Cloning and Analysis of Xyloglucan Endotransglucosylase/Hydrolase Gene in Cotton

Xianliang Li 1, a

1 School of Biology and Agriculture, Huanggang Polytechnic College, Huanggang, Hubei, China

libusher@sina.com

Abstract. Xyloglucan endotransglucosylase/Hydrolase (XTH) can loosening cell wall by breaking and re-forming the xyloglucan-cellulose crosslinks. Therefore, XTHs in cotton are related to elongation of cotton fibers. To investigate the relation, two XTH genes were cloned, respectively named as GhXTH1 (GenBank:AY189971) and GhXTH2 (GenBank:JN968478). GhXTH1 contained 885 nucleotides, encoding 289 aa. GhXTH2 consisted of 885 nucleotides, encoding 289 aa. GhXTH1 and GhXTH2 contained the conserved motif (DEIDFEFLG), characteristic of XTH family. Bioinformatic analyses showed that they possessed a transmembrane helice and signal peptide respectively. Phylogenetic analysis suggested that GhXTH1 was orthologous gene to AtXTH6, AtXTH7. GhXTH2 was orthologous gene to AtXTH9. In addition, we analysed the expression of GhXTH1 and GhXTH2, respectively in Gossypium hirsutum and Gossypium barbadense. The results showed that transcript levels were lower during cotton fiber development, and transcript level of GhXTH2 were higher than that of GhXTH1.

Keywords: Cotton fiber; Xyloglucan endotransglucosylase/Hydrolase; Cloning; expression analysis; elongation.

1. Introduction

Cotton fiber is developed from ovule epidermal cells, and mature cotton fiber length is one of the components of its quality. The length of mature fibroblasts can be 1000-3000 times of their diameter, and some can reach 35-40mm[1].The mature cotton fiber of Gossypium barbadense (Gb) is longer than that of Gossypium hirsutum (Gh), but 24 days after flowering (DPA), the fiber length of upland cotton is longer than that of island cotton[2]. The elongation of cotton fiber cells begins on the day of flowering[3-5], and the effective elongation of fiber cells is generally 24-28 days[1]. The maximum elongation rate of fiber cells occurs at 6-12 DPA and 15-20 DPA, and the elongation amount of these two stages can reach 80% of the final length[6].

The elongation of cotton fiber cells is driven by intracellular pressure and accompanied by cell wall relaxation. The increase of osmotic solute in the cell promotes the increase of osmotic pressure in the cell, which makes the cell absorb a lot of water, thus increasing the cell turgor. Soluble sugar, malate and potassium ion (K+) [7] were the main solutes that increased cell osmotic pressure. Phosphate pyruvate carboxylase (PEPC)[8], plasmodesmata[9-10], sucrose[11], and aquaporin[12] are related to the production or increase of cotton fiber cell swelling.

Cotton fiber is a dynamic network composed of cellulose, hemicellulose and structural proteins. Hemicellulose xylan is a structural polysaccharide of the primary wall of plant cells. The xylan chain is attached to the cellulose chain through covalent hydrogen bonds, and the adjacent cellulose microfibrils are crosslinked[13]. Xyloglucan cross-linking structure between microfilaments is one of the main limiting factors for cell elongation. This cross-linking structure can be uncoupled and reconnected by Xyloglucan endotrans glucosylase/Hydrolase (XTH), thus reducing the resistance to cell elongation. Therefore, XTH can promote cell elongation by relaxing cell wall. The recombinant Arabidopsis thaliana (At) XTH14 and XTH26 were added to the root system, showing a significant role in promoting root cell elongation[14]. Overexpression of AtXTH18,19,20 in Arabidopsis can stimulate hypocotyl elongation[15]. Overexpression of GähltH1 in cotton can enhance XTH activity and increase cotton fiber length by 15-20%[16]. The expression of XTH in cotton fiber was time specific and variety specific[17], and there was a
difference in the expression of XTH in sea island cotton and upland cotton fiber[18]. XTH expression mode is associated with cotton fiber length. In this study, two XTH genes were cloned and isolated from upland cotton, and their sequence analysis, phylogenetic analysis and expression analysis were carried out to lay the foundation for understanding the relationship between XTH gene and its family and cotton fiber elongation, so as to provide candidate genes for cotton fiber quality improvement, especially for length quality improvement.

