World Journal of Agricultural Sciences 3 (1): 130-139, 2007
ISSN 1817-3047

Comparative Sequence Analysis of Actin Related Gene Family Isolated from *Gossypium barbadense*

Osama A. Momtaz, Ahmed El-Fatih A. El-Doliefy, Ashraf H. Fahmy and Ayman A. Diab

Agricultural Genetic Engineering Research Institute (AGERI),
Agricultural Research Center (ARC), Ministry of Agriculture, Egypt

Abstract: Actin is a ubiquitous and highly conserved microfilament protein that is playing an important role in fiber quality and fiber cell developmental stages in cotton plants. Actin regulates microtubule's filaments and cellulose deposition orientation during cotton fiber cell development which directly affects fiber quality. Actins in plants are encoded by a multigene family that comprises dozens or even hundreds of actin genes that display high sequence homology and partially overlapping expression patterns. The aim of this work is to identify and characterize one of the actin related gene family in Egyptian cotton (*G. barbadense*) and compare its sequence with other actin sequences from *G. hirsutum*. In this work, P1-derived Artificial Chromosome (PAC) library was screened for actin gene using the GhACT2 degenerate primers. Fourteen actin positive PAC clones were identified. One of the clones was further characterized as a partial sequence of the cotton actin relating gene family (*GbACT*) by sequence analysis and a homology search of GenBank databases. Comparison of the deduced polypeptide of GbACT with other actins proteins from upland cotton species revealed several homologous regions. Sequence alignment of GbACT revealed 16 matches of upland cotton actins (*Gossypium hirsutum*). Multiple mega-alignment between the most close actins (*GhACT*; *GhACT2* and *GhACT5*) identified from the upland cotton database was performed using DNASTAR lasergene program software. The GbACT gene showed 93.8-97.7% sequence similarity, 83.6-85.3% at the open reading frame (ORF) level and 84.8% amino acid similarity were detected.

Key words: *Gossypium barbadense* · actin gene · sequence alignment · comparative aanalysis · cotton fiber · gene family

INTRODUCTION

Cotton fiber quality and/or fiber cell developmental stages are essentially based on cell wall biogenesis and highly affected by cytoskeleton. Cytoskeleton is a dynamic structure involved in many key processes including cell division, organelle movement and formation of cell wall. It is composed of three fibrous elements, the microtubules, actin filaments and intermediate filaments [1].

Actin filaments found in plant cells during all cell cycle stages and involved in transport of secretory vesicles, facilitate the delivery of membrane and cell wall components in these vesicles to the plasma membrane and, thereby, promote cell expansion. Actin cytoskeleton is essential for cell elongation and cell tip growth [2]. Due to its multifunctional roles in cell multiplication, growth and development, actin is one of the most abundant proteins in many eukaryotic cells and is conserved in many species [3].

Actin controls polar cell growth through its interaction with several actin-binding proteins [4]. Genetic studies showed that the actin cytoskeleton by interacting with ARP2/ARP3 protein complex, plays a pivotal role in controlling fiber cell shape and several other cell types [5].

Actins in plants are encoded by a multigene family that comprises dozens or even hundreds of actin genes that display high sequence homology and partially overlapping expression patterns [2]. Genes encoding actin have been identified and cloned from many species [6, 7], including partial actin sequence [8].

Corresponding Author: Dr. Ashraf H. Fahmy, Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), Ministry of Agriculture, Egypt

In Arabidopsis, the actin gene family contains 10 distinct members, of which eight are functional genes and two are pseudogenes [9]. Disruption of actin cytoskeleton during trichome development in Arabidopsis thaliana by any of actin interacting drugs resulted in randomly distorted trichomes with un-extended branches [10]. Mutations in actin 2 (ACT2) and actin 7 (ACT7) genes, which are responsible for the development of actin arrays resulted in dramatic reduction of root hair length [11].

In cotton fiber cells the predominant functionally expressed forms of actin genes are GhACT1, GhACT2, GhACT4, GhACT5 and GhACT11. Silencing of the GhACT1 gene in *G. hirsutum* showed short fibers, sterile ovules and small bolls which illustrate the role of GhACT1 in fiber maturation process. Moreover, in actin GhACT1-knocked out cotton plants, fiber elongation rate was slower by 3 folds, reflecting the role of actin on pollination, seed development, fiber quality and productivity [12].

In other plant species, the actin gene family also appears to have dozens of members [13-15]. Studies on actin sequences revealed that structural and functional divergence occurred within the gene family during evolution [9, 16]. The diversity of these functional roles is paralleled by diversity within plant actin gene families. Plant actins are very similar to other eukaryotic actins (<83-88% amino acid identity with most animal and fungal actins) [17]. Plant actins are encoded by gene families that are much more diverse than those in other eukaryotes.

Efforts have been made toward genomics and functional genomics of cotton fiber development. Actin genes in a few plant species such as Arabidopsis have been well-characterized, however our knowledge of cotton actin genes need to be explored. The objective of this work was to identify one of the actin related gene family in Egyptian cotton (*G. barbadense*) and compare its sequence with other actin sequences from *G. hirsutum.*

MATERIALS AND METHODS

Plant material: Seeds of the Egyptian extra long stable variety (Giza88) *Gossypium barbadense* L. were de-linted and planted in greenhouse. Giza88 seeds were kindly provided from Cotton Research Institute (CRI), Agricultural Research Center (ARC), Giza, Egypt.

