

Genome-wide Identification and Bioinformatics Analysis of Cotton KCR Gene Family

Hongyuan Xi ¹, Man Wu ², Shoulin Hu ^{1,*}

¹ College of Agriculture, Tarim University, Alar Xinjiang, 843300, China

² Institute of Cotton Research of Chinese Academy of Agricultural Sciences/National Key Laboratory of Cotton Bio-breeding and Integrated Utilization/Key Laboratory for Cotton Genetic Improvement, Ministry of Agriculture and Rural Affairs, Anyang, Henan 455000, China

* Corresponding author: Shoulin Hu (Email: hushoulinghu@163.com)

Abstract: 3-ketoacyl-CoA reductase (*KCR*) is a key enzyme gene in the biosynthesis pathway of ultra-long chain monounsaturated fatty acids. Combining the sequence of *KCR* protein in *Arabidopsis thaliana*, we conducted a whole-genome identification and systematic evolutionary study of the *KCR* gene in cotton, focusing on *Gossypium. arboreum*, *Gossypium. raimondii*, *Gossypium. barbadense*, *Gossypium. hirsutum*. We analyzed the expression pattern of the *KCR* gene in high oil and low oil materials using transcriptome data and quantitative real-time polymerase chain reaction (qRT-PCR). Four *KCR* genes were identified in *Gossypium. arboreum*, four in *Ramond cotton*, eight in *Sea Gossypium. barbadense*, and eight in *Gossypium. hirsutum*. Systematic evolutionary analysis divided the cotton *KCR* gene family members into three branches, and all branches had similar conserved motifs. Some promoter regions of *Gossypium. hirsutum KCR* genes contained light response and drought regulation elements, suggesting that the *Gossypium. hirsutum KCR* gene family may play a role in stress response. Four genes were identified that showed significant expression differences in the 15 DPA of high oil and low oil seed development, indicating that *GhKCR* gene family members may be related to lipid synthesis metabolism. The evolutionary relationship and related functions of *KCR* gene family members in cotton were analyzed and predicted, which provided a preliminary theoretical basis for the study of *KCR* gene function in cotton.

Keywords: Cotton; *KCR*; Gene Family; Very Long Chain Fatty Acids; Oil Content.

1. Introduction

Cotton (*Gossypium* spp) belongs to the malvaceae cotton genus, cotton because of its important social and economic value, making it one of the world's main crops. Cotton plays an important role in People's Daily life, it is not only the main source of raw materials for the textile industry, but also widely used in other fields such as food, feed, animal protection and medical treatment. In addition, cottonseed oil is a by-product of lint processing and is a high-quality plant energy source. Cottonseed oil is the world's fifth largest oilseed crop [1] after soybeans, palm oil, rapeseed/canola oil, and sunflower oil, so improving the cottonseed oil content and improving the nutritional quality of cottonseed oil can effectively alleviate China's dependence on imported soybeans and improve China's food security.

The synthesis of very long chain fatty acids mainly relies on the FAE complex formed by the four enzymes: 3-Ketoacyl-CoA synthase (*KCS*), 3-Ketoacyl-CoA reductase (*KCR*), 3-hydroxyacyl-CoA dehydrase (*HCD*), and enoyl-CoA reductase (*ECR*) to catalyze the process [2]. *KCR* is a key reductase in VLCFAS extension. *KCR* differs from the diversity of members of the *KCS* family, which has the fewest of the four ultra-long chain fatty acid synthases. At present, studies on *KCR* mainly focus on *Arabidopsis* [3], maize [4], sugarcane [5], cotton [6] and *Saccharomyces cerevisiae* [7]. First, two paralogous *KCRs* (*gl8a* and *gl8b*) were discovered in maize, both of which are involved in the synthesis of very long chain fatty acids. Furthermore, the double mutant *gl8a/gl8b* is lethal to the embryo development, indicating that *KCR* plays a crucial role in the development of plant embryonic tissues. In *Arabidopsis*, both *AtKCR1* and *AtKCR2* genes are expressed in flowers, stems, and leaves, but only the

suppression of *AtKCR1* leads to a decrease in VLCFAs in the cuticle wax, seed oil, and sphingolipids of *Arabidopsis*, suggesting that only *AtKCR1* is a functionally active *KCR* isoform involved in fatty acid elongation [3]. In addition, *Arabidopsis* mutant experiments have shown that loss of *AtKCR1* function can lead to seedling death [3]. In sugarcane, the growth rate, small conidial production and germination rate of the sugarcane mutant *FsKCR1* knocked out were significantly reduced, and *FsKCR1* regulated the synthesis of ultra-long chain fatty acids in *Fusarium sugarcane* [5]. In cotton, two genes encoding *KCR*, *GhKCR1* and *GhKCR2*, were discovered. The genetic complementation and biochemical characteristics of these genes in yeast suggest that they can function as NADPH-dependent 3-ketoacyl-CoA reductases, and that *GhKCRs* are involved in ER-related very long chain fatty acid elongation [6].

2. Materials and Methods

2.1. Genome-wide Identification of KCR Gene Family Members in Cotton

Two diploid cotton species, *Gossypium. arboreum* (A2, HAU) and *Raymond cotton* (D5, HAU), and two tetraploid cotton species, *Gossypium. hirsutum* (AD1, ZJU) and *sea Gossypium. barbadense* (AD2, ZJU), were downloaded from the CottonMD database (<http://yanglab.hzau.edu.cn/CottonMD/>) including their gene annotation GFF3 files and reference genome files [8].

