

RESEARCH

Open Access



Identification and functional characterization of bidirectional gene pairs and their intergenic regions in cotton

Jiangtao Yang^{1†}, Lihua Gao^{1†}, Xiaochun Zhang^{1†}, Ran Zheng¹, Xuan Liu¹, Yuxin Cui¹, Zhixing Wang^{1*} and Xujing Wang^{1*}

Abstract

Background In research to improve the quality of transgenic crops, it is often necessary to introduce multiple functionally related genes into recipient plants simultaneously to improve crop genetic traits effectively. Compared with unidirectional promoters, bidirectional promoters simultaneously regulate the expression of multiple genes and improve the efficiency of biotechnology. Therefore, in this study, bidirectional gene pairs were systematically analyzed in *Gossypium hirsutum* TM-1, and the structure, function and evolutionary relationships of the bidirectional genes were analyzed. The endogenous bidirectional promoters of cotton were mined, and their specific regulatory elements and biological functions were explored to provide useful promoter resources and a theoretical basis for cultivating new cotton germplasms with excellent fiber quality.

Results Using an improved search model, a total of 1,383 bidirectional transcript pairs were identified in the *Gossypium hirsutum* TM-1 genome, and their gene structure and functional annotations were systematically analyzed. Thirty bidirectional intergenic sequences were randomly screened for promoter activity analysis via a transient expression system, and 25 intergenic sequences were found to have bidirectional promoter activity. Comparative analysis of the bidirectional gene profiles of the four cotton subspecies revealed that these subspecies presented abundant bidirectional gene pairs with high homology and that the bidirectional genes in the cotton subspecies were more similar in terms of their molecular functions, cellular components and biological processes. In addition, parallel analysis of bidirectional genes in dicotyledons and monocotyledons revealed that abundant bidirectional gene pairs exist in different species. Although the total number of orthologous bidirectional genes was similar, there was a significant difference in the number of orthologous bidirectional gene pairs between dicotyledons and monocotyledons. This evolutionary analysis of the function and structure of homologous bidirectional gene pairs

[†]Jiangtao Yang, Lihua Gao and Xiaochun Zhang contributed equally to this work.

*Correspondence:
Zhixing Wang
wangcotton@126.com
Xujing Wang
wangxujing@caas.cn

Full list of author information is available at the end of the article



in different varieties and different subspecies of the same species revealed potential pathways by which these gene pairs originated, which may be necessary for the evolution of a new species.

Conclusion In this study, many bidirectional gene pairs in *Gossypium hirsutum* TM-1 were identified using computer programming, and systematic analysis was conducted to explore their functions and evolutionary relationships. In addition, the promoter activity of the bidirectional intergenic sequences was verified. The combination of computer programming screening, experimental validation and other methods is expected to provide preferred bidirectional promoters for transgenic breeding work via multigene cotransformation methods, and this information is valuable for genetic engineering research and applications.

Keywords *Gossypium hirsutum*, Bidirectional gene pairs, Bidirectional promoter, Gene function, Genome-wide

Introduction

With the rapid development of genome sequencing technology, this technique has rapidly become widely used in scientific research. By the end of 2020, whole-genome sequencing of 1,031 genomes of 788 different plant species was completed [1]. An increasing number of studies have revealed many types of genome structures and gene expression patterns, such as intron splicing, gene methylation, spatiotemporal gene expression and gene coexpression [2–5]. In recent years, the study of bidirectional gene pairs and their bidirectional promoters has attracted increasing attention from researchers. A bidirectional gene pair refers to two adjacent genes in a “head-to-head” arrangement with opposite transcriptional directions. Gene structure is a special gene arrangement in which a shared intergene sequence region regulates the transcription of both genes. The region between the transcription start site (TSS) of two genes is called the “bidirectional promoter”. Because of its ability to drive the expression of two genes simultaneously, the bidirectional promoter has great application potential for research on gene engineering and metabolic engineering related to multigene expression, so it has attracted increasing attention from researchers.

Bidirectional promoters are a class of structurally special promoters that can drive the expression of genes at both ends symmetrically or asymmetrically. Researchers hypothesize that the bidirectional transcriptional activity mechanism of the bidirectional promoter can be attributed mainly to two RNA polymerases simultaneously aggregating at the boundary of the nucleosome region and initiating transcription from two directions [6]. Specific sequence signals, chromatin modifications and three-dimensional structures of transcription sites facilitate unconventional yet tightly regulated transcription proceeding in both directions from bidirectional promoters [7]. There is no strict specific standard for the length of bidirectional promoters. There are significant differences in the lengths of bidirectional promoters among different species, with short bidirectional promoters reaching approximately 100 bp and long promoters reaching more than 2,000 bp. In animals, bidirectional

promoters generally exhibit a sequence length of 1,000 bp or less [8]. For example, in the human genome, more than 10% of genes are arranged in a “head-to-head” manner, the sequence length between bidirectional gene pairs is less than 1 kb, and these bidirectional gene pairs are driven by 1,352 bidirectional promoters [9]. In plants, many bidirectional promoters have sequences longer than 1,000 bp, extending up to 1,500 bp [10]. For example, in the *Arabidopsis thaliana* genome, many (13.3%) bidirectional gene pairs have been found, and many intergenic regions are 1,000 bp to 1,500 bp in length among these bidirectional gene pairs [11]. Some researchers believe that there are many *cis*-acting elements, such as expression regulatory elements and inducible expression elements related to tissue-specific expression in the bidirectional promoter sequences of plants; therefore, bidirectional promoters require longer sequences to meet the expression needs of the plant itself [12].

Since the discovery of the first bidirectional promoter in the chloroplast genome of maize and experimental verification, researchers have successively discovered bidirectional promoters in *Capsicum annuum*, *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, *Populus* and *Zea mays* [11, 13–15] and have conducted detailed studies on their functions. For example, the expression of anthocyanin biosynthesis-related genes (*ZmBz1*, *ZmBz2*, *ZmC1* and *ZmR2*) was driven by the maize embryo-specific bidirectional promoter PZmBD1 in maize seeds, and the first anthocyanin-rich purple maize variety was obtained [16]. An asymmetrically expressed bidirectional promoter (*At5g10300* and *At5g10290*) was screened in *Arabidopsis thaliana*, in which the expression of the bidirectional promoter was induced when the plant was subjected to pest bite injury. This bidirectional promoter can express multigene traits in a synergistic manner and can be used in plant biotechnology for pest control applications that require the stacking of defense genes [17]. The bidirectional promoter RPBSA was used to drive the expression of short RNAs in the Sleeping Beauty system to alter the original cell phenotype [18]. Using natural bidirectional promoters, regulatory elements are added or modified to synthesize green tissue-specific bidirectional promoters,

which exhibit strong promoter activity in both directions and enormous potential for utilization [19]. A high-astaxanthin maize germplasm was created using a maize seed-specific bidirectional promoter to drive the β -carotene hydroxylase- and ketoenzyme-encoding genes to extend the carotenoid synthesis pathway to the astaxanthin synthesis pathway in maize grains [20].

For transgenic improvement, it is occasionally necessary to transfer multiple exogenous genes into specific organisms, and these exogenous genes need to be expressed under the regulation of their respective promoter sequences [21]. Owing to the limited number of promoters used for genetic improvement, a specific promoter or a promoter with a similar sequence is often repeatedly used when multiple genes are transferred into an organism. Although this approach can achieve the goal of simultaneously transferring multiple genes, the construction of this type of vector is time-consuming and labor-intensive [22]. However, bidirectional promoters are able to drive gene expression in both directions, allowing the fusion of two or more genes at both ends when bidirectional promoters are used for transgenic purposes. This not only avoids the tedious steps of introducing different genes but also reduces the number of promoters introduced into the receiving organism and avoids concerns about the occurrence of transgenic silencing caused by the sequence homology between promoters introduced into the organism. Therefore, the mining of plant endogenous bidirectional promoters is a key goal in the application of plant endogenous bidirectional promoters for research on improving the variety of transgenic plants.

In this study, bidirectional gene pairs and their bidirectional promoters in *Gossypium hirsutum* TM-1 were systematically analyzed and verified using bioinformatics, qRT-PCR, and a plant transient expression system. The phylogenetic relationships of the bidirectional gene pairs and bidirectional promoters in three other cotton cultivars (*Gossypium raimondii*, *Gossypium herbaceum* and *Gossypium barbadense*) and two other plant species (*Arabidopsis thaliana* and *Oryza sativa*) were subsequently analyzed, and the results revealed high similarities between the different varieties of cotton and different plant species. In addition, evolutionary analysis of the functions and structures of orthologous bidirectional gene pairs in different plant species revealed their origins and potential evolutionary pathways. Mining the endogenous bidirectional promoters of cotton and exploring their specific regulatory elements and biological functions are helpful for rationally using transgenic technology to improve cotton fiber quality and provide rich promoter resources and a theoretical basis for cultivating new cotton germplasms with excellent fiber quality.

Results

Genome-wide screening of *Gossypium hirsutum* bidirectional gene pairs

The search model used in this study included a screening procedure for different transcripts of the same gene at both ends (Fig. 1C) on the basis of the search model of Liu (Fig. 1B) and screened bidirectional transcript pairs in the *Gossypium hirsutum* TM-1 genome. We wrote the screening procedure on the basis of a search pattern in which two adjacent genes may encode multiple transcripts. The screening parameters were set as follows: the distance between the transcription initiation sites of the bidirectional transcript pairs was $\leq 1,500$ bp; transcripts with short translated proteins or discontinuously translated proteins and transcripts that could not be translated into complete proteins were deleted; and the 5'-UTR sequence did not exceed 300 bp. Therefore, 1,383 bidirectional transcript pairs were screened from 76,943 transcripts in the whole genome of *Gossypium hirsutum* TM-1 (Supplementary Table 1), and the nucleotide sequences and encoded amino acid sequence information of these bidirectional transcripts were extracted.

Chromosomal distribution and gene structure analysis of bidirectional gene pairs in *Gossypium hirsutum*

Using MG2C analysis for the distribution of bidirectional transcript pairs, there were no significant differences in the At and Dt subgenomes of *Gossypium hirsutum*, which presented 522 and 514 bidirectional transcript pairs, respectively. However, 322 bidirectional transcript pairs were distributed on scaffolds, and the greatest numbers of these transcript pairs were distributed on At_chr7, Dt_chr9 and At_chr9 among the 26 chromosomes of *Gossypium hirsutum* (Fig. 2). Among these genes, 1,034 bidirectional genes could be linked to other genes in the cotton genome, and the associated genes were mainly unidirectional genes. Moreover, bidirectional gene pairs were distributed mainly in regions with high chromosome gene density, and the GC content in these regions was relatively high. The number of bidirectional gene pairs distributed on the chromosome was independent of the length of the chromosome, and they were densely distributed at specific locations on the chromosome (Fig. 3).

The gene structure was summarized and analyzed via bidirectional gene pair annotation information in the genomic data for *Gossypium hirsutum* TM-1. The results revealed that the length of each gene in the bidirectional gene pairs, the numbers of introns and exons, their lengths, and their distribution characteristics were random. No similarity in gene structure of the bidirectional gene pairs was observed (partial results Fig. 4, Supplementary Fig. 1).

