African Journal of Biotechnology Vol. 2 (6), pp. 165–168, June 2003
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2003 Academic Journals

# Isolation, characterization, and phylogenetic analysis of *copia*-like retrotransposons in the Egyptian cotton *Gossypium barbadense* and its progenitors

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

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

Accepted 16 May 2003

**We have used the polymerase chain reaction to analyze *copia*-like retrotransposons in the Egyptian cotton and its progenitors. All three cotton species studied contain reverse transcriptase fragments from *copia*-like retrotransposons. Sequence analysis of these reverse transcriptase fragments reveals that each is different from the others, with predicted amino acid diversities between 9 and 94%. The detection of stop codons and insertions/deletions in the derived amino acid sequences of the *Gossypium* RT clones, suggests that these clones represent defective retrotransposons. The presence of these sequences in *G. barbadense* progenitors, however, suggests the presence of active retrotransposons capable of producing new functional copies at an appropriate rate to compensate for the mutational loss of old ones. Phylogenetic analysis provided strong bootstrap support for a monophyletic origin of plant *copia*-like retrotransposons, yet showed high diversity within all species. Our results suggest that both vertical transmission of *copia*-like retrotransposons within *G. barbadense* lineages, and horizontal transmission between *G. barbadense* and its progenitors have played major roles in the evolution of *copia*-like retrotransposons in *Gossypium*.**

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

## INTRODUCTION

*Copia*-like group retrotransposons are one of the best characterized groups of plant retrotransposons (for review see Kumar and Bennetzen, 1999). They have been reported in a wide range of plant taxa, including angiosperms, gymnosperms, ferns, lycopods, and bryophytes (Konieczny et al., 1991; Flavell et al., 1992; Voytas et al., 1992; Friesen et al., 2001). Their ubiquity in the plant kingdom suggests that they are of very ancient origin (Bennetzen, 2000). In addition, their abundance has played a major role in plant genome structure and evolution (Bennetzen, 2002).

*Gossypium* L. (Malvaceae) has become a useful system for studying plant retrotransposons evolution (Abdel Ghany and Zaki, 2002; Zaki and Abdel Ghany, 2003). The phylogenetic relationships of the approximately 50 diploid and 5 allotetraploid species of *Gossypium* are well characterized (reviewed in Wendel and Cronn, 2002). The five allotetraploid *Gossypium* species (designated AF-genome) diverged from a single recent allopolyploidization event, and their parental diploids (Wendel, 1989). In this regard, *copia*-like retrotransposons were previously identified in *G. hirsutum* (Vanderwiel et al. 1993). In addition, fluorescent *in-situ* hybridization was used to study their chromosomal distributions (Hanson et al., 1999). In the current study, we isolated, cloned, and sequenced part of the reverse transcriptase (RT) domain of *copia*-like retrotransposons in the Egyptian allotetraploid cotton, *G. barbadense* cotton, and its progenitors. Our results revealed that all three cotton species studied here contain RT fragments from *copia*-like retrotransposons, suggesting that *copia*-like retrotransposons is a standard component of the *Gossypium* genome, and supporting the fact that *copia*-like retrotransposons represents a major component of the plant genome.

*Corresponding author; 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;** PCR: polymerase chain reaction, RT: reverse transcriptase gene.

## MATERIALS AND METHODS

### Plant materials and genomic DNA extraction

*Gossypium* species and cultivars, listed in Table 1, were kindly provided by Dr. Percival. Total DNA was extracted using Qiagen DNeasy kit (Qiagen, Germany).

**Table 1**. *Gossypium* species used in this study: isolated clones and their GenBank accession numbers.

| Species | Clone | Accession number |
|---|---|---|
| *G. arboretum* | Arb | U75244 |
| *G. barbadense* | | |
| Bahtim163 | Bah163 | U75221 |
| Bahtim185 | Bah185 | U75222 |
| Ashoumni | Ash | U75246 |
| *G. darwinii* | Dar | U75245 |

### PCR

Total DNA was subject to PCR with specific primers to amplify an approximately 280 bp region of the *copia*-like reverse transcriptase (5`-GGAATTCGAYGTNAARACNGCNTTYYT-3`) and (5`-GGGATC CAYRTCRTCNACRTANARNA`), where N= A+C+G+T, R= A+G, and Y= T+C (Voytas et al. 1992). 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 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 the competent *E. coli* strain TOPO 10. Plasmid DNA was isolated using QIA Spin mini-prep kit (Qiagen, Germany), and 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 2.1 (Kumar et al., 2001) 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.

