African Journal of Biotechnology Vol. 4 (11), pp. 1275-1280, November 2005
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2005 Academic Journals

Full Length Research Paper

Phylogenetic and molecular evolutionary analyses of Ty1-copia group retrotransposons in cultivated Egyptian cotton, *Gossypium barbadense* L.

Elsayed E. Hafez and Essam A. Zaki*

Nucleic Acids Research Department, Genetic Engineering & Biotechnology Research Institute, GEBRI, Research Area, Borg El Arab, Post Code 21934, Alexandria, Egypt.

Accepted 24 June, 2005

The Ty1-copia group retrotransposons have been characterized in cultivated Egyptian cotton, *G. barbadense* L., using degenerate PCR primers for their reverse transcriptase (RT) domains. Comparative nucleotide and amino acid sequences analyses showed that *G. barbadense* Ty1-copia RT sequences are heterogeneous and this heterogeneity is resolved into 11 distinct families. The high ratio of synonymous to nonsynonymous changes indicates that there is a strong selection for the RT domain of these families. Phylogenetic analysis revealed that two cultivated *G. barbadense* RT families are closely related to Ty1-copia group retrotransposons present in other plant species. In other words, these families mirror their own phylogenies rather than that of their host cultivars. On the other hand, the remaining *G. barbadense* RT families are closely related to their respective *Gossypium* species. These data show that Ty1-copia group retrotransposons tend to span species boundaries, suggesting that they existed early in plant evolution, and were diverged into heterogeneous sequences prior to modern plant species divergence.

Key words: Genome structure, *Gossypium*, repetitive DNA, polyploidy, sequence diversity, retrotransposons.

INTRODUCTION

The cotton genus, *Gossypium*, is a model system for examining many fundamental aspects of genome structure and evolution, plant development, and crop productivity. *Gossypium* consists approximately of 50 diploid and allopolyploid species distributed in arid to semi-arid regions of the tropic and subtropics (reviewed in Wendel and Cronn, 2003 and references therein). While all the diploid *Gossypium* species have the same chromosome number (n=13), their haploid genome sizes vary from 1 to 3.5 gb (Wendel and Cronn, 2003), as this variation is largely due to the different amounts of dispersed repetitive DNA (Zhao et al., 1998). *G. barbadense* was brought in Egypt in the late 19th century and this was followed by the Egyptian breeders' efforts to develop several elite *G. barbadense* cultivars, including the Giza cultivars (Anonymous, 1992). Egyptian cottons are worldwide renowned for being the highest fiber quality among the world's cottons (Anonymous, 1997).

In plants, Ty1-copia group retrotransposons have been extremely successful as evident to their abundance, and they constitute a major portion of the nuclear genomes (Kumar and Bennetzen, 1999). Their ubiquity in the plant kingdom suggests that they are of very ancient origin (Casacuberta and Santiago, 2003). In addition, their abundance has played a major role in plant genome structure and evolution (Freschotte et al., 2002).

We have previously isolated and characterized Ty1-copia group retrotransposons in the Egyptian cotton *G. barbadense* L. and its progenitors (Abdel Ghany and Zaki, 2003a). Our results revealed that both vertical transmission of Ty1-copia group retrotransposons within *G. barbadense* lineages, and horizontal transmission

*Corresponding author Email: gossypium@link.net, Tel: (+203) 459-3413, Fax: (+203) 459-3423.

Abbreviations: PCR, polymerase chain reaction; RT, reverse transcriptase gene.

between *G. barbadense* and its progenitors have played major roles in the evolution of Ty1-copia group retrotransposons in *Gossypium* (Abdel Ghany and Zaki, 2003a). Herein, we further elucidate the structure, function and evolution of Ty1-copia group retrotransposons in cultivated Egyptian cottons *G. barbadense* L.

MATERIALS AND METHODS

Plant materials and genomic DNA extraction

G. barbadense cultivars, employed in the current study, are listed in Table 1. Total DNA was extracted using Qiagen DNeasy kit (Qiagen, Germany).

Table 1. *G. barbadense* cultivars used in this study: isolated clones and their GenBank accession numbers.