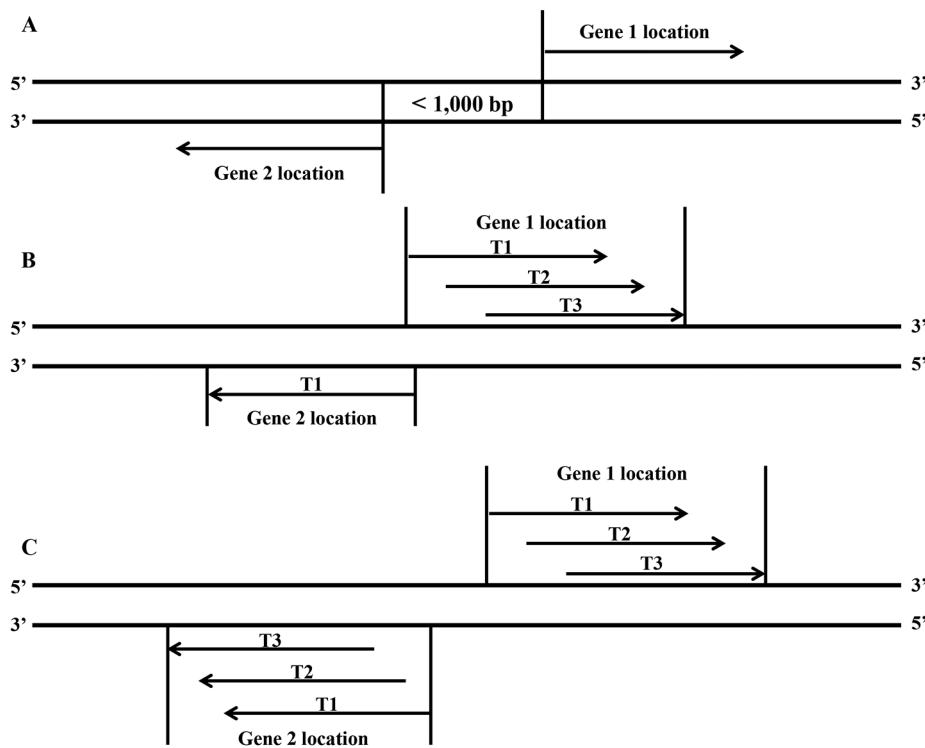


Fig. 1 Model of bidirectional gene pair screening in the whole genome. **A:** Bidirectional gene pair search pattern created by Trinklein. The gene location coordinates do not specify the location coordinates of different transcripts. **B:** The bidirectional transcript pair search model established by Liu. The search model uses the location coordinates of transcripts rather than unique gene location coordinates. For example, gene 1 has three transcripts, and the intergenic region between the three transcripts (T1, T2, T3) and the transcript of gene 2 can be considered a bidirectional promoter. The bidirectional promoter between T2-gene2 and T3-gene2 was not detected via Trinklein's search model because the transcriptional start site of the first transcript (T1) of gene 1 was considered the 5' end boundary of gene 1 overlapping with gene 2. **C:** The bidirectional transcript pair search model established in this study. The search model of this study is based on the Liu model, which involves adding the location coordinates of multiple transcripts of gene 2. For example, genes 1 and 2 have three transcripts, and the intergenic region between the three transcripts of gene 1 (T1, T2, T3) and the three transcripts of gene 2 (T1, T2, T3) can be considered a bidirectional promoter, making the search range more comprehensive

Gene expression profiling and functional annotation analysis of bidirectional gene pairs in *Gossypium hirsutum*

In a combined analysis with transcriptome data from seven different tissues (root, leaf, anther, stigma and fiber developmental stages (7 days post-anthesis (DPA), 14 DPA, and 26 DPA)) constructed in our laboratory [23], the intertissue expression levels of these bidirectional gene pairs were analyzed. In this study, tissue expression data for 1,891 bidirectional genes were obtained from 2,766 (1,383 pairs) bidirectional genes, among which 752 pairs had available tissue expression information in the transcriptome data. According to the trend of the expression level of each gene in seven different tissues of cotton, it was found that most bidirectional gene pairs (688, 91.5%) presented different expression patterns for each gene in different tissues, and only a small number of bidirectional gene pairs (64, 8.5%) presented the same expression patterns for each gene in different tissues. Therefore, the expression patterns of each gene in the bidirectional gene pairs were not interdependent (part of the results are shown in Fig. 5). To verify the reliability

of the transcriptome data, 10 pairs of bidirectional genes with significantly different expression levels among different tissues were selected for qRT-PCR. The results revealed that the selected genes presented significant differences in expression in different tissues of cotton, and the expression trend was consistent with the RNA-Seq results (Fig. 6).

Previous studies have shown that bidirectional transcript pairs may be functionally associated in vertebrates [9] and plants [11]. To clarify the biological function of bidirectional transcript pairs in the genome of *Gossypium hirsutum*, we conducted functional annotation analysis of the obtained bidirectional transcript pairs using the bioinformatics software blast2GO. The general analysis was performed on the basis of the three functional categories of cellular components, molecular functions and biological processes analyzed by GO enrichment. The advantage of blast2GO is that the functional information on the GO direct acyclic graph (DAG) was visualized and displayed in a combined graph. The color shading indicates the enrichment level of the bidirectional genes

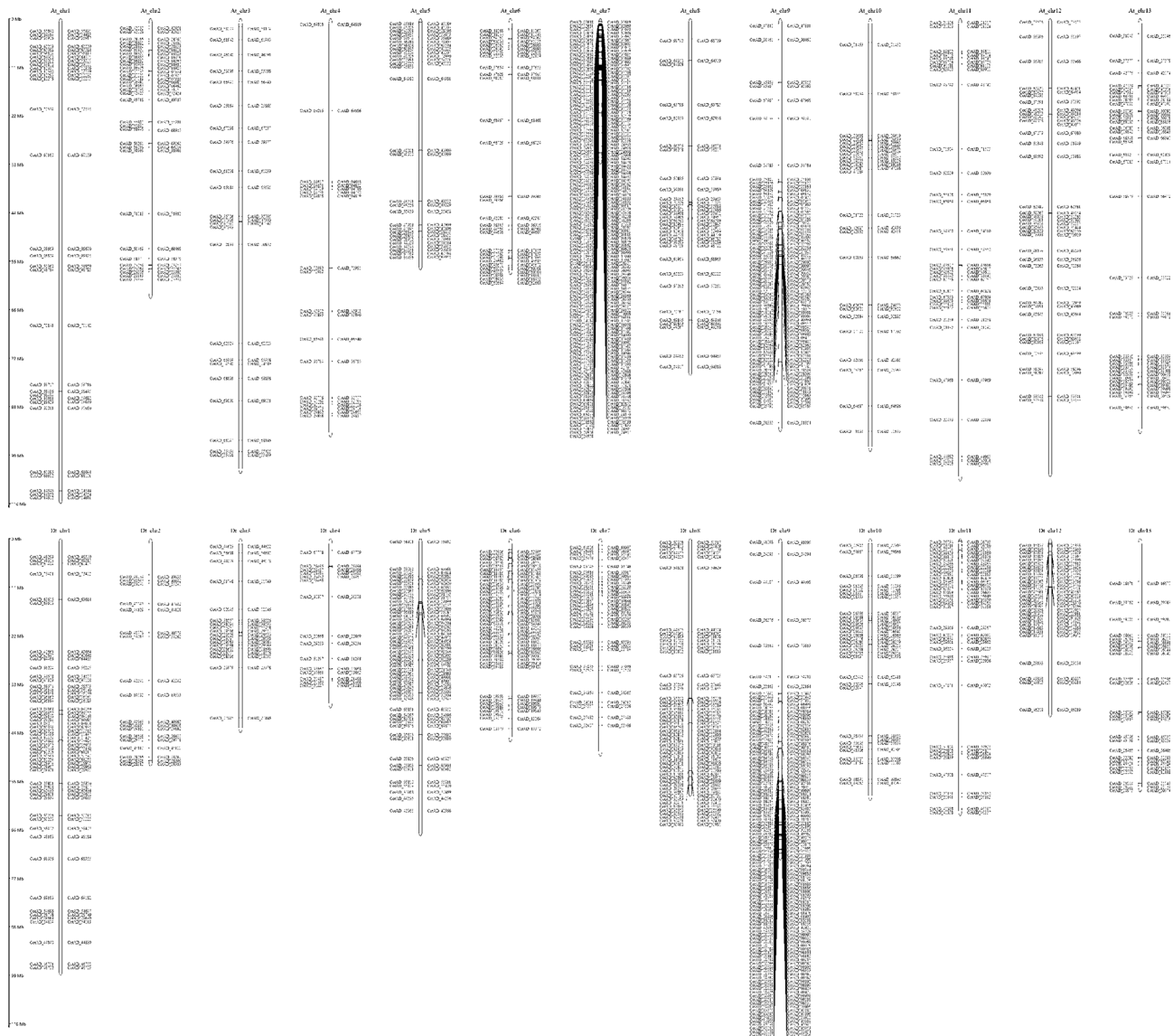


Fig. 2 Positional distribution of bidirectional gene pairs on the *Gossypium hirsutum* TM-1 chromosome. The left ruler is the chromosome length ruler, and the genes on the left and right sides of the chromosome at the same position represent a pair of bidirectional genes. The number of bidirectional gene pairs is not positively correlated with chromosome length, and the amount of gene pairs distributed on the chromosome is not related to the chromosome length

within the GO term, and the colored nodes indicate the places of direct annotation. To obtain a compact representation of bidirectional transcript pair information, we set the sequence filter to 10%, which means that only those nodes with at least 10% of the total sequence assignments are displayed, and the score filter is set to 100. Consequently, the number of annotated sequences does not exceed the parent nodes of their children, which will be omitted from the graph. These parameters ultimately led to the most concentrated DAG with detailed sequence annotations (Fig. 7B, D and E). This graph can provide more information about the most concentrated function of bidirectional transcript pairs. However, to

provide an optimal view of the dataset’s most relevant terms, the multilevel pie method was employed to ‘cut’ the GO DAG locally at different levels, and only the lowest number of GO terms per branch were displayed by setting the same sequence filter to 10%. Regarding cellular components, the expression products of bidirectional genes were mainly concentrated in cells and cell parts, intracellular and intracellular parts, organelles and membrane-bound organelles (Fig. 7A/B). Regarding molecular function, the main functions of these bidirectional gene products were related to catalytic activity, binding molecules, transferase activity, hydrolase activity and DNA binding (Fig. 7C/D). Regarding biological processes, the