Submit the *KCR* homologous protein to the HMMER (<http://www.hmmer.org/>) software for comparison, and save the results. At the same time, use BLASTP to search for similar amino acid sequences of *KCR* homologous protein in *Arabidopsis* (E-value < 1×10⁻⁵). Integrate the results obtained

from the two methods, take the intersection of the two, and this is the candidate gene for the *KCR* gene family. Use the InterProScan software to predict the protein structural domains of the candidate gene amino acid sequences. Submit the candidate genes to the ExPASy (<http://web.expasy.org/protparam/>) website to analyze the physical and chemical properties of the cotton *KCR* gene family.

2.2. Construction and Conserved Motif Analysis of Cotton KCR Gene Family Evolutionary Tree

The selected *KCR* family members were compared with Arabidopsis *KCR* family members through Clustal W, and the step length was set to 1000 by applying the Neighbor-Joining method in MEGAX software to construct the evolutionary tree. Finally, using the evolutionary tree tool Evolview (<https://evolgenius.info/evolview-v2/>) [9] of constructed phylogenetic trees to beautify the *KCR* gene family. TBtools[10] software was used to visualize the conserved motifs of selected *KCR* genes.

2.3. Chromosomal Localization and Collinearity Analysis of KCR Gene Family Members in Cotton

Gene replication events of *KCR* gene family members in four cotton seeds were analyzed using MCScanX software. Then, the gene annotation file (GFF3) and genome file (fasta) of the four cotton seeds were imported into TBtools software. Finally, the collinear results were visualized using TBtools tool.

2.4. Analysis of Expression Patterns of KCR Gene in Gossypium.hirsutum

Using cotton database of ZheJiang university (<http://cotton.zju.edu.cn/index.htm>) to download Gossypium.hirsutum transcriptome data of different organizations use early high oil material 3008 with low oil group 3012 different development periods of ovule

transcriptome data [11], in the end, Use the TBtools tool to visualize the presentation patterns.

2.5. Extraction and Real-Time Quantitative Polymerase Chain Reaction of Total RNA from Gossypium.hirsutum

Follow the instructions in the ChamQUniversal SYBR qPCR Master Mix (Q711) manufactured by Vazyme. See the instructions for details. Using His as the internal reference gene, the obtained fluorescence quantitative data were calculated by 2- $\Delta\Delta C_t$ method, and the histogram was drawn by GraphPad-prism8 software.

3. Results and Analysis

3.1. Identification of KCR Gene Family Members of Cotton Flower

The *KCR* gene family members in Gossypium.arboreum, Gossypium.raimondii, Gossypium.barbadense, and Gossypium.hirsutum were identified through two approaches. Firstly, the genomic databases of Gossypium.arboreum, Gossypium.raimondii, sea Gossypium.barbadense, and Gossypium.hirsutum were searched and compared using HMMER3.0 software and local BlastP, and the *KCR* gene in Arabidopsis thaliana was identified by the protein sequences of *KCR* genes AT1G67730.1 and AT1G24470.1. Through deweighting the protein sequences obtained from the two searches, 26 cotton *KCR* genes were acquired. Subsequently, the conserved domain structure was reconfirmed via the NCBI-CCD and MEME database, and the sequences that did not conform to the characteristic structure of the *KCR* gene family were eliminated. Eventually, a total of 24 gene family members with an intact conserved domain of cotton *KCR* were identified. Among them, there were 4 members of Gossypium.arboreum, 4 members of Gossypium.raimondii, 8 members of Gossypium.barbadense, and 8 members of Gossypium.hirsutum.

Table 1. Analysis of physical and chemical properties of KCR gene family members in four cotton

Gene ID	Numberof amino acid	Molecular weight	pI	Subcellular localization
GH A03G0536	320	36032.21	9.6	Chloroplast
GH A03G0544	320	35991.16	9.5	Chloroplast
GH A03G1517	273	29644.67	9.66	Chloroplast
GH A03G1518	330	36688.25	9.42	Chloroplast
GH D02G1694	306	34007.68	9.51	Chloroplast
GH D02G1695	265	28975.17	9.35	Chloroplast
GH D03G1424	320	35982.14	9.54	Chloroplast
GH D03G1431	320	35918.13	9.48	Chloroplast
GB D03G1446	320	35982.14	9.54	Chloroplast
GB A03G0532	320	36040.23	9.49	Chloroplast
GB D03G1452	320	35918.13	9.48	Chloroplast
GB A03G0521	320	35998.18	9.6	Chloroplast
GB A03G1591	307	34046.76	9.83	Chloroplast
GB D02G1753	306	34044.74	9.59	Chloroplast
GB A03G1592	330	36688.25	9.42	Chloroplast
GB D02G1754	1706	184636.42	7.06	Chloroplast
Garb 01G005860	320	36045.25	9.54	Chloroplast
Garb 01G005780	320	35954.07	9.59	Chloroplast
Garb 03G010230	307	34045.77	9.89	Chloroplast
Garb 03G010200	330	36603.18	9.35	Chloroplast
Grai 02G016450	320	35940.12	9.53	Chloroplast
Grai 02G016520	320	35946.19	9.53	Chloroplast
Grai 03G019580	311	34582.27	9.66	Chloroplast
Grai 03G019590	313	34442.42	9.19	Chloroplast

The amino acid quantity, molecular weight and isoelectric point of the protein encoded by these genes were analyzed. The results showed that the amino acid quantity of the protein in *Gossypium.hirsutum* was 265-330. Molecular weight of 29644.47-35991.6; The average equipotential point is 9.51. The amino acid quantity of protein in sea *Gossypium.barbadense* was 306-1706; The molecular weight is between 34044.74 and 184636.42; The average equipotential point is 9.25. The amino acid quantity of protein in *Gossypium.arboreum* is 307-330. The molecular weight is between 34045.77-36603.18; The average equipotential point was 9.59. The number of amino acids in *Gossypium.raimondii* seed protein is 311-320; Molecular weight: 3444.42-35946.19; The average equipotential point is 9.48. Subsequently, the amino acid sequences encoded by these genes were analyzed and the 24 *KCR* gene family members identified in four cotton species were predicted. The results showed that all the 24 *KCR* gene family members in four cotton species were located in chloroplasts.

