sequence analysis, yet it seems that they did not affect the coding information evident to the overall high amino acid

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).

A) DNA sequence:

```
CLUSTAL W (1.82) multiple sequence alignment

Sequence 1: GB1-AY257162      378 bp
Sequence 2: GB2-AY257163      374 bp

Alignment Score 75

GB1   --GATT---GCCCATACCGATGCTTAGTTGTCTTCCTTT-GGTCGGTGATGGCAGGTGGA 54
GB2   GGGACTTGGAACCATATGGCTGGACCACTGACGTCGAAGAGGACTATGAGCCCAAGGGGA 60
        ** *     *****  * **      ** * **     ** *  ***   ** * ***

GB1   GTTGTTGAAGCAGTCTTGA---------GCTTGTCTAATTTCCACTTGAGCAAAAAGTTC 105
GB2   AGGAACGAACGAGTTCCGATGAAGAGGAGGATCCTCCATCTCAACCTGGACGCACTCATG 120
             ***  ***  **          *  *     ** ** ** **  *  *    *

GB1   TGCTCCCTTAGATCGTCATTTTCCTCCTCAAGTTTCAACACTCTATGGCAAAGTTCCTGC 165
GB2   TGGAGTTCAAGAAGTCAAGCCTCCCTGACCATAGGAAGAAGCCTAAGACCTTGTGTATCC 180
      **       ***     *   ***    * *     *  *  *** * *   **   * *

GB1   TTACTCTCCTGCAAGAAGCGAGAAGGGATAAACAAGGTCTTAGGCTTCTTCCTATGGTCA 225
GB2   CTTCTCGCTTCTTGCAGGAGAGTAAGCAGGAACTTTGCCATAGAGTGTTGAAACTTGAGG 240
       * *** * *     * * *** * * *  ***   * * ***  *  *     * *

GB1   GGGAGGCTTGACTTCTTCAACTCCACATGAGTGTCTCCCGGTTGAGAT--GGCGGATCCT 283
GB2   AGGAAAATGACGATCTAAGGGAGCAGAACTTTTTGCTCAAGTGGAAATTAGACAAGCTCA 300
       ***   *     ***       ** *    * *      *  ** **    *     *

GB1   CCTCTTCATCGG-AACTCGTTCGTCCCTTCCCCCTTGGCTCATACTCCTCTTCCTCATCC 342
GB2   AGACTGGTTCAACAACTCCACCG-CCATCCCCACCAAAGGAAGAGAACTAAGCCTTGGGT 359
         **   **   ****    ** **   *** *       * *   **    ***

GB1   GTCCGCCCACCATATGGTTTCCAGGTCCCATATACC 378
GB2   ATGGGCAATCCCCTT-------------------- 374
       *  **   **   *
```

B) Amino acids sequence:

```
Sequence 1: GB1-AY257162    121 aa
Sequence 2: GB2-AY257163    121 aa

Alignment Score 72

GB1   --LPIPMLSCLPLVGDGRWSCSSLELVFPLEQKVLLPIVIFLLKFQHSMAKFLLTLLQEA 58
GB2   KGIAHTQGLVLFLWWGWRWSCTSLELVFPLEQKVLLPIVIFLLKFQHSMAKFLLTLLQEA 60
         .      * *  .  ****:**************************************

GB1   RRDKQGLRLLPMVREA-LLQLHMSVSRLRWRILLFIGTRSSLPPWLILLFLIRPPTIWFP 117
GB2   RRDTQGLRLLPMVREA-LLELHMSASRLRWRILLFIGTRSFLPLGLIVLFDVSGPAIWFQ 119
      ***.************ **:****.*************** **  **:** :  *:***

GB1   GPIY 121
GB2   --VH 121
        ::
```

Fig. (3): Comparative DNA nucleotide and amino acid sequences analysis of GB1 and GB2 using CLUSTALW.