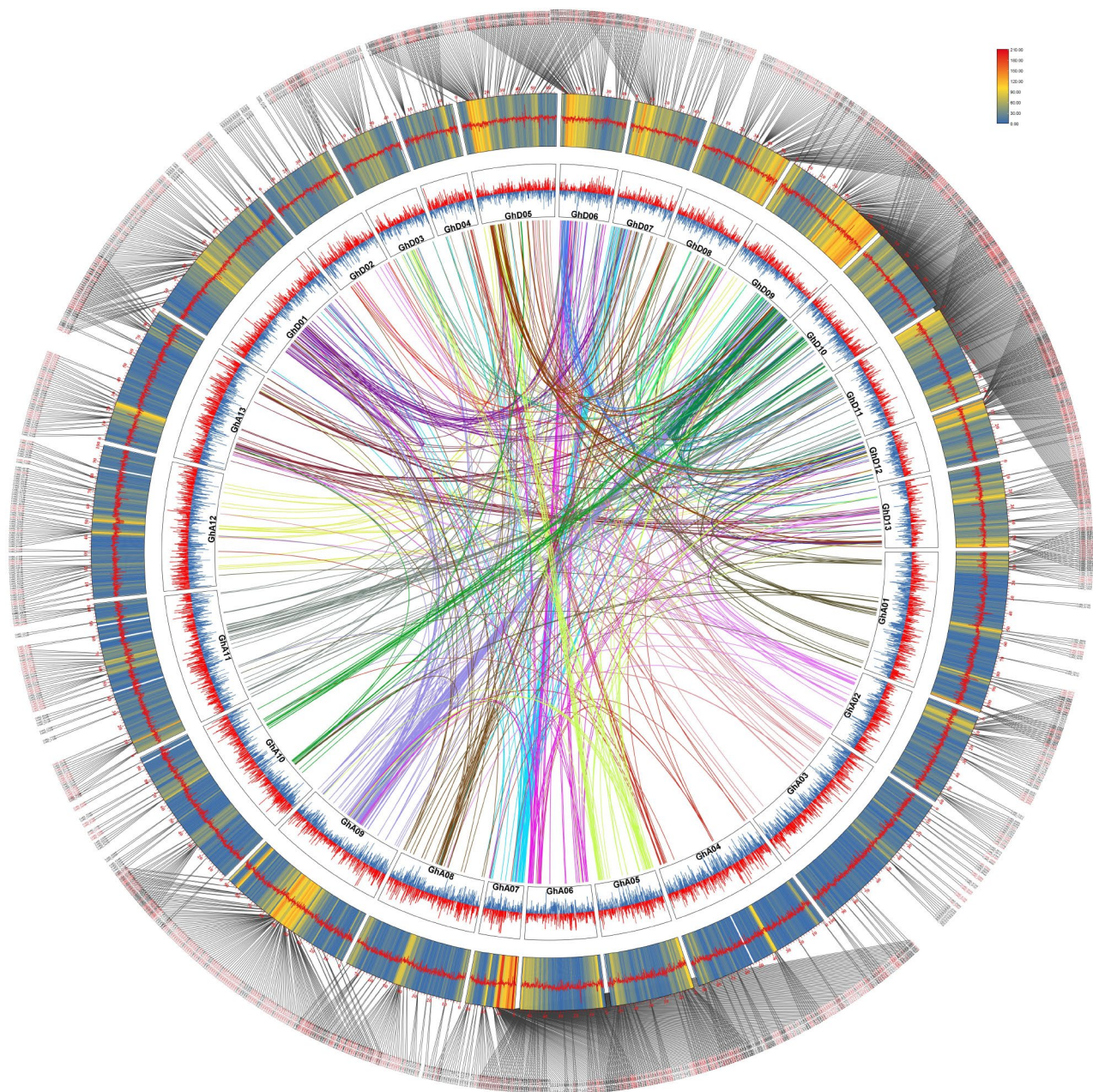


Fig. 3 Association analysis of bidirectional gene pairs in the cotton genome. In the circle, lines of the same color connect two genes with an association relationship. The inner circle is the GC skew of the chromosome sequence. The red polyline indicates the leading strand, which is a positive value; the blue polyline indicates the lagging strand, which is a negative value. The heatmap of the outer circle indicates the distribution of gene density. The number of genes per 100 Mb, with blue–yellow–red indicating that the number of genes increased. The broken line represents the GC content per 10 Mb of sequence length. The outermost gene is the distribution position of the gene in the bidirectional gene pair in the chromosome, the gene in red is the gene with an association in the genome, and the gene in black is the gene without an association in the genome

expression products of these bidirectional genes were found to be involved in a variety of metabolic processes, cellular processes and biosynthetic processes (Fig. 7E/F). The GO enrichment analysis results revealed that the associations of bidirectional transcript pairs with functional classes were not random. These genes are associated with a relatively limited set of cellular components,

molecular functions, and biological processes, suggesting that they play important roles in cotton.

Sequence characteristics of bidirectional promoters and analysis of *cis*-acting elements

Bidirectional promoters refer to the regional sequences between two adjacent gene transcription start sites in a “head-to-head” orientation with opposite transcriptional

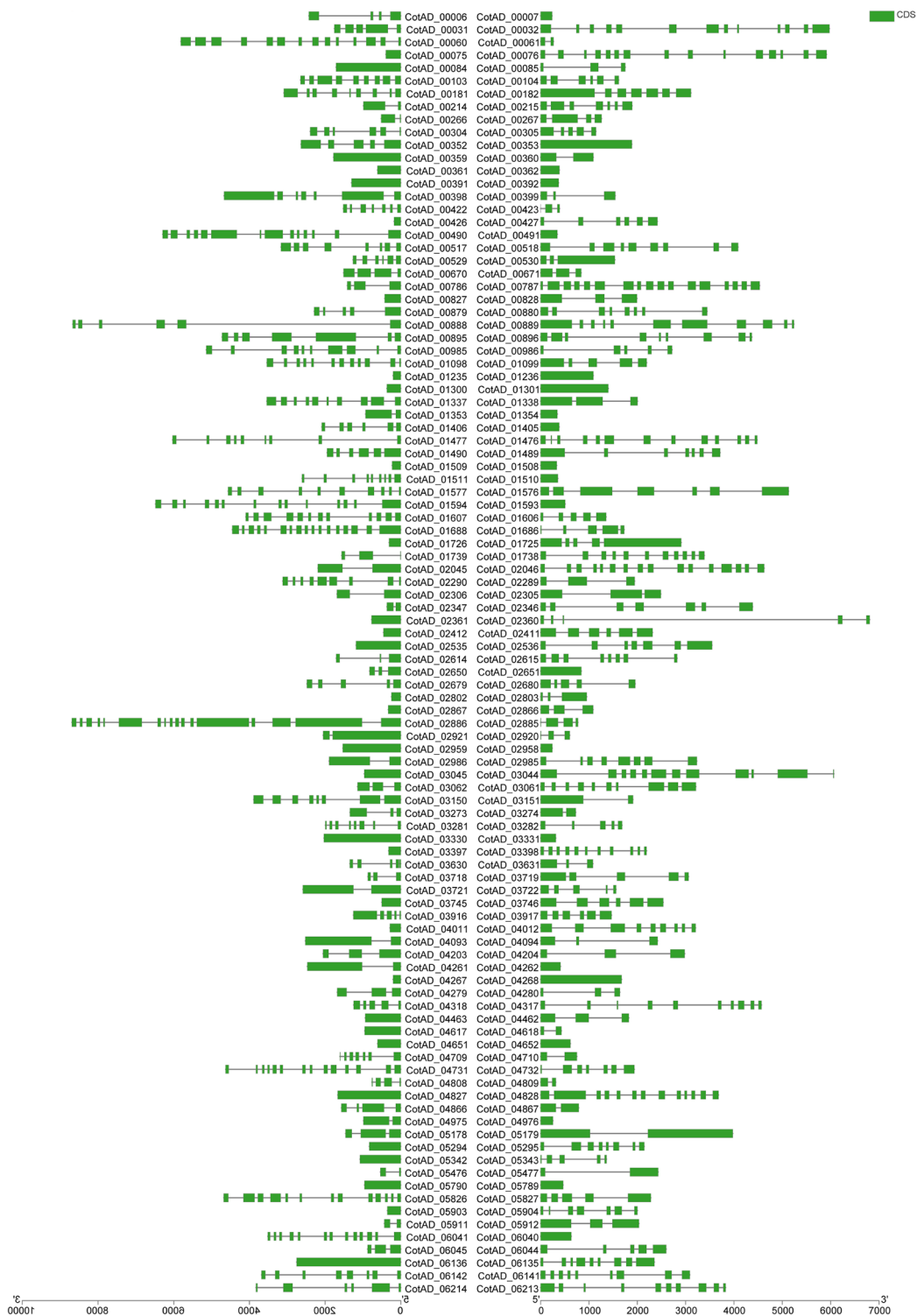


Fig. 4 Gene structure analysis of bidirectional gene pairs. The green column represents the exons, and the black solid line represents the introns. The length of each gene in the bidirectional gene pairs was not related, and the number of introns and exons in the bidirectional genes was not related

directions. One of the distinguishing features of bidirectional promoters is their high GC content. To verify whether the bidirectional promoter sequences in the genome of *Gossypium hirsutum* have this characteristic,

we analyzed the sequence characteristics of all the bidirectional promoters and 1,000 random unidirectional promoters in *Gossypium hirsutum* using a computer bioinformatics program (Python v3.9 and PlantCARE). The

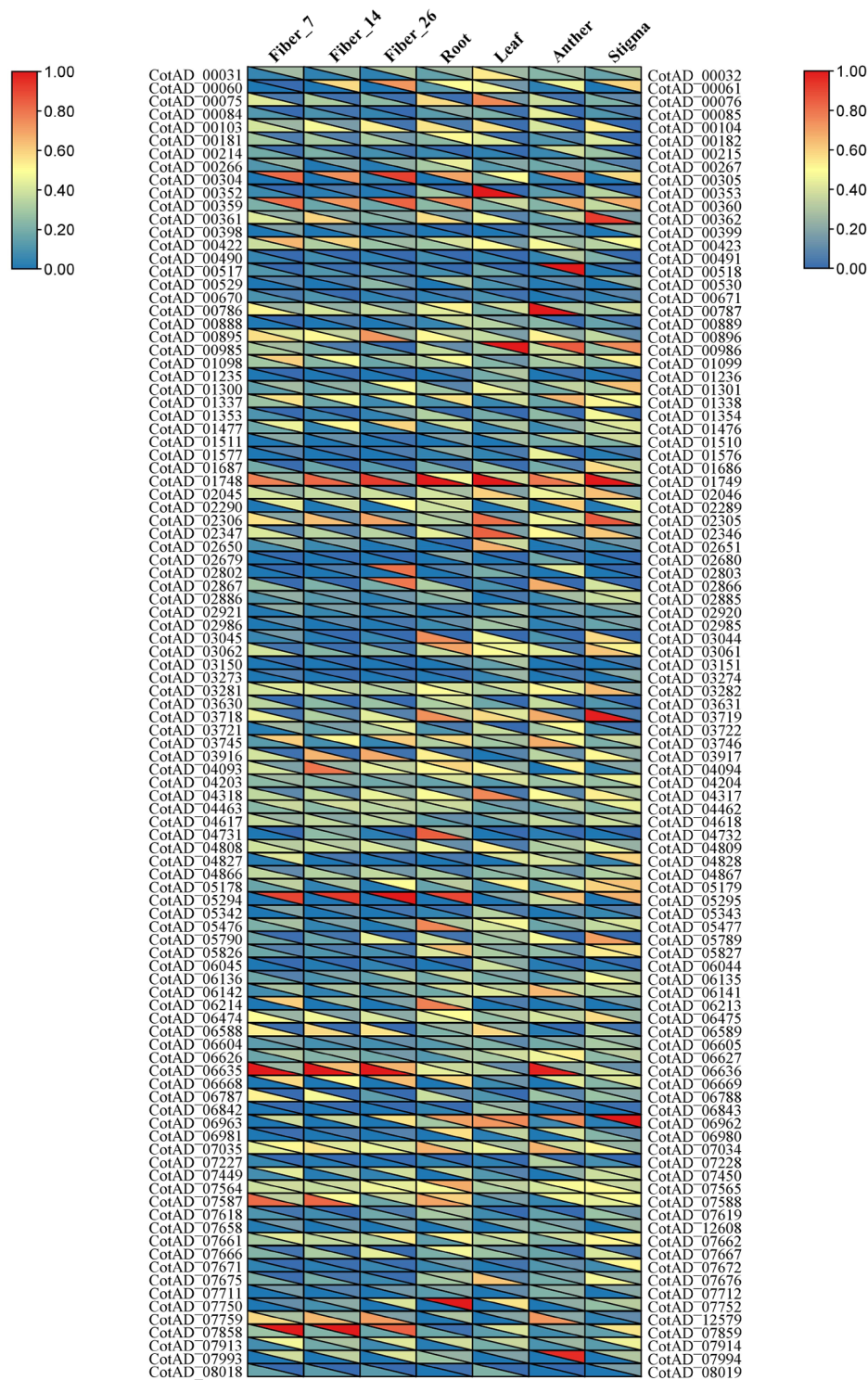


Fig. 5 Analysis of the tissue expression profiles of bidirectional gene pairs. The bottom left triangle represents the expression of the left genes in the bidirectional gene pairs, and the top right triangle represents the expression of the right gene in the bidirectional gene pairs. Fiber_7: 7 DPA fiber, Fiber_14: 14 DPA fiber, Fiber_26: 26 DPA fiber