3.2. Phylogenetic Analysis of Cotton KCR Gene Family

Phylogenetic trees of cotton *KCR* family members were constructed using 24 cotton *KCR* and 2 Arabidopsis *AtKCR* amino acid sequences. As depicted in Figure 1, all *KCR* could be classified into three branches, namely GroupI, GroupII, and GroupIII. The *KCR* family has the largest number of members in GroupIII, with 13 members. The smallest is Group I, having only six members. This result implies that *KCR* members are not evenly distributed among different evolutionary groups.

In *Arabidopsis thaliana*, *AtKCR1* and *AtKCR2* were respectively classified into GroupII and GroupIII subgroups. It can be deduced that the cotton gene in GroupII is closely related to *AtKCR2*, and the cotton gene in GroupIII is closely related to *AtKCR1*. However, the *KCR* family members of cotton in GroupI were not clustered with the *KCR* family members of *Arabidopsis thaliana*, indicating that the similarity between the *KCR* family members of cotton in this branch and those of *Arabidopsis thaliana* is relatively low.

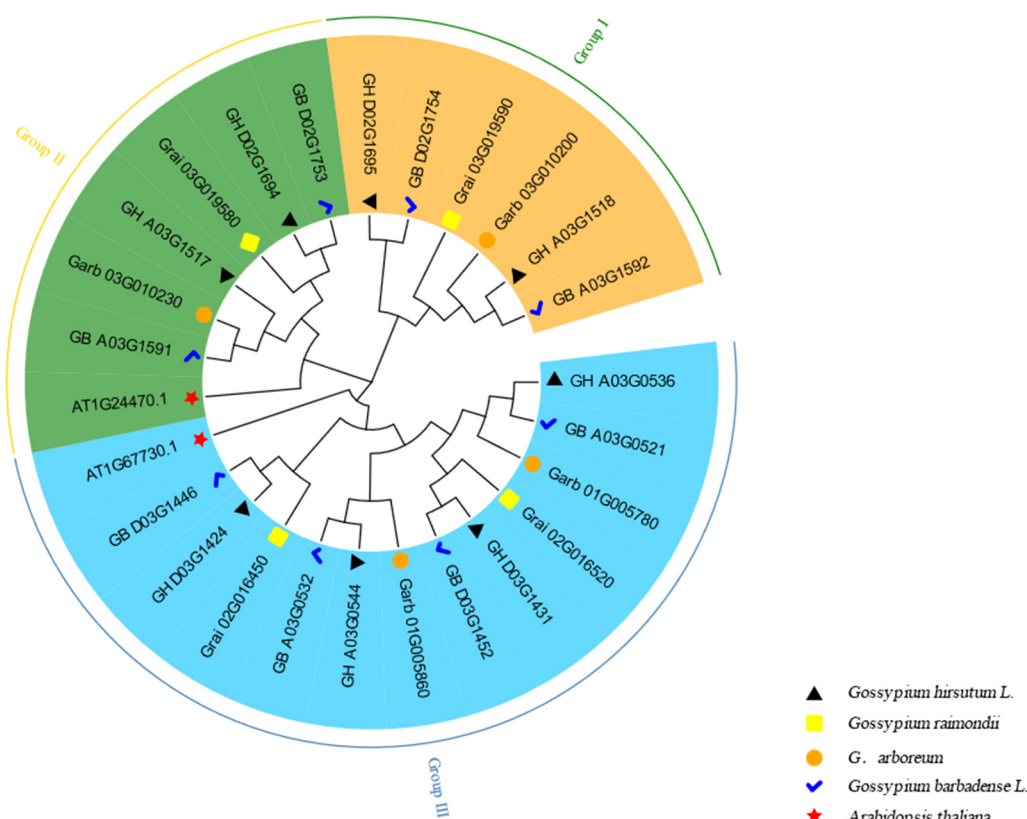


Figure 1. Shows the systematic analysis of the *KCR* gene in *Arabidopsis thaliana* and four species of *Gossypium* genus. The systematic tree was constructed using the neighbor joining (NJ) method.

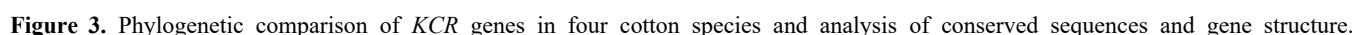
3.3. Gene Structure, Conserved Motif Analysis and Multiple Sequence Alignment of Cotton KCR

In order to study the sequence characteristics of *KCR* gene, 10 conserved motifs of 8 *KCR* proteins in cotton were analyzed using MEME online tool (Figures 2 and 3). The conserved motifs in Group I were basically the same. The distribution of Motifs of *KCR* protein members in Group I was completely consistent, and Motif 3 and 4 differences existed between *GH_A03G1517* and *GH_D02G1694* in GroupII. It is worth noting that motif 1, motif 2 and motif 7

are present in all cotton *KCR* proteins, which are highly conserved in evolution and may carry important functions, and their functions need to be further verified.

In order to further explore the gene structure of *KCR* gene family members, the intron-exon structural characteristics of *KCR* gene family members were analyzed in this study (Figure 3-B). As shown in the figure, the exon-intron number of *KCR* gene had little difference, ranging from 2 to 3, and the exon-intron distribution pattern of *KCR* gene in Group I subgroup was consistent, all containing 3 exons. *GH_A03G1517* lacks one intron compared with *GH_D02G1694* in the gene structure of *KCR* family members

similar conserved protein regions and motif compositions. These results not only effectively demonstrate the feasibility of phylogenetic evolutionary classification, but also indicate that they may have similar biological functions.