Nucleic acid isolation: Genomic DNA was isolated from young cotton leaves (Giza88) using Qiagen DNeasy™ Plant Mini kit (Cat. No. 69104) following the manufacture manual. Total RNA was isolated using SV total RNA Promega kit (Cat. No. Z3100) following the manufacture manual. Nucleic acids samples were stored at -80°C.

Construction and screening of PAC library: The pPACe4 (19.5 kb) vector used in the study was developed by Frengen *et al.* [18] and kindly provided by the Children's Hospital Oakland Research Institute (CHORI) (http://bacpac.chori.org/ppac4.htm). The vector was purified using plasmid Qiagen Max kit (Cat. No. 12163) following the manufacturer manual. pPACe4 vector was ligated to the high molecular weight (HMW) DNA as described by Momtaz *et al.* [19]. BAC library construction was performed as described previously [20]. About 500 of recombinant PAC clones were used as templates for PCR-based screening as described by Momtaz *et al.* [20]. The GhACT2 forward and reverse degenerate primers for actin gene family were used as described by Li *et al.* [12]. PCR-based screening reactions were fractionated on 4% TAE agarose gel. Fourteen actin positive PAC clones were identified.

Sequence alignment and data analysis: One of the fourteen positive actin clones designated as (GbACT) was sequenced using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE applied Biosystems, USA) in conjunction with ABI PRISM (310 Genetic Analyzer). Sequence alignment was performed using the Basic Local Alignment Search Tool (BLAST) programs from the National Center for Biotechnology Information (NCBI). Data analysis of the sequence on the levels of DNA, Open Reading Frame (ORF) and amino acid were performed using the DNASTAR lasergene software.

RESULTS AND DISCUSSION

PAC library construction and screening: The constructed PAC library produced 3000 clones. Of the clones, 10 % showed insert smaller than 50 Kb, 5 % showed insert size larger than 100 Kb and 85% of the clones showed 50-70 Kb insert. The library provides 0.3 haploid genome equivalents to (>88%) probability of finding any specific sequence [20]. The library was subjected to PCR-based screening using GhACT2 degenerate forward primer (TGCCCAGAAGTCCTCTT CCAG) and reverse primer (GCGCGGTCAAACTTCTG GGAAAAT) as described by Momtaz *et al.* [20]. Fourteen actin positive PAC clones were identified.

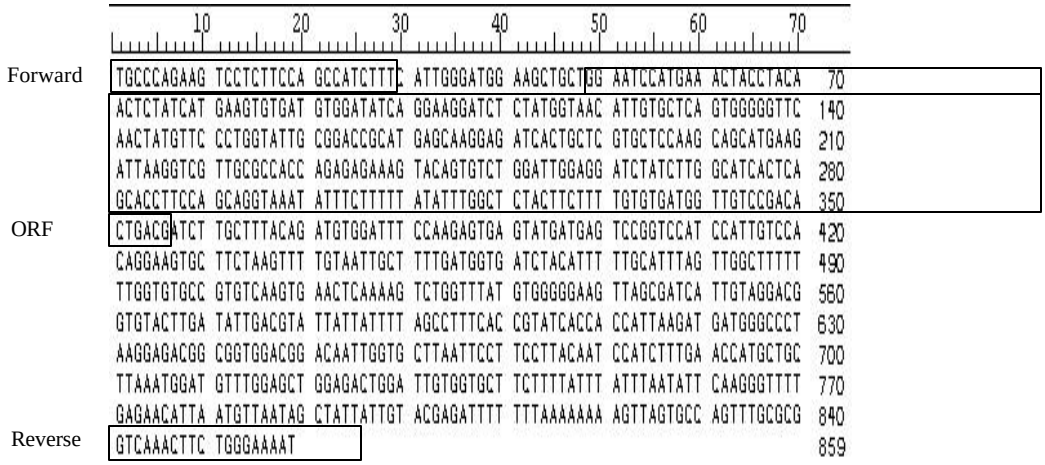

Fig. 1: DNA sequence of *GbACT*. First box in the start of the sequence represents the forward primer while the one in the end of the sequence represents the reverse primer. The middle box represents the Open Reading Frame (ORF) of the *GbACT* sequence. The free-box region of the sequence represents the intron (UTR) region in the *GbACT* sequence

