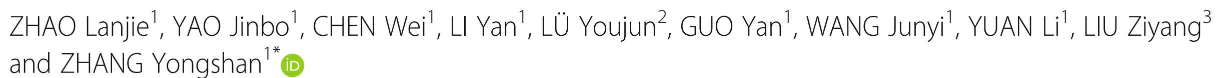
Journal of Cotton Research

RESEARCH

Open Access

A genome-wide analysis of *SWEET* gene family in cotton and their expressions under different stresses

ZHAO Lanjie[1], YAO Jinbo[1], CHEN Wei[1], LI Yan[1], LÜ Youjun[2], GUO Yan[1], WANG Junyi[1], YUAN Li[1], LIU Ziyang[3] and ZHANG Yongshan[1*]



Abstract

Background: The *SWEET* (Sugars will eventually be exported transporters) gene family plays multiple roles in plant physiological activities and development process. It participates in reproductive development and in the process of sugar transport and absorption, plant senescence and stress responses and plant-pathogen interaction. However, thecomprehensive analysis of *SWEET* genes has not been reported in cotton.

Results: In this study, we identified 22, 31, 55 and 60 *SWEET* genes from the sequenced genomes of *Gossypium arboreum*, *G. raimondii, G. hirsutum* and *G. barbadense*, respectively. Phylogenetic tree analysis showed that the *SWEET* genes could be divided into four groups, which were further classified into 14 sub-clades. Further analysis of chromosomal location, synteny analysis and gene duplication suggested that the orthologs showed a good collinearity and segmental duplication events played a crucial role in the expansion of the family in cotton. Specific MtN3_slv domains were highly conserved between *Arabidopsis* and cotton by exon-intron organization and motif analysis. In addition, the expression pattern in different tissues indicated that the duplicated genes in cotton might have acquired new functions as a result of sub-functionalization or neo-functionalization. The expression pattern of *SWEET* genes showed that the different genes were induced by diverse stresses. The identification and functional analysis of *SWEET* genes in cotton may provide more candidate genes for genetic modification.

Conclusion: *SWEET* genes were classified into four clades in cotton. The expression patterns suggested that the duplicated genes might have experienced a functional divergence. This work provides insights into the evolution of *SWEET* genes and more candidates for specific genetic modification, which will be useful in future research.

Keywords: *Gossypium*, Sugars will eventually be exported transporters (SWEETs), Gene expression patterns, Stress

Background

Sugar is a major carbon source and energy source for higher plants in their growth and development (Walmsley et al. 1998; Lalonde et al. 2004; Chen et al. 2010; Chen et al. 2012). Higher plants can use convert CO_2 into organic carbon in photosynthetic leaves (the source). The source is involved in the storage and transport of nutrients in plants (Ruan 2014; Rolland et al. 2002). However, sugar cannot be transported independently across the plant cell membrane system and requires the assistance of appropriate sugar transporters, such as MSTs (monosaccharide transporters) (Slewinski 2011), SUTs (sucrose transporters) (Kuhn and Grof 2010; Ayre 2011) and SWEETs (sugars will eventually be exported transporters) (Chen et al. 2010).

SWEETs are a new family of sugar transporters discovered in recent years, generally with seven transmembrane domains and two MtN3 motifs (Talbot 2010; Baker et al. 2012). SWEET shows the function of bidirectional reversible transport of sugar, and promotes the diffusion of sucrose to the apoplast pathway through transmembrane across the gradient of concentration on cell efflux (Baker et al. 2012; Lin et al. 2014; Eom et al. 2015). Since SWEET was first discovered using Forster resonance

* Correspondence: 13938698299@163.com
[1]State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, China
Full list of author information is available at the end of the article

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

energy transfer (FRET) by optical glucose sensors (Chen et al. 2010), *SWEET* family members have been identified by genome-wide analyses in different plant species, such as *Arabidopsis* (Chen et al. 2010), rice (Yuan and Wang 2013), tomato (Feng et al. 2015), soybean (Patil et al. 2015) and cucumber (Hu et al. 2017).

At present, researches on the functions of *SWEET* genes are carried out mainly in *Arabidopsis* and rice, while only a small part of them have been functionally characterized. The functions of transporting glucose and sucrose are identified in most well-studied *SWEETs* currently (Chen et al. 2010; Chen et al. 2012). For example, *AtSWEET1* is involved in the regulation of glucose uptake and efflux (Chen et al. 2010; Sonnewald 2011). *OsSWEET11* and *OsSWEET14* are the low affinity transport sucrose carriers that may be involved in the phloem sucrose loading process (Chen et al. 2012; Chen 2014). *AtSWEET1/4/5* may directly involve in the transport of sugar to regulate the osmotic active substances, or participate in cell wall sugar loading (Bauer et al. 2013). Other *SWEETs* have the function of transporting fructose and galactose (Klemens et al. 2013; Guo et al. 2014; Zhou et al. 2014). *AtSWEET16* and *AtSWEET17* are highly expressed in roots and involved in the transport of monosaccharides and polysaccharides across tonoplast (Klemens et al. 2013; Guo et al. 2014; Chardon et al. 2013). In addition, *SWEETs* participate in the transport of sugar and ions. *AtSWEET13* is involved in the regulation of aluminum ion balance in plants (Zhao et al. 2009). *OsSWEET11* forms a complex with two other copper transporter protein COPT1 and COPT5 in the plasma membrane (Yuan et al. 2009; Yuan et al. 2010), regulating the transport of copper ions and sugars (Yuan et al. 2010).

It has been found that some *SWEETs* take part in plant reproductive development also. *AtSWEET9* transports sucrose into apoplasts for nectar secretion (Lin et al. 2014). *AtSWEET8*, a glucose transporter expressed in tapetum and embryo sacs, participates in the development of the pollen and anther (Guan et al. 2008; Sun et al. 2013). The silenced *AtSWEET11* mutant showed lower pollen viability and even pollen sterility (Sonnewald 2011). Suppression of *OsSWEET11* resulted in pollen dysplasia, leading to male infertility (Yuan et al. 2009; Liu et al. 2011; Ge et al. 2000). *AtSWEET5* was highly expressed in the female gametophyte during the three-cell pollen stage (Yuan et al. 2010). Some researches have showed that *SWEETs* plays a role in the process of plant senescence and response to abiotic stresses. *AtSWEET15* can be induced by cold, salt and drought stress (Seo et al. 2011). Overexpressing *AtSWEET16* improved the tolerance to cold stress, osmotic stress and nitrogen availability in *Arabidopsis* (Klemens et al. 2013). In addition, *SWEETs* from other plants (e.g., *Hordeum vulgare*, tomato, etc.) are also involved in the regulation of abiotic stress (Yuan and Wang 2013).

Although some studies have been reported in cotton, such as, *SWEETs* are involved in sugar transport during fiber elongation and bacterial blight of cotton (Zhang et al. 2017; Cox et al. 2017; Phillips et al. 2017), the function of SWEET in cotton, especially in stress response and host-pathogen interaction, has not been identified until now. With the release of genomes sequences of two diploid cotton (A2, D5) and two allotetraploid cotton (AD1, AD2) (Phillips et al. 2017; Li et al. 2014; Paterson et al. 2012; Wang et al. 2012; Li et al. 2015; Zhang et al. 2015; Yuan et al. 2015; Liu et al. 2015) facilitates the survey of *SWEETs* in cotton. In this study, we identified the *SWEETs* in four cotton species by genome-wide analysis. This results will provide insights into the evolution of *SWEET* genes and more candidates for specific genetic modification, which will be useful in future research.

Results

Gene identification and conserved domain retrieval

The SWEET amino acid sequences reported in *Arabidopsis*, rice and cucumber were used as query sequences and blasted against sorghum, poplar, maize, cocoa, *Gossypium arboreum*, *G. raimondii*, *G. hirsutum,* and *G. barbadense* genome database with e-values of 1e-5. Twenty-three candidates in sorghum, 27 in poplar, 24 in maize, 21 in cacao, 31 in *G. arboreum*, 32 in *G. raimondii*, 60 in *G. hirsutum,* and 60 in *G. barbadense* were obtained, respectively, then, the conserved domain (IPR004316) was analyzed in the candidate *SWEET* gene family members by the PROSITE (http://prosite.expasy.org/) and InterProscan (http://www.ebi.ac.uk/interpro/) (Jones et al. 2014). Eventually, 23, 27, 22, 21, 22, 31, 55, and 60 genes were identified as *SWEET* family members, respectively (Table 1, Additional file 1: Table S1, Additional file 2: Table S2, and Additional file 3: Table S3). The high similarity of *SWEET* genes was found from two upland cotton genomes Nangjing Agri. Univ. version 1.1 (NAU version 1.1) and Beijing Genome Institute & Institute of Cotton Research of CAAS version 1.0 (BJI version 1.0)), and the genes from NAU, version 1.1 contained all the members from BJI version 1.0. Therefore, the *SWEET* genes from the NAU (Additional file 2: Table S2) were analyzed as samples from *G. hirsutum*. Since all *SWEETs* of *G. barbadense* on COTTONGEN came from scaffolds, we chose *SWEET* genes from *G. barbadense* on cottonFGD for analyses (Additional file 3: Table S3). The total numbers of *SWEET* genes identified in the two diploid cotton (*G. arboreum* and *G. raimondii*) were lower than that in allotetraploid (*G. hirsutum* and *G. barbadense*) cotton (Table 1, Additional file 1: Table S1, Additional file 2: Table S2 and Additional file 3: Table S3).