Cultivar	Clone	Accession no.
Giza7	G7	U75221
Giza31	G31	U75246
Giza36	G36	U75240
Giza67	G67	U75223
Giza70	G70	U75225
Giza71	G71	U75226
Giza75	G75	U75227
Giza77	G77	U75228
Giza83	G83	U75229
K101	K101	U75239
Brown Egyptian	BEG	U75235

PCR

Total DNA was subject to PCR with specific primers to amplify an approximately 280 bp region of the Ty1-copia group retrotransposons reverse transcriptase (RT) as described previously (Abdel Ghany and Zaki, 2003a). Briefly, 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 30 s, annealing at 47°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.

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 competent Esherichia 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 was generated using MEGA 3 (Kumar et al., 2004) from CLUSTALW alignments.

Accession numbers

DNA sequences, reported in the current study, were deposited in the NCBI nucleotide sequence database, GenBank, and are listed in Table 1.

Table 2. Amino acid similarities (%) between cultivated *G. barbadense* sequences and corresponding portions of T*y*1-*copia* RT domains from *Drosophila* (*copia*; Mount and Rubin, 1985), *Nicotiana* (Tnt1; Grandbastien et al., 1989), and *Arabidopsis* (Ta1-2; Voytas and Ausubel, 1988).

Clone	*Copia*	*Tnt*1	*Ta*1-2
G36	46	44	43
G67	43	44	45
G83	40	46	41
K101	43	43	43
G75	46	47	41
G7	40	42	43

RESULTS AND DISCUSSION

Using degenerate primers for the Ty1-copia group retrotransposons reverse transcriptase (RT) domain (Voytas et al., 1992), eleven putative RT clones from cultivated *G. barbadense* were obtained, and their DNA sequences were deposited in the NCBI nucleotide sequence database, GenBank (Table 1). Three additional RT sequences (Ash, Bah163, and Bah185) which were previously obtained from cultivated *G. barbadense* (Abdel Ghany and Zaki, 2003a) were added for further analysis. Blast search has shown that cultivated *G. barbadense* putative RT clones have nucleotide sequence similarities to the RT domains of other known plant Ty1-copia group retrotransposons (data not shown). Furthermore, derived amino acid sequences are compared to the same RT domains of known Ty1-copia group retrotransposons in Table 2. These results suggest that the aforementioned clones represent portion of the RT domain of the Ty1-copia group retrotransposons in cultivated Egyptian cottons.

Comparative nucleotide sequences analysis using ClustalW indicated that RT sequences in cultivated *G. barbadense* are different from each others with predicted nucleotide diversity from 6 to 100% (Figure 1A). Furthermore, pairwise comparison of amino acid sequences (Figure 2) revealed that these sequences can be classified into 11 distinct families (Table 3). The criterion for assignment to a family was ›90% amino acid