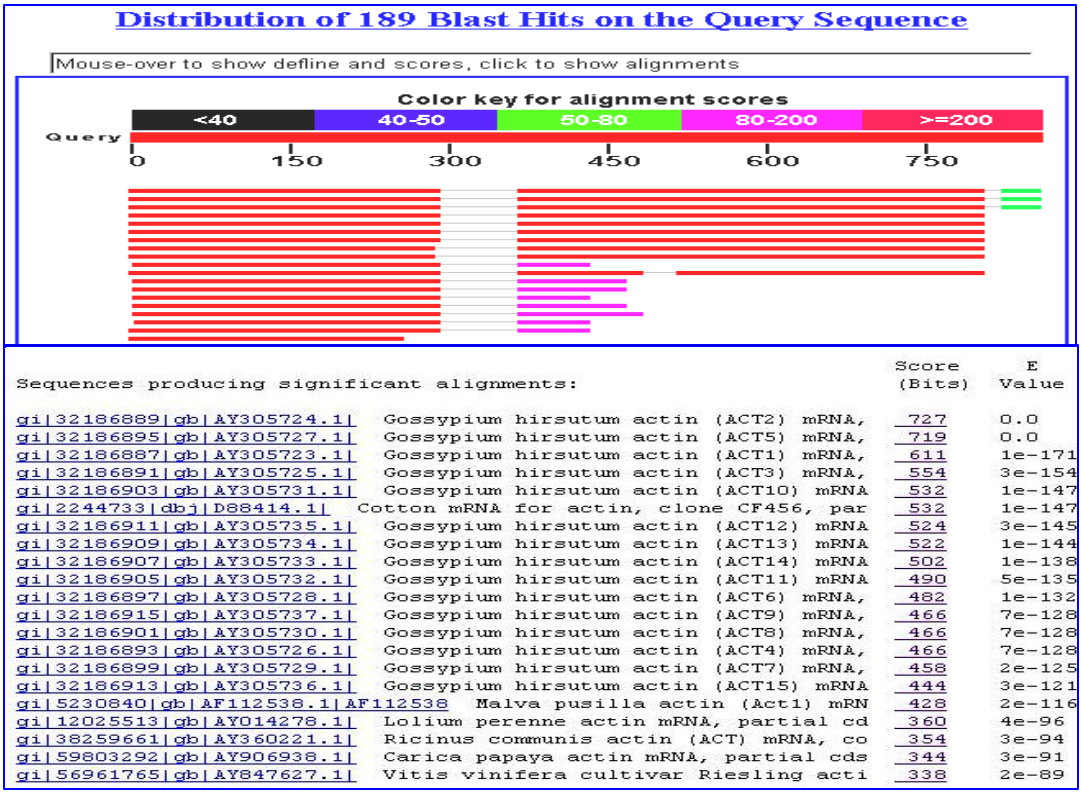

Fig. 2: *GbACT* DNA sequence alignment at the NCBI Database. The upper figure presents a schematic diagram of the closest actin genes, while the lower figure represents the closest sited genes and the accession number of these genes in gene bank

