African Journal of Biotechnology Vol. 2 (9), pp. 271-275, September 2003
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2003 Academic Journals

Full Length Research Paper

Phylogenetic and molecular evolutionary analyses of *gypsy* group retrotransposon families 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 & Biotechnology Research Institute, GEBRI, Research Area, Borg El Arab, Post Code 21934, Alexandria, Egypt.

Accepted 11 August 2003

***Gypsy* group retrotransposons in the Egyptian cotton, *Gossypium barbadense*, was examined by phylogenetic and molecular evolutionary analyses. DNA sequences of *gypsy* group retrotransposons in two *G. barbadense* cultivars revealed that these sequences are heterogeneous and represent two distinct families. Sequence variation between these families seems to preserve coding information of the reverse transcriptase domain. The high ratio of synonymous to nonsynonymous changes indicates that the reverse transcriptase domain of these families is evolving under purifying selection. Our phylogenetic analysis revealed that the closest relatives of cotton retroelements are found in other plants *gypsy* group retrotransposons. Cotton retroelements-encoded transcripts were detected in their related respective young seedlings using RNA slot-blot hybridization, suggesting their transcriptional activity. The wide distribution of *gypsy* group retrotransposons and the detection of their encoded transcripts illustrate their active role in the *Gossypium* genome.**

Key words: Evolution, *Gossypium*, *gypsy*, retrotransposons, reverse transcriptase, substitution rates, transcription.

INTRODUCTION

Gossypium L. contains 50 species whose phylogenetic relationships have been explored using multiple molecular data sets (reviewed in Wendel and Cronn, 2003). Data indicate that shortly after its origin, *Gossypium* experienced rapid divergence leading to modern monophyletic lineages, designated A through G, and K genomes, that vary in chromosome size and infertility (Wendel, 1989). The five natural polyploids in the genus are believed to have generated from a single polypliodization event 1.5 million years ago (MYA) (Senchina et al., 2003). They all represent the AD genome tetraploids combining an A-genome donated by the maternal diploid parent at the time of polyploidy formation and a D-genome from the pollen parent (Wendel and Cronn, 2003).

The genus *Gossypium* is a facile system for investigating the genomic organization and evolution of repetitive DNA sequences that become newly united in a common nucleus (Zhao et al., 1995). Cloning and characterization of the major repetitive DNA in the tetraploid (AD) cotton revealed that most dispersed repeat families are largely restricted to the A-genome diploid ancestor and are absent from the D-genome (Zhao et al., 1998). Some families of these dispersed

*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, **MYA**: million years ago, **ORF**: open-reading frame, **PCR**: polymerase chain reaction, **RT**: reverse transcriptase gene.

repeats are, however, found at low levels on chromosomes derived from the D-genome ancestor, suggesting that the repeats have spread since the formation of polyploid cotton (Zhao et al., 1998). A likely mechanism for spread of the dispersed repeats appears to be transposition (Zhao et al., 1998). This suggestion was supported by the fact that four of the dispersed repeats show sequence similarity to retroelements from other taxa (Zhao et al., 1998). It is well known that retroelements constitute an important fraction of the DNA content of plant genomes (Kumar and Bennetzen, 1999). Their abundance, dispersion across the nuclear genome, and their insertional activity indicate that they play a major role in plant genome structure and evolution (Bennetzen, 2000).

As part of a long-term program to understand the organization and evolution of the cotton genome, we describe the phylogenetic and molecular evolutionary analyses of *gypsy* group retrotransposons in the Egyptian cotton *Gossypium barbadense*. The current report complements our recent analysis of the characterization and distribution of *gypsy* and *copia* group retrotransposons in the Egyptian cotton (Abdel Ghany and Zaki, 2002, 2003, Zaki and Abdel Ghany, 2003).

Materials and methods

Plant materials, genomic DNA extraction and isolation of *gypsy* group retrotransposons in *G. barbadense.*

Gypsy group retrotransposons (Table 1) were isolated from *G. barbadense* as previously described (Zaki and Abdel Ghany, 2003).

Table 1. *Gossypium barbadense* cultivars used in the current study, isolated clones and their GenBank accession numbers.

