CELLULOSE SYNTHASE GENE FAMILY IN TETRAPLOID COTTONS: FURTHER STUDIES

Olga Yurchenko, Allan Zipf, and Khairy Soliman
Center for Molecular Biology, Plant and Soil Science
Alabama A&M University
Normal, AL
Aaron Jeffries
Agronomy Department
Iowa State University
Ames, IA
Pat Hogan
Section of Plant Biology
University of California-Davis
Davis, CA
Debby Delmer
Rockefeller Foundation
New York, NY

Abstract

Cotton, the world's most important natural fiber, is mostly cellulose. More higher strength cottons are needed as yarn and textile manufacturers adopt high speed production machinery. The enzyme responsible for cellulose production is the multimeric cellulose synthase, of which we know most about the catalytic subunit, *CesA*. The question is why does the simplest, fiberless, genome, *Arabidopsis*, contain 13 *CesA* loci while with fibered cotton only 5 have so far been identified. Genomic DNA was extracted from young leaves of *Gossypium hirsutum* and *G. barbadense* and used for PCR using a primer pair targeted towards the 5' region of the *CesA* gene. Amplified fragments were cloned and sequenced in the reverse and forward directions and the consensus sequences aligned and analyzed using web-based software. Phylograms, including known *G. hirsutum* and *Arabidopsis CesA* sequences, revealed that tetraploid cottons have at least one new *CesA* family member. Also, there were indications that *CesA* pseudogenes may also exist in both genomes.

Introduction

Cotton, the most important natural fiber in the world, is essentially cellulose, with the secondary wall being almost pure cellulose (Benedict et al., 1999). Due to global competition, many yarn and textile manufacturers have adopted high speed production machinery that has increased the demand for higher strength raw cotton fiber. Therefore, cotton fiber qualities must be continually improved to remain competitive in the markets and meet the needs of new spinning and weaving methods.

The strength of individual cotton fibers resides primarily in the wall thickness, and along with other quality measures such as fineness (micronaire) and length, depend not only on the rates but also the periods of synthesis, elongation and thickening (Benedict et al., 1999). The enzyme responsible for cellulose production is a member of the β-glycosyltransferase family (Richmond & Somerville, 2000), cellulose synthase (http://cellwall.stanford.edu/php/display_tree.php?family=CESA_only). Cellulose synthase is complexed in a membrane-bound, 6-lobed multimeric "rosette" (Eckardt, 2003), possibly complexed with sucrose synthase (Amor et al., 1995), that generates cellulose strands that then eventually associate as microfibrils. Only the catalytic subunit *CesA* has been identified and between 18 to 36 are thought to be present in a rosette (Gardiner et al., 2003).

CesA is, in itself, a gene family with each species having more than one locus [ex. *Arabidopsis* =13 (Holland et al., 2000), cotton=5 (http://cellwall.stanford.edu/php/sequence.php?superfamily=cesa&species=Gossypium+hirsutum&family=CESA)] *CesA* is thought to form hetero- or homo-dimers (Kurek et al., 2002) and more than one distinct type of *CesA* may be required for proper rosette assembly and function (Scheible et al., 2001).

In addition to the economic and biochemical importance of cellulose in cotton, scientists are trying to understand the diversity of the *CesA* gene family in *Gossypium* and whether or not the molecular diversity of CeSA in *Gossypium* reflects its evolutionary history. One question that is relevant to the study of the evolutionary and diversity of cotton is: Why does *Arabidopsis thaliana*, with the smallest genome in plants, have at least 13 *CeSA* genes and *G. hirsutum*, a tetraploid, has only 5 reported to date.

Materials and Methods

The overall scheme in Figure 1 was followed.

Young leaves were collected from greenhouse-grown *G. hirsutum* var TM-1 and *G. barbadense* 3-79 plants, placed on ice for transport to the lab. The leaves were frozen in liquid nitrogen, ground to a fine powder with a mortar and pestle and the genomic DNA (gDNA) extracted using the DNEasy minikit (Qiagen, Valencia, CA). DNA quality and yield was checked via spectrophotometer and electrophoresis on a 1% agarose gel.