Table 1 *SWEET* gene family in *G. arboreum*

Name	Gene ID	Chromosome number	Location	Length AA	MW/kDa	pI	Transmembrane domains	Predicted subcellular localization
GaSWEET1a	*Cotton_A_19609*	Chr5	16,241,997–16,243,716(−)	252	27.93	9.57	7	Plasma membrane (4.993)
GaSWEET1b	*Cotton_A_35416*	Chr10	25,530,678–25,532,070(+)	205	22.39	9.46	5	Plasma membrane (4.977)
GaSWEET7	*Cotton_A_13229*	Chr6	15,976,051–15,977,989(+)	254	28.24	9.58	7	Plasma membrane (4.986)
GaSWEET6	*Cotton_A_33804*	Chr3	87,659,800–87,661,744(+)	258	28.44	9.70	6	Plasma membrane (4.993)
GaSWEET2	*Cotton_A_03976*	Chr9	62,274,049–62,275,760(−)	233	25.74	9.13	7	Plasma membrane (4.995)
GaSWEET17b	*Cotton_A_40086*	Chr13	134,161,415–134,162,613(−)	239	26.57	9.2	6	Plasma membrane (4.996)
GaSWEET4	*Cotton_A_23444*	Chr12	81,146,814–81,149,352(+)	247	27.48	9.18	6	Plasma membrane (4.970)
GaSWEET11c	*Cotton_A_40170*	Chr9	58,836,260–58,837,517(−)	275	30.99	9.08	7	Plasma membrane (4.996)
GaSWEET16d	*Cotton_A_39774*	Chr1	38,174,341–38,176,348(+)	250	26.84	8.4	6	Plasma membrane (4.910)
GaSWEET16b	*Cotton_A_18817*	Chr9	21,965,412–21,967,080(+)	229	24.99	6.72	7	Plasma membrane (4.991)
GaSWEET11b	*Cotton_A_01942*	Chr6	28,940,090–28,941,462(−)	298	33.53	7.65	7	Plasma membrane (4.979)
GaSWEET16c	*Cotton_A_28964*	Chr4	112,187,698–112,189,654(+)	300	32.90	9.27	7	Plasma membrane (4.984)
GaSWEET11a	*Cotton_A_03658*	Chr1	67,395,426–67,398,120(+)	276	30.99	9.23	7	Plasma membrane (4.992)
GaSWEET10c	*Cotton_A_08012*	Chr6	115,372,468–115,373,820(−)	273	30.62	9.53	7	Plasma membrane (4.976)
GaSWEET15	*Cotton_A_10026*	Chr8	20,849,191–20,851,135(+)	286	31.91	7.58	7	Plasma membrane (4.845)
GaSWEET10b	*Cotton_A_16236*	Chr12	17,595,910–17,597,171(+)	278	31.30	9.44	7	Plasma membrane (4.984)
GaSWEET8	*Cotton_A_23442*	Chr12	81,161,147–81,162,097(+)	183	20.38	9.07	5	Plasma membrane (4.950)
GaSWEET16a	*Cotton_A_04853*	Chr7	42,633,580–42,634,115(−)	138	14.84	8.81	4	Plasma membrane (4.701)
GaSWEET17a	*Cotton_A_28959*	Chr4	112,303,189–112,305,084(+)	194	21.62	6.81	5	Plasma membrane (4.939)
GaSWEET10a	*Cotton_A_03661*	Chr1	67,444,715–67,445,549(+)	186	20.59	9.15	4	Plasma membrane (4.209)
GaSWEET5	*Cotton_A_02746*	Chr8	97,490,174–97,491,032(+)	140	15.93	8.79	3	Plasma membrane (4.975)
GaSWEET9	*Cotton_A_24220*	Chr13	34,028,951–34,031,462(+)	229	25.65	8.71	5	Plasma membrane (4.979)

These 168 *SWEET* genes of four cotton species were named according to their homologous genes in *Arabidopsis*. Each *AtSWEET* gene corresponded to approximately one to ten cotton *SWEET* genes, respectively. The naming rules were performed based on a published paper (Yang et al. 2017). Ga, Gr, Gh, Gb, and At were used as prefixes before the names of SWEET genes from *G. arboreum*, *G. raimondii*, *G. hirsutum*, *G. barbadense*, and *Arabidopsis*, respectively. "a", "b", "c", "d", "e", and "f" were appended to the gene names to distinguish the homologous genes (Table 1, Additional file 1: Table S1, Additional file 2: Table S2, and Additional file 3: Table S3). More than 88.69% of the 168 identified *SWEET* genes encode proteins ranging between 180 to 311 amino acids (AA), except for 19 genes with different lengths, i.e., less than 180 or more than 311 AA (Table 1, Additional file 1: Table S1, Additional file 2: Table S2, and Additional file 3: Table S3). The molecular weights (kDa) and isoelectric points (pI) of these predicted SWEET proteins ranged from 9.93 to 38.04 kDa, and from 5.47 to 10.08, respectively (Table 1, Additional file 1: Table S1, Additional file 2: Table S2, and Additional file 3: Table S3). Moreover, the protein subcellular localization prediction showed that the 168 SWEET proteins were located in the plasma membrane (Table 1, Additional file 1: Table S1, Additional file 2: Table S2, and Additional file 3: Table S3). The transmembrane domains (TMs) of 168 SWEET proteins were predicted by using the TMHMM Server v.2.0. The results showed that 101 SWEET proteins had 7 TMs. GrSWEET6a, GrSWEET10c and GhSWEET10c_Dt had 8 TMs. Fifty-eight SWEET proteins had 46 TMs. GaSWEET5, GhSWEET3a_Dt, GhSWEET3b_At, GbSWEET2c_A, and GbSWEET3b_D had 3 TMs. GbSWEET11_A had 2 TMs (Table 1, Additional file 1: Table S1, Additional file 2: Table S2, and Additional file 3: Table S3). The different numbers of transmembrane domains contained in these SWEET proteins indicated different functions.

Phylogenetic relationship analysis of SWEETs

To understand the evolutionary history of SWEET proteins among *Gossypium* and other species, phylogenetic analysis of 314 SWEET protein sequences (168 from cotton, 22 from maize, 19 from rice, 23 from sorghum, 27 from *Populus trichocarpa*, 21 from cocoa, 17 from *Arabidopsis thaliala*, and 17 from cucumber) was performed by

their sequence similarities with orthologs, using the neighbour-joining (NJ) method in MEGA 7. The result showed that SWEET proteins could be classified into four clades, namely, the I clade, II clade, III clade, and IV clade (Fig. 1). III clade, the largest one, contained 111 SWEET members, while I, II and IV clade contained 70, 71 and 62 members, respectively. To validate the phylogenetic tree constructed byNJ method, we also used the minimum evolution method to construct a tree. The results showed that SWEET proteins also were divided into four clades, almost consistent with the NJ method (Additional file 4: Figure S1). Although, there were differences between the topologies of the two trees constructed by the two methods, the members within the sub-clades and the topology within the sub-clades were relative stable, which indicated that the NJ tree could be used for further analysis.

To further study the evolutionary relationship and to predict the gene function, the SWEETs were divided into 14 sub-clades, named as α-ξ, respectively (Fig. 1). I clade included three sub-clades, II clade included three sub-clades, III clade included five sub-clades, and IV clade included three sub-clades. Interestingly, α, β, γ, δ, ε, ζ, μ, and ξ sub-clades were constituted of SWEETs from the monocotyledon species and dicotyledon species. The η, θ, ι, κ, and ν sub-clades were composed of SWEETs from dicotyledon species and the λ sub-clades were composed of SWEETs from the monocotyledon species. δ sub-clade contained the fewest members.

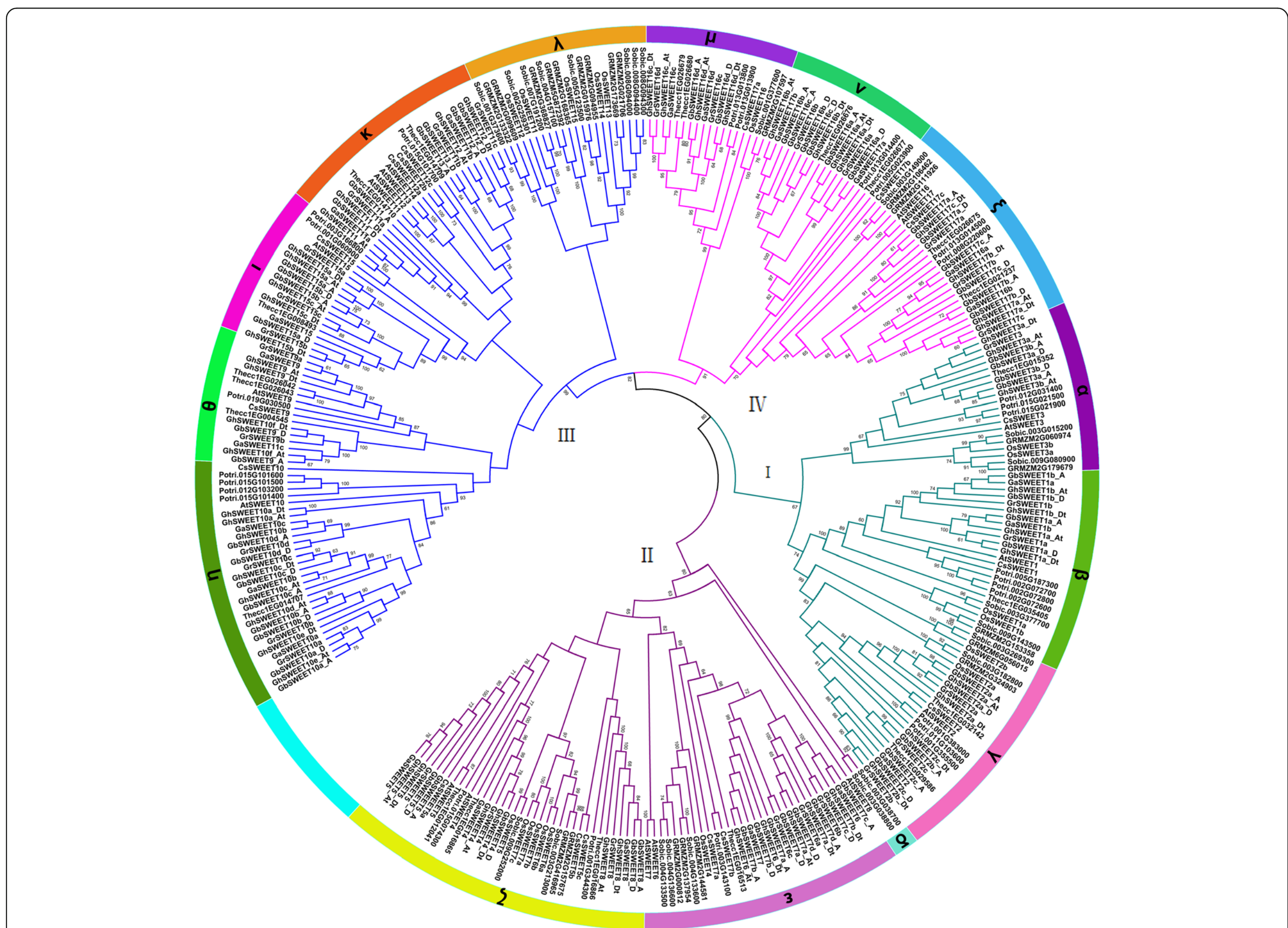


Fig. 1 Phylogenetic relationship of SWEETs from cotton, *Arabidopsis*, rice, cacao, poplar, sorghum, cucumber and maize. The phylogenetic tree was generated using MEGA 6.0 with the NJ method with 1 000 bootstrap replicates. MEGA 7.0 was used for constructing the NJ tree. The I clade, II clade, III clade, and IV clade is marked in cyan, purple, blue and pink, respectively. Each clade was classified into sub-clades, marked in different colors on the circle. α to γ represent the sub-clades in the I clade, δ and ζ represent the sub-clades in the II clade, η to λ represent the sub-clades in the III clade, and μ to ξ represents the IV clade. The prefixes Ga, Gr, Gh, Potri, At, Os, GRMZM, Sobic, and Thecc stand for *G. arboreum*, *G. raimondii*, *G. hirsutum*, *Populus trichocarpa*, *Arabidopsis thaliala*, *Oryza sativa*, *Zea mays*, *Sorghum bicolor*, and *Theobroma cacao*, respectively. The appendices At and Dt in the upland cotton indicate the A- and D-subgenome, respectively. The bootstrap values are shown near the nodes, and only those values greater than 50 are displayed