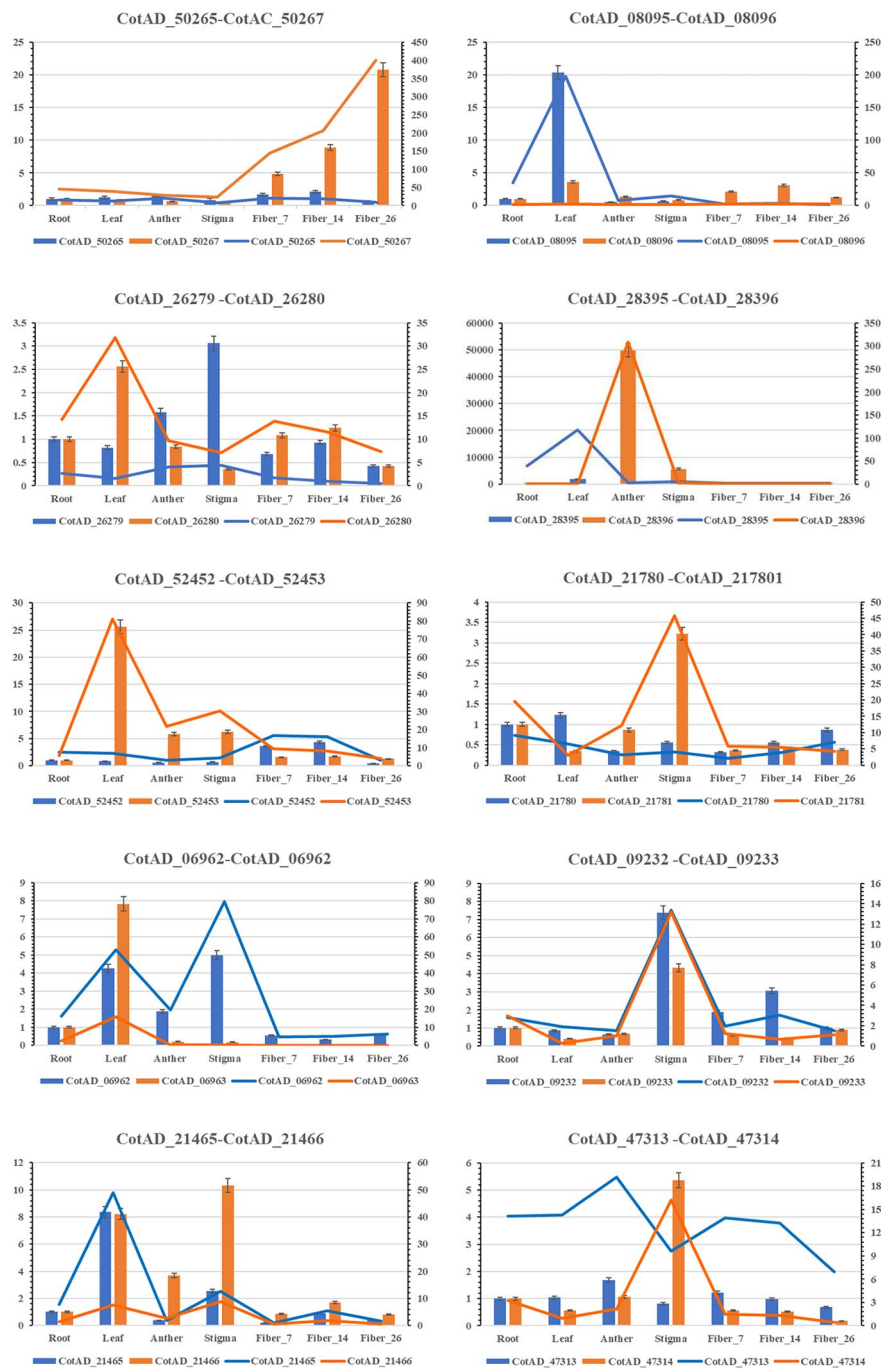


Fig. 6 qRT-PCR and RNA-seq analysis results of bidirectional gene pairs. The bars represent the qRT-PCR results, and the lines represent the RNA-seq results. The left vertical coordinate indicates the relative expression level, and the right vertical coordinate represents the FPKM value

length of the random unidirectional promoter sequences extracted was the average length of all bidirectional promoter sequences, which was 880 bp. The results revealed that the GC content in these bidirectional promoter sequences was as high as 35.05%, whereas the GC content in the randomly screened unidirectional promoters was 31.3% (Table 1). The core components of promoters, such as TATA-boxes, GC-boxes, and CAAT-boxes, were

subsequently analyzed, revealing that 6.58% of the bidirectional promoters had TATA-box elements, whereas 11% of the unidirectional promoters that were randomly screened had TATA-box elements. There was no symmetry in the distribution of the *cis*-acting elements in the bidirectional promoter sequence, and the distribution of the *cis*-acting elements in each bidirectional promoter was independent and had no regular arrangement

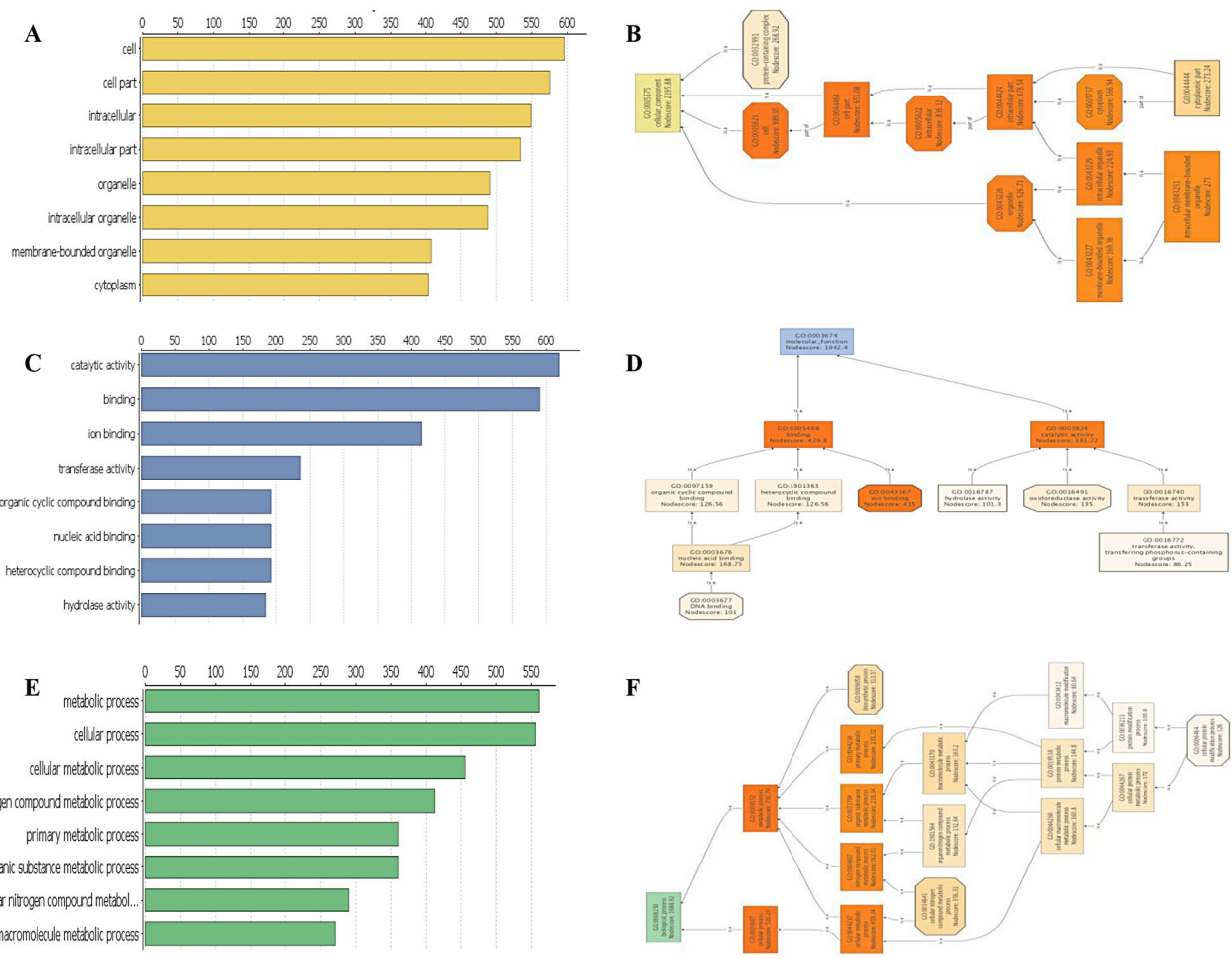


Fig. 7 GO enrichment analysis of the bidirectional transcript pairs of *Gossypium hirsutum*. **A and B, C and D, E and F:** Histogram charts and directed acyclic graphs of cellular components, molecular functions, and biological processes. A combo chart was generated by filtering annotation sequences together with a node sequence filter (more than 10% of the total number of sequences) and a score filter (> 100). The parental nodes that had more annotated sequences compared with their children were omitted from the diagram. A multilevel pie chart was derived from a combination chart that shows only the lowest GO item for each branch. Nodes colored by score values highlight the areas with the most concentrated annotations. The color shading indicates the enrichment level of the bidirectional genes in the GO term. The darker the color is, the more significant the enrichment. Red represents the most significant enrichment, followed by yellow, and no color indicates nonsignificant enrichment. The first line inside the box represents the number of GO terms

(Table 1; Fig. 8 and Supplementary Fig. 2). The lengths of the bidirectional promoters were subsequently colated and counted, and they were found to be distributed mainly in the range of 1,200 ~ 1,500 bp (Fig. 9).

Cloning and functional verification of bidirectional promoters

To verify whether the bidirectional promoters screened according to the Python programming search pattern presented promoter activity at both ends, 30 bidirectional promoters were randomly selected from bidirectional intergenic sequence intervals of different lengths for promoter activity functional analysis, and combined with transcriptome data from different tissues, tissue expression pattern data for these candidate bidirectional genes were obtained (Supplementary Table 2). Using the

double reporter (*gus* and *gfp*) pBD2G vector preserved for many years in our laboratory as the vector skeleton, a plant expression vector containing a bidirectional promoter sequence was constructed and transformed into cotton for verification using a transient expression system. The results revealed that 25 bidirectional gene intergenic sequences could drive the expression of the *gus* and *gfp* reporter genes in opposite directions, indicating that these intergenic sequences had promoter activity in both the positive and negative directions. These findings revealed that these bidirectional gene intergenic sequences are bidirectional promoters. The intergenic sequences of the other five bidirectional genes had promoter activity in a single direction. Specifically, for two of these, promoter activity was only detected in the GUS direction, and promoter activity in the GFP direction was