PCR amplification was performed using two primers derived from GhCesA1 (Figure 2) with restriction sites added to the 5' end. PCR 1A (Forward) 5'-ATCCGGAATTCGCCATGCTGACATTTGAATCTC-3' and PCR 2A (Reverse) 5'-AAAACTGCAGAACCAGTCGAGGAAGTTCATTTCC-3' were used in a PCR reaction method with The reaction components were 10X PCR buffer (without $MgCl_2$), 10mM of dNTPs, 10 µM each of forward and reverse primer, autoclaved H_2O and 0.5 units of High Fidelity Platinum *Taq* DNA polymerase (Invitrogen, Carslbad, CA) in a 50 µl reaction mix. The PCR reactions were performed in a Peltier (PTC 200) thermal cycler (MJ Research, Watertown, MA). The thermal cycler was programmed as follows: initial melting temperature was 94°C for 5 minutes, followed by 30 cycles of denaturing 94°C for 30 seconds, 50°C at 30 seconds for primer annealing and 72°C for 1 minute for DNA strand extension and a final extension for 10 min. at 72°C.

Amplified fragments were cloned using a TA Cloning Kit For Sequencing (Invtirogen) according to manufacturer's directions. Clones were screened for insert after isolation of plasmid DNA using a QIAPrep miniprep kit (Qiagen) using restriction enzyme digestion. Clones with inserts were sequenced backwards and forwards using a BigDye3 Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Reactions using M13 forward and M13 reverse primers [TOPO TA Cloning Kit for Sequencing (Invitrogen-Life Technologies, Carlsbad, CA, USA)] were prepared separately and sequence analyses of the individual clones were carried out using an ABI 3100 DNA Sequencing instrument (Applied Biosystems, Foster City, CA) at the Alabama A&M University Center For Molecular Biology. Three clones were sequenced at The University of Alabama Automated DNA Sequence Facility using M13-forward and M13-reverse sequencing primers.

Forward and reverse sequences from the same clone were used to determine a consensus sequence using Pairwise Blast (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html). Consensus sequences from all clones were initially submitted to a multiple sequence alignment program, CLUSTALW (http://www.ebi.ac.uk/clustalw/#), and the introns deleted and the sequences were trimmed at the 5' and 3' ends to reduce penalties for missing sequences. Comparison of obtained sequences was made to known, homologous *CesA* regions of *AtCesA* (*Arabidopsis thaliana*) genes 1-9, 13 and *GhCesA* (*G. hirsutum*) genes 1-4, obtained from GenBank mRNA/cDNA accessions, if available. Phylograms were generated from the re-aligned sequences using the embedded Neighbor-Joining analysis module in CLUSTALW with Gaps Ignored and Distance Corrected.

Results and Discussion

Genomic DNA was successfully isolated form both species and was of high quality as seen by 1% agarose gel electrophoresis (Figure 3). Single amplified gene products were obtained after PCR (Figure 4) but early attempts at direct sequencing were not successful, necessitating cloning before successful sequencing.

BLAST analysis of sequenced genomic amplified products showed the presence of an intron of ~80 bp within this amplified region. However, the length of this intron varied, indicating the possibility of at least two separate loci/alleles (data not shown). In order to compare with sequences representative of the different CesA families in GenBank, some of which were only known as mRNA/cDNA, it was necessary to remove the introns from all genomic DNA-derived sequences.

The sequences were trimmed at the 5' and 3' ends and manually realigned, if necessary, using the embedded Jalview software, available at the CLUSTALW website. Redundant sequences were also removed at this time. Aligned sequences were re-submitted and the phylogram generated using the embedded Neighbor-Joining method with Gaps Ignored and Distances Corrected.

Even within just this short region of DNA, the association of genes generally continued to reflect the homologies generated from longer stretches of DNA (Compare Figures 5 and 6). For example, *AtCesA4* and *GhCesA2* cluster together, *AtCesA2* and *AtCesA9* cluster together with the cluster of *AtCesA5* and *AtCesA6*, that then associates more with *AtCesA7*. However, *GhCesA4* does not associate with the expected *AtCesA1*, *AtCesA3* and *GhCesA3* (Lower right of Figure 6). Instead, *GhCesA4* clusters more with *AtCesA8* and *GhCesA1*, indicating this gene may have had an evolutionary story more complex than other family members.

Although some clone sequences clustered with known *CesA* genes from *G. hirsutum*, the majority formed their own distinct (super)cluster, indicating further divergence from known *CesA* family members in *G. hirsutum*. It was not possible to compare clone sequences with that of *GhCesA5* as this partial sequence did not span the region from which the primers were derived.

Therefore, there is the indication of at least one more *CesA* family member type in both *Gossypium hirustum* and *Gossypium barbadense*.