Among the various sub-clades of the phylogenetic tree, SWEETs from cotton were more closely related to those in cacao because they always clustered closely together with each other, except for the δ sub-clade and the λ sub-clade. In most cases, one cocoa gene corresponded to two or more than two homologous SWEET members from different cotton species. Among them, the two largest number of cotton homologous SWEET protein from different cotton species corresponding to one cacao protein in the β sub-clade and ε sub-clade were 12 SWEETs and 18 SWEETs, respectively (Fig. 1). There was the least cotton homologous SWEET members corresponding to one cacao gene in γ sub-clade, only had two genes (*GrSWEET2c* and *GbSEET2b_A*). These cotton SWEETs in almost all sub-clades showed that a tendency to cluster into the same sub-clade due to their relatively conserved functions. Almost all cotton orthologs from the A genome and At subgenome or from the D genome and Dt sub-genome tended to form an orthologous pairs at the ends of the clades, suggesting there was a closer relationship between the orthologs from At-A or Dt-D in cotton.

Genomic localization and duplication of *SWEET* genes in cotton

The identified 168 *SWEETs* from 4 cotton species were mapped to the corresponding chromosomes or scaffolds, indicating an uneven distribution (Fig. 2). 158 of 168 *SWEETs* were located to chromosomes, while the remaining 10 *SWEETs* from two allotetraploid cotton (*G. hirsutum* and *G. barbadense*) located in unmapped scaffolds. The identified 22 *SWEET* genes were assigned to 11 chromosomes from *G. arboreum* except A2_chr02 and A2_chr11 (Fig. 2a). Chromosomes A2_chr01, A2_chr06, A2_chr09, and A2_chr12 contained the most of *SWEET* genes, with 3 per chromosome. Chromosomes A2_chr03, A2_chr05, A2_chr07, and A2_chr10 contained 1 gene per chromosome, Other chromosomes contained 2 genes. Thirty-one *SWEET* genes were distributed on the 11 chromosomes in *G. raimondii*, no gene was found on chromosomes D5_chr06 and D5_chr10 (Fig. 2b). Chromosome D5_chr13 contained 5 genes. Chromosomes D5_chr02 and D5_chr09 had only 1 gene, and other chromosomes had 24 genes.

Forty-nine *SWEETs* were located in 23 chromosomes from *G. hirsutum*, while no gene was found on chromosome AD1_A06, AD1_D06, or AD1_D09 (Fig. 2c). Chromosomes AD1_A07, AD1_A11, AD1_D11, and AD1_D13 contained 4 *SWEETs* per chromosome, whereas chromosomes AD1_A01, AD1_A03, AD1_A04, AD1_A08, AD1_A09, AD1_A10, AD1_D01, AD1_D05, and AD1_D08 contained one gene per chromosome. In addition, other chromosomes contained 23 *SWEET* genes. Except for A01-D01, A08-D08, A11-D11, and A12-D12, the number of genes located on the chromosome in At subgenome was the same as that on the homologous chromosome in Dt subgenome, which implied that some genes were lost during the evolution or sequencing fault. In addition, *GhSWEET7a_At* was located on AD1_A02, *GhSWEET7a_Dt* was not found on chromosome AD1_D02, while it was located on AD1_D03. Similarly, *GhSWEET4_At* and *GhSWEET8_At* were located on AD1_A05, while they were not found on the corresponding AD1_D05, but on chromosome AD1_D03. Therefore, our results supported the hypothesis that a chromosome translocation occurred between AD1_A02 and AD1_D02, and also occured between AD1_A03 and AD1_D03 during the allotetraploidization.

Fifty-six *SWEET* genes were mapped on 22 chromosomes in *G. barbadense*, while no gene was found on chromosomes AD2_A06, AD2_D01, AD2_D05, or AD2_D06 (Fig. 2d). Chromosome AD2_D12 contained 5 genes, Chromosomes AD2_A03, AD2_A04, AD2_A09, AD2_D04, and AD2_D09 contained only 1 gene, respectively. Other chromosomes had 24 genes. As with upland cotton, except for A04-D04, A07-D07, A08-D08, A09-D09, and A10-D10, the number of genes located on the chromosome in At subgenome was the same as that of its homologous chromosome in Dt sub-genome, implied that some *SWEET* genes might have been lost during the evolution. In addition, chromosome translocation occurred between AD2_A02 and AD2_D03 in *G. barbadense*. For example, *GbSWEET1a-A* and *GbSWEET7d-A* were located in AD2_A02, while no gene was found on AD2_D02, and *GbSWEET1a-D* and *GbSWEET7d-D* distributed in AD2_D03.

To date, the mechanism of *SWEET* gene family expansion in cotton species remained unclear. Therefore, we studied the relationship between genetic differentiation and gene duplication within the *SWEET* gene family of four cotton species. A total of 51 pairs of *SWEET* genes were involved in segmental duplication events (defined as a method) by screening alignments of 158 genes on different chromosomes, while the tandem duplication event was not found (Fig. 3, Additional file 5: Table S4). The result showed that segmental duplication happened during the evolution and expansion of *SWEETs* in cotton. Nineteen pairs of genes with segmental duplication events were found in *G.hirsutum*, while 24 pairs in *G.barbadense*, 3 pairs in *G.arboreum* and 5 pairs in *G.raimondii*.

Analysis of gene structure and MtN3_slv domain location

To further investigate the diversification and evolution of the *SWEET* gene structure and conserved domain in cotton, we constructed the phylogenetic tree using MtN3_slv domain amino acid sequences from four cotton species and the *A. thaliala*, respectively (Fig. 4, Additional file 6: Figure S2, Additional file 7: Figure S3, and Additional file 8: Figure S4), the result showed that the genes were classified into 4 clades as above.

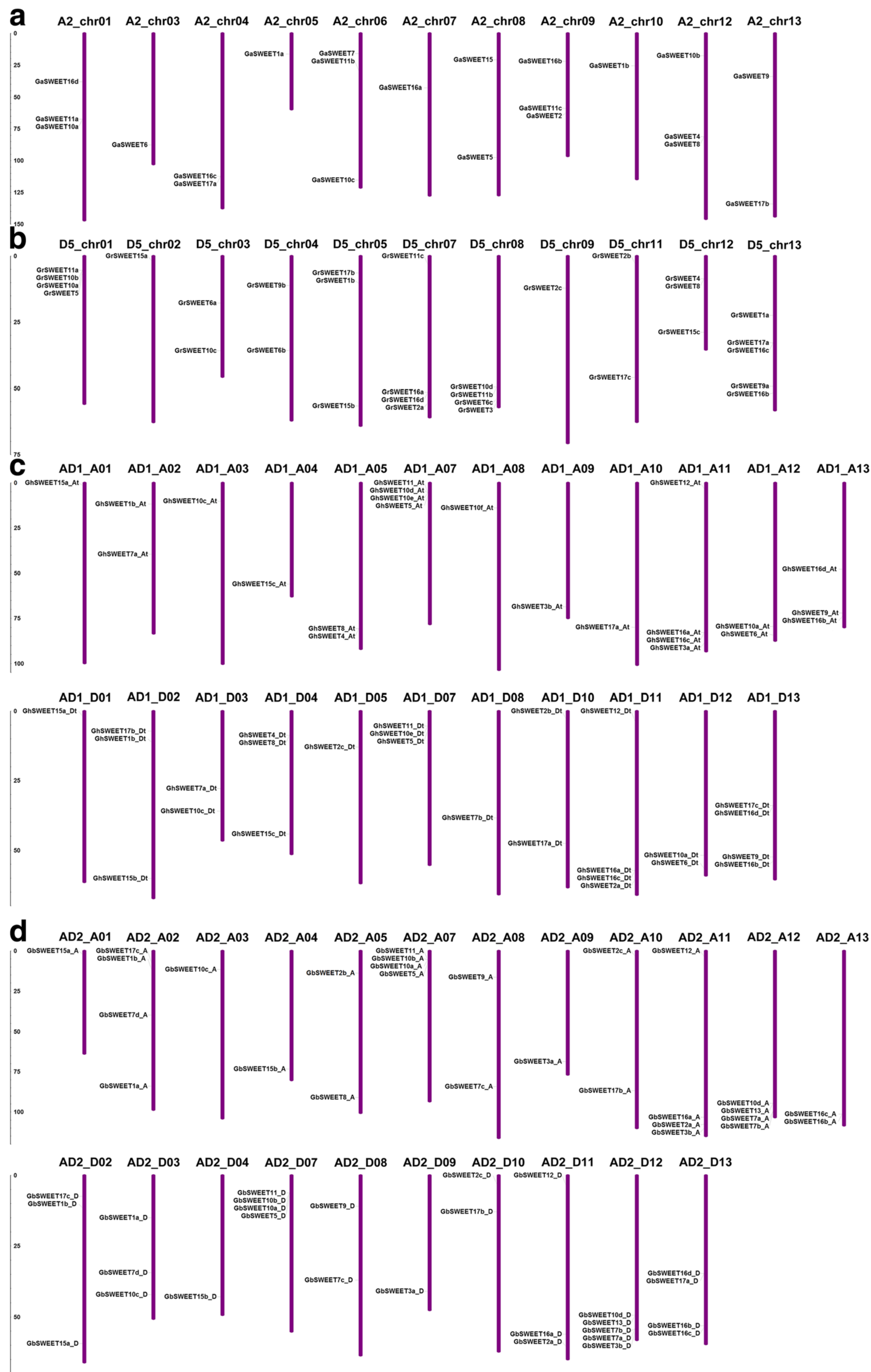


Fig. 2 The location of *SWEET* genes from four species cotton on Chromosomes. The scale represents megabases (Mb). The chromosome numbers of *G. arboreum* (A2_chr01 - A2_chr13) (**a**), *G. raimondii* (D5_chr01 - D5_chr13) (**b**), *G. hirsutum* (AD1_A01- AD1_A13, AD1_D01 - AD1_D13) (**c**), and *G. barbadense* (AD2_A01- AD2_A13, AD2_D01 - AD2_D13) (**d**) are indicated above each vertical bar. The purple vertical bar indicated the chromosome

23, 25, 10, 7 *paralogous* pairs of *SWEET* genes from *G. hirsutum*, *G. barbadense*, *G. raimondii* and *G. arboreum* in cotton were identified at the terminal nodes of the phylogenetic tree in contrast to 4 pairs in *A. thaliala*(Fig. 4a, Additional file 6: Figure S2a, Additional file 7: Figure S3a, and Additional file 8: Figure S4a).