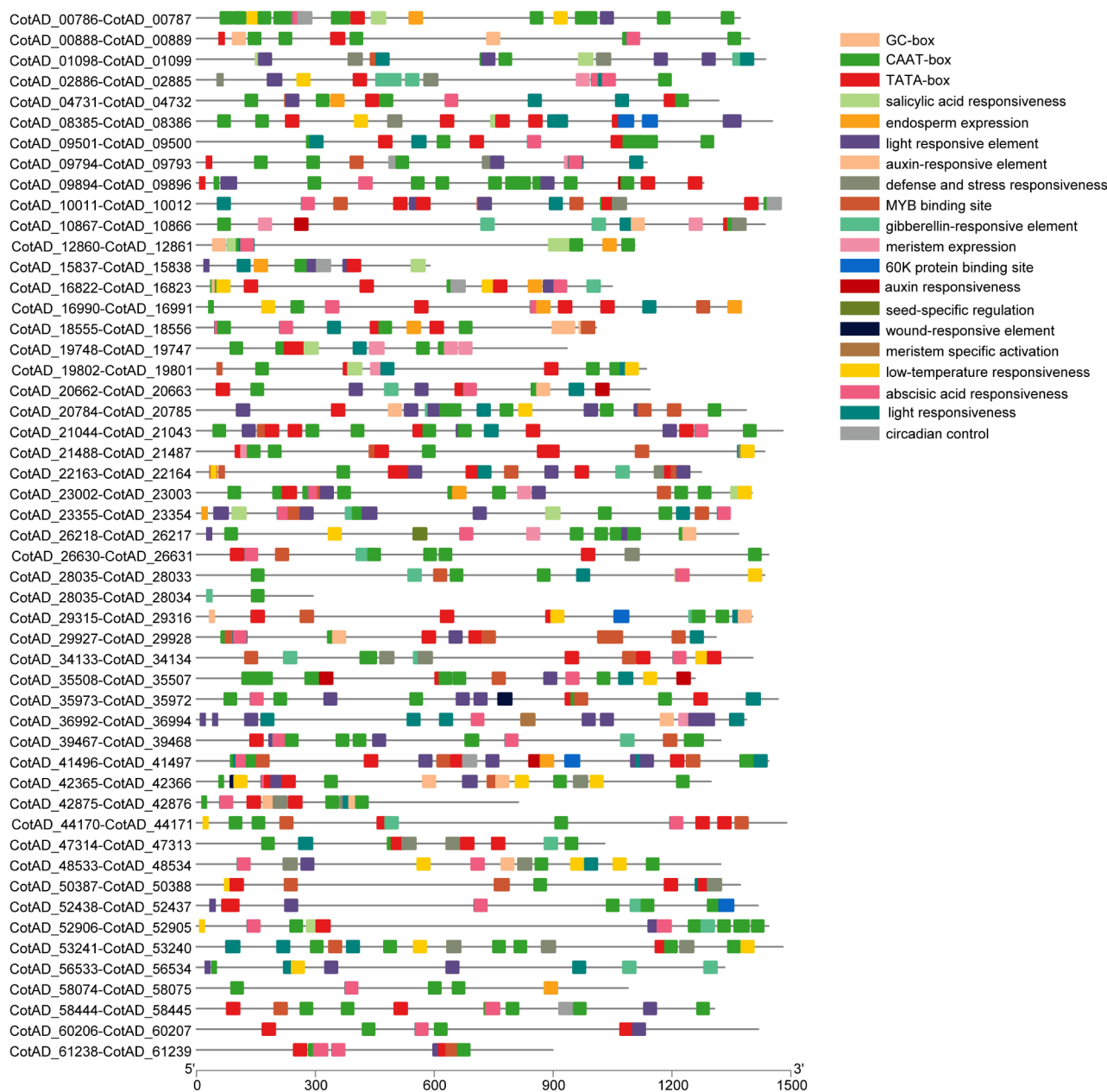


Fig. 8 Analysis of the distribution characteristics of *cis*-acting elements in partial bidirectional promoters. With the use of PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) online analysis software, the types, quantities, locations and other information of the *cis*-acting elements of the bidirectional promoter sequences were predicted. TBtools software was used to construct the graphs, and each square represents a *cis*-acting element. Different *cis*-acting elements are distinguished by different colors. The important *cis*-acting elements are listed in the graph so that the distribution of *cis*-acting elements in the promoter can be clearly seen

Table 1 Statistics of the GC content and core elements in bidirectional promoters and random promoters

Type of promoter	GC content	TATA-box	GC-box	CAAT-box
Random unidirectional promoters	31.3%	11%	0.23%	0.32%
Bidirectional promoters	35.05%	6.58%	2.96%	3.83%

exclusively detected in the remaining three (Fig. 10). In summary, bidirectional promoters were identified in the regional sequence between the bidirectional gene pairs screened in this study, further demonstrating that the bidirectional gene pairs screened in this screening mode actually existed in the genome of *Gossypium hirsutum* and were expressed.

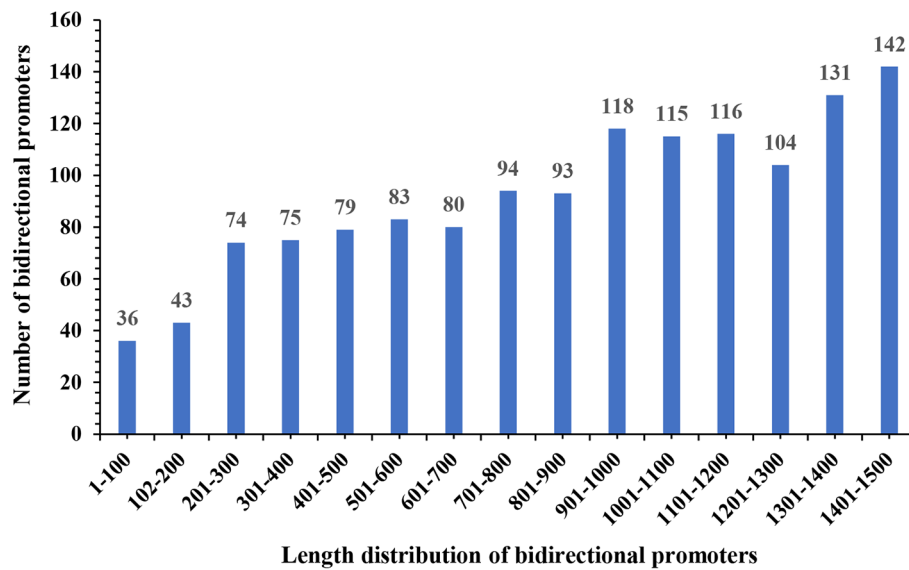


Fig. 9 Analysis of the length distribution of bidirectional promoter sequences. The number of bidirectional promoters within the range of 100 bp was counted. The abscissa shows the length distribution of the bidirectional promoters, and the ordinate shows the number of bidirectional promoters. The general trend was that the number of bidirectional promoters increased with increasing length

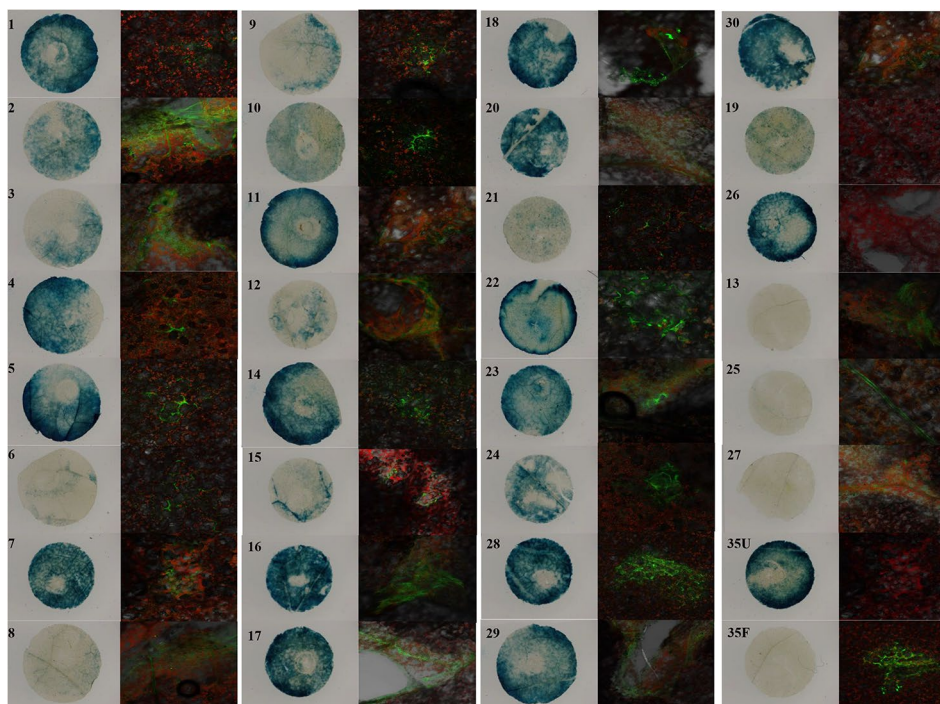


Fig. 10 Functional identification of bidirectional cotton promoters with dominant expression in cotton leaves. *Gossypium hirsutum* plants with 3–5 true leaves were selected. *Agrobacterium* was injected into the lower surfaces of the leaves, and the plants were cultivated for 48 h. Holes were punched in the infection site, and the diameter of the holes was 1.0 cm. Histochemical staining and green fluorescence were observed. The visible green fluorescence scale bar is 100 μm. The results revealed that the promoter activities of 19 and 26 could only be detected in the GUS direction; the promoter activities of 13, 25, and 27 could only be detected in the GFP direction; and the other genes presented promoter activity in both directions. 35U: 35 S-gus; 35 F: 35 S-gfp

Comparative analysis of bidirectional gene pairs and their bidirectional promoters in four cotton subspecies

This study revealed many bidirectional gene pairs and bidirectional promoters in the genome of *Gossypium hirsutum*; thus, to address the question of whether these special gene expression structure forms and their unique bidirectional promoters also widely exist in other subspecies of cotton, we screened bidirectional gene pairs and bidirectional promoters in the genomes of three cotton subspecies (*Gossypium raimondii*, *Gossypium herbaceum* and *Gossypium barbadense*) with available complete whole-genome sequences (Table 2). The results revealed many bidirectional gene pairs in the genomes of the four cotton subspecies, but the number of bidirectional gene pairs in the genome of *Gossypium hirsutum* was the greatest, reaching 1,383, followed by *Gossypium raimondii* and *Gossypium barbadense*, with 1,249 pairs and 1,091 pairs, respectively. The smallest number of bidirectional gene pairs was noted in the genome of *Gossypium herbaceum*, with 839 pairs. However, according to the ratio of bidirectional gene pairs among all genes in the genome, *Gossypium raimondii* had the highest ratio, accounting for 6.1% of all genes in this cotton subspecies. *Gossypium barbadense* had the lowest percentage, accounting for 2.7% of all genes. The above results indicated that many bidirectional gene pairs were present in the genomes of different subspecies of cotton, but the number of bidirectional gene pairs and their proportion relative to the number of all genes in the genome were not the same.

Functional annotation analysis of bidirectional genes in four cotton subspecies

To explore the relationships and differences in functional clustering between bidirectional gene pairs and bidirectional promoters in different cotton subspecies, this study conducted GO cluster and KEGG metabolic pathway analyses on all the bidirectional gene pairs in the genomes of *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium raimondii*. The results of the GO clustering analysis of the bidirectional genes in the four different cotton subspecies revealed that the most important molecular functions

of the bidirectional genes in all the cotton subspecies were the regulation of molecular binding, reaction catalytic activity, transfer activity and molecular functions (Fig. 11). Some bidirectional genes in *Gossypium herbaceum* and *Gossypium raimondii* also have antioxidant functions, but no bidirectional genes with this function were found in *Gossypium hirsutum* or *Gossypium barbadense*. Compared with those in the other three subspecies, some bidirectional genes in *Gossypium hirsutum* also presented increased transcription factor activity and protein binding. The greatest number of bidirectional gene products of the four cotton subspecies were related to cell and cell components, membrane and membrane components, organelles, and macromolecular complexes, organelles, and membrane-enclosed lumens. In terms of the clustering of biological processes, many differences exist between different subspecies of cotton. *Gossypium raimondii* has more bidirectional genes involved in the positive and negative regulation of biological processes and its own unique immune system processes. However, only bidirectional genes involved in the positive regulation of biological processes have been identified in *Gossypium hirsutum*. However, in contrast to *Gossypium hirsutum*, island cotton contains only bidirectional genes involved in negatively regulating biological processes. In general, more bidirectional genes in the four different subspecies of cotton were involved in biological processes such as metabolic processes, cell processes, single biological processes, localization, biological and process regulation, and cell component organization or biogenesis. By comparing the clustering degree of bidirectional genes in different subspecies of cotton in terms of cell composition, molecular function and biological process, we found that bidirectional gene enrichment in the two diploid cotton subspecies (*Gossypium herbaceum* and *Gossypium raimondii*) was relatively consistent, whereas *Gossypium hirsutum* and island cotton presented some differences compared with other subspecies but generally presented consistent patterns. This finding revealed that bidirectional gene pairs maintained a high degree of conservation during the evolution of different subspecies of cotton and played important roles in the growth and development of cotton.