2. Materials and methods

2.1 Test materials

The cotton materials used are land cotton Kezi 201, Huamian 99, and island cotton is Junhai No. 1. Take cotton petals 4 days (4 DPA), 9 days (9 DPA), 14 days (14 DPA), 19 days (19 DPA), 24 days (24 DPA) after flowering and put them into liquid nitrogen, and then store them in the refrigerator at -80 °C for standby. Escherichia coli strain DH5 for molecular cloning. For laboratory preservation, the intermediate vector pMD18-T was purchased from TaKaRa.

2.2 Extraction of total RNA and synthesis of cDNA

Total RNA of cotton fiber was extracted by hot boric acid method[19]. Then process the total RNA as shown in DNase I kit to remove DNA from it. Finally use 15μL DEPC ddH2O was dissolved, and the RNA concentration was measured by spectrophotometer. Then operate according to Promega reverse transcriptase system to synthesize the first strand of cDNA, and store the reverse transcribed cDNA at -20 °C.

2.3 GhXTH1 and GhXTH2 cloning

In NCBI, there is an XTH full length coding sequence (AY189971) and an EST sequence (DV848907). For AY189971, according to its full length sequence, primers were designed from both ends (GhXET-OE-F: AAAGTGGACATTCTCTCTCTCTCTCTCTGTTTA; GhXET-OE-R: AAAGGTACCTCCAGATGGGAGATGCAGACT), and 14 day cotton fiber cDNA was used as template for PCR amplification and sequencing. For DV848907, we compared it with the AtXTH sequence, and found that the EST sequence was highly homologous with the 5 'end of the AtXTH gene, and the homologous region contained the starting codon. Therefore, primers were designed according to this EST sequence to conduct 3' RACE (Rapid amplification of cDNA ends), and a DNA sequence was obtained by sequencing. This sequence was spliced with this EST sequence for translation. It was found that a complete protein sequence could be translated from the third base of the EST sequence. There are two start codons in this sequence, and the length of the first start codon translated into the protein sequence is relatively consistent with the length of the XTH protein sequence. The RACE method can be found in reference[20].

2.4 GhXTH bioinformatics analysis

ClusterW software is used to align and sort multiple sequences, and GenDOC and MEGA5.0 software are used to output the results of homology alignment and evolution tree construction[21]; SingaIP was used for signal peptide prediction[22]; The relative molecular weight and theoretical isoelectric point[23] of protein were calculated by ProtParam; TMHMM was used to predict the transmembrane domains of proteins.

2.5 GhXTH1 and GhXTH2 expression analysis

UBQ7 is an internal reference gene for RT-PCR analysis, and the designed primer sequence is GhUBI-RT-F, GAAGGCAATCCACCTGACCAAC; GhUBI-RT-F, TTTGACCTTCTTCTTTCTTG. GhXTH primer sequence: GhXTH1-RT-F, TCGTGACAGCAGATGAGAGATC; Gh.
3. Results and Analysis

3.1 Cloning and analysis of GhXTH1 and GhXTH2 full-length sequences

GhXTH1 has a full length sequence (AY189971) in NCBI. According to this sequence, primers were designed for PCR amplification, and the amplified product was subjected to agarose gel electrophoresis (Figure 1 (A)). The fragment size was between 750-1000bp, consistent with the full length sequence in NCBI. Sequencing of this fragment showed that there was a base difference between this sequence and AY189971, but the protein sequence was completely consistent. The gene was named GhXTH1.