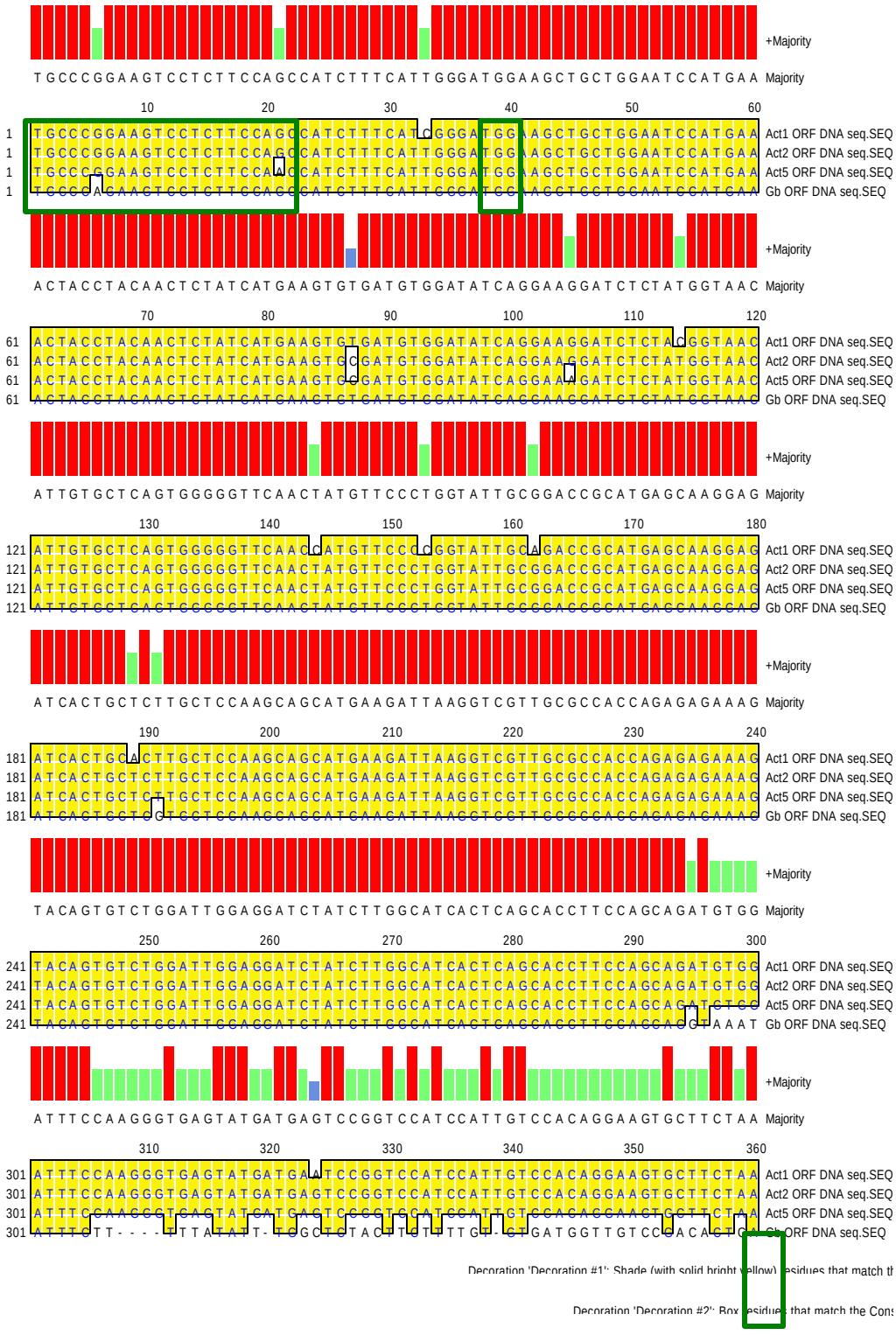

Fig. 3: Mega alignment between the *GbACT* and the ACT1, ACT2 and ACT5 on the level of the Open Reading Frames (ORFs). The alignment started from the Forward primer represented in long green box. The start codon (ATG) and the Stop codon (TAA and TGA) represented in small green boxes

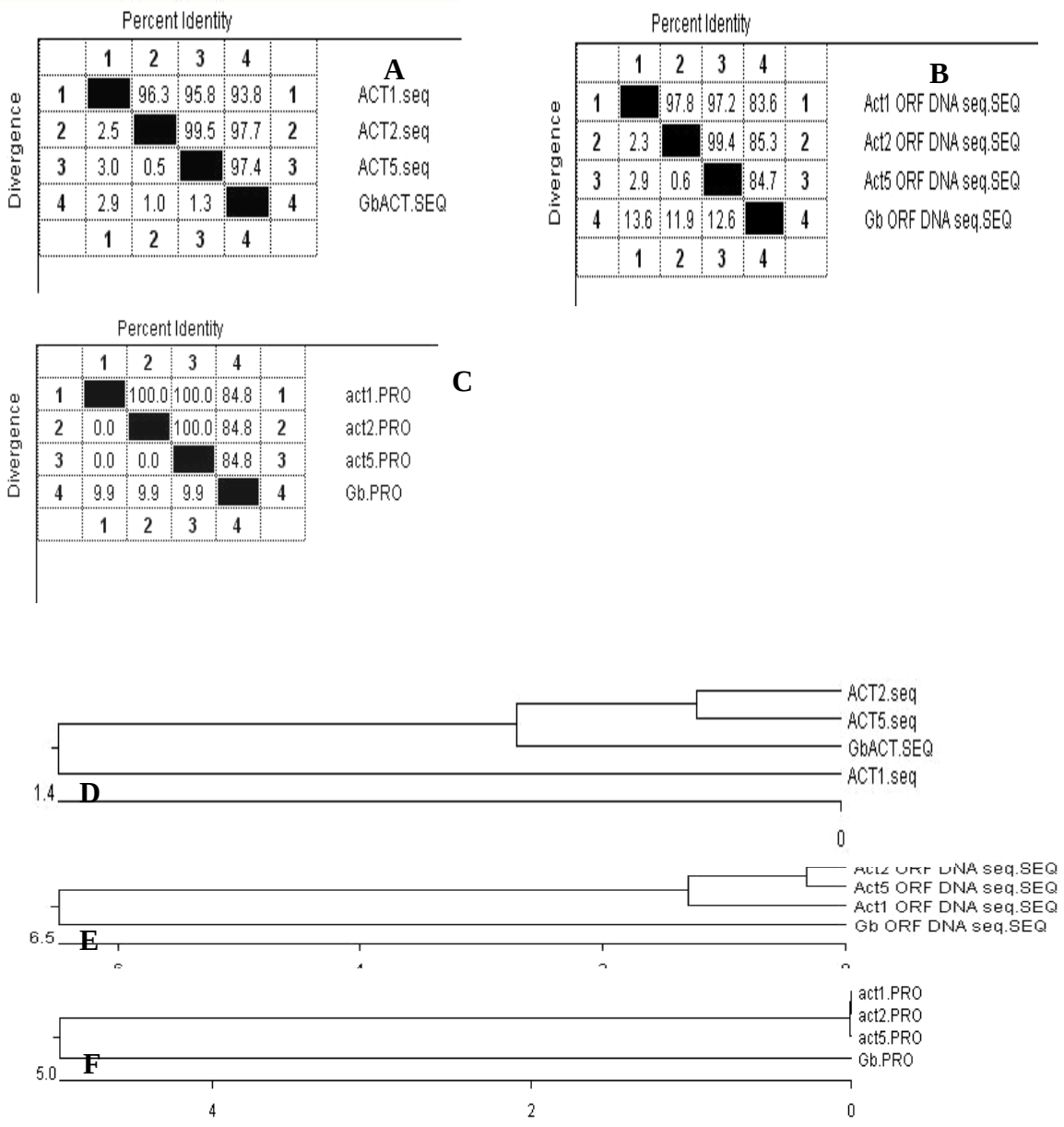

Fig. 4: (A) Percent of identity (similarity) and divergence between the three actins from *G. hirsutum* (ACT1, ACT2, ACT5) and the *GbACT* on the DNA sequence level. (B) Percent of identity (similarity) and divergence between the ACT1, ACT2, ACT5 and the *GbACT* on the Open Reading Frame (ORF) level. (C) Percent of identity (similarity) and divergence between the ACT1, ACT2, ACT5 and the *GbACT* on the amino acid sequence level. (D) A phylogeny tree summarizes the relationship between the *GbACT* and the ACT1, ACT2 and ACT5 on the level of DNA sequence. (E) A phylogeny tree summarize the relationship between the *GbACT* and the ACT1, ACT2 and ACT5 on the Open Reading Frame (ORF) level. (F) A phylogeny tree summarizes the relationship between the *GbACT* and the ACT1, ACT2 and ACT5 on the amino acid sequence level