Upon translational analysis of the cloned sequences, it was noted that a majority of the sequences were apparently "pseudogenes" as they terminated abruptly or had possible frameshifts (Figure 7). It is not known if these differing sequences were native or were due to polymerase errors during PCR and/or sequencing. Since the majority of presented sequences were sequenced in the reverse and forward directions and so represent actual sequence, this would seem to eliminate possible base-calling errors. Pseudogenes have been found in cotton before [*GhCesA2* pseudogenes in *G. arboreum*, *G. herbaceum* and A genome of *G. hirsutum* (Cronn et al., 1999)] but the ones we found were apparently more closely related to *GhCesA4/GhCesA1*. Further work is needed to confirm the presence of pseudogenes in cotton.

Conclusions

Predicted Product = 376 bp, Actual Product = ~450 bp => INTRON!

G. barbadense and *G. hirsutum* have *GhCesa1* orthologs.

G. barbadense and *G. hirsutum* have *GhCesa1* paralogs
(Meaning of second major cluster = Member of CesA1 family?
New CesA family?)

GhCesa1 gene family? Are sequence differences due to procedure
(*Taq*, ABI, etc. error) or real variation?

Acknowledgements

We acknowledge the efforts of Dr. Zhengdao Wu for assistance with the sequencing, Mr. James Bolton for training in DNA extraction and Mr. Tom Smith for greenhouse maintenance of the plants. Financial support was provided by Cotton Inc Grant # 99-681AL and a UC-Davis Cooperative Research Agreement supported by NSF Grant DBI 0110173.

References

Amor, Y., C.H. Haigler, S. Johnson, M. Wainscott, and D.P. Delmer. 1995. A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. Proc. Natl. Acad. Sci. USA. 92:93539357.

Benedict, C.R., R.J. Kohel and H.L. Lewis. 1999. Cotton fiber quality. *In*: COTTON: Origin, History, Technology and Production (C.W. Smith and J.T. Cothren, eds.). John Wiley & Sons, Inc. New York.

Cronn, R.C., R.L. Small, and J.F. Wendel. 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. USA. 96:1440614411.

Eckardt, N. 2003. Cellulose synthesis takes the CesA train. Plant Cell 15:1685-1688.

Holland, N., D. Holland,T. Helentjaris, K.S. Dhugga, B. Xoconostle-Cazares, and D.P. Delmer. 2000. A comparative analysis of the plant cellulose synthase (CesA) gene family. Plant Physiol. 123:1313–1324.

Kurek, I., Y. Kawagoe, D. Jacob-Wilk, M. Doblin, and D. Delmer. 2002. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. Proc. Natl. Acad. Sci. USA. 99:11109–11114.

Scheible, W.-R., R. Eshed, T. Richmond, D. Delmer, and C. Somerville. 2001. Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in Arabidopsis *Ixr1* mutants. Proc. Natl. Acad. Sci. USA. 98:10079–10084.

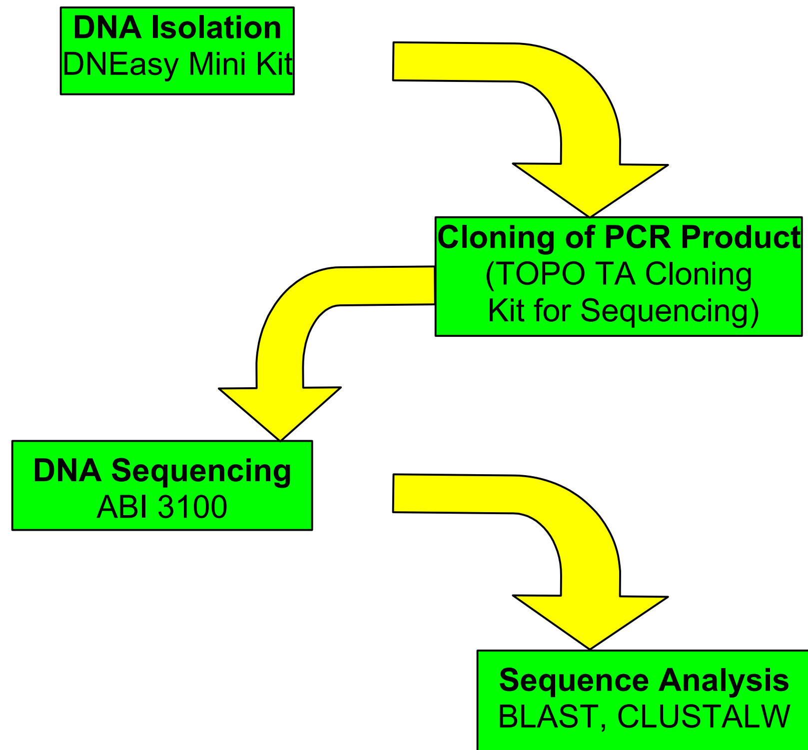

Figure 1. Overall Methodology for *GhirCesA*, *GbarCesA* Analysis.