There is an EST sequence DV848907 of GhXTH in NCBI. Comparing this sequence with the XTH sequence in Arabidopsis, it is found that this sequence covers the 5’ end of this gene. Therefore, only 3’-RACE is needed to obtain the full length sequence of this gene. According to the kit (SMART™ RACE cDNA Amplification Kit) instructions, carry out 3’- RACE, amplify the segment to carry out agarose gel electrophoresis (Figure 1 (B)), sequence the segment, splice the sequence obtained from sequencing with EST, translate the sequence obtained in Premier software, and find that a protein sequence can be translated from the third base. Starting from the first starting codon, a 289 amino acids (aa) can be translated. Using the DNA sequence encoding this sequence as the template to design a primer, and using this primer for PCR amplification, we can get a segment with the expected size (Figure 1 (C)). Sequencing this segment, the obtained sequence is completely consistent with the previous splicing sequence. The sequence has been submitted to the NCBI library (JN968478).

![Fig. 1](image-url)

**Fig. 1** Electrophoresis results of full-length fragments of GhXTH1 and GhXTH2. (A) full-length fragments of GhXTH1; (B) 3’-RACE fragments of GhXTH2; (C) full-length fragments of GhXTH2.

3.2 Analysis of GhXTH1 and GhXTH2 protein characteristics

The full length sequence of GhXTH1 contains 289 aa, and GhXTH2 contains 294 aa. XTH multiple sequence alignment was obtained by cloning XTH from Arabidopsis thaliana and several other plants. The results show that cotton XTH is relatively conservative with other XTHs, and all sequences have a conservative sequence DEIDFEFLG[24-25] that maintains XTH activity, but some sequences differ from it by 1-2 aa (Figure 2). The conservative sequence of GhXTH1 differs from this by 1 aa, and the third aa changes from isoleucine (I) to leucine (L); The conservative sequence of GhXTH2 differs from this by 2 aa, the first aa changes from aspartic acid (D) to asparagine (N),...
and the third aa changes from isoleucine to phenylalanine (F). Multiple XTH conserved sequences in Arabidopsis XTH family have 1-2 aa changes.

XTH is a cell wall protein. Therefore, it is helpful to understand the biological function of the two cotton XTHs by analyzing their signal peptide prediction, subcellular location prediction, isoelectric point and other protein characteristics. The Protparam program predicts that the theoretical molecular weight of GhXTH1 is 33.07 kDa, the theoretical isoelectric point (pI) is 6.38, the theoretical molecular weight of GhXTH2 is 33.61 kDa, and the theoretical pI is 5.40. Through the prediction of SignalP 4.1 Server, it is shown that GhXTH1 and GhXTH2 both have a signal peptide at the N-terminal (Figure 3), located at 1-25 aa. The function of the signal peptide is to locate the protein to the cell wall. TMHMM predicted that GhXTH1 and GhXTH2 both have a transmembrane structure (Figure 4), which is located at the N-terminal, behind the signal peptide in the protein sequence. N-glycosylation plays an important role in maintaining XTH activity[26]. Therefore, NetNGlyc 1.0 Server was used to analyze the N (asparagine) - glycosylation sites in GhXTH1 and GhXTH2. The results showed that GhXTH1 had one glycosylation site, GhXTH2 had two glycosylation sites, and GhXTH1 and GhXTH2 had one glycosylation site after the conservative sequence DEIDFEFLG.
Advances in Engineering Technology Research

ISSN: 2790-1688
Volume-5-(2023)