Based on the evolutionary relationship of phylogenetic tree, the detailed features of exon/intron and conserved

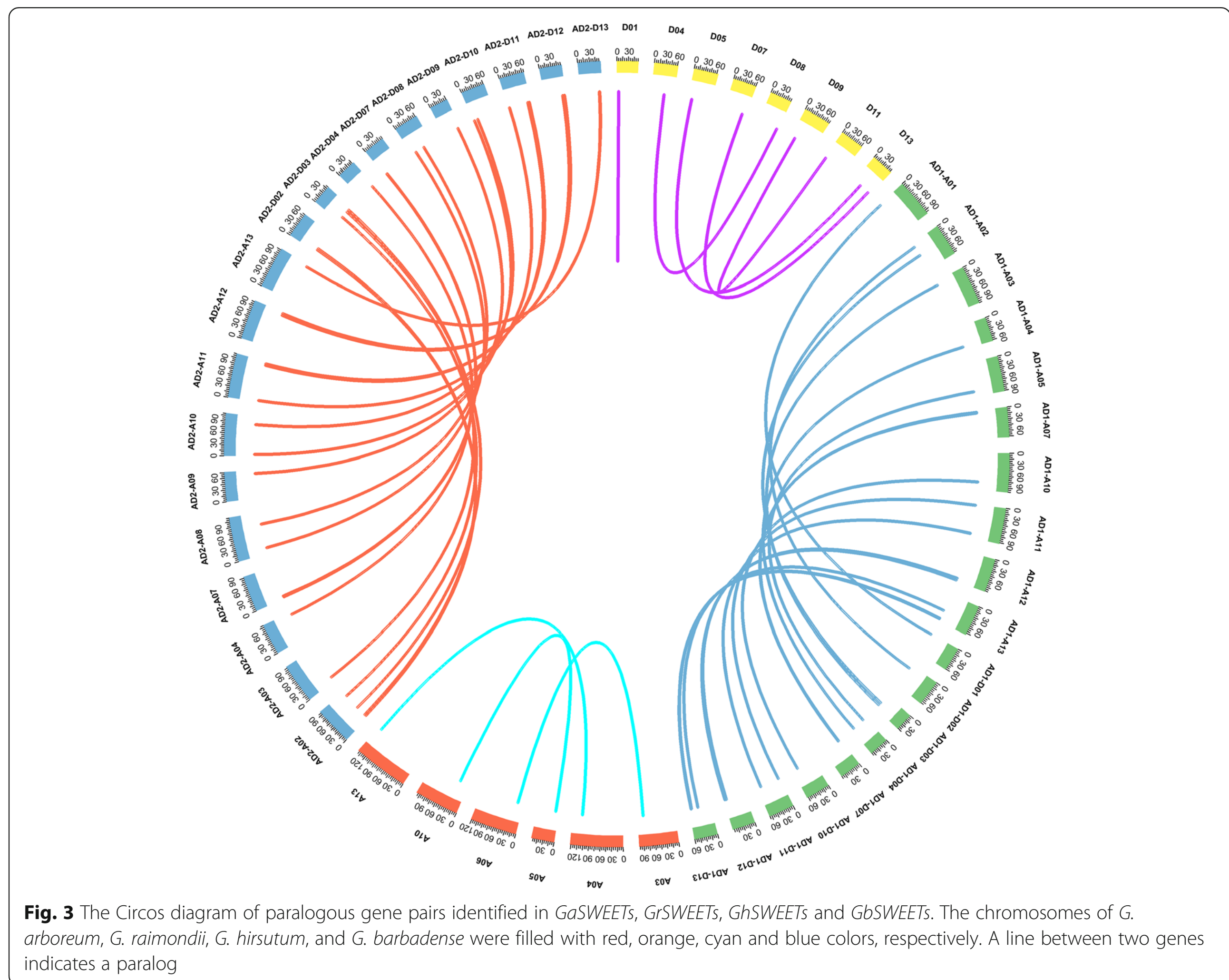


Fig. 3 The Circos diagram of paralogous gene pairs identified in *GaSWEETs*, *GrSWEETs*, *GhSWEETs* and *GbSWEETs*. The chromosomes of *G. arboreum*, *G. raimondii*, *G. hirsutum*, and *G. barbadense* were filled with red, orange, cyan and blue colors, respectively. A line between two genes indicates a paralog

domains of the *SWEET* genes were shown in Fig. 4a (Additional file 6: Figure S2a, Additional file 7: Figure S3a and Additional file 8: Figure S4a). Most *AtSWEETs* contained 6 exons and 5 introns, except for *AtSWEET6*(1 exon and no intron) and *AtSWEET7*(5 exons and 4 introns), In cotton, most *SWEET* genes contained 37 exons and 2–6 introns, except for *GbSWEET10a-A*and *GbSWEET10a-D* (9 exons and 8 introns, respectively) from *G. barbadense*. The *SWEET* genes within the same sub-clade exhibited similar exon/intron structures, especially in paralogous gene pairs, most of them shared a conserved exon/intron structure in terms of gene length or number of introns. Some variations of the exon/intron structure were found between *GaSWEET1a/1b*, *GaSWEET16a/16b*, *GaSWEET17a/17b*, indicating that loss of introns or gain events during evolution. The results indicated that the exon/intron structures were highly conserved in each sub-clade, and identical to the phylogenetic relationships.

The proteins of *SWEET* family were characterized by MtN3/saliva (MtN3_slv) conserved domain in previous studies (Talbot 2010; Baker et al. 2012). The typical conserved domains in the 168 SWEET proteins were identified in this study (Fig. 4c, Additional file 6: Figure S2C, Additional file 7: Figure S3C, and Additional file 8: Figure S4C). The result revealed that MtN3_slv were considerably conserved, with domain ranged from 117 to 279 aa. Most members of SWEET protein family contained two MtN3_slv domains, while 20 SWEET proteins only contained one, GhSWEET10a_At contained three MtN3_slv domains, (Fig. 4c, Additional file 6: Figure S2C, Additional file 7: Figure S3C, and Additional file 8: Figure S4C). The difference of the number of conserved domains in different SWEET protein suggested the diversity of their functions in cotton.

Expression patterns of *SWEET* genes in different tissues

We investigated the temporal and spatial transcription patterns and putative functions of different *SWEET*

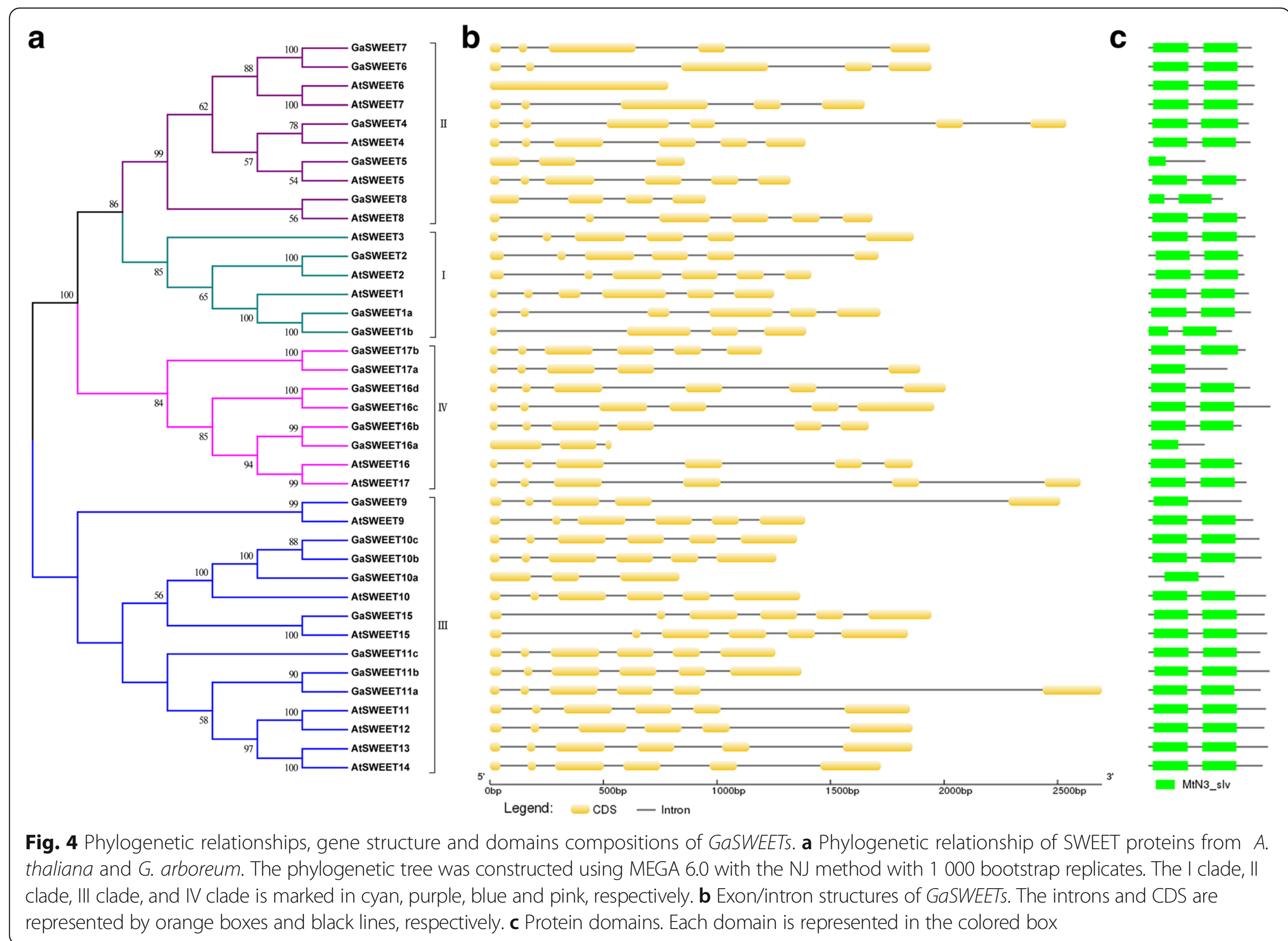


Fig. 4 Phylogenetic relationships, gene structure and domains compositions of *GaSWEETs*. **a** Phylogenetic relationship of SWEET proteins from *A. thaliana* and *G. arboreum*. The phylogenetic tree was constructed using MEGA 6.0 with the NJ method with 1 000 bootstrap replicates. The I clade, II clade, III clade, and IV clade is marked in cyan, purple, blue and pink, respectively. **b** Exon/intron structures of *GaSWEETs*. The introns and CDS are represented by orange boxes and black lines, respectively. **c** Protein domains. Each domain is represented in the colored box

genes during growth and development of *G. hirsutum* plant. The transcription levels in various tissues or organs of RNA-seq data from NCBI and COTTONFGD (http://www.cottonfgd.org/) were downloaded and analyzed (Yuan et al. 2015; Yang et al. 2017), including the vegetative (root, stem, and leaf) and reproductive (torus, petal, stamen, pistil, calycle, − 3 and − 1 days post anthesis (DPA) ovule, 0, 1, 3, 5, 10, 20, 25, and 35 days post anthesis (DPA) seed) tissues as well as in the fiber (5, 10, 20, and 25 DPA), and germinating seeds at 0 h, 5 h,10 h and from roots and cotyledons at 24 h, 48 h, 72 h, 96 h, and 120 h after imbibition. Their expression levels were varied (Figs. 5 and 6) indicating that *SWEETs* play different biological functions in different tissues of *G. hirsutum*.