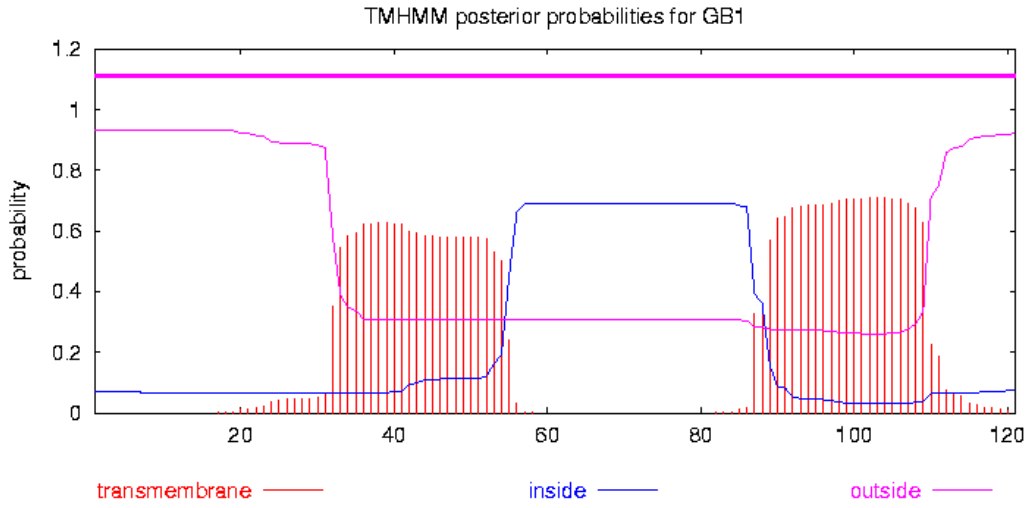
Synonymous and nonsynonymous nucleotide substitutions (d_S/d_N) in the putative *env* domain of GB1 and GB2 were studied in detail (Table 1). It is known that (d_S/d_N) can be informative with respect to the strength and direction of selection (Yang and Bielawski, 2000). Results from Table (1) yield no evidence of positive selection as d_S is slightly higher than d_N. The synonymous and nonsynonymous ratio is, therefore, high enough to infer that the *env* domain in *G. barbadense* has been under purifying selection.

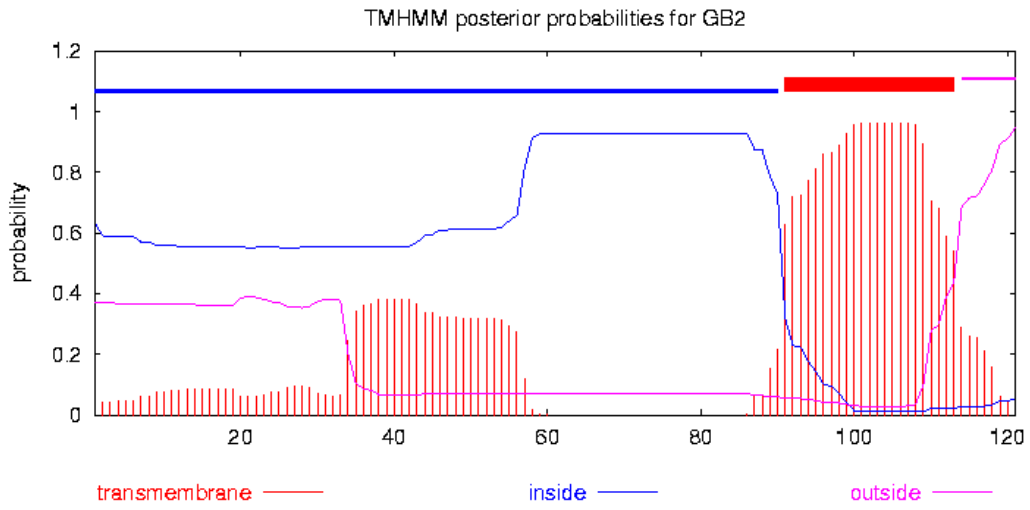
Table (1): Numbers of synonymous and nonsynonymous substitutions per site in the env domain of GB1 and GB2.

S: Synonymous substitutions	0.608	(± 0.054)
N: Nonsynonymous substitutions	0.580	(± 0.032)
d_S/d_N	1.7	(± 0.062)
s: No. of synonymous sites	80.167	(± 2.557)
n: No. of nonsynonymous sites	246.833	(± 2.652)

Numbers of synonymous and nonsynonymous substitutions and the standard errors (in parentheses) were respectively estimated according to Nei and Gojobori (1986).