Cultivar	Clone	Accession number
Giza 45	G45	U75247
Giza 84	G84	U75248

RNA slot-blot hybridization

PCR amplified probes were labelled with [α-^{32}P] dCTP using the random primer method (Feinberg and Vogelstein, 1983), and used for RNA slot-blot hybridization as described (Sambrook et al., 1989). Filters were hybridized overnight at 42°C in a solution containing (50% formamide, 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.

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 version 2.1 (Kumar et al., 2001) from CLUSTALW alignments.

RESULTS AND DISCUSSION

PCR primers specific for conserved domains of the reverse transcriptase (RT) genes of *gypsy* group retrotransposons amplified their corresponding gene in two *G. barbadense* cultivars: Giza 45 and 84 (Zaki and Abdel Ghany, 2003). These fragments were designated G45 and G84 respectively (Table 1). Using G45 and G84 as hybridization probes, it was revealed that *gypsy* group retrotransposons can be detected in wild type species of *Gossypium*, suggesting that *gypsy* group retrotransposons is a standard component of the *Gossypium* genome (Zaki and Abdel Ghany, 2003). Comparative amino acid sequences analysis of G45 and G84 using ClustalW program revealed homology of 51% (Figure 1), indicating sequence heterogeneity. The observed sequence heterogeneity suggests that G45 and G84 represent two distinct *gypsy* group retrotransposon families in *G. barbadense*. The criterion for assignment to a family was >90% amino acid identity in pairwise comparisons (Figure 1). This is consistent with previous studies that used a similar criterion in defining retrotransposon families (Konieczny et al., 1991, Flavell et al., 1992, Vanderwiel et al., 1993).

```
G45 RDSDVPKTAFRTRYGHYEFLVMPFGLTNAPAVFMDLMNRIFRQYLDRFVVVFID 54
G84 REGDEWKIAFKTKHSLYEWLVMPFGLTNTSSTFMRLMNHVLRAFIGKFCVVYFD 54
    *:.*  * **:*::. **:*********:.:.** ***::** ::.:* **::*

G45 DILVYSGDETEHAEHLRLVLQILRDKQLYAKFSKCEFWLREVSFLGHVV 103
G84 DILVYSRSLDDHLKHLRAVLDVLRKENLYANLKKCTFCSNQVVFLGFVV 103
    ****** .  :* :*** **::**.::***::.** *  .:* ***.**
```

Figure 1. Comparative amino acids sequence analysis of *G. barbadense* G45 and G84 using CLUSTALW.

Table 2. Numbers of synonymous and nonsynonymous substitutions per site in the RT domain of cotton retrotransposons.

S: Synonymous substitutions	0.665 (±0.079)
N: Nonsynonymous substitutions	0.405 (±0.036)
d_S/d_N	1.640 (±0.083)
s: No. of synonymous sites	67.250 (±2.981)
n: No. of nonsynonymous sites	217.750 (±2.802)

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

Synonymous and nonsynonymous nucleotide substitutions (d_S/d_N) in the RT domain of *G. barbadense gypsy* group retrotransposon families, G45 and G84, were studied in detail (Table 2). 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