3.4. Chromosome Localization, Gene Replication and Selection Pressure of KCR Gene Family Members in Cotton

In the diploid genome of *Gossypium.arboreum* and *Gossypium.raimondii*, *KCR* gene was located on Chr01 and Chr03 of *Gossypium.arboreum* and Chr02 and Chr03 of *Gossypium.raimondii*. In the tetraploid genome of sea *Gossypium.barbadense* and *Gossypium.raimondii*, *KCR* gene is distributed on 6 chromosomes, namely, A03, D02 and D03 of Sea *Gossypium.barbadense* and A03, D02 and D03 of *Gossypium.hirsutum*. In the diploid genome, the *KCR* in *Gossypium.arboreum* and *Gossypium.raimondii* is unevenly distributed on each chromosome (Figure 4). In the tetraploid genome, *KCR* genes in sea *Gossypium.barbadense* and sea *Gossypium.barbadense* were distributed on chromosomes A03, D02 and D03, suggesting that *KCR* genes in the evolution of tetraploid sea *Gossypium.barbadense* and sea *Gossypium.barbadense* are highly conserved.

The chromosomes of *Gossypium.hirsutum* and sea *Gossypium.barbadense* were obtained by crossing two diploid cotton species, *Gossypium.arboreum* and *Gossypium.raimondii*. Therefore, *Gossypium.hirsutum* and

sea *Gossypium.barbadense* are a typical allotetraploid species. The expansion of gene families depends mainly on tandem, segment and genome-wide replication. According to Figure 5 Cricos, it can be analyzed that the replication distance between the genes of *GhKCR* and *GbKCR* is relatively long, and most members of this family are repeated by segmentary replication (Figure 5). It is inferred that segmentary replication is likely to be important in the evolution of *KCR* gene family.

In order to further explore the evolutionary relationship of *KCR* gene family in cotton, *KCR* genes in four cotton species were analyzed by collinear analysis. As shown in Figure 3, a total of 3 homologous gene pairs were detected in *Gossypium.hirsutum*. The three homologous gene pairs in *Gossypium.hirsutum* belong to the A and D subgroups (*GH_A03G0536/GH_D03G1431*, *GH_A03G0544 /GH_D03G1424*, *GH_A03G1517/GH_D02G1694*), respectively. A total of 4 homologous gene pairs were detected in *Gossypium.barbadense* (*GB_A03G0532/GB_D03G1446*, *GB_A03G0521/GB_D03G1452*, *GB_A03G1591/ GB_D02G1753*, *GB_A03G1593/GB_D02G*), both belonging to subgroups A and D 1754). The Ka/Ks values of the above genes were all less than 0.5, indicating that the *KCR* gene was highly conserved and stable in the evolutionary process.

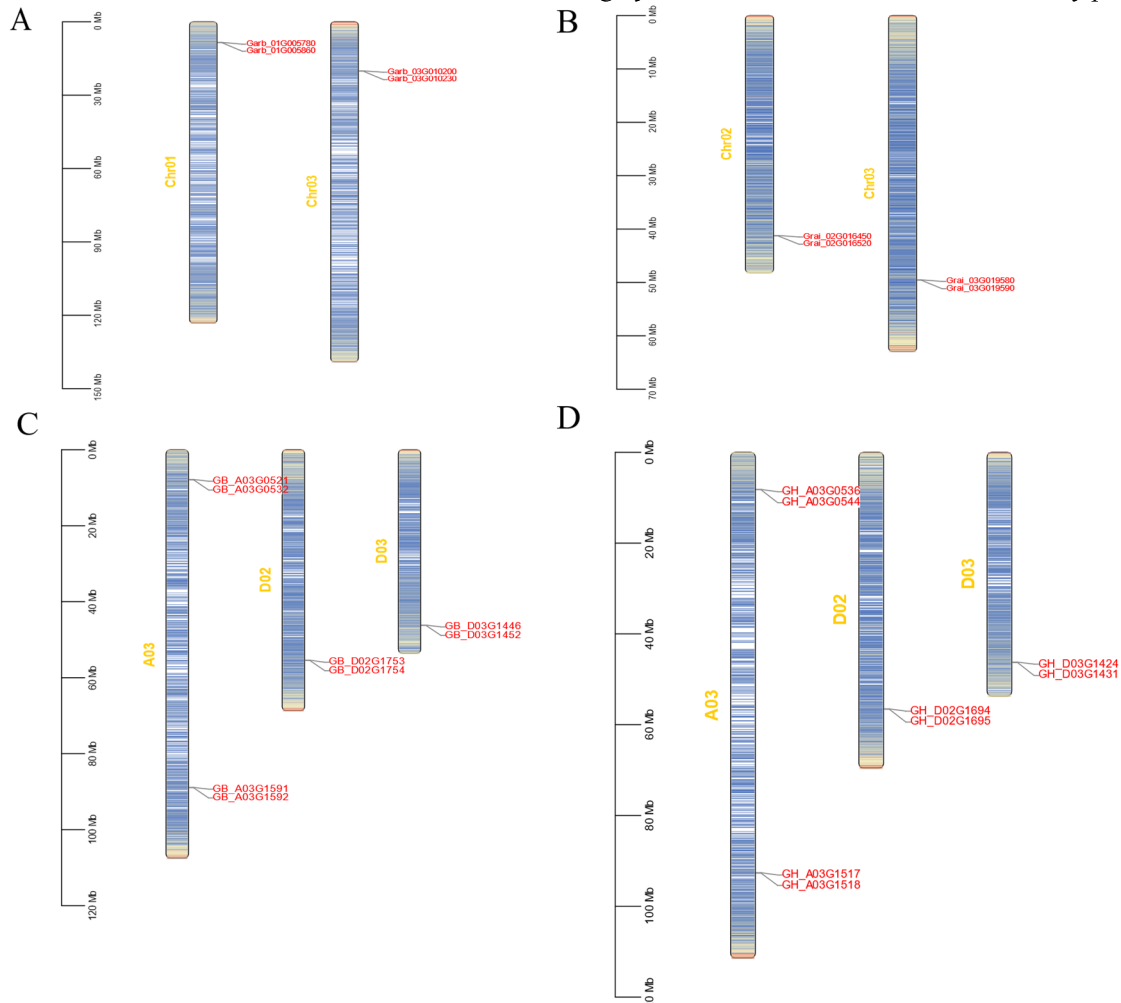


Figure 4. Chromosome distribution of four *KCR* members of cotton. A: *Gossypium.arboreum*; B: *Gossypium.raimondii*; C: *Gossypium.barbadense*; D: *Gossypium.hirsutum*;