GhSWEET5_At/Dt, *GhSWEET6_At*, *GhSWEET8_At*, and *GhSWEET10f_At* were not detected in roots, stems or leaves, *SWEET2a_At/Dt* expressed both in vegetative and reproductive tissues. *SWEET6a_At/Dt* expressed in reproductive tissues. Some *SWEETs* showed higher expression levels in reproductive tissues especially in floral organs, such as *SWEET1a_At/Dt* (petals, sepals, and 10–25 DPA fibers), *SWEET1b_At/Dt* (stamen, − 1 DPA ovule, 1DPA and 5DPA seeds, 10 DPA fiber), *SWEET2a_At/Dt* (petals and stamens), *SWEET7a_At/Dt* (petals and stamens). The results indicated that SWEETs also involved in the reproductive development of cotton, consistent with the results in *Arabidopsis* and rice (Yuan et al. 2009; Yuan et al. 2010; Guan et al. 2008; Sun et al. 2013; Liu et al. 2011; Ge et al. 2000). In addition, *SWEET2a_At/Dt*, *SWEET1b_Dt* and *SWEET4_At* were detected with high expression levels in germinating seed, cotyledons and roots during germination (Fig. 6). *SWEET1b_At* expressed in germinating seeds and from cotyledons after imbibitions. *SWEET10a_Dt*, *GhSWEET10e_At*, *GhSWEET10d_At* and *GhSWEET10b* expressed in cotyledons and roots after imbibition. *GhSWEET17a_At/Dt*, *GhSWEET17c_Dt*, *GhSWEET17b_Dt* and *GhSWEET3a_At/Dt* expressed in roots after imbibition. These *SWEETs* may be involved in the metabolic transport of sucrose during the germination of cotton seeds.

The expression patterns of orthologous genes between At and Dt in different tissues and organs were not always identical; for example, *GhSWEET10e_At* expressed in roots, receptacles, and carpels, while the expression

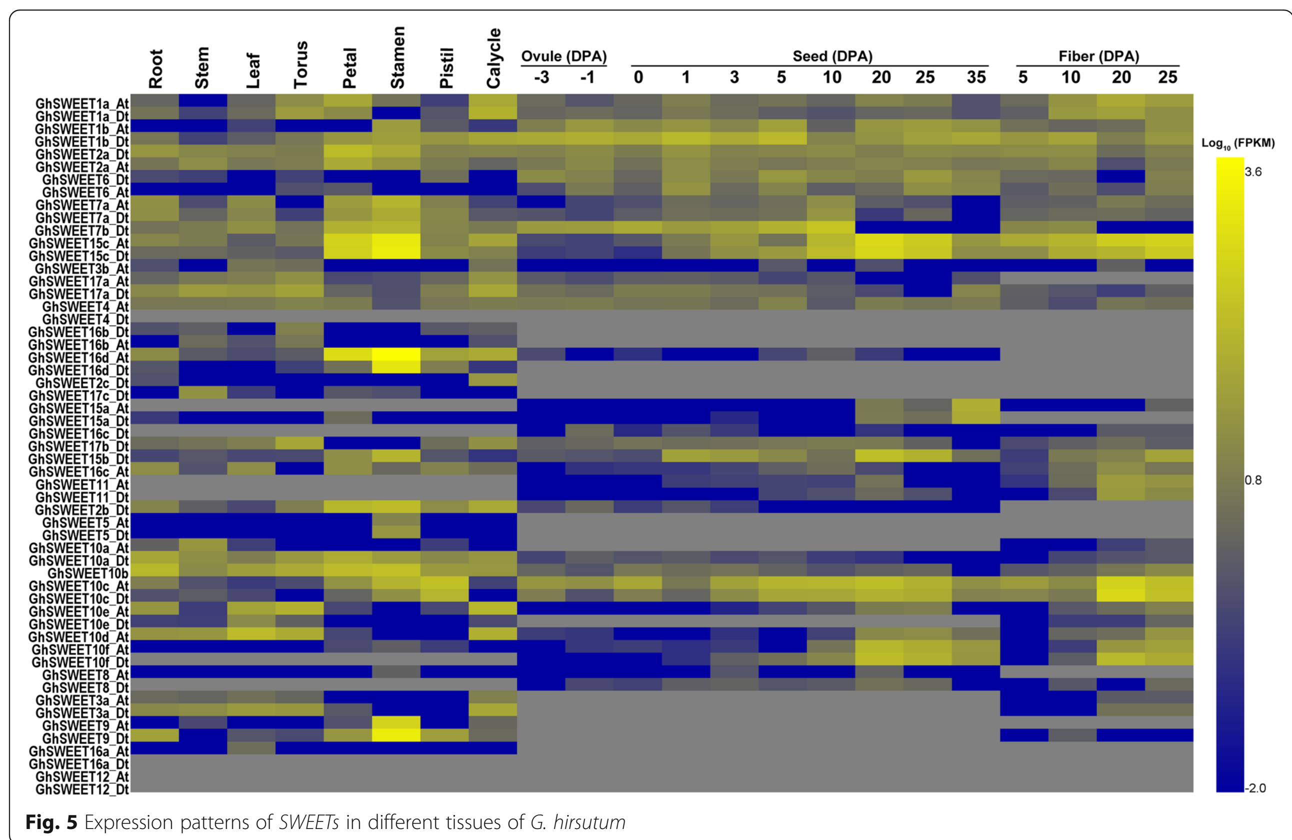


Fig. 5 Expression patterns of *SWEETs* in different tissues of *G. hirsutum*

of *GhSWEET10e_Dt* was not detected. *GhSWEET4_At* expressed in both germinated seeds and hypocotyls and roots, but the expression of *GhSWEET4_Dt* was undetectable.

Expression patterns of SWEET genes under multiple stresses

Cotton is frequently threatened by multiple abiotic stresses during growth and development. Therefore, we conducted a comprehensive analysis of *SWEETs* expression patterns under the conditions of cold, heat, PEG 6000 and NaCl multiple abiotic stresses from the RNA-seq data (Fig. 7).

Most *SWEET* genes were found to be down-regulated under salt and PEG 6000 conditions. *GhSWEET1a_At/Dt*, *GhSWEET1b_At/Dt* and so on showed a relatively stable expression response to multiple stress treatments. *GhSWEET2a_At/Dt*, *GhSWEET3a_Dt* and especially *GhSWEET2b_Dt* were strongly induced by cold stress, showing up-regulated expression. *GhSWEET4_At* and *GhSWEET10e_Dt* showed up-regulated expression under hot treatment. In addition, based on the RNA-seq data, we selected 9 genes (contains a pair of homologous genes) highly expressed under PEG 6000 or NaCl treatment and designed specific primers (Fig. 7; Additional file 9: Table S5) for qRT-PCR detection of leaves from plants treated with PEG 6000 or NaCl (Fig. 7). The expression pattern of *SWEET* genes detected by qRT-PCR was found to be coincided with the results of RNA-seq data (Fig. 8). These results indicated that they might take part in response to stress.

Discussion

SWEET gene family in cotton underwent enlargement during the evolution

SWEET genes family in *A. thaliala* (Chen et al. 2010), rice (Chen et al. 2012), tomato (Feng et al. 2015), Soybean (Patil et al. 2015) and cucumber (Hu et al. 2017) have been systematically analyzed. In this study, we performed a comprehensive investigation and analysis of *SWEET* genes from four cotton species, 22, 31, 55 and 60 *SWEET* genes were identified from *G. arboreum*, *G. raimondii*, *G. hirsutum*, and *G. barbadense*, respectively. Cotton contained more *SWEET* genes than species mentioned above, indicating that *SWEET* family genes undergo extensive expansion during cotton evolution. Although polyploidization was the main contributor in duplication, segment repeats also play an irreplaceable role in the expansion