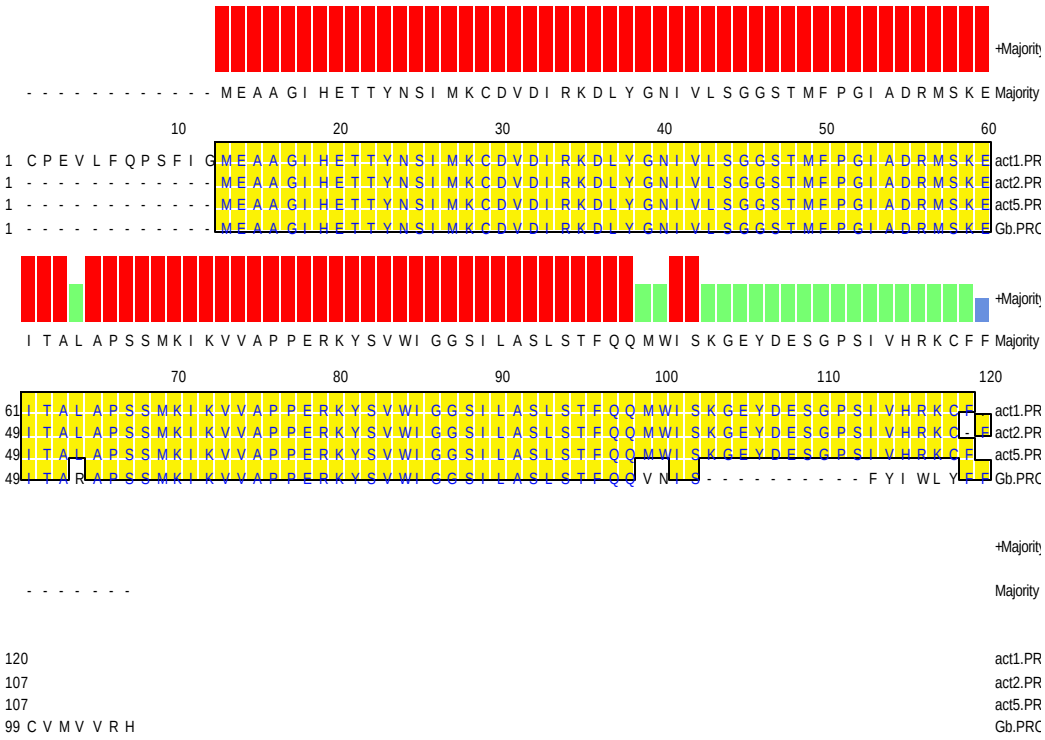

Fig. 5: Mega-alignment between the *GbACT* and the ACT1, ACT2 and ACT5 on the level of amino acid sequence. The alignment started from the methionin start codon (M) and ended with a stop codon

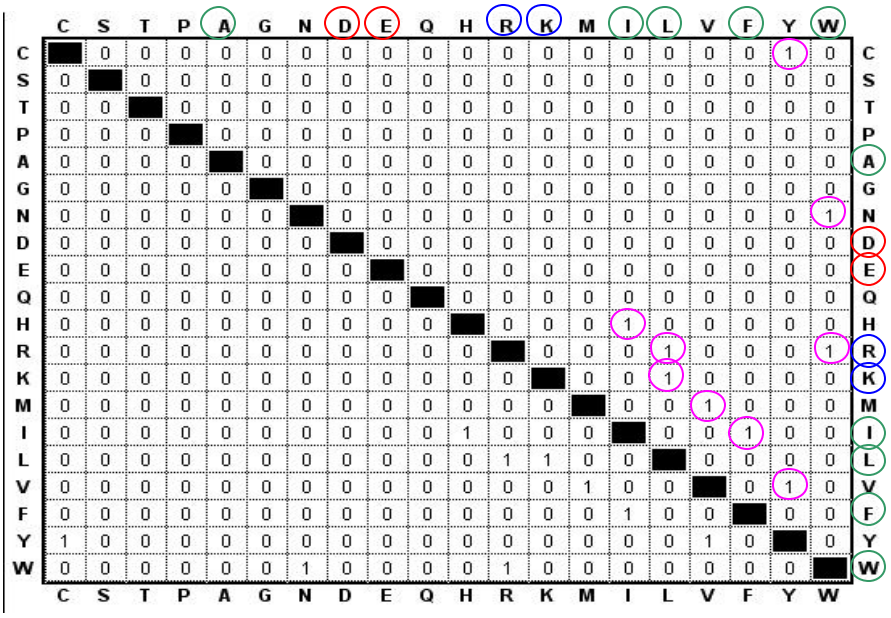

Fig. 6: Table represents the residue substitutions of the amino acid sequences of *GbACT* compared to that residue in the amino acid sequence found in the ACT1, ACT2 and ACT5 genes. Acidic residues circled in red, basic residue circled in blue and the hydrophobic residues circled in green. Number and the kind of residue substitutions that found between the aligned amino acid sequences circled in purple