We subsequently conducted a cluster analysis of the KEGG metabolic pathways of bidirectional gene pairs in different subspecies of cotton. The results revealed that bidirectional genes in different subspecies of cotton were relatively concentrated in metabolic pathways such as protein translation and processing, signal transduction, carbohydrate metabolism, transportation and catabolism, energy metabolism, environmental adaptation, cell growth and death and other metabolic pathways, but slight differences were noted among the different subspecies (Fig. 12). This finding revealed that bidirectional

Table 2 Search for bidirectional gene pairs in the genomes of four cotton subspecies

Species	Genome size (Mb)	Total genes	Total BDGPs	Total BDGs/total genes
<i>Gossypium hirsutum</i>	2,430	76,943	1,383	3.6%
<i>Gossypium barbadense</i>	2,570	80,876	1,091	2.7%
<i>Gossypium herbaceum</i>	1,556	43,952	839	3.8%
<i>Gossypium raimondii</i>	775.2	40,976	1,249	6.1%

Note BDGPs: bidirectional gene pairs; BDGs: bidirectional genes

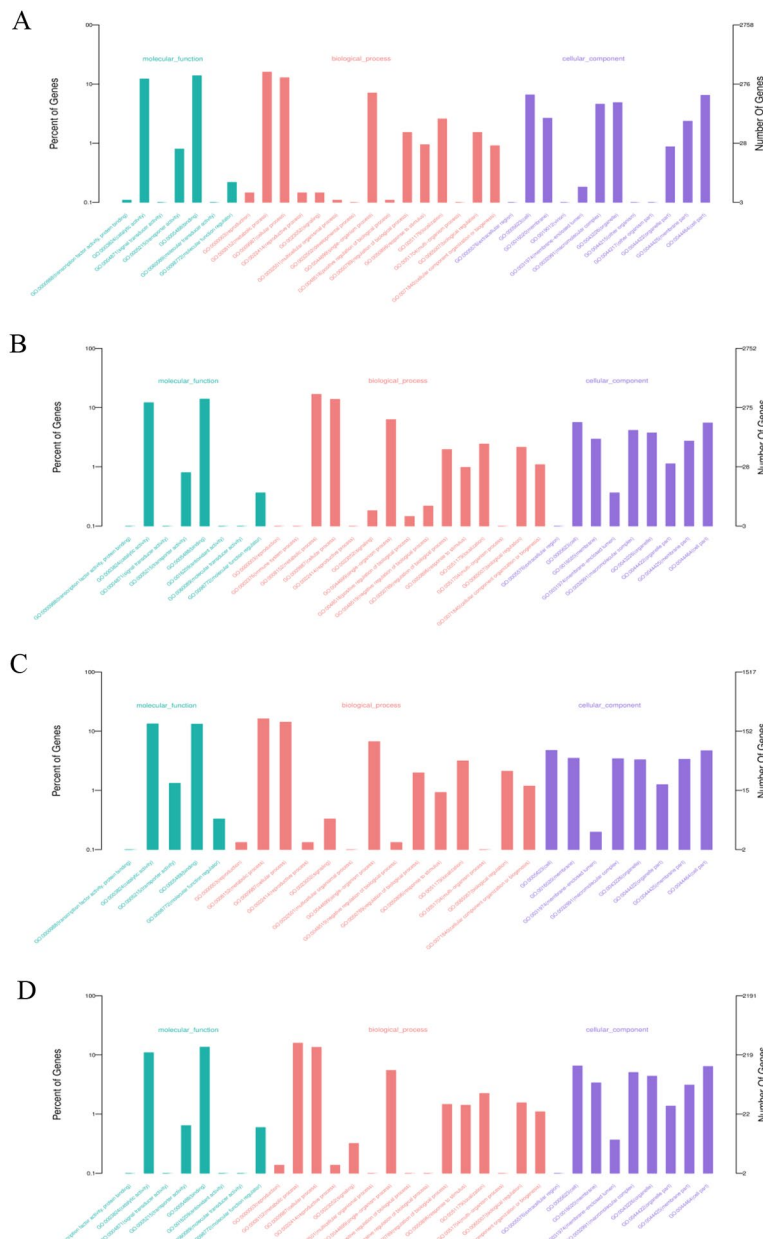


Fig. 11 GO cluster analysis of bidirectional genes in different cotton subspecies. All the bidirectional genes in each cotton subspecies were analyzed for GO functional clustering. The functions of the bidirectional genes were studied at the following three levels: molecular function, biological process and cellular component. **A:** *Gossypium hirsutum*; **B:** *Gossypium barbadense*; **C:** *Gossypium herbaceum*; **D:** *Gossypium raimondii*

gene pairs maintained a high degree of conservation during the evolution of different subspecies of cotton and played an important role in the specific metabolic pathways involved in cotton growth and development.

Analysis of the positional distribution and sequence characteristics of bidirectional promoters in four cotton subspecies

The above results indicated that the bidirectional genes in the four different cotton subspecies presented some similarities in many functional characteristics, leading

to the question of whether the bidirectional promoter sequences also exhibited special commonalities. Therefore, bidirectional intergenic sequences screened from the four different cotton subspecies were compared and analyzed in this study. Owing to the high GC content of bidirectional promoters, we analyzed the GC content of bidirectional intergenic sequences in different cotton subspecies. The results revealed that the GC contents of the bidirectional intergenic sequences in *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium raimondii* were 32.6%, 33.4% and 29.8%, respectively, whereas

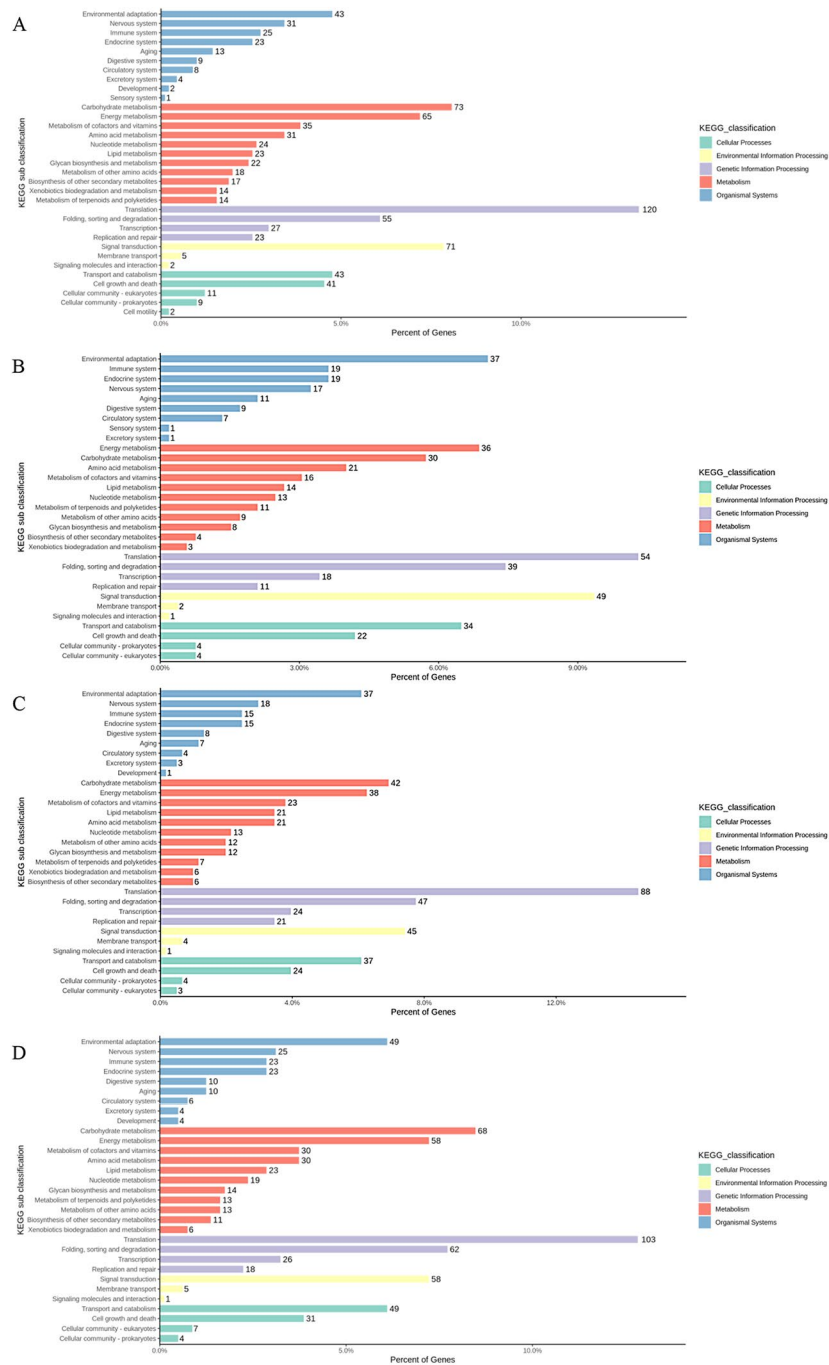


Fig. 12 Cluster analysis of the KEGG metabolic pathways of bidirectional genes in different cotton subspecies. All bidirectional genes in each cotton subspecies were analyzed for KEGG metabolic pathways to elucidate the functions of bidirectional genes from different metabolic pathways in cotton. **A:** *Gossypium hirsutum*, **B:** *Gossypium barbadense*, **C:** *Gossypium herbaceum*, **D:** *Gossypium raimondii*

the GC contents in the random promoters were 30.1%, 30.9% and 28.3%, respectively (Fig. 13A). These results indicate that the GC content of the bidirectional promoters in *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium raimondii* was generally greater than that of the random promoters, which was consistent with the

GC content of the bidirectional promoters in the *Gossypium hirsutum* genome.

The chromosomal distribution of bidirectional gene pairs in the genome was subsequently statistically analyzed. No significant differences in the distribution of bidirectional gene pairs were noted in the At and Dt

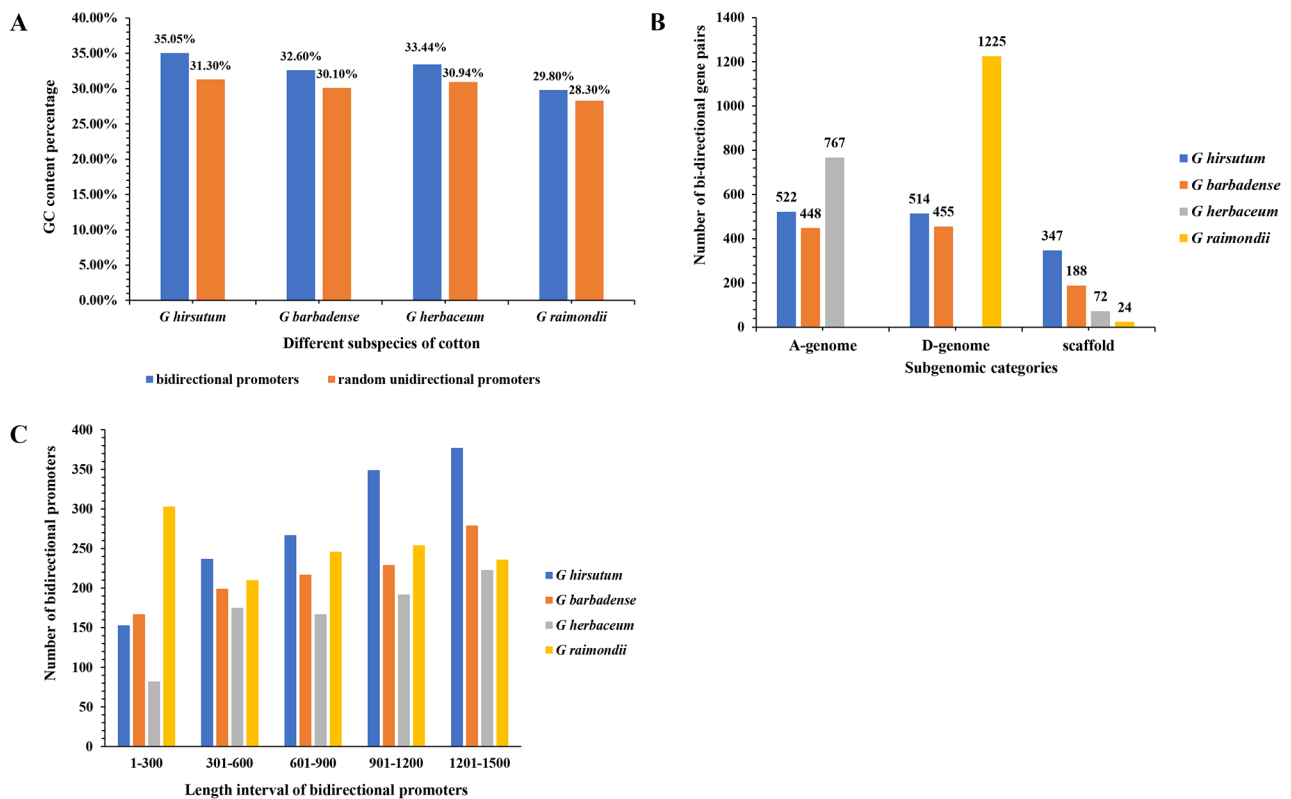


Fig. 13 Sequence characteristics of bidirectional promoters in cotton subspecies. **A:** Statistical analysis of the GC contents of bidirectional promoter sequences in four cotton subspecies. The blue column represents the average GC content of bidirectional promoter sequences, and the red line represents the average GC content of random promoter sequences. **B:** The distributions of bidirectional gene pairs in the A-genome and D-genome were statistically analyzed separately. **C:** Statistical analysis of bidirectional promoter sequence length in four cotton subspecies. The number of bidirectional promoters distributed in different length intervals exhibited different distribution trends, and the distribution trends of *Gossypium hirsutum*, *Gossypium barbadense* and *Gossypium herbaceum* were consistent, with the highest percentage occurring in the interval of 1,201–1,500 bp. The distribution trends of *Gossypium raimondii* had the greatest percentage in the range of 1–300 bp