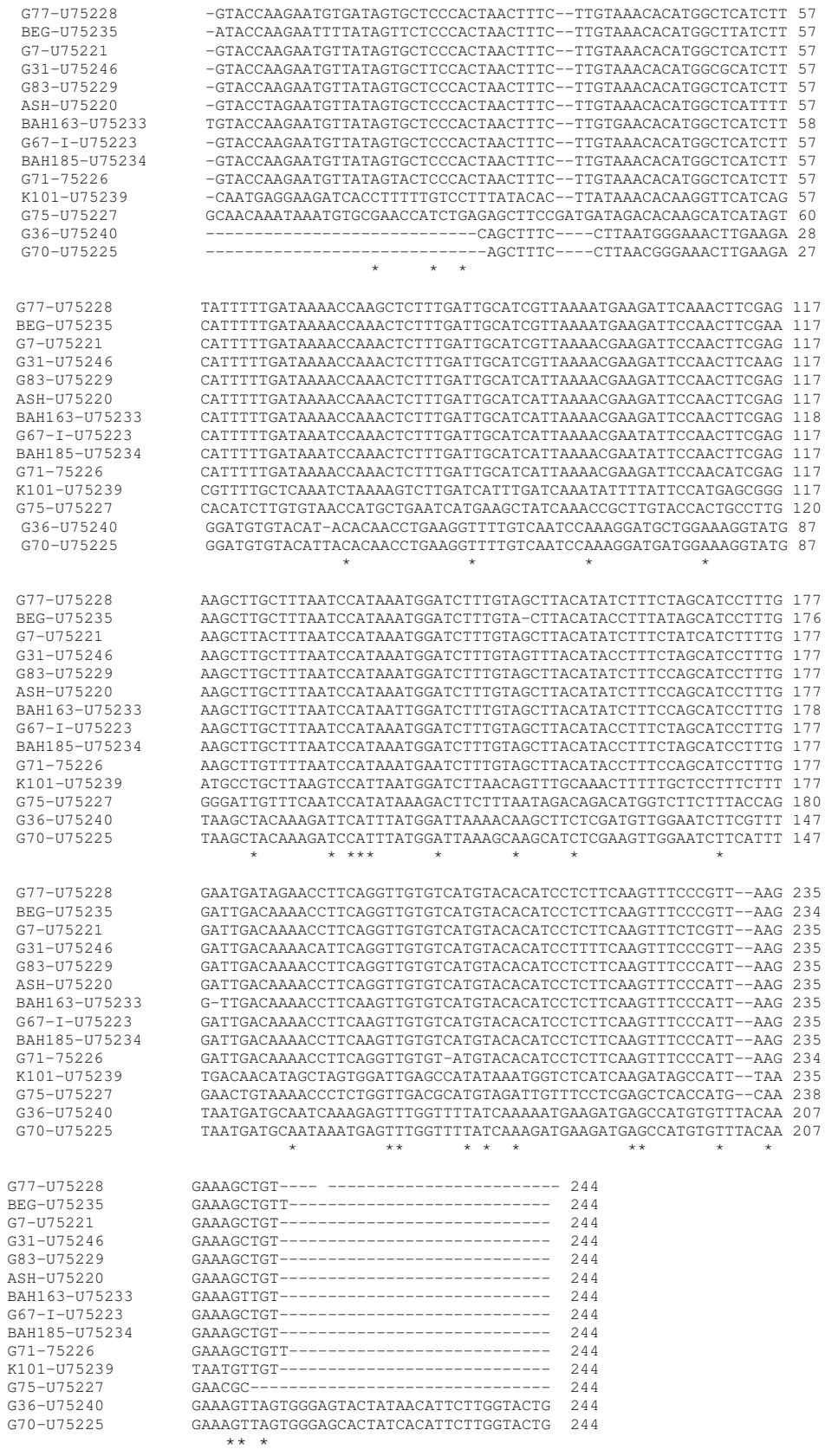

Figure 1. Comparative nucleotide sequences analysis of cultivated *G. barbadense* RT clones using ClustalW. Identical nucleotides are designated by (*).

Table 3. Cultivated *G. barbadense* RT clones arranged in families based on amino acid similarity.

Element family	Average amino acid divergence between clones based on pairwise comparisons (%)
Family 1	N/A
Family 2	N/A
Family 3	9
Family 4	N/A
Family 5	N/A
Family 6	N/A
Family 7	7
Family 8	6
Family 9	N/A
Family 10	N/A
Family 11	9

N/A not applicable.

Table 4. Numbers of synonymous (d_s) and nonsynonymous (d_n) substitutions per site in the RT domain for cultivated *G. barbadense* T*y1-copia* group retrotransposon families.

Family	No. of clones	d_s	d_n	d_s /d_n
1	1	N/A	N/A	N/A
2	1	N/A	N/A	N/A
3	2	0.081 ± 0.046	0.046 ± 0.018	1.8 ± 0.032
4	1	N/A	N/A	N/A
5	1	N/A	N/A	N/A
6	1	N/A	N/A	0.000
7	2	0.063 ± 0.023	0.029 ± 0.007	2.2 ± 0.010
8	2	0.000	0.000	0.000
9	1	N/A	N/A	N/A
10	1	N/A	N/A	N/A
11	1	0.028 ± 0.029	0.024 ± 0.013	1.1 ± 0.042

identity in pairwise comparison, which is consistent with previous studies that used similar criterion in defining families (Konieczny et al., 1991; Flavell et al., 1992; Vanderwiel et al., 1993). It is noteworthy that the same criterion was also employed for Ty3-gypsy group retrotransposons characterization in the Egyptian cottons (Abdel Ghany and Zaki, 2003b).