961 g**gccatgctg acatttgaat ctc**tagtaga aacagccgac tttgcaagaa agtgggttcc
PCR 1A (Forward)
1021 attctgcaaa aaattttcca ttgaaccccg ggcacctgag ttttacttct cacagaagat
1081 tgattacttg aaagataaag tgcagccctc ttttgtaaaa gaacgtagag ctatgaaaag
1141 agattatgaa gagtacaaaa ttcgaatcaa tgctttagtt gcaaaggctc agaaaacacc
1201 tgatgaagga tggacaatgc aagatggaac ttcttggcca ggaaataacc cgcgtgatca
1261 ccctggcatg attcaggttt tccttggata tagtggtgct cgtgacatcg aa**ggaaatga**
1321 **acttcctcga ctggtt**tacg tctctagaga gaagagacct ggctaccaac accacaaaaa
PCR 2A (Reverse)

Figure 2. Partial *GhCesA1* cDNA sequence showing location of PCR primers used in this study. Both primers had restriction enzyme sequences added to the 5' ends for cloning purposes.

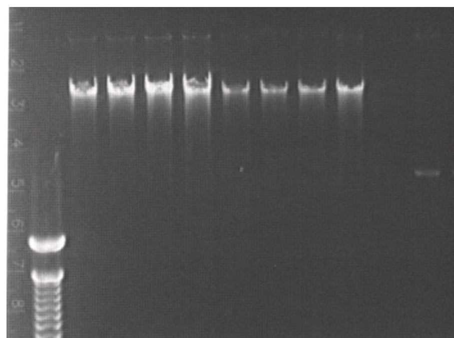

Figure 3. Electrophoresis of representative cotton genomic DNA on 1% agarose gel.

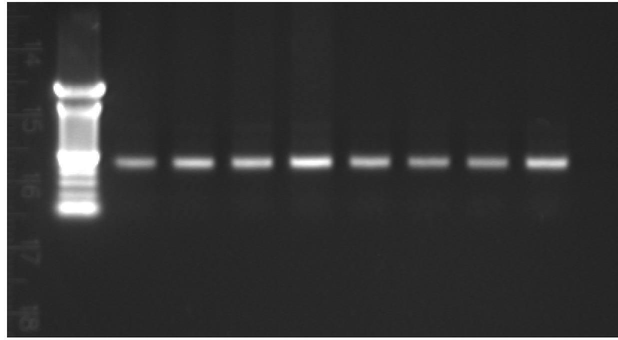

Figure 4. Gel electrophoresis of PCR products using primer pair 1A/2A.