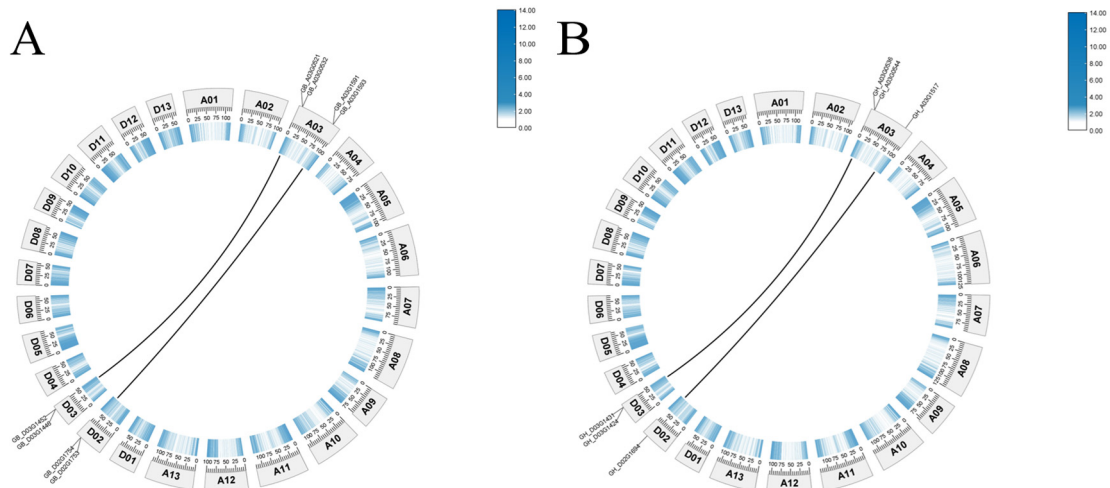


Figure 5. Cricos analysis of *Gossypium. barbadense* and *Gossypium. hirsutum*. A: *Gossypium. barbadense* B: *Gossypium. hirsutum*;

3.5. Analysis of Cis-acting Elements of KCR Gene in *Gossypium.hirsutum*

Promoters play an important role in the regulation of gene expression levels. In order to better understand the function of *KCR* gene in cotton, 2000 bp promoters upstream of each *KCR* gene in *Gossypium.hirsutum* were extracted. The 2000 bp promoter regions upstream of 8 *KCR* genes were predicted by PlantCARE. High temperature stress (CCAAT-box, AT-rich) was found in *KCR* gene promoter region of *Gossypium.hirsutum* element), plant hormones (TGACG-motif, CGTCA-motif, TGA-element, ARE-Motif, ABRE), drought stress (MBS, MYC), light response (G-box, AE-box, CTT-Motif, GATA-motif) Etc. (Figure 6). The high temperature stress element (CCAAT-box) is located in the

promoter region of *GH_A03G0536*, *GH_A03G0544* and *GH_D03G1424*, and these three genes are located in the Group III subgroup in the evolutionary tree, suggesting that the members of this subgroup may play a key regulatory role in the process of high temperature response. Drought resistance-related elements (MBS and MYC) are the two most abundant transcription factors in the *KCR* family of *Gossypium.hirsutum*, and most *KCR* genes in *Gossypium.hirsutum* contain regulatory factors for light and drought resistance in their promoter region. In summary, most promoter regions of *KCR* gene in *Gossypium.hirsutum* contain photoresponsive and drought regulatory elements. These elements are highly conserved in the *KCR* gene family, suggesting that *KCR* protein is an important transcription factor involved in cotton stress resistance.

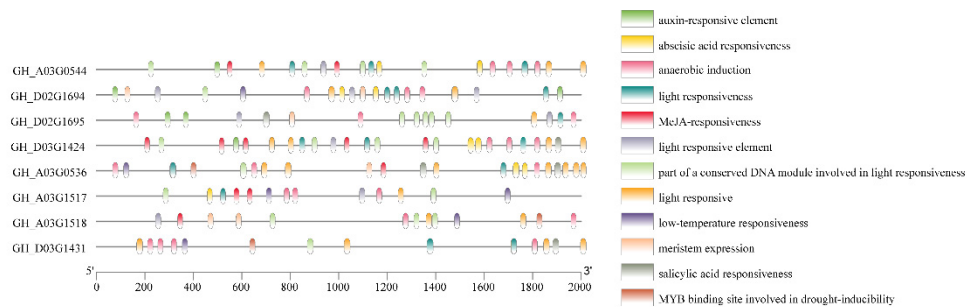


Figure 6. Analysis of cis-acting elements of *KCR* gene in *Gossypium. Hirsutum*

3.6. Analysis of KCR Gene Family Expression Patterns in *Gossypium.Hirsutum*

The transcriptome data of *Gossypium.hirsutum* from Zhejiang University were used to analyze the physiological role of *KCR* gene in cotton growth and development. The expression patterns of 8 *KCR* genes in ovule, anther, calyx, leaf, petal, pistil, root, sepal, stem and receptor at different developmental stages were analyzed. As shown in Figure 7, according to the expression pattern of each gene at the same stage and in different time periods, it was found that *GH_A03G1517* had the highest expression in ovule of 3DPA and 5DPA (Days Post anthesis). *GH_D02G1694* gene was almost not expressed in 10 different tissues and organs such as ovule, epicycles, receptus and sepals. The expression of