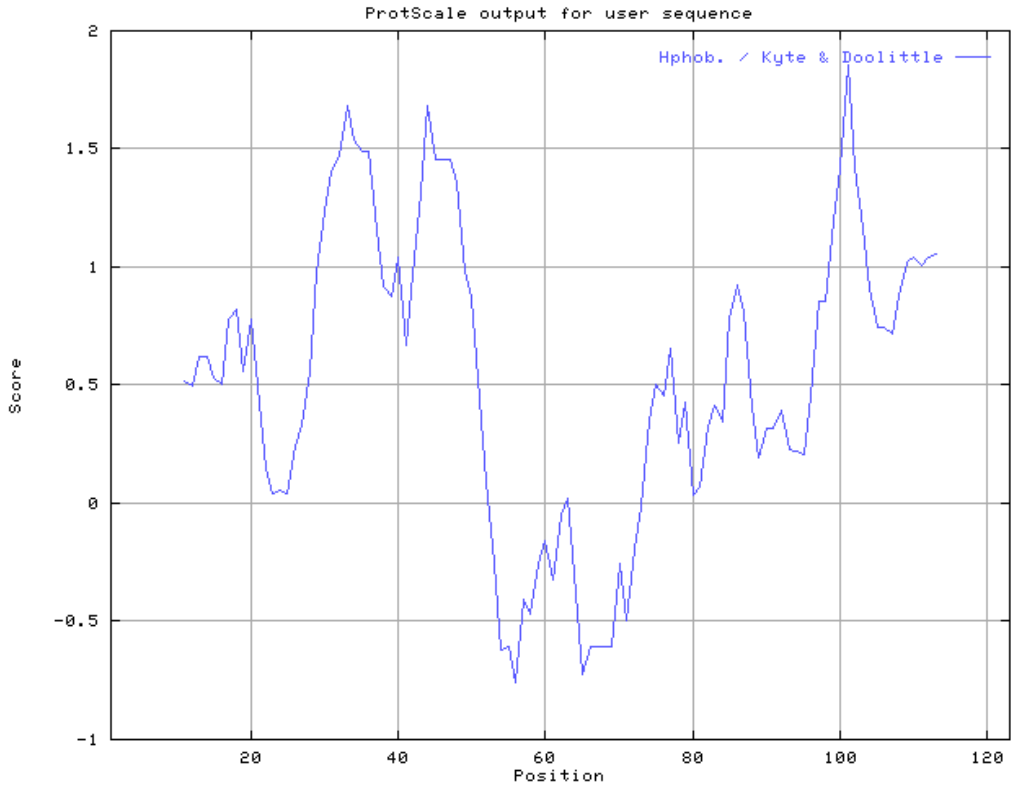
The predicted GB1 and GB2 *env* sequences were examined for diagnostic motifs found in the *env* genes. Retroviral *env* proteins are typically transported through the endomembrane system, where they are proteolytically cleaved to generate surface (SU) and transmembrane (TM) proteins prior to being released on the cell surface (Coffin *et al*., 1977). A structural predication algorithm effective for the TM domain (Rost *et al*., 1995; Hirokawa *et al*., 1998) strongly predicted the presence of hydrophobic, membrane spanning helices in the putative GB1 and GB2 translation products (Fig. 4), as expected for *env* genes. Moreover, the program TMpred assigned scores of 1239 and 1740 for GB1 and GB2, respectively (scores above 500 are considered significant; Hofmann and Stoffel, 1993).