There are 33 members of Arabidopsis XTH family, some of which have clear biological functions. In order to understand the evolutionary relationship between GhXTH1, GhXTH2 and Arabidopsis XTH, phylogenetic analysis was conducted (Figure 6). AtXTHs can be divided into three categories. Most of the Type I XTH members contain four exons, Type II XTH members have two or three exons, Type III XTH members have four or five exons, and there is a feature sequence[25] at the C end of Type 3 XTH. However, with the analysis of the superposition system of AtXTHs and rice XTHs (OsXTHs), it was found that there was no obvious difference between class I and class II XTH members[27]. Class III showed xylanase hydrolase activity rather than transglycosylase activity[28]. However, enzyme activity analysis showed that not all III enzymes had hydrolase activity, and one III XTH (SIXTH5) in tomato showed transglycosylase activity[29]. Therefore, there is no relationship between XTH phylogenetic classification and its activity. GhXTH1 is closely related to AtXTH6 and AtXTH, and GhXTH2 is closely related to AtXTH9, both of which belong to class I/II XTH.

Fig. 4 Analysis of transmembrane domain of GhXTH1 and GhXTH2

Fig 5 Prediction of N-glycosylation of GhXTH1 and GhXTH2

3.3 Phylogenetic analysis of GhXTH1 and GhXTH2

There are 33 members of Arabidopsis XTH family, some of which have clear biological functions. In order to understand the evolutionary relationship between GhXTH1, GhXTH2 and Arabidopsis XTH, phylogenetic analysis was conducted (Figure 6). AtXTHs can be divided into three categories. Most of the Type I XTH members contain four exons, Type II XTH members have two or three exons, Type III XTH members have four or five exons, and there is a feature sequence[25] at the C end of Type 3 XTH. However, with the analysis of the superposition system of AtXTHs and rice XTHs (OsXTHs), it was found that there was no obvious difference between class I and class II XTH members[27]. Class III showed xylanase hydrolase activity rather than transglycosylase activity[28]. However, enzyme activity analysis showed that not all III enzymes had hydrolase activity, and one III XTH (SIXTH5) in tomato showed transglycosylase activity[29]. Therefore, there is no relationship between XTH phylogenetic classification and its activity. GhXTH1 is closely related to AtXTH6 and AtXTH, and GhXTH2 is closely related to AtXTH9, both of which belong to class I/II XTH.

Fig. 4 Analysis of transmembrane domain of GhXTH1 and GhXTH2

Fig 5 Prediction of N-glycosylation of GhXTH1 and GhXTH2

3.3 Phylogenetic analysis of GhXTH1 and GhXTH2

There are 33 members of Arabidopsis XTH family, some of which have clear biological functions. In order to understand the evolutionary relationship between GhXTH1, GhXTH2 and Arabidopsis XTH, phylogenetic analysis was conducted (Figure 6). AtXTHs can be divided into three categories. Most of the Type I XTH members contain four exons, Type II XTH members have two or three exons, Type III XTH members have four or five exons, and there is a feature sequence[25] at the C end of Type 3 XTH. However, with the analysis of the superposition system of AtXTHs and rice XTHs (OsXTHs), it was found that there was no obvious difference between class I and class II XTH members[27]. Class III showed xylanase hydrolase activity rather than transglycosylase activity[28]. However, enzyme activity analysis showed that not all III enzymes had hydrolase activity, and one III XTH (SIXTH5) in tomato showed transglycosylase activity[29]. Therefore, there is no relationship between XTH phylogenetic classification and its activity. GhXTH1 is closely related to AtXTH6 and AtXTH, and GhXTH2 is closely related to AtXTH9, both of which belong to class I/II XTH.
3.4 Expression analysis of GhXTH1 and GhXTH2 in cotton fiber development

In order to obtain the expression of two XTHs in different cotton fibers at different development stages, primers were designed by cloning two XTH sequences, and semi quantitative analysis was carried out to understand the expression pattern of two XTHs in cotton fiber development of upland cotton and island cotton (Figure 7). The results showed that the expression of GhXTH1 and GhXTH2 in upland cotton and island cotton decreased with the development of cotton fiber. At the primary cell wall stage of cotton fiber development, the cells elongate faster, and the large expression of XTH can relax the primary cell wall and release the resistance caused by cell elongation. It can be seen from Figure 7 that GhXTH2 expression in upland cotton and island cotton is higher than that in corresponding development period.

4. Discussion and conclusion

Two XTH genes, GhXTH1 and GhXTH2, were isolated from cotton fiber by traditional PCR and RACE techniques. Sequence alignment found that they all have a conservative motif DEIDFELGL in the XTH family, which is not only conservative in the XTH family, but also relatively conservative in the GH16 family to which XTH belongs[30-31]. According to the conserved sequence of XTH family in Arabidopsis and rice, the motif has some changes at some
sites. Therefore, GhXTH1 and GhXTH2 have 1-2 amino acid residues on this motif, which does not affect their becoming members of XTH family.

Arabidopsis XTH family consists of 33 members, which can be divided into 3 types[32] from the perspective of phylogeny. Phylogenetic analysis showed that GhXTH1 and GhXTH2 belong to Group I/II. Type I XTH has a common feature. XTH gene consists of four exons, and its conservative motif is located on three exons. However, phylogenetic classification has nothing to do with XTH activity[29]. In rice, two XTHs (OsXTH19, OsXTH20) in class III only show hydrolase activity, while OsXTH1 in class I has endotransferase activity and hydrolase activity[33]. Phylogenetic analysis showed that GhXTH1 was closely related to AtXTH6 and AtXTH7, while GhXTH2 was closely related to AtXTH9. Comparing the three-dimensional structure of GhXTH1 with that of a XTH (TmNGX1) with hydrolase activity, we can see that they are different in the conservative ring structure responsible for hydrolysis activity. Therefore, GhXTH1 is mainly transglycosylase activity[17]. AtXTH9 activity has not been reported yet, but from phylogenetic analysis, AtXTH9, GhXTH2 and GhXTH1 are closely related, and they should be similar in enzyme activity. AtXTH9 expression was positively regulated by farred light[34]. Therefore, GhXTH2 expression may be positively regulated by infrared light.

XTH, as an enzyme protein, its activity is the basis of its physiological function. In fact, XTH has many physiological functions, including cell growth[15-16, 35], fruit softening[29, 36], organ shedding[37-38], vascular formation[39-42], etc. The three XTH enzyme activity types in rice are different. Overexpression or inhibition of their expression in rice cannot significantly change the rice phenotype, which indicates that the three XTHs have functional redundancy[33]. GhXTH1 and GhXTH2 are highly expressed in cotton fiber, which is related to cotton fiber development. GhXTH1 is up-regulated in cotton varieties with longer cotton fiber or island cotton[17], indicating that GhXTH1 is related to cotton fiber elongation. More direct evidence is that after GhXTH1 is overexpressed in cotton, cotton fiber glycosylase activity increases and mature fiber length increases. GhXTH2 expression in cotton fiber is higher than GhXTH1, therefore, GhXTH2 may play a stronger role in cotton fiber elongation than GhXTH1. On the other hand, it cannot be ruled out that GhXTH1 and GhXTH2 have functional redundancy in cotton fiber elongation.

Two full-length cDNAs of the full-length xylan transfer/hydrolase genes GhXTH1 and GhXTH2 were obtained. GhXTH2 is a new member of the GhXTH family. GhXTH1 and GhXTH2 both contain signal peptide and transmembrane structure, which indicates that they are secreted proteins located on the plasma membrane. GhXTH1 and GhXTH2 both contain the XTH family's conservative motif DEIDFEFLG, and there is a glycosylation site after the motif, which is necessary for XTH to produce activity. This indicates that the two sequences obtained have the general characteristics of the family. Phylogenetic analysis showed that GhXTH1 was closely related to Arabidopsis XTH6 and 7, and GhXTH2 was closely related to Arabidopsis XTH9. GhXTH1 and GhXTH2 decreased with the development of cotton fiber, and GhXTH2 was higher than GhXTH1.

Reference