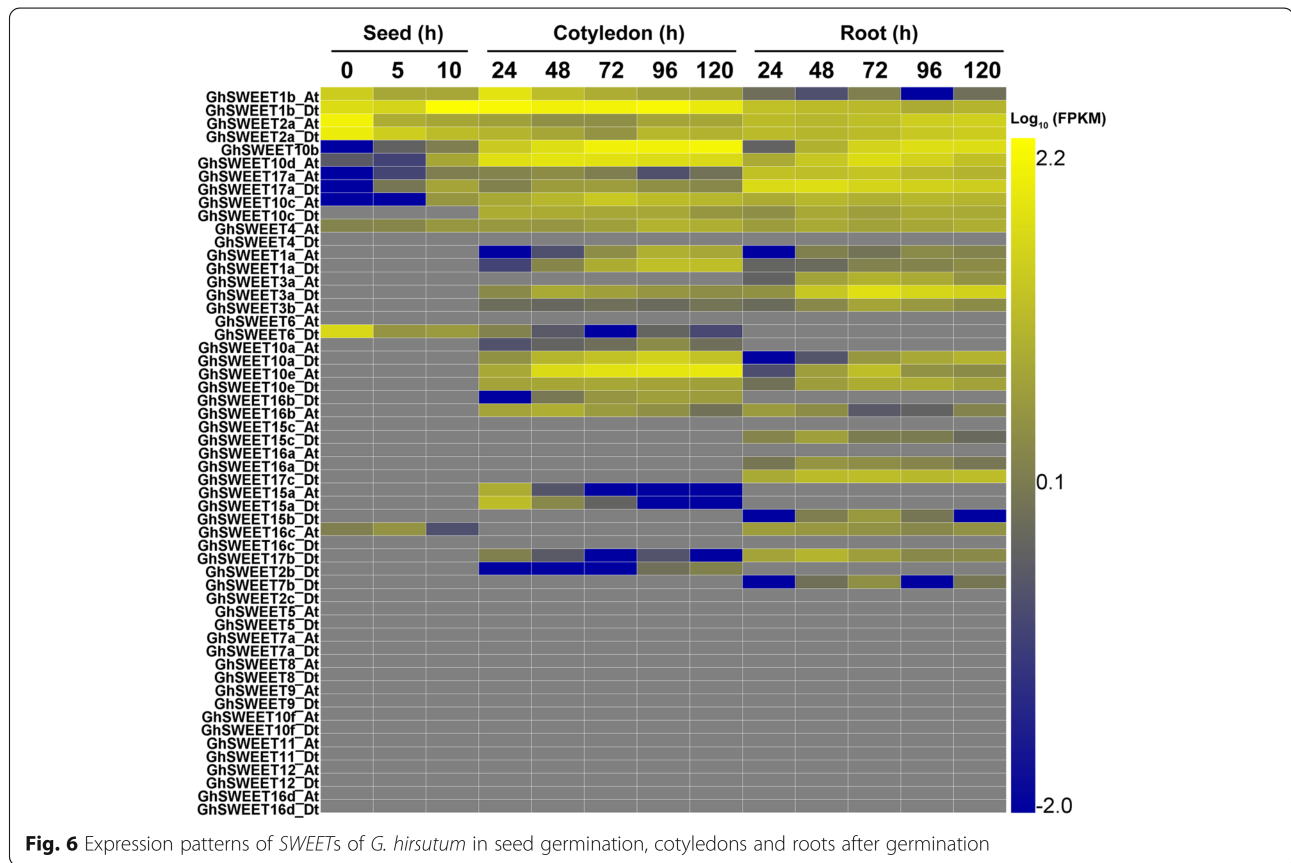


Fig. 6 Expression patterns of *SWEET*s of *G. hirsutum* in seed germination, cotyledons and roots after germination

of gene families in the genome (Paterson et al. 2012; Li et al. 2015). Fifty-one segment repeat pairs in 4 cotton species were found in our study, suggesting that segment duplication further promote the expansion of the *SWEET* family.

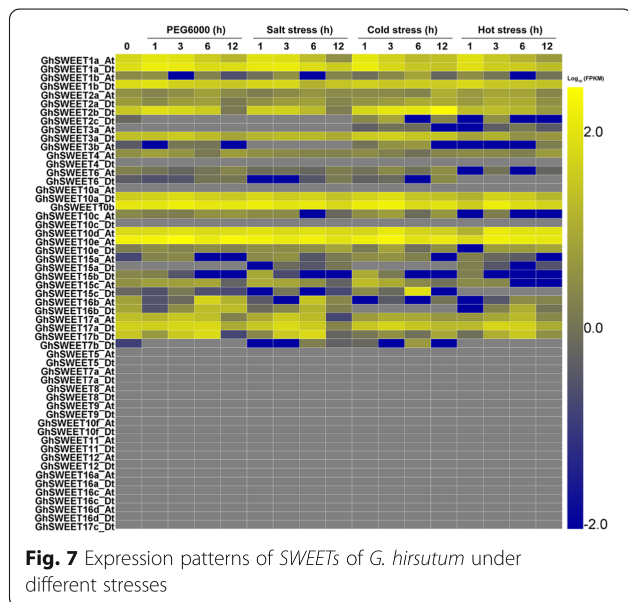


Fig. 7 Expression patterns of *SWEETs* of *G. hirsutum* under different stresses

Cotton SWEETs have been highly conserved during the evolution

The OsSWEET2b protein is the only eukaryotic SWEET protein with resolved three-dimensional structure so far (Tao et al. 2015). SWEET proteins generally contain seven transmembrane domains and two MtN3_slv domains (Chen et al. 2010; Chen et al. 2012; Talbot 2010; Baker et al. 2012), the N-terminus and C-terminus are outside and inside the cytoplasm of the cell, respectively. Each MtN3_slv contains 3 TMs, that is, TM1-TM3-TM2 arranged in the form of a triple-helix-bundles (THB) (Chen et al. 2010; Chen et al. 2012; Talbot 2010; Baker et al. 2012). The topology structure of SWEET proteins is clearly different from that of other sugar transporters, but their function is the same. In this study, most of SWEET proteins contain two MtN3_slv domains, though one or three domains are also founded in some genes, suggesting that the function is highly conservative in the evolutionary process. Most of the known sugar transporter proteins are located on the plasma membrane and are involved in the transport of sugars (Lalonde et al. 2004). In this study, all members of the SWEET family identified from cotton were located on the plasma membrane, which is consistent with previous research of SWEETs localization (Chen et al. 2012). However, recently, some study also showed that AtSWEET16/17 is located on the tonoplast involved

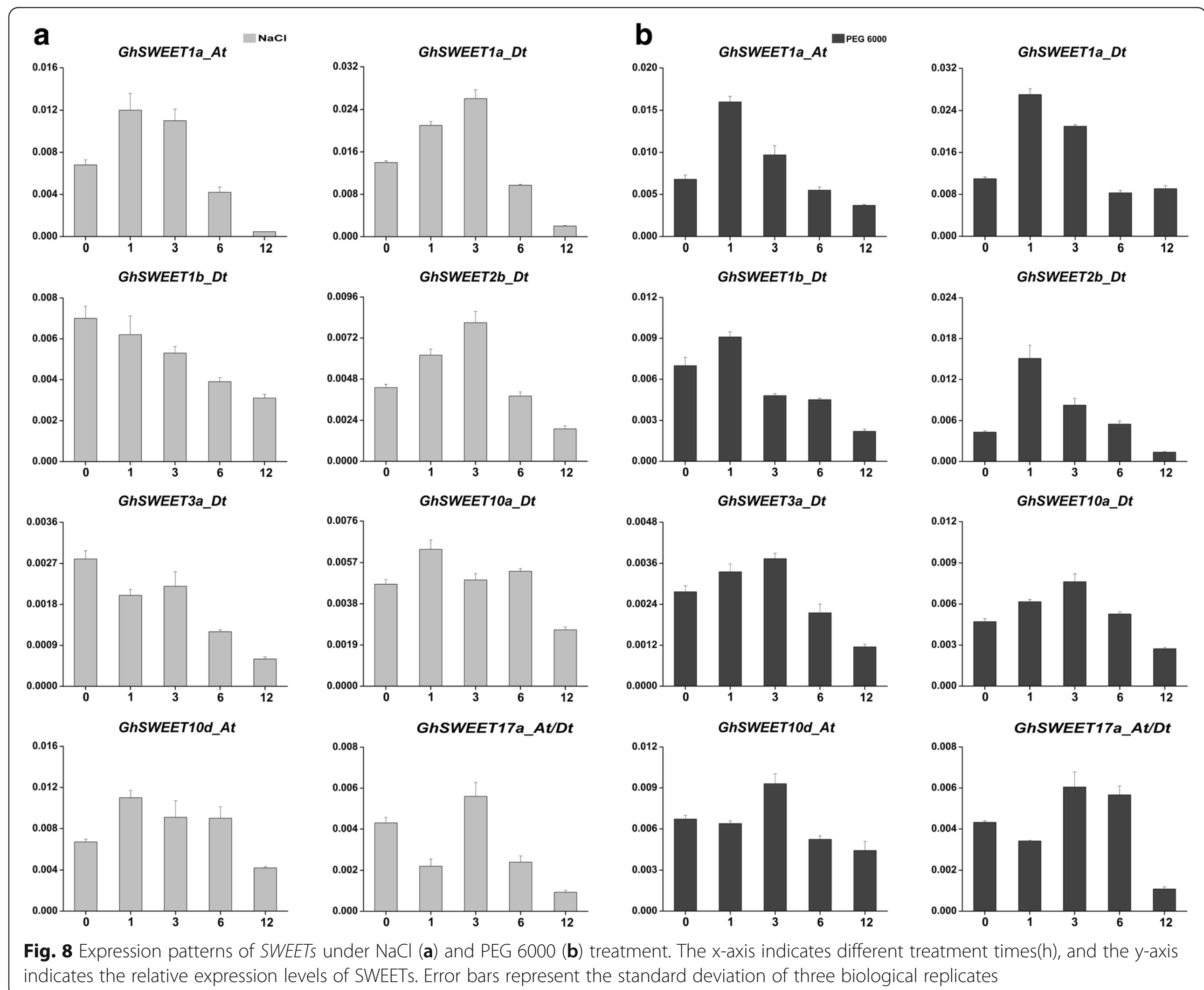


Fig. 8 Expression patterns of *SWEETs* under NaCl (**a**) and PEG 6000 (**b**) treatment. The x-axis indicates different treatment times(h), and the y-axis indicates the relative expression levels of SWEETs. Error bars represent the standard deviation of three biological replicates

in sugar transport. In addition, the length of SWEET protein varies significantly and the predicted isoelectric point is significantly different, suggesting that different SWEET proteins may play roles in different microenvironments.

Expression and putative functions of *SWEETs*

Previous studies have shown that the *SWEET* genes regulates the transport, distribution and storage of carbohydrates in plants and is involved in many important physiological processes in plants including phloem loading, reproductive development, disease-resistance, stress response, host-pathogen interactions and so on. SWEET family gene expression patterns in upland cotton were analyzed and found that their expression patterns differed significantly. *GhSWEET1a_At/Dt*, *GhSWEET2a_At/Dt*, *GhSWEET5_At/Dt*,*GhSWEET7a_At/Dt*, *GhSWEET9_At/Dt*, *GhSWEET15c_At/Dt* highly expressed in floral organs, indicating different members of the *GhSWEETs* family in cotton expressed in different parts of the flower and in different developmental stages. *AtSWEET1/4/5/7/8*, *AtSWEET13/14/15* have a relatively high level of expression in floral organs of *Arabidopsis* (Feng et al. 2015; Engel et al. 2005; 47 Wellmer et al. 2006). *OsSWEET1a/2a/3a/4/5/15* in rice also showed relatively high expression levels at different developmental stages of flowers and panicles (Feng et al. 2015). This shows that *SWEETs* play a universal role in plant reproductive development.