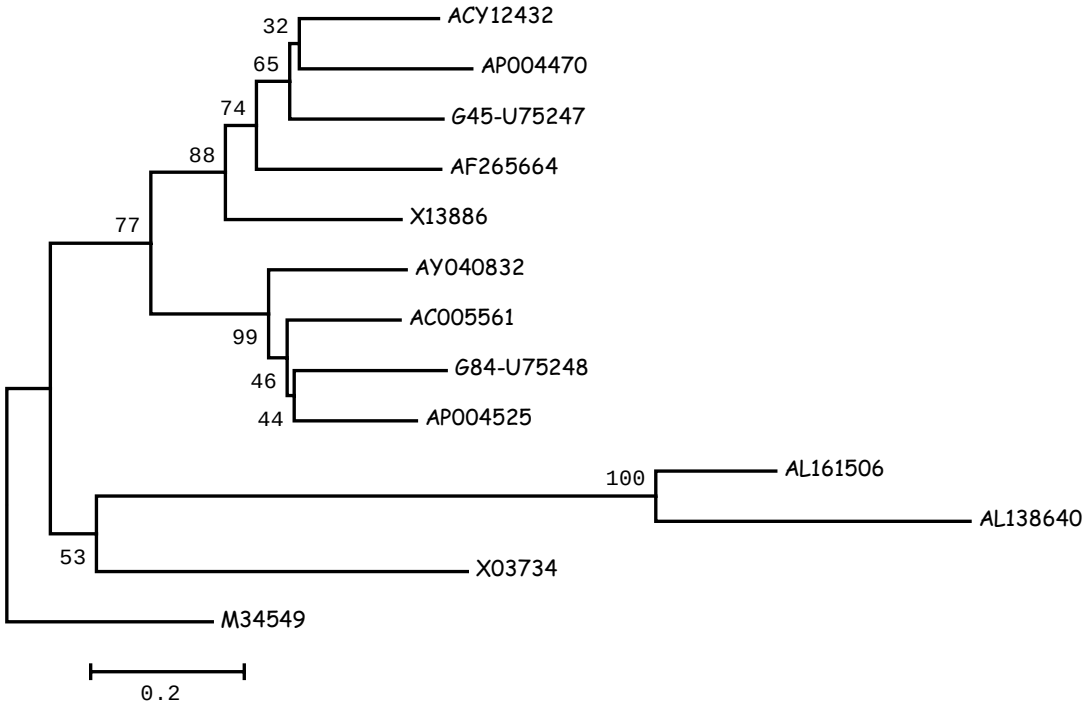

Figure 2. Phylogenetic tree showing relationship between reverse transcriptase nucleotide sequences of *G. barbadense* G45 and G84 and plant *gypsy* retrotransposons. The Neighour-Joining method (Saitou and Nei, 1987) was used to construct the tree. The numbers on the branches represent bootsrap support for 1,000 replicates. Names refer to the accession number of the nucleotide sequences that encode the corresponding reverse transcriptase genes. Sequences used and their GenBank accession numbers include: ACY12432 *Ananas comosus gypsy* group retrotransposon (Thomson et al., 1998); AP004470 *Lotus japonicus* genomic DNA Chromosome 4 27319-27627 (Sato et al., 2001); AF25664 *Solenum tuberosum* resistance gene cluster 130667-130975 (Vossen et al., 2000); X13886 *L. henryi del gypsy* group retrotransposon (Smyth et al., 1989); X03734 *Drosophila melanogaster gypsy* retrotransposon (Yuki et al., 1986); AL161506 *Arabidopsis thaliana* Chromosome 4 182355-182050 (Marin and Llorens, 2000); AL138640 *A. thaliana* Chromosome 3 37138-36830 (Marin and Llorens, 2000); AC005561 *Oryza sativa osr31/rire7* T*y*3/*gypsy* LTR-retrotransposon (McCarthy et al., 2002); AP004525 *L. japonicus* genomic DNA Chromosome 5 65809-66114 (Sato et al., 2001); AY040832 *Hordeum vulgare gypsy* group retrotransposon *cereba* (Hudakova et al., 2001), and M34549 yeast T*y*3 retrotransposon (Hansen et al., 1986).

from Table 2 yield no evidence of positive selection as d_S is greater than d_N. The synonymous and nonsynonymous ratio is, therefore, high enough to infer that the RT domain has been under purifying selection.

We sought to study the evolutionary relationships of the identified retroelements in *G. barbadense*. G45 and G84 were compared and aligned with other RT genes of plant *gypsy* group retrotransposons (accession numbers are shown on the tree) and T*y*3 as the outgroup (Figure 2). The neighbour-joining phylogram provided strong bootstrap support for a monophyletic origin of plant *gypsy* group retrotransposons, yet showed high diversity within all species. G45 has the strongest affinity with *Lotus japonicus* genomic DNA, chromosome 4, and *Ananas comosus gypsy* group retrotransposon (Sato et al., 2001, Thomson et al., 1998) with 75% and 74% amino acids identity respectively. On the other hand, G84 has the strongest affinity with *L. japonicus* genomic DNA, chromosome 5, *Hordeum vulgare cereba* and *Oryza sativa osr31/rire7 gypsy* group retrotransposons (Sato et al., 2001, Hudakova et al., 2001, McCarthy et al., 2002) with 75% amino acids identity.

To determine whether G45 and G84 are transcribed in *G. barbadense*, we performed RNA slot-blot hybridization using ^{32}P-labeled PCR amplified probes (Figure 3). Using two different total RNA concentrations, G45 and G84-encoded transcripts were detected, as evident by the detection of similar hybridization intensities in Giza 45 and 84 cultivars respectively. To normalize for RNA loading and eliminate that differences in expression were due to differences in G45 and G84 RT sequence diversity, genomic DNA from Giza 45 and 84 was subjected to DNA slot hybridization using the same ^{32}P-labeled probes. A similar degree of hybridization was detected (Figure 3B), indicating that sequence diversity did not affect the results of RNA slot-blot hybridization.

Gypsy group retrotransposons are present within all higher plant divisions as large highly heterogeneous populations (Kumar and Bennetzen, 1999). Phylogenetic analyses have shown that these populations are resolved

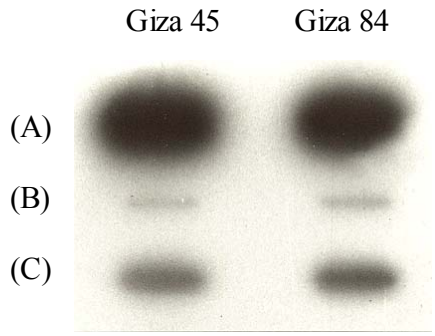

Figure 3. RNA slot-blot hybridization of two cotton cultivars, Giza 45 and Giza 84, performed by use of their respective related probes G45 an d G84 probes. (A): genomic DNA (100ng), (B) and (C): (500ng) and (5µg) of total RNA respectively.

into diverse families, which span species boundaries, such that the closest homologue of one family is often from a different species (Eickbush and Malik, 2002). DNA sequences of *gypsy* group retrotransposons in two *G. barbadense* cultivars revealed that these sequences are heterogeneous and represent two distinct families. Sequence variation between these families seems to preserve coding information of the RT domain. The high ratio of synonymous to nonsynonymous changes indicates that the RT domain of these families is evolving under purifying selection. Moreover, RT sequences in cotton have evolved under functional constraints and likely to play a role in the life cycle of these elements. Our results contribute and exemplify the increasingly reports of 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). These examples, taken from across the phylogenetic spectrum, illustrate that sequence conservation is a general property of retrotransposons.

Gypsy group plant retrotransposons with *envelope (env)*-like genes have been reported (Zaki, 2003). Phylogenetic analysis of the RT domain of plant *gypsy* group retrotransposons indicated that they resolve into two lineages: one universally lacking and the other containing *env* genes (Vicient et al., 2001). Our phylogenetic analysis revealed that G84 RT sequence is clustered with *env*-containing plant retrotransposons. This suggests that G84 represents an *env*-containing *gypsy* group retrotransposon family in *G. barbadense*. It is noteworthy that *env*-like sequences, *GM*5 and *GM*6, were previously reported in *G. barbadense* (Abdel Ghany and Zaki, 2002). Currently, it is unknown whether G84, *GM*5 and *GM*6 represent the same retrotransposon family. Further experimental analysis is required to address this question.

The phylogenetic analysis of the RT domain provides the evolutionary relationships among *gypsy* group retrotransposons to be inferred (Malik et al., 2000; Eickbush and Malik, 2002). Our phylogenetic analysis revealed that the closet relatives of G45 and G84 are found in other *gypsy* group RT of plants (*A. comosus* and *H. vulgare*) than to each other. These evolutionary relationships suggest either an ancient origin of plant retrotransposons (vertical transmission), or horizontal transmission, in which these retrotransposons have jumped the species-gap (Eickbush and Malik, 2002). The observation that branch lengths separating plant retrotransposons are usually similar, indicating a similar evolutionary distance, disagrees with the horizontal transmission hypothesis, and supports the existence of a diverse group of retrotransposon families in the progenitor of plants. This suggestion is supported by the fact that *gypsy* group retrotransposons are detected in all *Gossypium* species examined (Zaki and Abdel Ghany, 2003).

Plant retrotransposons are known to be transcriptionally silent in most plant tissues during development, suggesting transcriptional control is a major mechanism of control for their retrotransposition (Kumar and Bennetzen, 1999). Their expression and transposition are, however, inducible by stresses such as protoplast isolation and tissue culture (Grandbastien et al., 1997). Detection of their transcripts under ordinary growth conditions has also been reported (Suoniemi et al., 1996, Pearce et al., 1997). In this regard, G45 and G84-encoded transcripts were detected in their related respective young seedlings using RNA slot-blot hybridization, suggesting that G45 and G84 are transcrptionally active retrotransposons. However, the presence of either stop codons or insertions/deletions that have caused frame shifts in G45 and G84 derived amino acid sequences, suggests that these clones represent defective retrotransposons. Nevertheless, the detection of G45 and G84-encoded transcripts, intermediates in the retrotransposition process (Kumar and Bennetzen, 1999), suggests that subsets of these molecules are competent for retrotransposition.

Acknowledgements

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 216:351-353.

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

Bennetzen JL (2000). Transposable elements contributions to plant gene and genome evolution. Plant Mol. Biol. 42:351-369.

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

Feinberg AP, Vogelstein B (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6-13.
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.
Friesen N, Brandes A, Heslop-Harrison J (2001). Diversity, origin and distribution of retrotransposons in conifers. Mol. Biol. Evol. 18:1176-1188.
Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta TM (1997). The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica 100:241-252.
Hansen LJ, Chalker DL, Sandmeyer, SB (1988). T*y*3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol. Cell. Biol. 8: 5245-5256.
Hudakova S, Michalek W, Presting GG, Hoopen R, Santos K, Jasencakova Z, Schubert I (2001). Sequence organization of barley centromeres. Nucleic Acids Res. 29:5029-5035.
Konieczny A, Voytas DF, Cummings MP, Ausubel FM (1991). A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics 127:801-809.
Kumar A, Bennetzen JL (1999). Plant retrotransposons. Annu. Rev. Genet. 33:479-532.
Kumar SK, Tamura K, Jakobsen IB, Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244-1245.
Malik HS, Henikoff S, Eickbush TH (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. 10:1307-1318.
Marin I, Llorens C (2000). T*y*3/*Gypsy* retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. Mol. Biol. Evol. 17:1040-1049.
Matsouka Y, Tsunewaki K (1999). Evolutionary dynamics of T*y*1-*copia* group retrotransposons in grass shown by reverse transcriptase domain analysis. Mol. Biol. Evol. 13:1384-1392.
McCarthy EM, Liu J, Lizhi G, McDonald JF (2002). Long terminal repeat retrotransposons of *Oryza sativa*. Genome Biol. 3:1-11.
Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418-426.
Pearce SR, Harrison G, Heslop-Harrison P, Flavell AJ, Kumar A (1997). Characterization and genomic organization of T*y*1-*copia* group retrotransposons in rye. Genome 40:617-625.
Saitou N, Nei M (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.
Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Sato S, Kaneko T, Nakamura Y, Asamizu E, Kato T, Tabata S (2001). Structural Analysis of a *Lotus japonicus* Genome. GenBank Accessions (AP004525 and AP004470), NCBI, Bethesda, U.S.A.
Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003). Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 20:633-643.
Suoniemi A, Narvanto A, Schulman AH (1996). The BARE-1 retrotransposon is transcribed in barely from an LTR promoter active in transient assays. Plant Mol. Biol. 31:295-306.
Smyth DR, Kalitsis P, Joesph J, Sentry JW (1989). Plant retrotransposon from *Lilium henryi* related to T*y*3 of yeast and *gypsy* group of *Drosophila*. Proc. Natl. Acad. Sci. USA 86:5013-5019.
Stuart-Rogers, C, Flavell, AJ (2001). The evolution of T*y*1-*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.
Thomson 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.
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.
Vossen EA, Voort JN, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM (2000). Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. Plant J. 5:567-76.
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 (1989). New world cottons contain old world cytoplasm. Proc. Natl. Acad. Sci. USA. 86:4132-4136.
Wendel JF, Cronn R (2003). Polyploidy and the evolutionary history of cotton. Advances in Agronomy 78:139-186.
Wilhelm M, Wilhelm FX (2001). Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. 58:1246-1262.
Yang Z, Bielawski JP (2000). Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15:469-503
Yuki S, Ishimaru S, Inouye S, Saigo K (1986). Identification of genes for reverse transcriptase-like enzymes in two *Drosophila* retrotransposons, *412* and *gypsy*; a rapid detection method of reverse transcriptase genes using YXDD box probes. Nucleic Acids Res. 11:3017-3030.
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.
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.
Zhao X, Wing RA, Paterson AH (1995). Cloning and characterization of the majority of repetitive DNA in cotton (*Gossypium* L.). Genome 38:1177 1188.