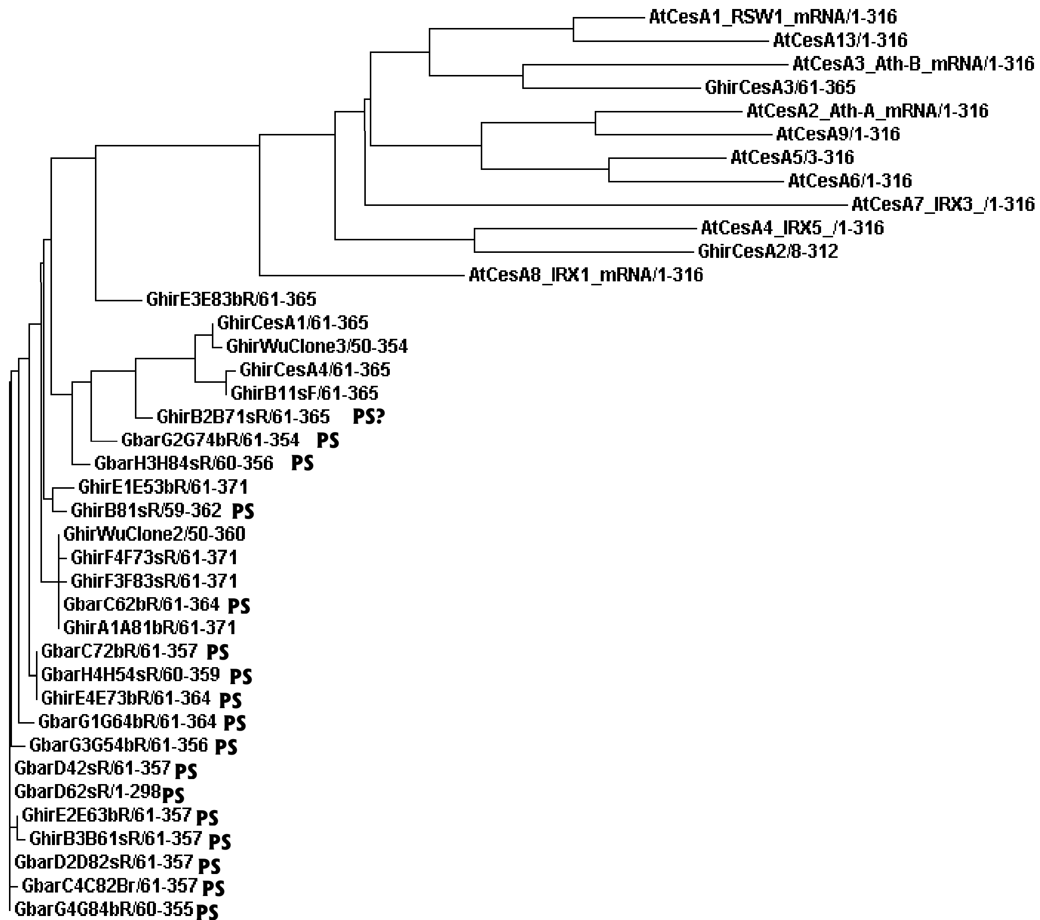

Figure 5. Phylogram generated using Neighbor-Joining method with No Gaps and Distance Corrected (available at CLUSTALW website) on non-redundant, intron-deleted, 5'- and 3'-trimmed sequences from *Gossypium hirsutum* var. TM-1, *G. barbadense* var. 3-79, *GhCesA1-4* and *AtCesA1-9, 13* aligned using CLUSTALW (website http://www.ebi.ac.uk/clustalw/#). **PS** indicates possible pseudogene.

Figure 6. Unrooted, Bootstrapped Tree of the *CesA* Family http://cellwall.stanford.edu/php/display_tree.php?family=CESA_only

```
 21    G  F  H  F  A  K  S  S  L  L  N  P  G  H  P  S  F  T  S  H
 21   W  V  P  F  C  K  K  F  S  I  E  P  R  A  P  E  F  Y  F  S
 21  V  G  S  I  L  Q  K  V  L  Y  *  T  Q  G  T  R  V  L  L  L
 61  GTGGGTTCCATTTTGCAAAAAGTTCTCTATTGAACCCAGGGCACCCGAGTTTTACTTCTC
     |||||||||||| |||||||| || || ||||||||  ||||||| ||||||||||||||
 61  GTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCACGGGCACCTGAGTTTTACTTCTC
 21    G  F  H  S  A  K  N  F  P  L  N  H  G  H  L  S  F  T  S  H
 21   W  V  P  F  C  K  K  F  S  I  E  P  R  A  P  E  F  Y  F  S
 21  V  G  S  I  L  Q  K  I  F  H  *  T  T  G  T  *  V  L  L  L

 41    R  S  L  I  T  *  R  I  S        L  I  C  E  R  T  *  S
 41   Q  K  L  D  Y  L  K  D  I        P  H  L  *  K  N  V  E  Q
 41  T  E  A  *  L  L  E  G  Y.........P  S  F  V  K  E  R  R  A
121  ACAGAAGCTTGATTACTTGAAGGATAT-------CCCTCATTTGTGAAAGAACGTAGAGC
     ||||||| ||| ||||||||| ||||        ||||| ||||| ||||||||||||||
121  ACAGAAGATTGTTTACTTGAAAGATAAAGTGCAGCCCTCTTTTGTAAAAGAACGTAGAGC
 41    R  R  L  F  T  *  K  I  K  C  S  P  L  L  *  K  N  V  E  L
 41   Q  K  I  V  Y  L  K  D  K  V  Q  P  S  F  V  K  E  R  R  A
 41  T  E  D  C  L  L  E  R  *  S  A  A  L  F  C  K  R  T  *  S
```

Figure 7. DNA and Protein Sequence Comparison of GhirE2E63bR *CesA* pseudogene vs. **GhirCesA4-intron deleted (bold)**.