## RESULTS AND DISCUSSION

PCR amplification with degenerate primers for the *copia*-like reverse transcriptase (RT) domain (Voytas et al., 1992) produced 5 putative RT clones: 3 from *G. barbadense*, and 1 clone from *G. arboretum* and *G. darwinii* respectively (Table 1). Blast search confirmed the RT nature of the cloned products. The high amino acid similarities, observed in the Blast search, supports the interpretation that the 5 sequences generated in this study represent portions of *copia*-like retrotransposons RT genes.

Extreme sequence diversity amongst RT genes of the *copia*-like retrotransposons has been observed both within and between plant species (Flavell et al., 1992). Furthermore, they are more closely related to elements present in other plant species (Voytas et al., 1992). A similar pattern of sequence heterogeneity is observed in *G. barbadense* and its progenitors. Comparative nucleotides and amino acids sequences using CLUSTALW were performed (Figure 1). In addition, pairwise comparisons (Table 2) showed amino acid diversities of 28% (Bah163/Bah185), and 26% (Bah163/Ash) within *G. barbadense* cultivars, and 9% (Arb/Dar), and 10% (Arb/Bah163) between *Gossypium* species.

**Table 2**. Amino acids pairwise comparisons of *copia*-like putative RT sequences in *G. barbadense* and its progenitors.

| Clone | Arb | Bah163 | Bah185 | Ash |
|---|---|---|---|---|
| Bah163 | 10 | | | |
| Bah185 | 26 | 28 | | |
| Ash | 26 | 26 | 94 | |
| Dar | 9 | 93 | 33 | 33 |

The detection of either stop codons and insertions/deletions that have caused frame shifts in the derived amino acid sequences of the *Gossypium* RT clones, suggests that these clones represent defective retrotransposons. In addition, a number of the retrotransposons in the Egyptian cotton and its progenitors obviously are not functional and currently must be evolving as pseudogens. The long-term survival of these sequences, however, suggests the presence of active retrotransposons capable of producing new functional copies at an appropriate rate to compensate for the mutational loss of old ones. It is noteworthy that the majority of plant *copia*-like retrotransposons are thought to be rarely active (Kumar and Bennetzen, 1999).

Relationships among the derived amino acid sequences of the 5 clones with each other and other


**Figure 1.** Comparative nucleotides and amino acids sequence analysis of *Gossypium* RT clones using CLUSTALW. Key: 1: Bah163, 2: Bah185, 3: Ash, 4: Arb, 5: Dar.

retrotransposons were investigated by constructing a neighbor-joining tree (Saitou and Nei, 1987), with accession numbers are shown on the tree, and T*y*1 as the outgroup (Figure 2). The neighbour-joining phylogram provided strong bootstrap support for a monophyletic origin of plant *copia*-like retrotransposons, yet showed high diversity within all species. Bah163 from *G. barbadense* has the strongest affinity with Dar from *G. darwinii* with 93% amino acids identity. On the other hand, Bah185 from *G. barbadense* has the strongest affinity with Ash from *G. barbadense* with 94% identity.

This study aimed to use the PCR technique to study the evolution of *copia*-like retrotransposons in the Egyptian cotton and its progenitors. We have shown that the amplified fragments are comprised of a very heterogeneous collection of RT sequences. These results are consistent with previous which concluded that *copia*-like retrotransposons are present as highly heterogeneous populations within all higher plants (Flavell et al., 1992, Voytas et al., 1992). Such sequence heterogeneity contrasts strongly with the limited diversity seen in such retrotransposons in yeast and *Drosophila* (Peterson-Burch and Voytas, 2002). Plant genomes seem to be inherently more susceptible to the generation of sequence diversity in copia-like retrotransposons than the genomes of *Drosophila* and yeast (Bennetzen, 2002). This has been contributed to the distinctive strategy adopted by plants for determination of their germlines, and their great tolerance to chromosomal alterations (Flavell et al., 1992). 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 *copia*-like retrotransposons sequence diversity in *G. barbadense*. Further experimental data such as copy number determination, chromosomal distribution, and sequencing of large contiguous regions of the cotton will significantly add to fundamental knowledge about the population heterogeneity of *copia*-like retrotransposons in the *Gossypium* genome.