GhSWEET7b_Dt, *GhSWEET15c_At/Dt*, *GhSWEET10c_At/Dt*, *GhSWEET15b_ Dt*, and *GhSWEET10f_At/Dt* highly expressed in seeds of different developmental stages, *GhSWEET1b_At/Dt* has not only higher expression levels at different stages of seed formation but also during seed germination. The mutant of *AtSWEET17* plants are dwarfed and have a low seed yield, suggesting that *AtSWEET17* plays a role in the carbon distribution of plants (Chardon et al. 2013). *OsSWEET14* deletion mutation to the decrease of plant seed size and development delay. *OsSWEET14* deletion homozygous mutant plants

reproduction developed 30 days later than that of the heterozygous mutant (Braun et al. 2014), suggesting that *SWEETs* are also involved in the development of plant seeds. SWEETs are reported to be involved in sugar transport during fiber elongation (Cox et al. 2017). The high expression of *GhSWEET1a_At/Dt*, *GhSWEET1-b_At/Dt*, *GhSWEET15c_At/Dt*, *GhSWEET10c_At/Dt*, *GhSWEET15b_Dt*, and *GhSWEET10f_At/Dt* in fiber indicates that they may be involved in the fiber development of cotton and probably could be the candidate genes for further study of cotton fiber development.

Under stress, plants can maintain the balance of cell osmotic potential by regulating the redistribution of soluble sugar in vivo, which helps the plants to maintain normal growth under stress (Slewinski 2011; Kuhn and Grof 2010; Eom et al. 2015). Many *SWEET* genes in different plants are the key factors that regulate the redistribution of soluble sugar and respond to many abiotic stresses at the transcriptional level, indicating that they may be closely related to plant stress response (Yuan and Wang 2013; Klemens et al. 2013; Seo et al. 2011). The mutants have defects of *AtSWEET11* and *AtSWEET12* affects freezing tolerance in *Arabidopsis* (Hir et al. 2015). In our study, some of the *GhSWEETs* show an up-and down-regulated expression under the stress of salts and PEG 6000. The *GhWEET2a_At/Dt, GhWEET3a_Dt* and *GhWEET2b_Dt* are Up-regulated under cold treatment, while *GhWEET10e_Dt* and *GhWEET4_At* are Up-regulated under hot treatment. These results indicate differential expression suggested that the genes might have experienced functional divergence, and the study of SWEET function helps to artificially control the distribution of plant carbohydrates and has very significant potential value in improving crop yield, quality and cultivating new resistant varieties.

Conclusions

This study conducted a comprehensive analysis of *SWEET* gene family in the sequencing genomes of four cotton species for the first time. The *SWEET* family genes were classified into 4 groups in the phylogenetic tree. The SWEET genes are highly conserved among cotton and other plant species. A chromosomal location and gene duplication analysis revealed that segment repeat events promoted the expansion of the *SWEET* gene family in cotton. The duplicated genes may have undergone functional divergence in cotton because they showed different expression patterns in different tissues and organs. In addition, some members of the *SWEET* gene family may be involved in the regulation of stress response. This results promoted the understanding of the evolution of cotton *SWEET* genes, were helpful in further studies on the function of cotton *SWEET* family genes in future.

Methods

Gene retrieval and genome-wide identification analysis

The genome sequence data of four cotton species, *G. arboreum* (BJI, version 2.0), *G. raimondii*, (JGI, version 2.1), *G. hirsutum* (NAU, version 1.1, BJI, version 1.0), and *G. barbadense* acc. XinHai-21 (NAU, version 2.1) were retrieved from the CottonGen website (Yu et al. 2014) and the CottonFGD website (Zhu et al. 2017).

The rice (version 7.0), sorghum (version 3.1), cacao (version 1.1), poplar (version 3.0) and maize (version 1.1) genome sequence data were used from JGI (https://phytozome.jgi.doe.gov/pz/portal.html). The cucumber (version 2.0) genome sequence data (http://www.icugi.org/) (Li et al. 2011) were used.

To identify potential SWEET proteins in four cotton species, the published amino acid sequences of AtSWEETs (http://www.arabidopsis.org), OsSWEETs and CsSWEETs were used as query sequences. The obtained candidate genes and identified by performing a BLAST (E-value 1e-5) searches individually against the four cotton species genome databases. Then, the MtN3_slv domain was searched from the obtained candidate genes by InterProScan (Jones et al. 2014), and final the *SWEET* sequences were identified. The SWEET amino acid sequences from rice, cacao, poplar, tomato, sorghum, cucumber and maize were used and identified using the same method as employed for cotton species genome databases. Furthermore, the ExPASy tool (http://web.expasy.org/) was used to analyze the physicochemical parameters (i.e., length, molecular weight, and isoelectric point) of SWEET proteins of cotton that were identified from the currently available genomic database. The subcellular localization of each gene was predicated by the CELLO v2.5 server (Yu et al. 2004). The number of TM domains was predicted using the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM).

Multiple sequence alignment and phylogenetic analysis

Full-length or domain amino acid sequences of SWEET proteins were multiple aligned using ClustalX 2.0. The phylogenetic tree was constructed using the NJ method of MEGA7 with the pairwise deletion option and Poisson correction model (Kumar et al. 2016). For the reliability of interior branches, the bootstrap tests were performed with 1 000 replicates. To confirm the phylogenetic tree, constructed using the NJ method, the minimum-evolution method was also used.

Chromosome location and collinearity analysis

The physical chromosome locations of all *SWEET* genes were obtained from the genome sequence databases. The chromosomal location image was generated by Mapinspect 1.0 software. The predicted SWEET proteins were first aligned by ClustalW 2.0 at EMBL-EBI

(http://www.ebi.ac.uk/Tools/msa/clustalw2/) prior to a gene duplication analysis. Gene duplication events were defined according to the following conditions: the alignment region covered more than 80% of the longer gene and the identity of the aligned regions was over 80% (Li et al. 2017). The collinearity pairs of SWEET family were mapped using Circos software (Krzywinski et al. 2009).

Gene structure analysis and conserved domain sequence prediction

Arabidopsis and the four cotton species (*G. arboreum*, *G. raimondii*, *G. hirsutum*, and *G. barbadense*) *SWEETs* sequences were aligned by ClustalX 2.0, respectively; and MEGA 7.0 (Kumar et al. 2016) was used to construct an NJ tree using the method and parameters as described above. The exon/intron organization of the individual *SWEET* genes from *Arabidopsis* and cotton were performed using the Gene Structure Display Server (GSDS, http://gsds1.cbi.pku.edu.cn/) (Hu et al. 2014). Then, InterProScan was used to analyze the SWEET protein conserved domain of the four cotton species (Jones et al. 2014).

Transcriptome data analysis of SWEET gene expression from heat-map

The RNA-seq data was downloaded from the NCBI Sequence Read Archive (SRA: PRJNA248163, http://www.ncbi.nlm.nih.gov/sra/?term=PRJNA248163) and CottonFGD website (Yuan et al. 2015; Zhu et al. 2017). The fragments per kilobase million (FPKM) values denoting the expression levels of *SWEET* genes were isolated from a comprehensive profile of the TM-1 transcriptome data (Trapnell et al. 2012). A heat-map analysis was performed using HemI 1.0 software (Deng et al. 2014).

qRT-PCR

Cotton seeds of TM-1 were obtained from Shihezi University. The cotton (TM-1) seeds were germinated on a wet germinated disc for 3 days at 28 °C, and then transferred to a liquid culture medium (Yang et al. 2014). The seedlings were treated with 10% PEG 6000 and 300 mmol•L^{-1} NaCl at the 34 leaf stage. The true leaves were collected at 0, 1, 3, 6, and 12 h after the treatment and were immediately frozen in liquid nitrogen for RNA extraction. Total RNA was extracted from the seedlings. cDNA was synthesized by using an EASY-spin Plus Plant RNA Kit (Aidlab) with gDNA Eraser (Takara). The qRT-PCR reactions were conducted using a SYBR Green I Master mixture (Roche, Basel, Switzerland) according to the manufacturer's protocol on a Light Cycler 480II system (Roche, Switzerland). The cotton histone (*His*) gene (GenBank accession no. AF024716) was used as a standard control.

Additional files

Additional file 1: Table S1. The members of *SWEET* gene family in *G. raimondii*. (XLS 26 kb)

Additional file 2: Table S2. *The members of SWEET* gene family in *G. hirsutum*. (XLS 33 kb)

Additional file 3: Table S3. *The members of SWEET* gene family in *G. barbadense*.
(XLS 36 kb)

Additional file 4: Figure S1. Phylogenetic tree of *SWEET* genes indicating that *SWEET* genes could be divided into four clades. MEGA 7.0 was used for constructing the tree using the minimum-evolution method. The bootstrap values are shown near the nodes, and only those values greater than 50 are displayed. (TIF 2299 kb)

Additional file 5: Table S4. The paralogous pairs of *SWEET* genes in *G. arboreum*, *G. raimondii*, *G. hirsutum*, and *G. barbadense*. (XLS 17 kb)

Additional file 6: Figure S2. Phylogenetic relationships, gene structure and domain compositions of the *GrSWEET* genes. (a) Phylogenetic relationship between *A. thaliana* and *G. raimondii*. The phylogenetic tree was constructed using MEGA 7.0 with the jNJ method with 1 000 bootstrap replicates. The I clade, II clade, III clade, and IV clade is marked in cyan, purple, blue and pink, respectively. (b) Exon/intron structures of *GrSWEETs*. The introns and CDS are represented by orange boxes and black lines, respectively. (c) Protein domains. Each domain is represented in the colored box. (TIF 875 kb)

Additional file 7: Figure S3. Phylogenetic relationships, gene structure and domain compositions of the *GhSWEET* genes. (a) Phylogenetic relationship between *A. thaliana* and *G. hirsutum*. The phylogenetic tree was constructed using MEGA 7.0 with the NJ method with 1 000 bootstrap replicates. The I clade, II clade, III clade, and IV clade is marked in cyan, purple, blue and pink, respectively. (b) Exon/intron structures of *GhSWEETs*. The introns and CDS are represented by orange boxes and black lines, respectively. (c) Protein domains. Each domain is represented in the colored box. (TIF 1529 kb)

Additional file 8: Figure S4. Phylogenetic relationships, gene structure and domain compositions of the *GbSWEET* genes. (a) Phylogenetic relationship between *A. thaliana* and *G. barbadense*. The phylogenetic tree was constructed using MEGA 6.0 with the NJ method with 1 000 bootstrap replicates. The I clade, II clade, III clade, and IV clade is marked in cyan, purple, blue and pink, respectively. (b) Exon/intron structures of *GbSWEETs*. The introns and CDS are represented by orange boxes and black lines, respectively. (c) Protein domains. Each domain is represented in the colored box. (TIF 1078 kb)

Additional file 9: Table S5. Primers used for qRT-PCR in this study. (XLS 25 kb)

Abbreviations

BJI version 1.0: Beijing Genome Institute & Institute of Cotton Research of CAAS version 1.0; BLAST: Basic local alignment search tool; DPA: Days post anthesis; FPKM: Fragments per kilobase of transcript per million mapped fragments; FRET: Forster resonance energy transfer; *G. arboreum*: *Gossypium arboreum*; *G. hirsutum*: *Gossypium hirsutum*; *G. raimondii*: *Gossypium raimondii*; MSTs: Monosaccharide transporters; MW: Molecular weight; NAU version 1.1: Nangjing Agri. Univ. version 1.1; NJ: Neighbour joining; pI: Isoelectric point; qRT-PCR: Quantitative real-time polymerase chain reaction; SUT: Sucrose transporter; SWEET: Sugars will eventually be exported transporters

Funding

This work was supported by the The National Key Research and Development Program of China (2016YFD0101400, 2017YFD0101600).