GH_A03G1518 was the highest in the ovule of 3DPA and 5DPA, but almost not expressed in the medium tissues of epicalyx, receptor and sepal. The expression of *GH_D03G1431* was high in the ovule of 3DPA and 5DPA, and relatively high in the calyx of different tissues. *GH_A03G0544* was highly expressed in ovule of 3DPA and 5DPA, and relatively high in sepal and calyx of different tissues. The expression level of *GH_A03G0536* was relatively high in ovule of 3 and 5DPA, and the highest in calyx of different tissues, followed by the highest in petal, sepal and receptor. The expression of *GH_D03G1424* was relatively high in calyx, receptacle and petal. *GH_D02G1695* was not expressed in ovule or in different tissues.

Using the transcriptome data of high-oil material 3008 and low-oil material 3012 previously published by our research

group, we analyzed the expression patterns of cotton *KCR* gene in different ovule development stages, as shown in Figure 8. A total of four genes, *GH_A03G1518*, *GH_A03G0544*, *GH_D03G1431* and *GH_D03G1424*, were found to have significant differences between high-oil materials and low-oil materials. At 20 DPA, the expression of *GH_A03G1518* in high-oil material 3008 is 3 times that in

low-oil material 3012. At 10 DPA, the expression of *GH_A03G0544* in high-oil material 3008 is twice that in low-oil material 3012. At 10 DPA, the expression of *GH_D03G1431* and *GH_D03G1424* in high-oil material 3008 was significantly different from that in low-oil material 3012, suggesting that these four genes may be involved in the anabolic process of cottonseed oil.

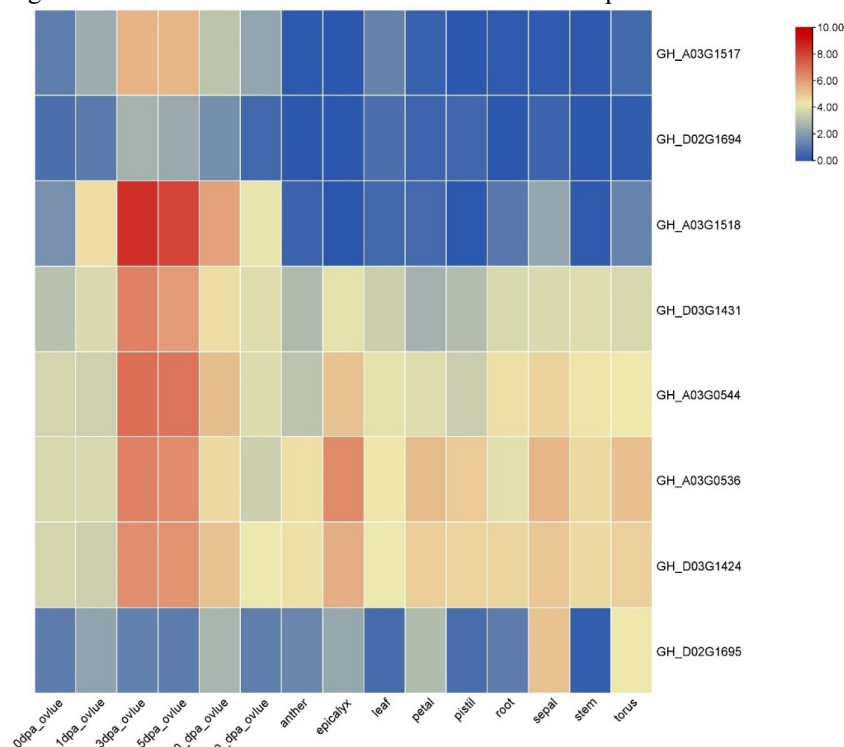


Figure 7. Analysis of *KCR* gene family expression patterns in *Gossypium. hirsutum*.