subgenomes of the *Gossypium barbadense* genome, which contained 448 and 455 bidirectional gene pairs, respectively, and 188 bidirectional genes were located on the scaffolds. In the *Gossypium barbadense* genome, bidirectional gene pairs were distributed mainly on the 26 chromosomes of A05, A11 and D05 (Supplementary Fig. 3). In the genome of *Gossypium herbaceum*, bidirectional promoters are distributed mainly on the A05, A11 and A09 chromosomes (Supplementary Fig. 4). In the genome of *Gossypium raimondii*, bidirectional promoters are distributed mainly on the D09, D07 and D04 chromosomes (Supplementary Fig. 5). The distributions of bidirectional gene pairs in the A-genome and D-genome were statistically analyzed separately. The results revealed that in the A-genome, *Gossypium herbaceum* presented the most bidirectional gene pairs, followed by *Gossypium hirsutum*, whereas in the D-genome, *Gossypium raimondii* presented the most bidirectional gene pairs, followed by *Gossypium hirsutum* (Fig. 13B). These findings indicate that during the process of diploid cotton hybridization, chromosome reconstruction, chromosome

displacement, and human domestication to form allotetraploid cotton, significant changes in the position and structure of bidirectional gene pairs occurred, leading to variations in the number of genes in the A-genome and D-genome. The statistical results for the bidirectional promoter sequence length revealed that the distribution of bidirectional promoters in *Gossypium barbadense* and *Gossypium herbaceum* were essentially consistent with that in *Gossypium hirsutum*, which increased with length, with the highest percentage in the range of 1,200–1,500 bp. In the genomes of *Gossypium raimondii*, bidirectional promoters were most common in the range of 1–300 bp (Fig. 13C).

Phylogenetic analysis of bidirectional genes

The importance of gene function is reflected mainly by conservation during the evolutionary process. There are many bidirectional gene pairs in animals, plants and microorganisms, and whether their functions are conserved in different species needs to be explored. Therefore, we analyzed the homologous genes of bidirectional

genes in the genomes of *Gossypium hirsutum* and three other subspecies of cotton, the monocotyledon species rice and the dicotyledon species *Arabidopsis thaliana*. The comparison analysis of homologous genes was performed using BLAST software. In total, 2,711 unigenes of 1,383 pairs of bidirectional gene pairs based on bidirectional intergenic sequences of less than 1,500 bp in *Gossypium hirsutum* were searched for homologous genes in the genomes of *Gossypium barbadense*, *Gossypium herbaceum*, *Gossypium raimondii*, *Arabidopsis thaliana* and *Oryza sativa*; specifically, 2,468, 2,412, 2,341, 2,032 and 2,112 homologous genes were searched, respectively. The evolutionary relationships of bidirectional genes in different species were identified by calculating the distances between these homologous gene TSSs, and five length thresholds of 0.5 kb, 1.0 kb, 1.5 kb, 3.0 kb and 5.0 kb were set (Table 3). The proportion of bidirectional gene pairs formed by homologous genes from different subspecies of cotton, such as *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium raimondii*, and the dicotyledon *Arabidopsis thaliana* was greater than that of the monocotyledon *Oryza sativa*. In the genomes of three different cotton subspecies, *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium raimondii*, 812, 822 and 838 pairs of homologous genes conform to the characteristics of bidirectional gene pairs, respectively. In total, 894 pairs were present in the dicotyledon *Arabidopsis thaliana*, whereas only 127 pairs exist in rice. Given that the 2,766 bidirectional genes of *Gossypium hirsutum* presented a relatively close overall number of homologous genes among the above five species, the number of bidirectional genes present in the genomes of the above five species was also relatively large. It was inferred that the bidirectional genes present in the genome of *Gossypium hirsutum* exhibited a certain degree of conserved function, and their sequence characteristics were also conserved.

Materials and methods

Identification and screening of bidirectional gene pairs

The genome-wide data and functional annotation information of *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium raimondii* were obtained from the CottonGen (<https://www.cottongen.org/>) website. The genome annotation data

for *Arabidopsis thaliana* and *Oryza sativa* were downloaded from the Phytozome13 (<https://phytozome-next.jgi.doe.gov/>) website. The species genome data contain much gene annotation information, including exons, introns, 5'-UTRs, 3'-UTRs, intergene sequences, gene locations, forward and reverse strands and other information. Therefore, useless redundant information was filtered before screening bidirectional gene pairs and their intergenic region sequences for subsequent analysis. We used the Python program compiled by Liu [15] and modified it to screen bidirectional gene pairs in the genome of *Gossypium hirsutum* (Supplementary script 1). The following screening conditions were employed. First, the basic annotation information of “genes” or “mRNAs” in the genome was determined, and information regarding the chromosome, starting position, ending position, and positive and negative strands was obtained. The screening parameters were set as follows on the basis of the definition of bidirectional gene pairs: (1) The two “head-to-head” genes must be on the same chromosome. (2) There must be a single gene on the positive and negative strands. (3) The starting point value of the positive-strand gene minus the starting point value of the negative-strand gene should be set within the range of 0~1,500 bp. These screened gene pairs were considered bidirectional gene pairs.

Functional annotation and clustering analysis of bidirectional genes

The distribution of bidirectional transcript pairs on the *Gossypium hirsutum* TM-1 chromosome was systematically collated using MG2C (http://mg2c.iask.in/mg2c_v2.0/) online analysis software and TBtools mapping software. GO enrichment analysis and gene functional annotation of the obtained bidirectional gene pairs were performed using blast2GO bioinformatics software, and the bidirectional gene pairs were analyzed in detail based on three categories: cell composition, molecular function and biological process. By setting the corresponding analysis parameters, the main functions of gene enrichment were displayed via blast2GO (v2.2.31+) software. The metabolic pathways associated with the obtained bidirectional gene pairs were analyzed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and the main metabolic pathways associated with gene

Table 3 Orthologous genes of bidirectional gene pairs of *Gossypium hirsutum* in five other plant species

Species	Total	Distance between TSSs of gene pairs				
		0.5 kb	1.0 kb	1.5 kb	3.0 kb	5.0 kb
<i>Gossypium barbadense</i>	2,468	236	574	812	932	1,165
<i>Gossypium herbaceum</i>	2,412	272	568	822	948	1,136
<i>Gossypium raimondii</i>	2,341	302	584	838	967	1,096
<i>Arabidopsis thaliana</i>	2,032	387	612	894	1,068	1,220
<i>Oryza sativa</i>	2,112	32	68	127	146	162

enrichment were mapped to identify the specific processes involved in the growth and development of cotton.

Tissue expression profile analysis and qRT-PCR validation of bidirectional gene pairs

Upland cotton K312 was planted in a greenhouse. The roots and leaves were collected at 15, 25, and 35 days after germination, and the surfaces of the materials were cleaned with ddH₂O. The root materials from the three periods were mixed together as the root tissue material and wrapped with aluminum foil, and the leaf material was subjected to the same procedure. The material was frozen in liquid nitrogen for 5 min and then transferred to a -80 °C refrigerator for storage. Anthers were collected from cotton at -3~0 DPA. Stigmas were collected on the day of flowering, and fibers were collected at 7, 14, and 26 DPA (removing ovules). These samples were quickly placed in liquid nitrogen for freezing treatment for approximately 10 min and then transferred to a -80 °C freezer for storage. Total RNA was extracted from each sample via the RNeasy Pure Plant Kit (Polysaccharides & Polyphenolics-rich) (TIANGEN, Beijing, China) following the manufacturer's protocol.

In this study, we completed transcriptome sequencing of seven different tissues (roots, leaves, anthers, stigmas and fibers at 7, 14, and 26 DPA) of *Gossypium hirsutum* K312 [23], and the tissue expression profiles of bidirectional gene pairs were obtained by combining the transcriptome data for different cotton tissues. Ten pairs of bidirectional genes were randomly screened, and specific PCR primers were designed with the Primer3 online tool (<http://bioinfo.ut.ee/primer3-0.4.0/>). cDNA synthesis was performed from 1 µg of total RNA in a 20-µL reaction mixture using a PrimerScript RT kit (TAKARA, Dalian, China) according to the manufacturer's instructions. The 20 µL reactions were performed with 10 µL of SYBR Premix Ex Taq II (TLi RaseH Plus) (TAKARA, Dalian, China), 0.8 µL of 10 mM forward and reverse primers each, 7.4 µL of ddH₂O and 1.0 µL of cDNA template, after which amplification reactions were conducted. The cotton *Ghsad1* gene (NCBI reference sequence: NM_001327106.1) was used as an internal reference gene [23]. The qRT-PCR conditions were as follows: 95 °C for 2 min followed by 40 cycles of 95 °C for 5 s and 62 °C for 34 s. The tissue expression profile was analyzed via qRT-PCR to verify the reliability of the tissue expression profile of the bidirectional gene.

Identification of orthologous bidirectional gene pairs

BLAST (v2.13.0+) software was used to analyze the orthologous genes (E value < 10⁻¹⁰) of bidirectional gene pairs in the genomes of *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium herbaceum*, *Gossypium raimondii*, *Arabidopsis thaliana* and *Oryza sativa*. Bidirectional

gene pairs in the genomes of *Gossypium barbadense*, *Gossypium herbaceum*, *Gossypium raimondii*, *Arabidopsis thaliana* and *Oryza sativa* were screened and analyzed with a Python program. The phylogenetic relationships of the bidirectional gene pairs were analyzed by comparing four cotton subspecies and three plant species.

Identification of bidirectional promoters and analysis of their sequence characteristics

According to the position annotation information for the bidirectional gene pairs in the genome of *Gossypium hirsutum*, Python programming was used to extract the sequences between the transcription start sites of the two genes in the bidirectional gene pair from the genome sequence. Random unidirectional promoter screening requirements: First, 2,766 bidirectional genes were removed from the list of all cotton genes (76,943) with annotation information, and the remaining genes (74,177) were considered unidirectional genes. One thousand unidirectional promoter sequences were randomly extracted. Second, the length of random unidirectional promoter sequences was the average length of all bidirectional promoter sequences. Finally, the values of the unidirectional promoter sequence features were the averages of three random screens of unidirectional promoters. Using programs to calculate the GC content, TATA boxes, GC boxes and other components in bidirectional promoters and unidirectional promoter sequences, we can predict whether these sequences contain the unique sequence characteristics of bidirectional promoters.

Cloning and functional analysis of the bidirectional promoters

Thirty bidirectional promoters were randomly screened as targets for the verification of promoter activity, and specific primers were designed according to the sequences of these bidirectional promoters (Supplementary Table 3). *EcoRI* and *HindIII* digestion sites were added at both ends of the primers for PCR amplification under the following conditions: 95 °C for 5 min; 30 cycles of 95 °C for 30 s, 52–62 °C for 30 s, and 72 °C for 30–60 s; and 72 °C for 10 min. After PCR amplification, sequencing was performed. The restriction endonucleases *EcoRI* and *HindIII* were subsequently used to double digest the cloned bidirectional promoter and pBD-2G plasmid containing both the *gus* and *gfp* reporter genes, which were linked with T4 ligase, and transformed into *E.coli* DH5α competent cells using the heat shock method. The plasmid was extracted for sequencing and double digestion verification, and the plasmid was successfully transformed into *Agrobacterium tumefaciens* LBA4404. Cotton was transiently transformed using the method of Yang et al. for transforming cotton into a transient expression system [26], and the expression of reported

genes was tested using histochemical staining and green fluorescence observation to judge the feasibility and reliability of using this retrieval mode to screen bidirectional promoters. An Axio LSM 700 laser-scanning confocal microscope (Zeiss Co., Ltd., Jena, Germany) was used for GFP observations.