Synonymous and nonsynonymous nucleotide substitutions (d_S/d_N) and throughout in the RT domain of cultivated *G. barbadense* Ty1-copia group retrotransposons were investigated (Table 3). Numbers of d_S and throughout d_N substitutions were estimated according to Nei and Gojobori (1986). It is known that (d_S/d_N) are indicative to the strength and direction of selection (Yang and Bielawski, 2000). Table 3 shows that dS is greater than dN and (d_S/d_N) ratio suggests that the RT domain in cultivated *G. Barbadense* has been under purifying selection. These results are consistent with our previous studies for Ty1-copia and Ty3-gypsy group retrotransposons in *Gossypium*, respectively (Abdel Ghany and Zaki, 2003a,b). Furthermore, they are in accordance with other studies for strong selection for RT sequences (Konieczny et al., 1991; Flavell et al., 1992; Voytas et al., 1992; Matsuoka and Tsunewaki, 1999; Friesen et al., 2001; Stuart-Rogers and Flavell, 2001).

Evolutionary relationships among the RT derived amino acid sequences in cultivated *G. Barbadense* with each others and other Ty1-copia group retrotransposons were investigated by constructing a neighbor-joining tree (Saitou and Nei, 1987), with accession numbers are shown on the tree, and Ty1 as the outgroup (Figure 2). Based on the branching pattern, 11 families are recognised on the tree, which is consistent with family classification in Table 3. The neighbour-joining phylogram showed that cultivated *G. barbadense* RT families 1 and 2 are closely related to Ty1-copia group retrotransposons present in other plant species (Figure 3). On the other hand, the remaining RT families are

```
G7-U75221          -YQECYSAPTNFLVN-TWLIFIFDK---TKLFDCIVKTKIPTSRSLLSINGSL-LTYLSI 54
G31-U75246         -YQECYSASTNFLVN-TWRIFIFDK---TKLFDCIVKTKIPTSRSLLSINGSL-FTYLSS 54
G77-U75228         -YQECDSAPTNFLVN-TWLIFIFDK---TKLFDCIVKMKIQTSRSLLSINGSL-LTYLSS 54
G67-U75223         -YQECYSAPTNFLVN-TWLIFIFDK---SKLFDCIIKTNIPTSRSLLSINGSL-LTYLSS 54
BAH185-U75234      -YQECYSAPTNFLVN-TWLIFIFDK---SKLFDCIIKTNIPTSRSLLSINGSL-LTYLSS 54
G83-U75229         -YQECYSAPTNFLVN-TWLIFIFDK---TKLFDCIIKTKIPTSRSLLSINGSL-LTYLSS 54
ASH-U75220         -YLECYSAPTNFLVN-TWLIFIFDK---TKLFDCIIKTKIPTSRSLLSINGSL-LTYLSS 54
G71-75226          -YQECYSTPTNFLVN-TWLIFIFDK---TKLFDCIIKTKIPTSRSLFSINESL-LTYLSS 54
BEG-U75235         -YQEFYSSPTNFLVN-TWLIFIFDK---TKLFDCIVKMKIPTSKSLLSINGSLYLHTFIA 55
K101-U75239        -NEEDHLFVLYTLIN-TRFISVLLK---SKSLDHLIKYFIPAGCLLKSINGSQ-FANFLL 54
G36-U75240         SFP--WETRGCVHTQPEGFVNPKDAGKVCKLQRFIYGLKQASRCWNLRFNDAI--KEFGF 56
G70- U75225        AFLNGKLEEDVYITQPEGFVNPKDDGKVCKLQRSIYGLKQASRSWNLHFNDAI--NEFGF 58
BAH163-U75233      -----VPRMLCSHLSCEHMAHLHFN---QTLLHHNEDSNFEK---LALIHNWIFVAYISF 49
G75-U75227         -QQINVRTIELPMIDTSIIVTSCVT----MLNHEAIKPLVPLPWGLFQSIRLL--TDMVF 53
                                    .                                       :

G7-U75221          IF-WIDKTFRLCHVHI-LFKFLVKESC 79
G31-U75246         IL-WIDKTFRLCHVHI-LFKFPVKESC 79
G77-U75228         IL-WNDRTFRLCHVHI-LFKFPVKESC 79
G67-U75223         IL-WIDKTFKLCHVHI-LFKFPIKESC 79
BAH185-U75234      IL-WIDKTFKLCHVHI-LFKFPIKESC 79
G83-U75229         IL-WIDKTFRLCHVHI-LFKFPIKESC 79
ASH-U75220         IL-WIDKTFRLCHVHI-LFKFPIKESC 79
G71-75226          IL-WIDKTFRLCMYTS-SSSFPLRKAV 79
BEG-U75235         SF-GLTKPSGCVMYTS-SSSFPLRKAV 80
K101-U75239        LS-FDNIASGLSHING-LIKIAI--CC 77
G36-U75240         IK-NEDEPCVYKKVSGSTITFLVL--- 79
G70-U75225         IK-DEDEPCVYKKVSGSTITFLVL--- 81
BAH163-U75233      QHPLVDKTFKLCHVHI-LFKFPIKESC 75
G75-U75227         FT-RNCKTLWLTHVDC-FLELTMQER- 77
                             .          :  :
```