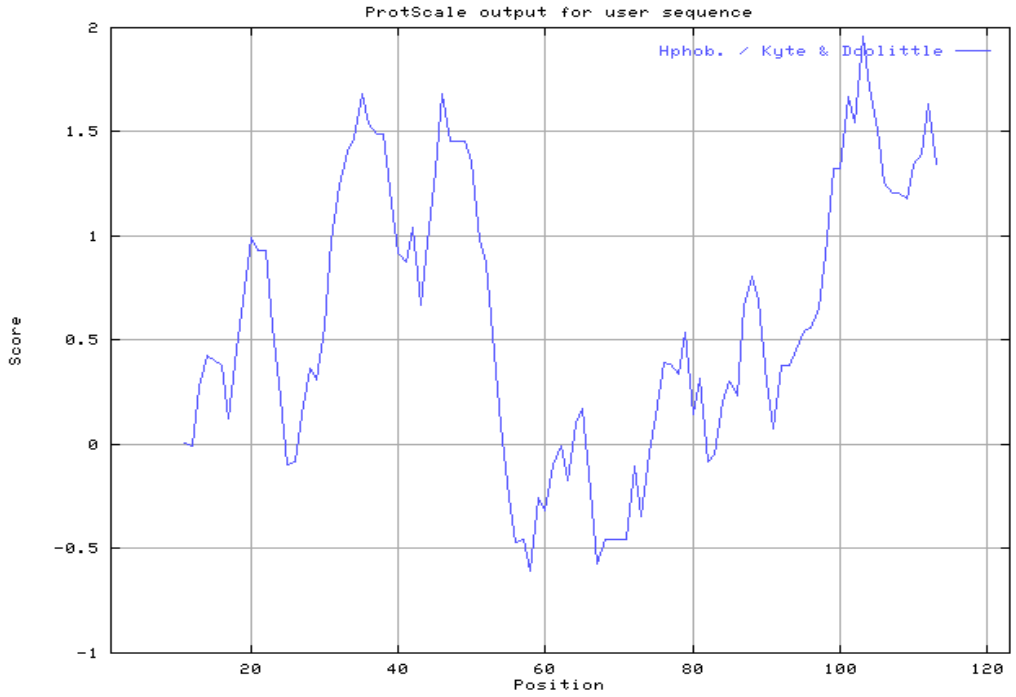


A) Probability plot for occurrence of transmembrane domains.

i- GB1



ii- GB2



B) Hydrophilicity plot for the predicted *env* protein.

Fig. (4): Conserved motifs of the putative env domain of G. barbadense GB1 and GB2.

Relationships among *G. barbadense 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 (Fig. 5). The phylogenetic analysis revealed high level of amino acid sequences diversity as evident to the branch lengths which are proportional to the degree of divergence. In addition, plant *env*-like sequences group together, suggesting their monophyletic origin. *G. barbadense env*-like sequences are, however, more closely related to elements present in other plant species. GB1 and GB2 closest homologue is that of barley *Bagy*-2 retroelement.

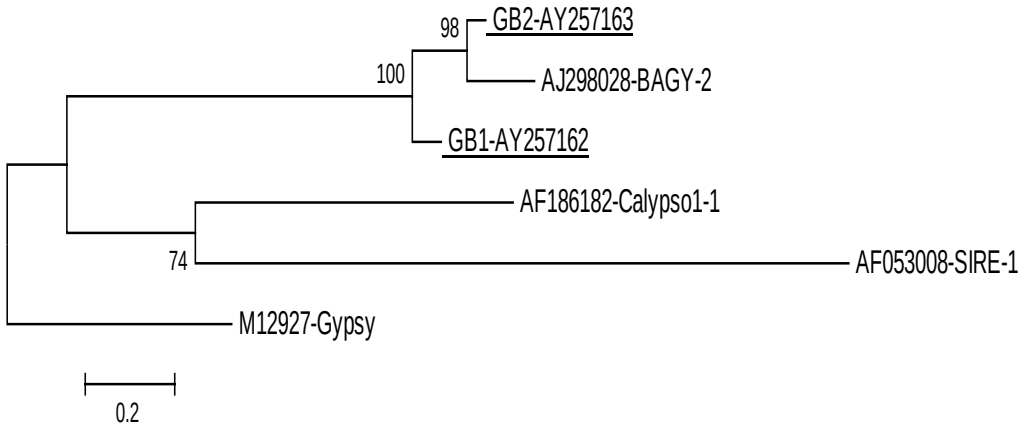


Fig. (5): Phylogenetic tree showing relationship between envelope domain amino acid sequences of G. barbadense (underlined), Ty3/gypsy plant and Drosophila gypsy group retrotransposons. The Neighbor-Joining method (Saito and Nei, 1987) was employed to construct the tree, and the Drosophila gypsy as the outgroup. 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.

In this study, we investigated the presence of *Bagy*-2 *env* domains in the Egyptian cotton. This was carried out using modified PCR program. The modified PCR program is based on providing *env* primers sufficient time to seek their corresponding sequences, and thus allowing the formation of a stable *env* complex substrate for DNA polymerase. Accordingly, the modified PCR program promotes the opportunity in the filed of evolutionary genetics for cloning such sequences across taxonomic groups. Comparative DNA sequence of cotton *env* clones revealed the presence of sequence length polymorphisms and deletions. Nevertheless, the coding information seems to be preserved. The ratio of synonymous to nonsynonymous changes indicates that the *env* domain in cotton is evolving under purifying selection. Moreover, *env* sequences in cotton have evolved under functional constraints and likely to play a role in the life cycle of these elements. 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).

Acknowledgements

This work was supported by a grant from the US-Egypt Science & Technology Foundation, National Academy of Science to E.A. Zaki.

REFERENCES

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

Boeke, J.D., Eickbush, T.H., Sandmeyer, S.B. and Voytas, D.F. (2000). Metaviridea. In: Murphy FA (eds) Virus Taxonomy: ICTV VIIth Report, New York: Springer-Verlag.