***GbACT* alignment at the DNA and ORF levels:** GbACT sequence (Fig. 1) was aligned using the (BLAST) programs from the (NCBI). The database showed that the DNA sequence of the GbACT matches 189 different actin genes from different plant species. Sixteen of them were cotton *G. hirsutum* actin genes as shown in Fig. 2 and Fig. 4A. All the close homologue matches of cotton were of *G. hirsutum* and were of actin mRNAs. The identified GbACT sequence was a DNA sequence whereas the database contains the actin sequences in the mRNAs form. Therefore, the alignment shows some intervening regions (introns) in the middle of the sequences aligned in Fig. 2. The first intron was of size of ~74 nucleotides and was found in common with the all aligned actins, while the second intron was in size of ~16 nucleotides and was common in the hits of *G. Hirsutum* mRNAs designated ACT2, ACT5 and ACT1. For the *G. hirsutum* actins (ACT1, ACT2 and ACT5) the E Value was (0.0), which is highly significant, shown in Fig. 2. Multiple sequence alignment against GbACT revealed 39-97% identity between sequences (Fig. 4A and 4D). The other mRNAs from *G. hirsutum* actins (ACT3, ACT4, ACT6, ACT7, ACT8, ACT9, ACT10, ACT11, ACT12, ACT13, ACT14 and ACT15) showed different similarity ranges from 91%-98% identity. On the other hand, the *Malva pusilla* actin (Act1) mRNA showed 93% identity and the Glycine max actin (Soy86) gene showed 89% identity. These results showed that these family members are so close to each other.

Analysis of the Open Reading Frames (ORF) of ACT1, ACT2, ACT5 and GbACT starting from the forward primer sequence and ending with the terminal stop codon, represented a new insight inside the translated region and clearly demonstrates the main differences on the DNA sequences that lead to the major changes in the protein expressed from these sequences. The predicted GbACT open reading frame was shorter than *Gossypium hirsutum* ACT open reading frames, with substitutions at the 3'end of GbACT in comparison to GhACT1, GhACT2 and GhACT5.

The base substitution in the ACT1, ACT2, ACT5 and GbACT at positions 87, 105, 114, 144, 153, 162, 184, 191 and 295 were slightly acceptable because GbACT was similar to at least one of the other actins in these substitutions The only exceptions were in the position (191) where the base (T) in the three actins (ACT1, ACT2, ACT5) was substituted with (G) in GbACT. While in the base position (295) the base (A) in the three actins was substituted with (G) in GbACT. The base substitution was varied differentially and considerably starting from the position (297) in the alignment ended with the terminal stop codon. The stop codon was also substituted from (TAA) in the three actins (ACT1, ACT2 and ACT5) to the terminal stop codon (TGA) in the GbACT (Fig 3). The 3' end differences in the ORF was further assessed during alignment of the translated amino acid sequence. Noteworthy, the alignments also showed that GbACT sequence is shorter than *Gossypium hirsutum* ACT1, ACT2 and ACT5.

The data of similarity and divergence, estimated by the DNASTAR software program, is shown in Fig. 4B. The percentage of similarity between *GbACT* and the other three actins (*ACT1*, *ACT2* and *ACT5*) ranged from 85.3 to 83.6%. The ACT2 gene showed the closest sequence similarity to the *GbACT* (85.3%) followed by *ACT5* (84.7%) and *ACT1* that showed 83.6% sequence similarity. This data goes with that of similarity and divergence conducted from the alignment of the core DNA sequence (Fig. 4A). The phylogeny tree showed in Fig. 4E summarizes the data of alignment in the ORF level and confirmed the data that obtained and represented by Fig. 3.

Alignment at the Amino acid level: Since the similarity was very strong on the DNA sequence level between the three actins (ACT1, ACT2 and ACT5) and the GbACT, a more sensitive alignment was performed using the translated amino acid sequence for the ORF. Using amino acid sequences in multiple alignment allows for a more sensitive comparison, revealing third position mutations, subsititutions between amino acid with conserved properties, frameshifts, conserved motifs and domains.

The translated ORFs were used in comparing the amino acid sequences from the three actins of *G. hirsutum* (ACT1, ACT2, ACT5) aligned against GbACT. The alignment results showed a very conservative amino acid sequence, with some substitutions at the C-terminal (Fig. 5). The hydrophobic amino acid L (Leu) that was conserved in the three actins is substituted with the basic and charged amino acid Arg in GbACT at position 64, while at position 99 the amino acid methionine is substituted with the hydrophobic amino acid Val and in the position 100 the hydrophobic amino acid (Trp) is substituted with the polar amino acid (Asn) in the GbACT.