Availability of data and materials

The RNA-seq analyses for *SWEETs* are available in the Sequence Read Archive (SRA) (SRA: PRJNA248163, http://www.ncbi.nlm.nih.gov/sra/?term=PRJNA248163). All another data generated or analyzed during this study are included in this published article and its Additional files.

Authors' contributions

ZYS, ZLJ, YB conceived and designed the research. ZLJ, CW and LY performed the experiments. GY, LZY and WJY prepared the materials. ZLJ, LYJ and YL analyzed the data. ZLJ wrote the paper. ZYS and CW revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, China. [2]Anyang Institute of Technology, Anyang 455000, Henan, China. [3]University of Saskatchewan, Saskatoon S7N 5A5, SK, Canada.

Received: 27 April 2018 Accepted: 17 July 2018
Published online: 02 August 2018

References

Ayre BG. Membrane-transport systems for sucrose in relation to whole-plant carbon partitioning. Mol Plant. 2011;4(3):377–394.

Baker RF, Leach KA, Braun DM. SWEET as sugar: new sucrose effluxers in plants. Mol Plant. 2012;5(4):766–768.

Bauer H, Ache P, Wohlfart F, et al. How do stomata sense reductions in atmospheric relative humidity. Mol Plant. 2013;6(5):1703–706.

Braun DM, Wang L, Ruan YL. Understanding and manipulating sucrose phloem loading, unloading, metabolism, and signalling to enhance crop yield and food security. J Exp Bot. 2014;65(7):1713–735.

Chardon F, Bedu M, Calenge F, et al. Leaf fructose content is ontrolled by the vacuolar transporter *SWEET17* in *Arabidopsis*. Curr Biol. 2013;23(8):697–702.

Chen LQ. SWEET sugar transporters for phloem transport and pathogen nutrition. New Phytol. 2014;201(4):1150–55.

Chen LQ, Hou BH, Lalonde S, et al. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature. 2010;468:527–32.

Chen LQ, Qu XQ, Hou BH, et al. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science. 2012;335(6065):207–11.

Cox KL, Meng FH, Wilkins KE, et al. TAL effector driven induction of a *SWEET* gene confers susceptibility to bacterial blight of cotton. Nat Commun. 2017;8:15588.

Deng WK, Wang YB, Liu ZX, et al. Heml: a toolkit for illustrating heatmaps. PLoS One. 2014;9(11):e111988.

Engel ML, Holmesdavis R, McCormick S. Green sperm. Identification of male gamete promoters in *Arabidopsis*. Plant Physiol. 2005;138(4):2124–33.

Eom JS, Chen LQ, Sosso D, et al. *SWEETs*, transporters for intracellular and intercellular sugar translocation. Curr Opin Plant Biol. 2015;25:53–62.

Feng CY, Han JX, Han XX, Jiang J. Genome-wide identification, phylogeny, and expression analysis of the *SWEET* gene family in tomato. Gene. 2015;573(2):261–72.

Ge YX, Angenent GC, Wittich PE, et al. *NEC1*, a novel gene, highly expressed in nectary tissue of *Petunia hybrida*. Plant J. 2000;24(6):725–34.

Guan YF, Huang XY, Zhu J, et al. *RUPTURED POLLEN GRAIN1*, a member of the *MtN3/saliva* gene family, is crucial for exine pattern formation and cell integrity of microspores in *Arabidopsis*. Plant Physiol. 2008;147(2):852–863.

Guo WJ, Nagy R, Chen HY, et al. *SWEET17*, a facilitative transporter, mediates fructose transport across the tonoplast of *Arabidopsis* roots and leaves. Plant Physiol. 2014;164(2):777–89.

Le Hir R, Spinner L, Klemens PA, et al. Disruption of the sugar transporters *AtSWEET11* and *AtSWEET12* affects vascular development and freezing tolerance in *Arabidopsis*. Mol Plant. 2015;8(11):1687–90.

Hu B, Jin JP, Guo AY, et al. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–97.

Hu LP, Zhang F, Song SH, et al. Genome-wide identification, characterization, and expression analysis of the *SWEET* gene family in cucumber. J Integr Agric. 2017;16(7):1486–501.

Jones P, Binns D, Chang HY, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–240.

Klemens PA, Patzke K, Deitmer JW, et al. Overexpression of the vacuolar sugar carrier *AtSWEET16* modifies germination, growth, and stress tolerance in *Arabidopsis*. Plant Physiol. 2013;163(3):1338–352.

Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–645.

Kuhn C, Grof CP. Sucrose transporters of higher plants. Curr Opin Plant Biol. 2010;13(3):288–98.

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–74.

Lalonde S, Wipf D, Frommer WB. Transport mechanisms for organic forms of carbon and nitrogen between source and sink. Annu Rev Plant Biol. 2004;55(1):341–372.

Li FG, Fan GY, Lu CR, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524–30.

Li FG, Fan GY, Wang KB, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet. 2014;46(6):567–72.

Li XH, Liu GY, Geng YH, et al. A genome-wide analysis of the small auxin-up RNA (*SAUR*) gene family in cotton. BMC Genomics. 2017;18:815.

Li Z, Zhang ZH, Yan PC, et al. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. BMC Genomics. 2011;12(1):540.

Lin IW, Sosso D, Chen LQ, et al. Nectar secretion requires sucrose phosphate synthases and the sugar transporter *SWEET9*. Nature. 2014;508(7497):546–49.

Liu QS, Yuan M, Zhou Y, et al. A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. Plant Cell Environ. 2011;34(11):1958–69.

Liu X, Zhao B, Zheng HJ, et al. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Sci Rep. 2015;5:14139.

Paterson AH, Wendel JF, Gundlach H, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature. 2012;492(7429):423–27.

Patil G, Valliyodan B, Deshmukh R, et al. Soybean (*Glycine max*) *SWEET* gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. BMC Genomics. 2015;16(1):520.

Phillips AZ, Berry JC, Wilson MC, et al. Genomics-enabled analysis of the emergent disease cotton bacterial blight. PLoS Genet. 2017;13(9):e1007003.

Rolland F, Moore B, Sheen J. Sugar sensing and signaling in plants. Plant Cell. 2002;14(suppl):s185–s205.

Ruan YL. Sucrose metabolism: gateway to diverse carbon use and sugar signaling. Annu Rev Plant Biol. 2014;65(1):33–67.

Seo PJ, Park JM, Kang SK, et al. An *Arabidopsis* senescence-associated protein SAG29 regulates cell viability under high salinity. Planta. 2011;233(1):189–200.

Slewinski TL. Diverse functional roles of monosaccharide transporters and their homologs in vascular plants: a physiological perspective. Mol Plant. 2011;4(4):641–62.

Sonnewald U. SWEETS-the missing sugar efflux carriers. Front Plant Sci. 2011;2:7.

Sun MX, Huang XY, Yang J, et al. *Arabidopsis RPG1* is important for primexine deposition and functions redundantly with *RPG2* for plant fertility at the late reproductive stage. Plant Reprod. 2013;26(2):83–91.

Talbot NJ. Cell biology: Raiding the sweet shop. Nature. 2010;468(7323):510–11.

Tao YY, Cheung LS, Li S, et al. Structure of a eukaryotic *SWEET* transporter in a homotrimeric complex. Nature. 2015;527(7577):259–263.

Trapnell C, Roberts A, Goff O, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protoc. 2012;7(3):562–78.

Walmsley AR, Barrett MP, Bringaud F, Gould GW. Sugar transporters from bacteria, parasites and mammals: structure-activity relationships. Trends Biochem Sci. 1998;23(12):476–81.

Wang KB, Wang ZW, Li FG, et al. The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet. 2012;44(10):1098–103.

Wellmer F, Alves-Ferreira M, Dubois A, et al. Genome-wide analysis of gene expression during early *Arabidopsis* flower development. PLoS Genet. 2006;2(7):e117.
Yang Z, Gong Q, Qin WQ, et al. Genome-wide analysis of *WOX* gene in upland cotton and their expression pattern under different stresses. BMC Plant Biol. 2017;17:113.
Yang ZR, Zhang CJ, Yang XJ, et al. *PAG1*, a cotton brassinosteroid catabolism gene, modulates fiber elongation. New Phytol. 2014;203(2):437–48.
Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 2004;13(5):1402–406.
Yu J, Jung S, Cheng CH, et al. CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Res. 2014;42(Database issue): 1229–236.
Yuan DJ, Tang ZH, Wang MJ, et al. The genome sequence of sea-island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. Sci Rep. 2015;5:17662.
Yuan M, Chu ZH, Li XH, et al. Pathogen induced expressional loss of function is the key factor in race-specific bacterial resistance conferred by a recessive *R* gene *xa13* in rice. Plant Cell Physiol. 2009;50(5):947–55.
Yuan M, Chu ZH, Li XH, et al. The bacterial pathogen *Xanthomonas oryzae* overcomes rice defenses by regulating host copper redistribution. Plant Cell. 2010;22(9):3164–76.
Yuan M, Wang SP. Rice *MtN3/saliva/SWEET* family genes and their homologs in cellular organisms. Mol Plant. 2013;6(3):665–74.
Zhang TZ, Hu Y, Jiang WK, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531–37.
Zhang ZY, Ruan YL, Zhou N, et al. Suppressing a putative sterol carrier gene reduces plasmodesmal permeability and activates sucrose transporter genes during cotton fiber elongation. Plant Cell. 2017;29(8):2027–46.
Zhao CR, Ikka T, Sawaki Y, et al. Comparative transcriptomic characterization of aluminum, sodium chloride, cadmium and copper rhizotoxicities in *Arabidopsis thaliana*. BMC Plant Biol. 2009;9(1):32.
Zhou Y, Liu L, Huang W, et al. Overexpression of *OsSWEET5* in rice causes growth retardation and precocious senescence. PLoS One. 2014;9(4):e94210.
Zhu T, Liang CZ, Meng ZG, et al. CottonFGD: an integrated functional genomics database for cotton. BMC Plant Biol. 2017;17:101.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions