Identification of chalcone synthase genes and their expression patterns reveal pollen abortion in cotton

Xiangjun Kong, Aziz Khan, Zhiling Li, Jingyi You, Fazal Munsif, Haodong Kang, Ruiyang Zhou

Abstract

Chalcone synthase (CHS) is a key enzyme and producing flavonoid derivatives as well play a vital roles in sustaining plant growth and development. However, the systematic and comprehensive analysis of CHS genes in island cotton (G. barbadense) has not been reported yet especially response to cytoplasmic male sterility (CMS). To fill this knowledge gap, a genome-wide investigation of CHS genes were studied in island cotton. A total of 20 GbCHS genes were identified and grouped into five GbCHSs. The gene structure analysis revealed that most of GbCHS genes consisted of two exons and one intron, and 20 motifs were identified. Twenty five pairs duplicated events (12 GbCHS genes) were identified including 23 segmental duplication pairs and two tandem duplication events, representing that GbCHS gene family amplification mainly owned to segmental duplication events and evolving slowly. Gene expression analysis exhibited that the GbCHS family genes presented a diversity expression patterns in various organs of cotton. Coupled with functional predictions and gene expression, the abnormal expression of GbCHS06, 10, 16 and 19 might be associated with pollen abortion of CMS line in island cotton. Conclusively, GbCHS genes exhibited diversity and conservation in many aspects, which will help to better understand functional studies and a reference for CHS research in island cotton and other plants.

Keywords:
Chalcone synthase
Gene family
Cytoplasmic male sterility
Island cotton

1. Introduction

Chalcone synthase (CHS) enzyme (commission number: E. C.2.3.1.74) is a key enzyme in the flavonoid biosynthetic pathway of plants, which catalyzes p-coumaroyl-CoA and three malonyl-CoA molecules to form phenyl styrene ketone (chalcone) (Koes et al., 1994). Chalcone is the precursor in the synthesis of a wide range of flavonoid derivatives, such as flavone, flavanol, anthocyanin and glycosides (Zhang et al., 2017). In plants, several physiological and biological process is strongly allied with the CHS gene. These include: the formation of anthocyanin in Matthiola incana (Hemleben et al., 2004), disease resistance in sorghum (Cui et al., 1996). Up to now, many CHS genes have been identified in angiosperms. The majority of these CHS genes in different species shows more than 60% homologous sequence and encoded a 40–45 kDa subunits, which include a Cys-His-Asn catalytic triad (CHN) in their active sites (Jiang et al., 2008). In addition, CHS genes have different expression pattern and tissues especially. For instance, CalCHS1, CalCHS2, and CalCHS3 were expressed in cassia alata root (Supachai et al., 2002). In Pisum sativum, PsCHS1 and PsCHS2 expressed in root and floral, but PsCHS6 and PsCHS7 especially expressed in root (Ito et al., 1997).

With development of bioinformatic and completion of plant genome draft, genome-wide identification and expression profile of CHS gene family have become feasible. With this method, ten CHS genes were identified in citrus, which contained a novel CHS gene and identified the expression pattern in different tissues and developmental stages (Wang et al., 2018). Moreover, 14 CHS genes were identified, and these genes presented tissue-specific expression patterns and differentially responded to MeJA treatment in Salvia miltiorrhiza (Deng et al., 2018). Although the whole genome sequencing of island cotton (G. barbadense) has been completed (Tao et al., 2017), the characterization, expression, and function of CHS are still elusive.
Cotton is an important fiber producing plant and exhibits yield improving heterosis in specific hybrid combinations (Zhu, 2016). Cytoplasmic male sterility (CMS) is an imperative player of heterosis utilization in plants. However, due to a few types of cotton CMS, its heterosis utilization is limited. Therefore, it is important to study the molecular mechanisms of cotton CMS for germplasm innovation and utilization of cotton heterosis. Previous studies indicated that mutated or abnormal expression of CMS line of Raphanus sativus, the expression of CMS was drastically suppressed at later stage of anther development in CMS line (Yang and Terachi, 2008). However, in island cotton, the expression pattern and function of CMS gene response to pollen abortion is still mysterious.

In the present study, a total of 20 GbCHS genes were identified in island cotton genome. Subsequently, the characterization of GbCHS genes were investigated, including protein length, molecular weight, chromosome location, phylogenetic, gene structure, conserved motifs, and gene duplication. In addition, expression analysis of GbCHS family genes in different organs and various developmental stages of pollen abortion in CMS line H276A were explored. This work will provide insight for further function study of GbCHS gene family and the mechanisms of cotton CMS.

2. Method and materials

2.1. Plant materials

Cotton CMS (H276A) line and its maintainer (H276B) line were produced by our research team (Kong et al., 2017). These lines were grown in the experimental field of Guangxi University under natural conditions. At anthesis, root, stem, leaf, calyx, and petal of H276B were sampled. For both lines, various development stages of anther (Pollen mother cell stage (3–4 mm), Tetrad stage (4–5 mm), Early uninucleate stage (5–6 mm) and Later uninucleate stage (6–7 mm)) were collected and stored at −80 °C for RNA isolation.

2.2. Identification of GbCHS genes

Genome data of island cotton were downloaded from Cotton Functional Genomics Database (https://cottonfgd.org/). Predicted CHS proteins from the island cotton genome were scanned using HMMER 3.0 (Finn et al., 2011) with the Hidden Markov model (HMM) corresponding to the Pfam (Finn et al., 2014) Chal_sti-/synt_N (PF027979) and Chal_sti_synt_C (PF00195) domains. Cotton specific Chal_sti_synt_N HMM and Chal_sti_synt_C HMM were constructed using hmmbuild from HMMER 3.0 with predicted CHS proteins and the Chal_sti_synt_N HMM and Chal_sti_/synt_C HMM. The specific Chal_sti_synt_N HMM and Chal_sti_synt_C HMM were used, and all proteins with an E-value lower than 1e–5 were selected. Meanwhile, all identified GbCHS proteins were used as queries to search against the island cotton protein database using default parameters. With the help of CDD (http://www.ncbi.nlm.nih.gov/cdd/), InterPro (http://www.ebi.ac.uk/interpro) and PFAM databases, only the sequences with Chal_sti_synt_N and Chal_sti_synt_C domains were considered as GbCHS proteins and used for further analyses.

2.3. Chromosomal mapping and phylogenetic analysis

All GbCHS genes were mapped to island cotton chromosomes based on genome annotations. The map was drafted using MapGene2Chrom web V2.0 (http://mg2c.iask.in/mg2c_v2.0/). GbCHS proteins were used for phylogenetic analysis using MEGA-X (https://www.megasoftware.net/). The neighbor-joining (NJ) method was applied to generate an unrooted phylogenetic tree with the pairwise detection option and 1000 bootstrapping replicates.

2.4. Gene structures, Gene Ontology annotation, conserved motifs, and gene duplication analysis

MEME (Multiple Em for motif elicitation) V5.0.5 program (http://meme-suite.org/tools/meme) was employed to identify the conserved motifs with default parameters, except for the maximum number of motifs (20). GSDS (Gene structure display server) V2.0 (http://gsds.cbi.pku.edu.cn/) was employed to identify the gene structure of GbCHS genes. In addition, the secondary structure of Medicago sativa CHS have been comprehensively analyzed (Ferrer et al., 1999). Hence, its sequence was used as a template to align the sequences of the GbCHS using the online servers DBD (http://www.rcsb.org/) and ESPript (http://esprit.bprc.fr/ESPript/ESPript/). The Blast2GO V5.2 (https://www.blast2go.com/) was used to investigate Gene Ontology (GO) annotation of GbCHS using with the amino acid sequences and performed with default parameter. Gene duplication events of GbCHS genes were investigated. We defined the gene duplication using the following criteria: 1) the alignment of whole protein length covered >80% of the longest gene, 2) the aligned region had an identity >80%. To determine the evolutionary pressure acting on duplicated genes, Ka and Ks values were calculated using Ka/Ks calculator 2.0 (Zhang et al., 2006).

2.5. RNA isolation, cDNA synthesis and quantitative RT-PCR (qRT-PCR)

Total RNA from different cotton tissues were isolated with the EasyPure Plant RNA Kit (Trans, China). RNA integrity and concentrations were confirmed using 1% agarose gels and a NanoDrop 2000 (Thermo, USA). cDNA of each sample was synthesized with 1 μg total RNA followed the Transcript One-step gDNA Removal and cDNA Synthesis SuperMix (Trans, China). Gene expression profile was assessed by real-time qRT-PCR with TransStart Tip Green qPCR SuperMix (Trans, China) with a C1000 Touch™ Thermal Cycler (Bio-Rad, USA). 18S served as the internal reference gene. All primers used in this study were designed using Primer 5.0 and are shown in Supplement 1. Conditions used for qRT-PCR were: 95 °C for 3 min, 95 °C for 5 s, 60 °C for 30 s, 40 cycles. The melt curve was generated by heating to 95 °C with an increment of 0.5 °C for 5 s. The relative expression level of the various sample were calculated using 2−ΔΔCt method (Livak and Schmittgen, 2001) with three replicates. The data of gene expression between H276A and H276B of cotton were statistically analyzed with SPSS 18.0

3. Results

3.1. Identification and chromosomal location of the GbCHS family genes

Genome-wide identification of CHS genes was conducted to explore the characteristics of CHS family genes of island cotton. A total of 70 candidate CHS protein sequences were obtained by searching the protein database of island cotton using Hidden Markov Model (HMM) profile. Then, 50 candidate CHS protein sequences were abandoned as the absent of Chal_sti_C and Chal_sti_/N domains. Finally, twenty non-redundant GbCHS genes were identified and renamed from GbCHS01 to GbCHS20 based on their
Table 1
The related information of the chalcone synthase family genes in island cotton.

| Gene Name | Translation Product | Size (aa) | Mw (kDa) | PI   |
|-----------|---------------------|-----------|----------|------|
| GbCHS01   | GOBAR_AA24378.1     | 389       | 42.54    | 5.98 |
| GbCHS02   | GOBAR_AA24046.1     | 384       | 42.45    | 8.92 |
| GbCHS03   | GOBAR_AA15712.1     | 259       | 28.9     | 8.79 |
| GbCHS04   | GOBAR_AA15713.1     | 393       | 43.9     | 7.54 |
| GbCHS05   | GOBAR_AA15714.1     | 393       | 43.49    | 8.02 |
| GbCHS06   | GOBAR_AA15715.1     | 363       | 40.22    | 7.51 |
| GbCHS07   | GOBAR_AA00712.1     | 349       | 38.14    | 6.56 |
| GbCHS08   | GOBAR_AA15716.1     | 389       | 42.6     | 6.12 |
| GbCHS09   | GOBAR_AA15717.1     | 389       | 42.6     | 6.12 |
| GbCHS10   | GOBAR_AA15718.1     | 399       | 44.21    | 5.37 |
| GbCHS11   | GOBAR_AA15719.1     | 389       | 42.51    | 5.98 |
| GbCHS12   | GOBAR_AA15720.1     | 392       | 42.84    | 7.11 |
| GbCHS13   | GOBAR_AA15721.1     | 295       | 33.09    | 8.42 |
| GbCHS14   | GOBAR_AA15722.1     | 393       | 43.63    | 6.62 |
| GbCHS15   | GOBAR_AA15723.1     | 326       | 35.47    | 5.72 |
| GbCHS16   | GOBAR_AA15724.1     | 386       | 42.79    | 7.48 |
| GbCHS17   | GOBAR_AA15725.1     | 389       | 42.64    | 6.12 |
| GbCHS18   | GOBAR_AA15726.1     | 389       | 42.67    | 6.12 |
| GbCHS19   | GOBAR_AA15727.1     | 360       | 39.48    | 5.62 |
| GbCHS20   | GOBAR_AA15728.1     | 389       | 42.65    | 5.72 |

Fig. 1. Chromosome distribution and gene duplication of GbCHS genes. A and D represents A and D sub-genome of G. barbadense, respectively. The left side shows the scale of a chromosome length. The tandem duplication gene clusters are indicated by a red line, and segmental duplication genes are linked by dashed lines.

Fig. 2. Phylogenetic and gene structure analysis of GbCHS genes. a. The phylogenetic tree of the GbCHS genes was constructed using MEGA-X. Bootstrap values (more than 50) are presented at each branch. b. Intron-exon structure of 20 GbCHS genes was explored using GSDS. The black boxes and black lines represented exons and introns, respectively. The scale of gene length are presented at the bottom.
chromosomes order. The nucleotide and protein sequences of GbCHSs are presented in Supplement 2, 3. Chromosome distribution profile showed that ten and nine GbCHS genes unevenly located on A sub-genomes (A02, 05, 08, 09, 10 and 12) and D sub-genomes (D02, 05, 08, 10 and 12), respectively. Also, the GbCHS20 was located on the Gb_scaffold 0063 (Fig. 1). Physical and chemical properties of GbCHS proteins are listed (Table 1), including protein length, molecular weight, and theoretical isolectric point.

3.2. Phylogenetic and structural analysis of GbCHSs

To understand the phylogenetic and evolutionary relationship among each GbCHS genes, an unrooted phylogenetic tree was built using the protein sequences of GbCHS with neighbor-joining (NJ) method. All GbCHSs were clustered into five major categories according to the tree topology and bootstrap values. The class I occupied the highest number of nine GbCHS genes, followed by class III with five GbCHS, whereas class II, IV, and V have only two genes in each group (Fig. 2a). In addition, with sequences alignment between CDS and genomic sequences of GbCHS genes, their exon-intron structure analysis was also conducted. Most of GbCHS family members (13) contained one intron and two exons (Fig. 2b). While GbCHS15 included only one exon, and rest had multiple exons and introns. Moreover, the same group of GbCHSs had a similar intron-exon organization structure.

In addition, the secondary structure of GbCHSs were analyzed using M. sativa CHS as a template (Fig. 3). The “gatekeeper"
phenylalanines connected with CoA-binding (Austin, 2003) at positions 215 and 265 are conserved. At these positions, almost of GbCHS are conserved except GbCHS03, 13 at position 215, and GbCHS02, 12, 05, 13 and 14 which contains a tyrosine substituted for phenylalanine at position 265 (Fig. 3), which potentially caused remarkable functional diversity, such as the choice of the initial substrates. The catalytic triad and the CHS family-specific Pro375 residues (Austin, 2003) were also conserved in almost of all GbCHSs. These results indicated that GbCHSs presented a high similarity in sequences with MsCHSs, suggesting that the CHS family is conserved during the evolutionary process.

3.3. Gene duplication of GbCHS gene family

To investigate the expansion mechanism of GbCHS gene family, we studied the gene duplication of GbCHS family genes such as segmental and tandem duplication events. A total of 25 pairs of duplicated genes (12 GbCHS genes) were identified. Among these, only

| Paralogous genes | Ka  | Ks  | Ka/Ks | Selective pressure |
|------------------|-----|-----|-------|-------------------|
| GbCHS01&GbCHS09  | 0.012 | 1.025 | 0.012 | Purity selection |
| GbCHS01&GbCHS11  | 0.001 | 0.035 | 0.032 | Purity selection |
| GbCHS01&GbCHS15  | 1.012 | 0.960 | 1.054 | Positive selection |
| GbCHS01&GbCHS17  | 0.012 | 0.945 | 0.013 | Purity selection |
| GbCHS01&GbCHS18  | 0.014 | 0.930 | 0.015 | Purity selection |
| GbCHS04&GbCHS05  | 0.100 | 0.327 | 0.307 | Purity selection |
| GbCHS04&GbCHS14  | 0.047 | 0.099 | 0.472 | Purity selection |
| GbCHS05&GbCHS14  | 0.075 | 0.359 | 0.209 | Purity selection |
| GbCHS09&GbCHS11  | 0.011 | 1.025 | 0.011 | Purity selection |
| GbCHS09&GbCHS15  | 1.042 | 0.860 | 1.211 | Positive selection |
| GbCHS09&GbCHS17  | 0.001 | 0.043 | 0.031 | Purity selection |
| GbCHS09&GbCHS18  | 0.003 | 0.059 | 0.050 | Purity selection |
| GbCHS10&GbCHS19  | 0.035 | 0.097 | 0.364 | Purity selection |
| GbCHS11&GbCHS15  | 1.013 | 0.957 | 1.058 | Positive selection |
| GbCHS17&GbCHS15  | 1.034 | 0.884 | 1.171 | Positive selection |
| GbCHS17&GbCHS11  | 0.011 | 0.948 | 0.012 | Purity selection |
| GbCHS18&GbCHS11  | 0.012 | 0.941 | 0.013 | Purity selection |
| GbCHS18&GbCHS15  | 1.036 | 0.879 | 1.179 | Positive selection |
| GbCHS18&GbCHS17  | 0.001 | 0.052 | 0.025 | Purity selection |
| GbCHS20&GbCHS01  | 0.025 | 0.869 | 0.028 | Purity selection |
| GbCHS20&GbCHS09  | 0.017 | 1.070 | 0.016 | Purity selection |
| GbCHS20&GbCHS11  | 0.024 | 0.871 | 0.027 | Purity selection |
| GbCHS20&GbCHS15  | 1.025 | 0.909 | 1.128 | Positive selection |
| GbCHS20&GbCHS17  | 0.016 | 0.966 | 0.017 | Purity selection |
| GbCHS20&GbCHS18  | 0.017 | 0.943 | 0.018 | Purity selection |

![Fig. 4](image_url) The sketch map of conserved motif distribution in 20 GbCHS proteins. The scale of protein length was presented at the bottom.
two tandem duplication events (GbCHS04/05 and GbCHS17/18) were detected. Data indicated that GbCHS gene family amplification mainly owned to segmental duplication events. Moreover, the combined analysis with a phylogenetic tree revealed that most duplication gene pairs derived from the same category. For instance, GbCHS10/19 in class V and GbCHS04/05 in class VI (Supplement.4, Fig. 1). The evolutionary selection of duplicated genes was conducted by the Ka/Ks ratio (Table.2). A ratio of Ka/Ks more than one represented positive selection, a ratio of Ka/Ks less than one represented purifying selection. A total of 6 duplicated gene pairs had a Ka/Ks ratio greater than one and all were associated with GbCHS15. The present data suggested that minority genes underwent positive selection and GbCHS15 was the original gene in GbCHS gene family. All other duplicated genes pairs were purity selection, indicating that most GbCHS genes were slowly evolved.

3.4. Gene ontology annotation and conserved motif analysis of the GbCHS gene family

We performed GO classification enrichment of 20 GbCHS genes with Blast2GO. Cellular components, molecular functions, and biological processes were confirmed (Supplement.5). Regarding cellular components, only GbCHS16 was identified and located on the nucleus (GO:0005634). In the term of biological processes, all the genes have the function of transferase acyl_groups other than amino-acyl groups (GO: 0016747). The analysis of biological process revealed that all CHS genes were associated with the biosynthetic process (GO:0009058). Also, using the protein sequences of all GbCHSs to search against the Arabidopsis databases with STRING (https://string-db.org). After BLAST analysis, the highest scoring proteins were identified (supplement.6). Data showed that all members of group I, II and III have a high identity with TT4 gene and Group IV and V have a high identity with LAP5 and LAP6, respectively.

To obtain more comprehend diversity of motif compositions, the MEME web server was used to assess conserved motifs in the GbCHS protein sequences. A total of 20 motifs were identified (Fig. 4). The same class of GbCHSs according to their phylogenetic relationships shared similar motif compositions and order. Some motifs were peculiar among different classes. Such as motif 15 of class IV, motif 19 of class V and motif 20 of class VI. However, many motifs were identified in most GbCHS members. For instance, motif 1, 2, 3, 4, 5, 6 and 7. Thus, the components of the motif of GbCHS reveals the conservation and diversity function of the GbCHS family. Also, all motifs were submitted to the Pfam web server for conserve domain checking. The results showed that motifs 1, 3, 5 and 6 belong to Chal_sti_synt_N domain and motifs 2, 4, and seven belong to Chal_sti_synt_C domain (Table 3). These data suggested that these motifs may play an imperative role in the gene function of GbCHS.

3.5. Tissues-specific expression profiling of GbCHS genes

The tissues-specific expression analysis of 20 GbCHS genes was investigated using the qRT-PCR method. Relative expression levels of 20 GbCHSs in root, stem, leaf, calyx, petal, and anther were displayed with a heat map (Fig. 5). Gene expression data were normalized with internal control 18S, and relative expression level more than 0.05-fold relative to that of 18S was deemed to be detected (Tao et al., 2017) (Supplement.7). Members of group I have shown diversity in expression patterns. For instance, GbCHS09/17, GbCHS08 were expressed in petal, GbCHS18 were expressed in stem and petal, GbCHS01/11 were universally expressed in all the tissues.

Table 3

| Motif | Length | Best Possible Match | Domain |
|-------|--------|---------------------|--------|
| 1     | 50     | EWQGPKISKTHLVFCCTTSGVDMPCADYQLTKLGLRPSV娶NKMYQQGF | Chal_sti_synt_N |
| 2     | 50     | WNSLFWVAPGCGPAILDQVEAEILLALPKERGTRHSLVEYGNMSSACVFL | Chal_sti_synt_C |
| 3     | 50     | KRRVMTELTEIKENP/VEYMAPSLLDAQRVMYVVEPKLKEAATRAIK | Chal_sti_synt_N |
| 4     | 43     | VSAQITQIPDSGDAIDGHLEBGVTCLHKLHDPGLSSKNIES | Chal_sti_synt_N |
| 5     | 41     | AQRCAQTATVLAAGITSTTIPNCVQDSTPYDDYFRRINTSEHK | Chal_sti_synt_N |
| 6     | 41     | NNKARVLCVCSEITAVTFQGTDHTLDLSVGQALFDGDAA | Chal_sti_synt_N |
| 7     | 29     | QTTRGGELEWVLFGGFLGTVETVTVLHIS | Chal_sti_synt_C |
| 8     | 11     | TVLVAKGLAE | * |
| 9     | 15     | ELKFRDMKCRMSMI | * |
| 10    | 15     | AVIGCSIPPEKEKP | * |
| 11    | 11     | DEMRRKSRREGC | * |
| 12    | 11     | VEAFQPLGSD | * |
| 13    | 8      | MTVVEVR | * |
| 14    | 11     | KLERLCKTTTV | * |
| 15    | 15     | MATENNL | * |
| 16    | 13     | MSKIDNNNAPWHR | * |
| 17    | 11     | CERLAMTVGLT | * |
| 18    | 8      | VMRYMREE | * |
| 19    | 11     | MGSEEPEKDF | * |
| 20    | 11     | FETLIFICGVD | * |

* Indicated that no domain sequences were contained in conserved motifs.
expressed, whereas GbCHS20, GbCHS15, and GbCHS07 were not over expressed in leaf. Genes of group II were lowerly expressed in most of the tissues except root. For group III, all GbCHSs (GbCHS13/05, GbCHS03, and GbCHS04/14) were highly expressed in root and stem but did not express in the calyx. Members of group IV and V were universally expressed in different tissues, GbCHS10/19 expressed highly in stem, leaf, and anther, whereas expression of GbCHS06/16 was higher in root, stem and anther. In summary, the GbCHS family genes presented a diverse expression patterns in various tissues of cotton, suggesting close association with the growth and development.

3.6. Gene expression profile of GbCHS genes during pollen abortion

Pollen abortion is a complicated process and related to the flavanone synthesis pathway. Tissues specific expression showed that 16 GbCHS genes were expressed in anther. These genes were selected to detect the expression patterns in response to pollen abortion of cotton (Fig. 6). Gene expression profile was performed at pollen mother cell (PMC), tetrad stage (Td), early uninucleate (early Uni) and late uninucleate (late Uni). At the PMC stage, the expression of most GbCHS was down-regulated in H276A than that of in H276B, except GbCHS01/11 and GbCHS04/14. Likewise, all GbCHS genes occupied down-regulated expression in H276A at the Td stage excluding GbCHS10/19 which was highly up-regulated. At the early uninucleate stage, GbCHS01/11, GbCHS20, GbCHS05/13, and GbCHS04/14 showed similar expression level between both materials, whereas the rest of GbCHS genes excluding GbCHS10/19 showed down-regulation expression in H276A. Interesting, unlike to other stages, all the GbCHS genes occupied up-regulated expression at late Uni stage except GbCHS07 and GbCHS20. The present results provide a diverse pattern of expression level in GbCHS genes and hence provide multi-function potential for pollen abortion in cotton. Moreover, we also found that most of the GbCHS gens expressed lower before and at the initial stage of the pollen abortive in H276A. These data indicated that abnormal expression of GbCHS genes might be associated with pollen abortion.

4. Discussion

Cotton genome data exploits further understanding of functional and regulatory mechanisms of the gene family. CHS is a key enzyme in flavonoid derivatives production and is an important...
player of plant growth and development. Whereas, the diverse functions of GbCHS genes, particularly in CMS system remained unclear. Therefore, the aim of this was a systematic understanding of GbCHSs’ diversity roles and the regulatory mechanism; a global genome overview of the GbCHS gene family. In this study, 20 CHS genes were identified according to the island cotton genome, while 8, 14 and 27 CHS genes were described in beans, Z. mays and O. sativa, respectively (Ryder et al., 1987; Han et al., 2016; Han et al., 2017). Which indicated the members of CHS genes in the plant is divergent. In general, multi-gene families in large plant genome originated from whole genome duplication and domestication. Cotton genome is passing through a series of duplications during its evolutionary processes, like those of A. thaliana (Raes et al., 2003) and O. sativa (Goff et al., 2002). In this study, 25 pairs duplicated events (12 GbCHS genes) were identified, including two tandem duplication events and 23 segmental duplication pairs, indicating about 60% of GbCHS genes arose from the duplicated chromosomal regions. This data is similar to O. sativa (Han et al., 2017), suggesting a vital role of CHS genes expansion in plant evolution. The higher frequency of segmental duplication relative to tandem duplication illustrated a major contribution of CHS gene family expansion. However, the segmental duplications contributed to the gene disperse and causing gene gradual evolving (Andrew et al., 2003; Yu et al., 2015). Also, the Ka/Ks ratios of 19 pairs of duplicated GbCHSs less than one, which indicated that the GbCHS gene family mainly went through purifying selection and gradual evolving.

The molecular evolution analysis revealed that, most CHS genes were classified into two or more subfamilies (Durbin et al., 2000). In this study, all GbCHSs were divided into five classes according to phylogenetic relationships. The gene structure analysis and the motif distribution were consistent with the phylogenetic relationship. It was reported that most of the CHS genes consist of two exons and one intron (Durbin et al., 2000). In the present study, 65% (13/20) GbCHS genes were consist with two exons and one intron, while some GbCHS genes shows a diverse compositions. For example, GbCHS15 had only one exon, and GbCHS08 had four and three exons. This diversity of gene structures are important for the evolution of gene families (Philippe et al., 2012). A total of 20 conservative motifs were identified in GbCHS gene family and among these, motifs 1, 3, 5 and 6 belons to Chal_sti_snt_N domain whereas motifs 2, 4, and 7 belong to Chal_sti_snt_C domain. These two domains are related to actyl transfer activity and transferase activity (Stefan et al., 2008) and wide spread in all GbCHS genes, which indicates that GbCHS genes function in catalyzing the formation of polyketone compounds. Some classes had specific motifs, like motif 15 of class IV, motif 19 of class V and motif 20 of class VI. The multifority of gene structure and conserved motif distribution contribute to the diversity of function of GbCHS family genes.

Members of the CHS gene family expressed diversity in different tissue of the plant. For instance, the most of the GmCHSs were highly expressed in leaves, whereas the GmCHS6, GmCHS7, GmCHS8, GmCHS10, and GmCHS11 were abundant in roots compared with other tissues (Vadivel et al., 2018). In this study, all GbCHS genes were expressed in petal, and 12 GbCHS genes were highly expressed in root and stem. Gene expression patterns are associated with their functions, and differential expression analysis can supply critical information for gene family research (Jingkang et al., 2008). In this study, 16 GbCHS genes were detected in anther and their expression patterns in the pollen abortion process were investigated. The pollen abortion of cotton CMS line H276A initial from tetrad stage and thoroughout abortive at late uninucleate (Kong et al., 2017). Majority of GbCHS genes expression were inhibited at pollen mother cell and tetrad stages. This indicating that, gene expression occurs before the phenotype and a similar expression pattern suggestingthat genes may work simultaneously. The inhibition of CHS and the other flavonoid biosynthetic genes expression were associated with nuclear-dependent male sterility and CMS (Yang and Terachi, 2008). GbCHS06, 10, 16, and 19 highly expressed in anthers. In addition, sequences alignment showed that their high identity with LAP5 and LAP6. LAP5 and LAP6 encode anther-specific proteins with homology to chalcone synthase and related to anther exine development. lap5 and lap6 mutations reduced the accumulation of flavonoids, which resulted in abnormal of anther exine and consequently male sterility (Dobritsa et al., 2010). Collectively, Our data suggested that GbCHS06, 10, 16 and 19 might be associated with pollen abortion of cotton CMS.

5. Conclusion

In this study, a total of 20 GbCHS genes were identified in the genome of island cotton. The phylogenetic relationships, gene structures, chromosomal locations, functional predictions and gene expansion revealed the diversity of GbCHS family genes. In addition, combing function prediction and gene expression respond to cotton CMS. We concluded that GbCHS06, 10 and 19 might be associated with CMS. This work will provide a systematic and comprehensive function and evolution of GbCHS gene family.

Author contributions

R Z conceived, designed and supervised the study. X K performed the experiments and drafted the manuscript. A K, Z L, J Y and H K participated in the experiments. A K and F M revised the manuscript and inserted useful suggestion.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of China (Grant No. 31363048) and the Project funded by China Postdoctoral Science Foundation (Grant No. 2019M653809XB).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2020.08.013.

References

Andrew, B., Steven, C., Russ, S., Georgiana, M., 2003. Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics 165, 309–319.

Austin, M.B., 2003. The chalcone synthase superfamily of type III polyketide synthases. Cheminform 20, 79–110.

Cui, Y., Magill, J., Frederiksen, R., Magill, C., 1996. Chalcone synthase and phenylalanine ammonia-lyase mRNA levels following exposure of sorghum seedlings to three fungal pathogens. Physiol. Mol. Plant Pathol. 49, 187–199.

Deng, Y., Li, C., Li, H., Lu, S., 2018. Identification and characterization of flavonoid biosynthetic enzyme genes in Salvia miltiorrhiza (Lamiaceae). Molecules 23, 1467.

Dobritsa, A.A., Zhentian, L., Shuh-Ichi, N., Ewa, U.W., Huhman, D.V., Daphne, P., et al., 2010. LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in Arabidopsis. Plant Physiol. 153, 937–955.

Durbin, M.L., McCaig, B., Clegg, M.T., 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol. Biol. 42, 79–92.