Compared to *G. hirsutum* actins, GbACT was significantly shorter. The shorter region reflects a deleted N-terminus. The Actin functional domain is present over the entire stretch of GhACT1, GhACT2 and GhACT5

while the domain is obviously trunctated at GbACT's N-terminal (using RPS-BLAST). This may indicate that GbACT is an isoform of actin. GbACT missing half of the actin domain and any known actin domain signature motifs (results not shown), still maintains the highly conserved actin sequence. The homology is mantained in GbACT from position 1-92 of the amino acid sequence.

The lack of the actin domain N-terminal in GbACT is interesting since all identified actin isoforms maintain the N-terminus. Mutation studies on the actin N-teminus showed disruption of the polymerization process, with the exception of yeast actin [21]. The N-terminal is present in all isoforms, however its sequence is not strictly conserved. The variability in actin isoform N-terminal sequences has been correlated to functional adaption of the different actin forms [21]. The complete lack of the N-terminal in GbACT makes this an interesting discovery for its expression, obvious conservation, yet lacking major actin features. This suggests that GbACT is an 'actin-like' isoform at best.

Another distinguishing feature for all actin isoforms, is a characteristic acidic N-terminus [22]. The beginning of GbACT's N-terminus consists of mostly hydrophobic residues, however, the overall charge on the N-region of the amino acid sequence is negative. This uncanny resemblance supports the evidence characterising GbACT as an 'actin-like' isoform.

A big gap was observed in the interval position 103-112 and six continuous amino acid substitutions were observed in the positions (113-118). Where the hydrophobic amino acid I is substituted with a hydrophobic F, the hydrophobic V is substituted with the polar Y and the charged H is substituted with the hydrophobic I. While the charged and basic R is substituted with the hydrophobic W, the charged and basic K is substituted with the hydrophobic L and the charged and polar C substituted with the charged and polar Y successively (Fig. 5). These amino acids substitutions may reflect structural and functional differences between the proteins at the C-terminal.

Profilin binds at the actin C-terminal. Profilin is a small eukaryotic protein that binds monomeric actin preventing polymerization. It has been shown that profilin plays an important role in cell wall synthesis and deposition stages during cotton fiber elongation. Actin binding affinity to profilin increases with N-terminal deletions. The deletions associated with increased profilin binding would occur in either or all actin domain residue positions 176, 177, 316-372 [12]. These deletions are roughly observed in GbACT, indicating GbACT may have a stronger affinity to profilin.

Variations within the actin gene family may represent substitutions in less constrained regions of the sequence or isoform-specific functional adaptation. The variations found the C-terminal most likely represents the latter, where the substitutions in the C-terminus is not likely to be a random sequence variation from GhACT1, GhACT2, or GhACT5. The substitutions represent grossly contrasting amino acid physiochemical properties, indicating a change in the physiochemical property of the C-terminus extremity.

In the sequence RKHYCDE, at the C-terminal, the polar amino acids are substituted with hydrophobic amino acids at positions 99, 115, 115 and 117. Five of the six amino acids are aromatic. Substitutions for aromatic amino acids are rare and are usually associated with functional or structural adaptations or severe mutations [23]. Notably, the acidic (DE) amino acids in this region were conserved, as shown in Fig. 6.

Overall the sequence similarity between the three noted *G. hirsutum* actins and GbACT was 84.8% (Fig. 4C), indicating a high level of homology between the sequences and structural and functional similarity.

Phylogenetic analysis using DNA sequences of actin gene family members have commonly indicated that gene families are more conserved between than within species [21]. This was evidenced in Fig. 4D, where GhACT2 and GhACT5 showed stronger homology to GbACT than GhACT1. This however was not evidenced when evolutionary distances were assessed using ORFs or protein. In fact, evolutionary distance increased respectively. This may be due to the conserved gene architecture between GhACT2 and ACT5 with GbACT which may be attributed to the conserved intron sequences in ACT2, ACT5 and GbACT. Removal of intron (in ORFs) leads to increased evolutionary distance of GbACT. In Fig. 4F, the phylogenetic tree supports amino acid sequence alignment evidence for diverged structure and function of GbACT.

CONCLUSIONS

Sequence analysis and comparison of GbACT revealed that it is not a just a homolog of either of *G. hirsutum's* actins, but may be a novel 'actin-like' isoform present, until now, uniquely in *G. barbadense*. GbACT C-terminal variation from GhACT1, GhACT2 and GhACT5, may represent difference in profilin binding affinity and thus cotton fiber quality and length between *G. barbadense* and *G. hirsutum*. Further work is needed to determine this conclusively. Structural and functional

studies on GbACT may be insightful in determining the effect of having an incomplete actin domain at the N-terminal especially that specific actin interaction sites are missing. This will be useful in determining how the actin subdomains interact during folding and actin polymerization. Furthermore, the characterization and sequence analysis of the GbACT gene from Egyptian cotton provides a useful tool to identify transcription regulators confirming its fiber-specific expression and to direct potential target genes for fiber quality improvement. In addition, the work represented in this paper consider a start point towards cloning and isolating the full length cDNA coding for actin for the strategic development of cotton plants with improved fiber length and quality by using genetic transformation tools.

ACKNOWLEDGMENTS

The Authors would like to thank Maryam Elshafei, teaching assistance in Faculty of Biotechnology - Modern Science and Art University (MSA) for her assistance with the manuscript and valuable discussion.