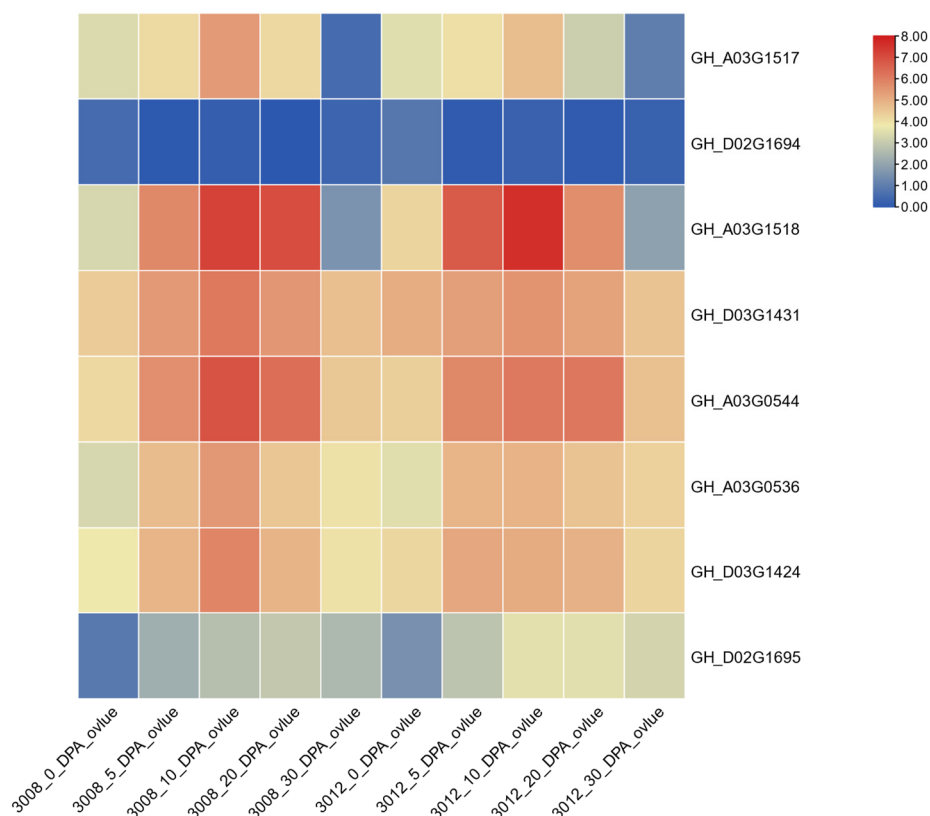


Figure 8. Analysis of expression patterns of *KCR* gene family in *Gossypium.hirsutum* in high-oil material 3008 and low-oil material 3012

3.7. Expression Pattern Analysis of GhKCR Gene during Ovule Development in High Oil and Low Oil Materials

The effects of *GhKCR* gene on cottonseed development in high oil material (N0409) and low oil material (N6940) were further analyzed by qRT-qPCR. The effects were 0 DPA, 5 DPA, 10 DPA, 15 DPA, 20 DPA, 25 DPA, 30 DPA, 35 respectively Spatial and temporal expression patterns of DPA (Figure 9). The results showed that during ovule development, the expression levels of four *KCR* candidate genes in high-oil

materials were all higher than those in low-oil materials (FIG. 9), especially in the critical period of seed oil accumulation, when the difference was most significant at 15 DPA, and the expression levels of both of them showed a trend of first increasing and then decreasing with the development of ovule in the two extreme materials. *GH_A03G15118* gene had the most significant difference at 15 DPA during the oil synthesis phase, with high expression in high oil materials and low expression in low oil materials. These studies suggest that members of the *GhKCR* gene family may regulate lipid anabolism.

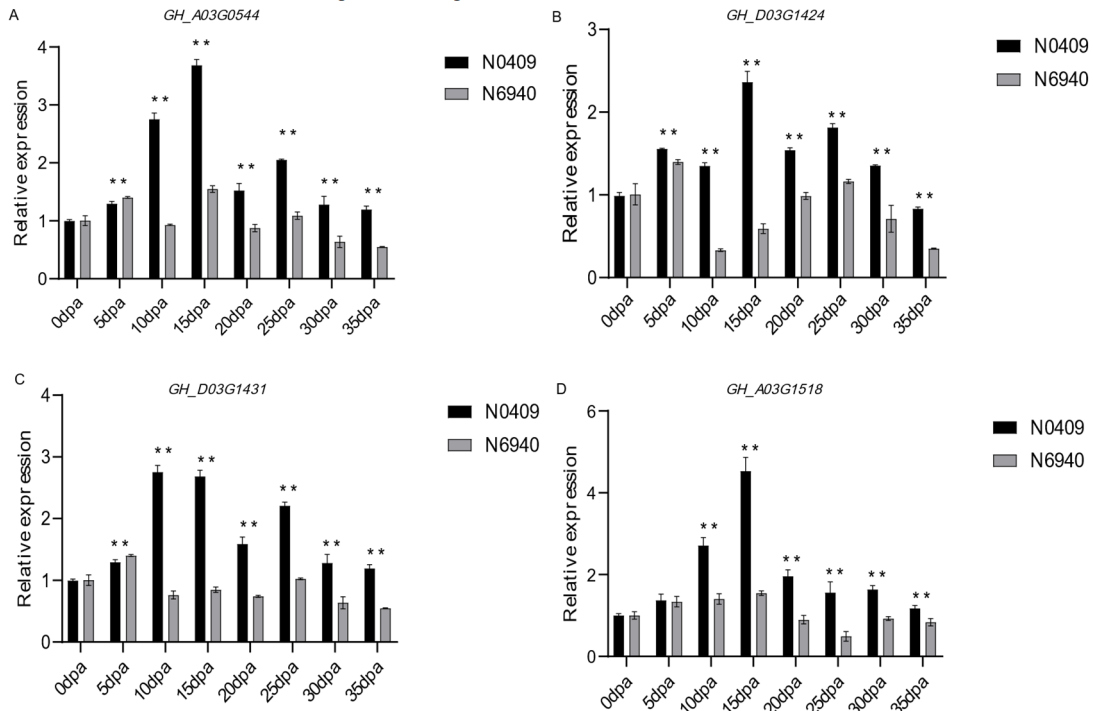


Figure 9. Expression pattern analysis of *GhKCR* gene family during ovule development in high-oil and low-oil materials. **Note:** A is GH_A03G0544, B is GH_D03G1424, C is GH_D03G1431, and D is GH_A03G1518

4. Discuss

In recent years, with the completion of the cotton genome sequencing project, the study of gene family and gene function has been greatly promoted. As an important cash crop in China, cotton can not only provide natural fiber, but also an important source of vegetable oil and biofuel. How to increase the content of cottonseed oil and improve its fatty acid components has become an important topic in current cotton breeding. VLCFAs are involved in the synthesis of lipids such as phospholipids, sphingolipids, cutin wax, xylosin, and triacylglycerol, so VLCFAs play a crucial role in the entire life process of plants, such as forming cell membranes, protecting plants from stress, and storing oil in seeds. The biosynthesis of VLCFAs is controlled by FAE complex composed of *KCS*, *KCR*, *HACD* and *ECR*. *KCS*, as a rate-limiting enzyme in VLCFAs synthesis, has received extensive attention. However, studies on *KCR* in plants mainly focus on *Arabidopsis*, sugarcane, corn and cotton, and no studies on oil and fatty acids in cotton have been reported. In this paper, based on cotton genome data, members of *KCR* gene family were identified by bioinformatics methods, and their protein motifs, phylogenetic relationships and expression patterns of ovule in high oil and low oil materials at different developmental stages were analyzed.