Coffin, J.M., Hughes, S.H. and Varmus, H.E. (1997). Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Eickbush, T.H. and Malik, H.S. (2002). Origins and evolution of retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington, D.C., pp 1111-1144.

Feschotte, C., Jiang, N. and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329-41.

Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14:378-379.

Hofmann, K. and Stoffel, W. (1993). TMbase, a database of membrane spanning protein segments. Biol Chem 347:166.

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. and 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, S.K., Tamura, K., Jakobsen, I.B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244-1245.

Latin, H.M., Majumdar, A. and Gaucher, E.A. (1998). *SIRE*-1, a *copia*/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc Natl Acad Sci USA 95:6897-6902.

Laten, H.M., Havecker, E.R., Farmer, L.M. and Voytas, D.F. (2003). SIRE-1, an endogenous family from *Glycine max*, is highly homogenous and evolutionary young. Mol Biol Evol 20:1-13.

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

Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.

Peterson-Burch, B.D., Wright, D.A., Laten, H.M. and Voytas, D.F. (2000). Retroviruses in plants? TIG 16:151-152.

Rost, B., Casadio, R., Farselli, P. and Sander, C. (1995). Transmembrane helices predicted at 95% accuracy. Protein Sci 4:521-533.

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

Song, S.U., Gerasimova, M., Kurkulos, M., Boeke, J.D. and Corces, V.C. (1994). An *env*-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. Genes Dev 8:2046-2057.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTALW: 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, C.M., Kalendar, R. and Schulman, A.H. (2001). *Envelope*-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res 11:2041-2049.

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

Wright, D.A. and Voytas, D.F. (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.

Yang, Z. and Bielawski, J.P. (2000). Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:469-503.

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

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

الملخص العربي

عزل ، توصيف وتحديد العلاقات الوراثية الجزيئية لمشابه العامل الو راثي المتنقل *Bagy-2* في القطن المصري

عبدالغني عبد الحكيم عبد الغني* وعصام زكي**
*معهد الكفاية الانتاجية- جامعة الزقازيق- الزقازيق- مصر
**قسم بحوث الاحماض النووية-معهد بحوث الهندسة الوراثية و التكنولوجيا الحيوية- مدينة مبارك للأبحاث العلمية-اسكندرية-مصر

بالرغم من أن العامل الوراثي المتنقل *Bagy-2* يعتبر من أكثر العوامل الوراثية المتنقلة *gypsy*-group retrotransposons انتشارا داخل الجينوم الخاص بالنباتات وحيدة الفلقة Monocots مثل الشعير والارز والذرة، الا ان *Bagy-2* لم يتم حتي الان عزله وبالتالي دراسته داخل الجينوم الخاص بالنباتات ثنائية الفلقة Dicots وذلك للدرجة العالية من التغيير الجزيئي sequence heterogeneity الخاصة بهذا العامل الوراثي ، لذلك فان الهدف من الدراسة الحالية هو التوصل الي طريقة باستخدام أساليب الهندسة الوراثية لعزل مشابهه العامل الوراثي *Bagy-2* من الجينوم الخاص بالقطن المصري.
أوضحت الدراسة انه قد تم التوصل الي طريقة جديدة لعملية التفاعل البلمري المتسلسل(PCR) لعزل مشابهه العامل الوراثي *Bagy-2* من الجينوم الخاص بالقطن المصري، متيحا الفرصة الي أستخدام هذة الطريقة الجديدة لعزل عوامل وراثية تتمييز بدرجة عالية من التغيير الجزيئي sequence heterogeneity من كائنات مختلفة داخل السلم التطوري.
وبتحديد التتابع النيكلوتيدى DNA sequence لعدد أثنين من مشابهه العامل الوراثي *Bagy-2* من الجينوم الخاص بالقطن المصري وجد أنه يوجد درجة عالية من التماثل علي مستوى التتابع النيكلوتيدى والاحماض الامنية مع العامل الوراثي *Bagy-2*، وأن هذة المشابهات تتمييز بالنوع التطوري purifying selection ، وعند تحديد العلاقات الوراثية الجزيئة وجد ان العامل الوراثي المتنقل *Bagy-2* الخاص بنبات الشعير يعتبر من أكثر العوامل الوراثية المتنقلة المرتبطة وراثيا وتطوريا بالعومل الوراثية الخاصة بنبات القطن.