REFERENCES

1. Anthony, R.G. and P. Hussey, 1999. Dinitroaniline herbicide resistance and the microtubule cytoskeleton. Trends in the Plant Science, 4: 112-116.
2. Kost, B., Y.-Q. Bao and N.-H. Chua, 2002. Cytoskeleton and plant organogenesis. Philosophical Transactions of the Royal Society, Series B, 357: 777-789
3. Sheterline, P., and J.C. Sparrow, 1994. Biological characteristics of actin. Protein Profile, 1: 1-12.
4. Fu, Y., H. Li and Z. Yang, 2002. The ROP2 GTPase controls the formation of cortical fine F-actin and the early phase of directional cell expansion during Arabidopsis organization. Plant Cell, 14: 777-794.
5. Hussey, P.J., E.G. Allwood and A.P. Smertenko, 2002. Actin-binding proteins in the Arabidopsis genome database: propertiesof functionally distinct plant actin-depolymerizing factors/cofillins. Phil. Tans. R. Soc., 357: 791-798.
6. Xue-Bao Li, Xiao-Ping Fan, Xiu-LanWang, Lin Cai and Wei-Cai Yang, 2005. The Cotton ACTIN1 Gene Is Functionally Expressed in Fibers and Participates in Fiber Elongation. The Plant Cell, 17: 859-875.
7. Morrissette, N.S. and L.D. Sibley, 2002. Disruption of microtubules uncouples budding and nuclear division in *Toxoplasma gondii*. J. Cell Sci. 115: 1017-1025.
8. Lew, A.E., A.R. Dluzewski, A.M. Johnson and J.C. Pinder, 2002. Myosins of Babesia bovis: Molecular characterization, erythrocyte invasion and phylogeny. Cell Motility and the Cytoskeleton, 52: 202-220.
9. McDowell, J.M., S. Huang, E.C. Mckinney, Y.Q. An and R. Meagher, 1996. Structure and evolution of the actin family in *Arabidopsis thaliana*. Genet. Socity of Amer., Genetics, 142: 587-602.
10. Mathur, J., P. Spielhofer, B. Kost and N.H. Chua, 1999. The actin cytoskeleton is required to elaborate and maintain spatial patterning during trichome cell morphogenesis in *Arabidopsis thaliana*. Development, 126: 5559-5568.
11. Gilliland, L.U., L.C. Pawloski, M.K. Kandasamy and R.B. Meagher, 2003. Arabidopsis actin gene ACT7 plays an essential role in germination and root growth. Plant J., 33: 319-328.
12. Li, X-B., F. Xiao-Ping, W. Xiu-Lan, C. Lin and Y. Wei-Cai, 2005. The Cotton *ACTIN*1 Gene is Functionally Expressed in Fibers and Participates in Fiber Elongation. The Plant Cell, 17: 859-875.
13. Baird, W.V. and R.B. Meagher, 1987. A complex gene superfamily encodes actin in petunia. EMBO J., 6: 3223-3231.
14. Thangavelu, M., D. Belostotsky, M.W. Bevan, R.B. Flavell and H.J. Rogers *et al*., 1993. Partial characterization of the Nicotiana tabacum actin gene family: evidence for pollen specific expression of one of the gene family members. Mol. Gen. Genet., 240: 290-295.
15. Meagher, R.B. and R.E. Williamson, 1994. The plant cytoskeleton in Arabidopsis, Cold spring Harbor laboratory press, Cold spring Harbor, NY., pp: 1049-1084.
16. Meagher, R.B., 1994. The impact of historical contingency on gene phylogeny: plant actin diversity, in evolutionary biology, Vol. 28, plenum Press, N.Y.
17. Meagher, R.B., 1991. Divergence and differential expression of actin gene families in higher plants. Int. Rev. Cytol., 125: 139-163.
18. Frengen, E., D. Weichenhan, B. Zhao, K. Osoegawa, G. Van and P.J. Jong, 2000. Modular bacterial artificial chromosome vectors for transfer of large Inserts into mammalian cells. Genomics, 58: 118-126.
19. Momtaz, O.A., M.I. Shehata, S. El-Khodary and H.Z. Hassan, 2006. Idntification and isolation of smHSPs gene family from *Gossypium barbadense* L. c.v. Giza90 Egyptian cotton variety. Arab J. Biotech., 9: 203-215.

20. Momtaz, O.A., A.E.A. El-Doliefy, A. Bahieldin and E.H. Hassanein, 2007. Idntification of actin related gene family from *Gossypium.* barbadense. Giza88 cultivar using PCR-based positional cloning. Arab J. Biotech. 10, (1): (in press)
21. Hennessey, E.S., D.R. Drummond and J.C. Sparrow, 1993. Molecular genetics of actin function. Biochemistry Journal, 282: 657-671.
22. Zeche, K. and K. Weber, 1978. Actins from mammals, bird, fish and slime mold characterized by isoelectric focusing in polyacrylamide gels. Eur. J. Biochem., 89: 105-112.
23. Henikoff, S. and J.G. Henikoff, 1992. Amino acid substitution matrices from protein blocks. Proceeding of the National Academy of Science, 89: 10915-10919.