Figure 2. Comparative amino acid sequence analysis of cultivated *G. barbadense* RT clones using ClustalW. Conserved substitutions are designated by (:); and semiconserved substitutions are designated by (.).

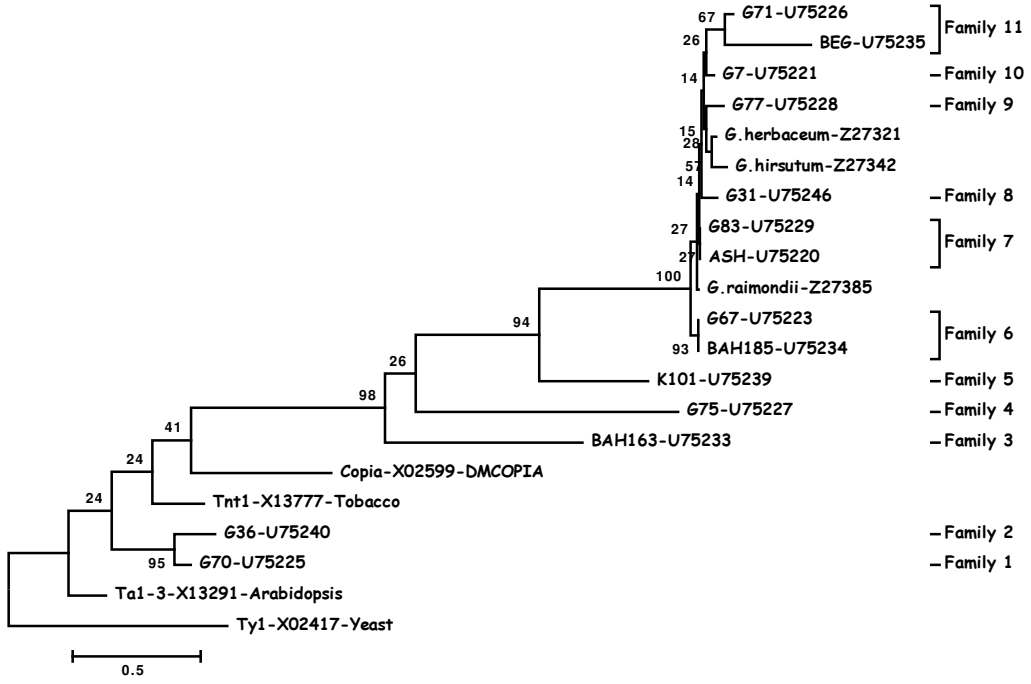

Figure 3. Phylogenetic tree showing evolutionary relationship between reverse transcriptase amino acid sequences of cultivated *G. Barbadense* and plant, yeast, and *Drosophila* T*y*1-*copia* group retrotransposons. The Neighbor-Joining method (Saitou and Nei, 1987) was used to construct the tree. The numbers on the branches represent bootsrap support for 1,000 replicates. Names refer to the accession number of the nucleotide sequences that encode the corresponding reverse transcriptase genes.

clustered and therefore closely related to their respective *Gossypium* species. These results suggest that RT families 1 and 2 mirror their own phylogenies rather than that of their host cultivars. Furthermore, the fact that the aforementioned families span species boundaries suggests that they existed early in plant evolution prior to modern plant species divergence (Figure 3).

Our phylogenetic analysis revealed that cultivated *G. barbadense* Ty1-copia group retrotransposon families 1, 2, 3, 4 and 5 possess long branches. On the other hand, the remaining families (6 to 11) are characterized with short branches. It is known that branch lengths are proportional to sequence divergence and may be the result of faster sequence evolution brought about in part by the error-prone nature of RT (Eickbush and Furano, 2002). Our amino acid substitution model, however, indicates that there is a strong selection for RT sequences. This is supported by the fact that despite RT sequence diversity in cultivated *G. barbadense*, there is a high degree of RT similarity with different plant species, which suggests strong selective constraints. These results suggest that RT plays a major role in the retrotransposition of these elements and therefore evolves slowly (Capy et al., 1994).

The data presented here enhance our knowledge and understanding of the diversity and evolution of Ty1-copia group retrotransposons in *Gossypium*. As they replicate and insert at multiple sites into the *Gossypium* genome, individual retrotransposons can potentially serve as markers of diversity. In this regard, Ty1-copia group retrotransposons are currently being developed as retrotransposon-based molecular markers for genetic diversity and evolution in *Gossypium*.

ACKNOWLEDGEMENTS

This work was supported by a grant from the US-Egypt Science & Technology Foundation to E.A. Zaki. This article is dedicated to the soul of A.M. Zaki.

REFERENCES

Anonymous (1992). Egyptian cotton varieties report. Cotton Res. Inst. Agric. Res. Center, Min. of Agric. Giza, Egypt.. pp. 2-8.

Anonymous (1997). In: “Agricultural Statistics Board”. NASS, USDA.www.usda.gov/nass.

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

Abdel Ghany AA, Zaki EA (2003a). Isolation, characterization, and phylogenetic analysis of Ty1/copia-like retrotransposons in the Egyptian cotton *G. barbadense* and its progenitors. Afr. J. Biotech. 2(6): 165-168.

Abdel Ghany AA, Zaki EA (2003b). Phylogenetic and molecular evolutionary analyses of gypsy group retrotransposon families in the Egyptian cotton *Gossypium* barbadense. Afr. J. Biotech. 2(10): 271-75.

Capy P, Anxolabehere D, Langin T (1994). The strange phylogenies of transposable elements: Are horizontal transfers the only explanation? TIG 10: 7-12.

Casacuberta JM, Santiago N (2003). Plant LTR-retrotransposons and MITEs: Control of transposition and impact on the evolution of plant genes and genomes. Gene 311: 1-11.

Eickbush TH, Furano AV (2002). Fruit flies and humans respond differently to retrotransposons. Curr. Opin. Genet. Dev. 12: 669-674.

Flavell AJ, Dunbar R, Anderson R, Pearce SR, Hartley R, Kumar A (1992). Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res. 20: 3639-3644.

Feschotte C, Jiang N, Wessler SR (2002). Plant transposable elements: Where genetics meets genomics. Natl. Rev. Genet. 3: 329-41.

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

Grandbastien MA, Spielmann A, Caboche M (1989). Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337: 376-380.

Konieczny A, Voytas DF, Cummings MP, Ausubel FM (1991). A superfamily of Arabidopsis thaliana retrotransposons. Genetics 127: 801-809.

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

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

Mount SM, Rubin GM (1985). Complete nucleotide sequence of the Drosophila transposable element copia: Homology between copia and retroviral proteins. Mol. Cell Biol. 5: 1630-1638.

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

Stuart Rogers C, Flavell AJ (2001). The evolution of Ty1-copia group retrotransposons in gymnosperms. Mol. Biol. Evol. 18: 155-163.

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.

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, Ausubel FM (1988). A copia-like transposable element family in Arabidopsis thaliana. Nature 336: 242-244.

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 (2003). Polyploidy and the evolutionary history of cotton. Adv. Agron. 78: 139-186.

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

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.