The origin and evolutionary relationships of *copia*-like retrotransposons still remain very interesting, yet, debatable, and controversial (Eickbush and Furano, 2002). Previous studies have suggested three possible mechanisms, not mutually exclusive, may account for the evolution of copia-like retrotransposons in plants (Stuart-Rogers and Flavell, 2001, Peterson-Burch and Voytas, 2002). First, there may have been an explosive radiation of retrotransposons in the common ancestor of all plants, with most of these elements maintained with only limited further divergence. Second, multiple horizontal transfers of retrotransposons may have occurred between physically and phylogenetically distant populations or taxa. Finally, the constraints on retrotransposons evolution may be such that these elements have reached near-identical ranges of sequence diversity in widespread modern plant taxa. In this regard, according to our phylogenetic analyses, *copia*-like retrotransposons in *G. barbadense* and its progenitors is dominated by germ line vertical transmission. Our data, however, revealed that Bah163 of *G. barbadense* closest homologue is that of *G. darwinii*. Horizontal transfer of retrotransposons has been suggested as an explanation of this phenomenon in plants (Friesen et al., 2001). This suggestion is supported by the fact that *G. darwinii* is believed to have diverged from a common ancestor about 4-11 million years prior to being united in a common polyploid nucleus (Wendel, 1989). A more comprehensive survey of *copia*-like retrotransposons in *Gossypium* species is certainly

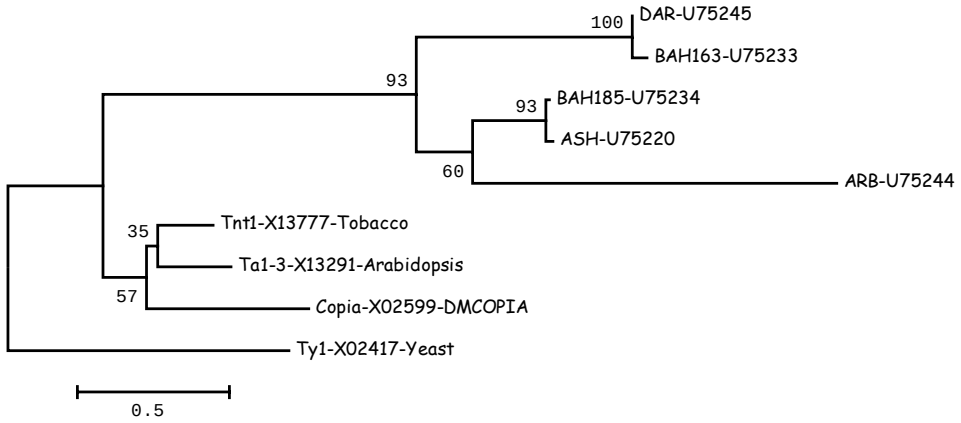


**Figure 2.** Phylogenetic tree showing relationship between reverse transcriptase amino acid sequences of *G. Barbadense* and its progenitors and plant, yeast, and *Drosophila copia*-like retrotransposons. The Neighbor-Joining method (Saitou and Nei, 1987) was used to construct the tree. The numbers 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.

required to further clarify their evolutionary relationships. In conclusion, we suggest that both vertical transmission of *copia*-like retrotransposons within *G. barbadense* lineages, and horizontal transmission between *G. barbadense* and its progenitors have played major roles in the evolution of *copia*-like retrotransposons 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 AA, Zaki EA (2002). Cloning and sequencing of an *envelope*-like gene in *Gossypium*. Planta 216:351-353.

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. 42:347-353.

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). 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.

Hanson RE, Islam-Faridi N, Crane CF, Zwick MS, Czeschin DG, Wendel JF, McKnight TD, Price HJ, Stelly DM (1999). T*y*1-*copia*-retrotransposons behaviour in a polyploid cotton. Chromosome Res. 8:73-76.

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

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

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

Peterson-Burch BD, Voytas DF (2002). Genes of the Pseudoviridea (Ty1/copia Retrotransposons). Mol. Biol. Evol. 19:1832-1845.

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 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.

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

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