In this study, 24 *KCR* genes were identified from four

cotton seeds and divided into three branches by phylogenetic analysis, each with the same conserved motif. It suggests that there may be similar evolutionary relationships or functions between them. In *Arabidopsis thaliana*, *AtKCR1* and *AtKCR2* were divided into GroupII and GroupIII subgroups, respectively. It can be inferred that the cotton gene in GroupII is closely related to *AtKCR2*, and the cotton gene in GroupIII is closely related to *AtKCR1*. In *Arabidopsis thaliana*, *AtKCR1* and *AtKCR2* genes are expressed in flowers, stems and leaves, and the inhibition of *AtKCR1* leads to the decrease of VLCFAs in cuticle wax, seed oil and sphingoid of *Arabidopsis thaliana*, indicating that the cotton *KCR* gene in GroupIII is the same as *AtKCR1*. Possibly a functional *KCR* isomer involved in fatty acid elongation [3]. Most of the *KCR* genes contain PLN02780 conserved domain, indicating that this conserved domain is likely to be a necessary element for the normal function of *KCR* gene family members.

By analyzing the expression patterns of *KCR* gene family in *Gossypium.hirsutum*, it was found that there were differences in the expression patterns of this gene family, indicating that members of this gene family may play different biological functions and participate in different biological processes. Analysis of the expression patterns of *KCR* *Gossypium.hirsutum* gene in different ovule and different tissues showed that *GH_A03G1517*, *GH_D02G1694* and *GH_D02G1695* were almost not

expressed in *Gossypium.hirsutum* expression patterns. Four genes *GH_A03G1518*, *GH_A03G0544*, *GH_D03G1431* and *GH_D03G1424* with significant expression differences were selected for fluorescence quantitative verification, and the results were consistent with the previous transcriptome data. It is speculated that these four genes play an important role in the process of oil accumulation.

5. Conclusion

The functional analysis of *KCR* gene in many plants has been reported, but there are few studies on *KCR* gene family analysis, and there is no identification and functional analysis of *KCR* gene family in cotton. In this study, a total of 24 *KCR* family members were identified from 4 *Gossypium. arboreum*, 4 *Gossypium. raimondii*, 8 *Gossypium. barbadense* and 8 *Gossypium. hirsutum*, which were divided into 3 clades through phylogenetic analysis, among which 3 different clades had similar conserved motifs. KaKs analysis showed that *KCR* gene family evolution was relatively conservative. *GH_A03G1518*, *GH_A03G0544*, *GH_D03G1431*, *GH_D03G1424*, which were significantly different between high oil materials and low oil materials, were screened through the RNA-Seq transcriptome data of high oil materials and low oil materials. RT-qPCR further confirmed that the differences of the above four genes were still significant. These results laid a foundation for further research on the biological functional characteristics of *KCR* gene in cotton.

Acknowledgments

This work was supported in part by a grant from National Key R&D Program (2023YFD2301201&2023YFD2301200).

References

- [1] ZIA M A, SHAH S H, SHOUKAT S, et al. Physicochemical features, functional characteristics, and health benefits of cottonseed oil: a review [J].
- [2] XU Y, SINGER S D, CHEN G. Protein interactomes for plant lipid biosynthesis and their biotechnological applications [J]. *Plant Biotechnol J*, 2023, 21(9) : 1734-44.
- [3] FRÉDÉRIC, BEAUDOIN, XIANZHONG, et al. Functional characterization of the Arabidopsis beta-ketoacyl-coenzyme A reductase candidates of the fatty acid elongase [J]. *Plant Physiology*, 2009.
- [4] DIETRICH C R, PERERA M A D N, YANDEAUKMELSON M D, et al. Characterization of two GL8 paralogs reveals that the 3-ketoacyl reductase component of fatty acid elongase is essential for maize (*Zea mays* L.) development [J]. *Plant Journal*, 2010, 42(6).
- [5] Quan YU, Zeng Quan, Huang Zhenxin, et al. Function of β -ketoacyl-CoA reductase gene *FsKCR1* in *Fusarium sugarcane* [J]. *Genomics and Applied Biology*, 2023, (012) : 042.
- [6] YONGMEIQIN, FRANCOISMAPUJOL, YONGHUISHI, et al. Cloning and functional characterization of two cDNAs encoding NADPH-dependent 3-ketoacyl-CoA reductase from developing cotton fibers [J]. *Cell research*: 2005.
- [7] BEAUDOIN F, GABLE K, SAYANOVA O, et al. A *Saccharomyces cerevisiae* Gene Required for Heterologous Fatty Acid Elongase Activity Encodes a Microsomal β -Keto-reductase [J]. *Journal of Biological Chemistry*, 2002, 277(13) : 11481.
- [8] YANG Z, WANG J, HUANG Y, et al. CottonMD: a multi-omics database for cotton biological study [J]. *Nucleic Acids Res*, 2023, 51(D1) : D1446-d56.
- [9] BALAKRISHNAN S, SHENGHAN G, LERCHER M J, et al. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees [J]. *Nucl Acids Research*, 2019, (W1) : W1.
- [10] CHEN C, WU Y, LI J, et al. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining [J]. *Mol Plant*, 2023, 16(11) : 1733-42.
- [11] SONG J, PEI W, WANG N, et al. Transcriptome analysis and identification of genes associated with oil accumulation in upland cotton [J]. *Physiologia plantarum*, 2022, 174(3) : e13701.