Discussion

In research on improving the quality of transgenic crops, it is difficult to change the genetic characteristics of a target trait via the introduction of a single gene. It is often necessary to introduce multiple functionally related genes or even genes related to the entire metabolic pathway into recipient plants simultaneously to effectively improve the genetic traits of crops for research purposes. Recently, the use of transgenic technology to introduce multiple genes into plants at the same time to achieve simultaneous improvement of multiple traits has become a major research focus. However, simultaneous expression of multiple foreign genes in transgenic plants often requires highly active promoters [24]. Previous studies have shown that if a homologous sequence between two promoter sequences is introduced into plants, it may cause “coinhibition” of gene expression in plants, which subsequently leads to partial gene silencing [25]. Therefore, the selection of highly expressed bidirectional promoters is crucial in plant genetic engineering. Compared with unidirectional promoters, bidirectional promoters can simultaneously regulate the expression of multiple genes and improve the efficiency of biotechnology. Bidirectional promoters will likely become the best promoters for transgenic breeding via multigene cotransformation methods in the future and have inestimable value in the field of genetic engineering research and application.

With the successful completion of human whole-genome sequencing, researchers have sorted and analyzed the obtained data in detail. In the plant genome, mining and functional studies of bidirectional promoters at the genome-wide level have focused mainly on *Arabidopsis thaliana* [11], *Oryza sativa*, *Populus* [13] and *Zea mays* [15], and there has been no research on bidirectional promoters in cotton species. In this study, the bidirectional gene pairs and bidirectional promoters in the terrestrial cotton genome were systematically studied using the discovered bidirectional gene pairs *Ghrack1* and *Ghuhrf1* as the starting points [26]. Trinklein et al. [9] first screened and studied bidirectional gene pairs and their bidirectional promoters at the human genome-wide level and compiled a set of procedures that can achieve bidirectional promoter screening in the whole genome, and this method has also been widely used in the screening of bidirectional promoters in plant genomes (Fig. 1A). Given that the research objects in this case were genes, it

was impossible to complete the screening of transcripts. However, a gene can be differentially spliced because of variability or multiple transcription start sites to form multiple transcripts, which results in the loss of relevant data. Liu et al. [15] added a screening procedure for transcripts to the program written by Trinklein, thus increasing the reliability of the screening of bidirectional gene pairs (Fig. 1B). The search model employed in this study was based on the screening model of bidirectional gene pairs in the maize genome developed by Liu et al. [15], and the location coordinates of multiple transcripts were added. For example, gene 1 and gene 2 had three transcripts, and the intergenic region between the three transcripts of gene 1 (T1, T2, T3) and the three transcripts of gene 2 (T1, T2, T3) could be considered a bidirectional promoter, making the search range more comprehensive (Fig. 1C). On the basis of previous research results on bidirectional promoters combined with the characteristics of plant promoters, the threshold value of the length for screening bidirectional intergenic sequences was determined to be 1,500 bp. The modified screening model was used to screen and perform functional analysis of bidirectional gene pairs in the *Gossypium hirsutum* genome. The results were consistent with the results of previous reports in other plants, and many (1,383) bidirectional gene pairs were identified in the genome of *Gossypium hirsutum*. The functional cluster analysis of these genes was performed using blast2GO software. These bidirectional genes were enriched mainly in cell and cell parts, cells, organelles and membrane-bound organelles and participated in a variety of metabolic processes, cell processes and biosynthetic and other biological processes, mainly involving molecular functions such as catalytic activity, binding, transfer activity, hydrolase activity and DNA binding. This result was consistent with previous research reports showing that most bidirectional genes are housekeeping genes in species and were relatively conserved during the process of evolution [7, 15].

For bidirectional promoters, the most important orthologous role involves the sharing of transcriptional regulatory elements and RNA polymerase II binding sites [27]. When bidirectional promoter sequences are short, all *cis*-acting elements involved in transcription are shared to achieve the functions necessary for transcription. Therefore, bidirectional promoters drive the symmetrical expression patterns of genes at both ends [28]. Thus, at a relatively close distance, interactions at the orthologous transcription level become quite important. However, when the length of the bidirectional promoter sequence is more than 400 bp, there are more regulatory elements in the promoter. Both ends can use their own unique elements to drive the expression of downstream genes, resulting in fewer and fewer types of shared regulatory elements. Thus, the expression patterns of

bidirectional gene pairs differ across different tissues, resulting in an asymmetric expression pattern [28, 29]. In summary, no interdependence of the expression pattern of each gene in the bidirectional gene pair was observed, which is consistent with the results of Liu and Ahmad et al. [7, 30], indicating that the data obtained via the screening model applied in this study are highly reliable. Given that the expression patterns of bidirectional gene pairs differ, the specific mechanisms whereby bidirectional promoters drive gene expression in both directions need to be further studied. In addition to promoter length and transcriptional regulatory elements, other factors may affect the regulatory functions of bidirectional promoters in the bidirectional promoter structure [2, 7]. Future research will also require numerous experiments to verify the functions of these genes in transgenic plants to confirm the expression pattern of bidirectional transcription gene pairs, which will play an important auxiliary role in revealing the coexpression mechanism of gene pairs with bidirectional transcription.

A high GC content is an important feature of bidirectional promoters. Previous studies have shown that 70.8% of human bidirectional promoters exhibit a GC content of more than 60%, and the average GC content is 66%, which is much greater than the GC content of 53% observed among randomly selected promoters [31]. Similarly, the GC contents of bidirectional promoters in sorghum, rice, soybean, *Arabidopsis thaliana* and maize are 55%, 48.2%, 31%, 34% and 49.9%, respectively, and these values are also significantly greater than the reported GC contents of randomly selected promoters in these species, which are 46.5%, 44.3%, 29.8%, 32.1% and 44.5%, respectively [13, 15]. On the basis of the genome sequence information of *Gossypium hirsutum*, we extracted the interregional sequences between these bidirectional genes and analyzed the GC content of these sequences. The results revealed that the GC content in these bidirectional promoter sequences was as high as 35.05%, whereas the GC content in the randomly screened unidirectional promoters was 31.3%. These study results were consistent with those of previous studies. Moreover, bidirectional gene interregional sequence lengths mainly range from 1,200 to 1,500 bp, verifying the hypothesis of Mitra et al. that there are many *cis*-acting elements, such as expression regulatory elements and inducible expression elements related to tissue-specific expression in the bidirectional promoter sequences of plants, such that bidirectional promoters require longer sequences to meet the expression needs of the plant itself [12].

To verify the bidirectional promoter activity of the screened bidirectional intergenic sequences, 30 pairs were randomly selected for intergenic sequence cloning and transient cotton expression verification. The results revealed that 25 bidirectional intergenic sequences had

bidirectional promoter activity, and the other 5 had promoter activity in one direction. Among these pairs, only 2 showed detectable promoter activity in the GUS direction, and the other 3 only showed detectable promoter activity in the GFP direction. This finding revealed that bidirectional promoters were present in the inter-interval sequences of the screened bidirectional gene pairs, but some pseudobidirectional promoters were also present. In this work, we used a computer program to screen bidirectional promoters in cotton. In the results of big data screening, ensuring 100% accuracy is difficult, and deviations in the selected bidirectional promoters may exist [6]. It is possible that the bidirectional intergenic region sequences were too short to include the basic *cis*-acting elements that each end of the bidirectional promoter is expected to include, resulting in an inability to drive reporter gene expression [32]. It is also possible that these promoters are nonleaf tissue-specific expression promoters expressed in nonleaf tissues of cotton such that some bidirectional promoters, or one end thereof, cannot drive the expression of reporter genes in the transient expression system of cotton leaves [33]. The situation described in this article also applies to maize. Liu et al. noted in the "Identification and functional characterization of bidirectional gene pairs and their intergenic regions in maize" that "Five of the MA (9) candidates and seven of the RS (18) candidates were transcribed in only one direction. One of the MA candidates showed no promoter activity in either direction." The bidirectional promoters screened on the basis of big data showed similar results, with no promoter activity at one or both ends [15]. In summary, these bidirectional promoters with promoter activity in both directions can be used as promoter resources in research on cotton fiber quality improvement.

In total, 1,091, 839 and 1,249 bidirectional gene pairs were identified in the other subspecies of cotton. The genome size of *Gossypium herbaceum* is more than twice that of *Gossypium raimondii*, but the number of bidirectional gene pairs is only two-thirds that of *Gossypium raimondii*. The reason for this phenomenon may be that the genome of *Gossypium herbaceum* underwent a large-scale inversion at the overall level during evolution [34], leading to a large expansion of the genome, destruction of the original intergenic structure, and separation of its original bidirectional gene pairs. We subsequently compared and analyzed the bidirectional genes in the genomes of *Gossypium hirsutum* in the *Gossypium barbadense*, *Gossypium herbaceum*, *Gossypium raimondii*, *Arabidopsis* and *Oryza sativa* genomes, identified many homologous genes, and identified and analyzed the TSSs among these genes. The results revealed that the number of homologous bidirectional genes in different subspecies of cotton and monocotyledonous and dicotyledonous plants was similar to that in *Gossypium hirsutum*.

However, the proportion of these homologous genes forming bidirectional gene pairs in different subspecies of cotton and dicotyledons was much greater than that in the monocotyledon *Oryza sativa*, which was consistent with the results obtained by Liu et al. for monocotyledons [15]. Green algae are the smallest eukaryotic organisms currently known and are the species of origin for green plant evolution. Species that evolved from green algae retain almost all of these homologous genes [35]. By comparing the bidirectional genes among *Zea mays*, *Sorghum bicolor*, *Glycine max*, *Oryza sativa* and *Arabidopsis*, Liu et al. reported that these genes presented a high degree of similarity in structure and function and further reported that the bidirectional genes were conserved both functionally and structurally by comparing the bidirectional genes among *Zea mays*, short-stemmed grass and green algae; however, the specific genetic composition of the bidirectional gene pairs changed constantly [15]. Therefore, it is hypothesized that in the process of species evolution, although the genes of the original species are selectively retained, new genes also develop, and the genome capacity expands as the gene population expands, resulting in the formation of a new species after a certain stage [36]. However, species with different levels of evolution retain different numbers of homologous genes relative to the original species. Species with low evolutionary rates retain a greater number of homologous genes of their original species compared with species with high evolutionary rates, and species with more similar evolutionary rates share a greater number of homologous genes and a greater frequency of homologous genes in pairs [37]. This is one of the evolutionary pathways and manifestations from lower plants to higher plants. Most bidirectional genes are housekeeping genes that are relatively conserved in the process of species evolution [7, 15]. These results demonstrated that bidirectional genes are conserved in the process of evolution and that their structures and functions are conserved to some extent.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05548-w>.

Supplementary Material 1
 Supplementary Material 2
 Supplementary Material 3
 Supplementary Material 4
 Supplementary Material 5
 Supplementary Material 6
 Supplementary Material 7
 Supplementary Material 8
 Supplementary Material 9

Acknowledgements

Not applicable.

Author contributions

Conceptualization, J.Y., X.W. and Z.W.; data curation, J.Y., L.G. and X.Z.; methodology, J.Y. and X.Z.; software and validation, J.Y. and L.G.; formal analysis and investigation, R.Z., X.L. and Y.C.; writing—original draft preparation, J.Y. and L.G.; writing—review and editing, X.W. and Z.W.; project administration, X.W.; funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by grants from the Major Project of Agricultural Biological Breeding (2023ZD04062 and 2022ZD0402003) and the Major Project of China on New Varieties of GMO Cultivation (2016zx08005-003).

Data availability

All data generated or analysed during this study are included in this published article. The raw sequence data from this study have been deposited at the NCBI Sequence Read Archive (BioProject accession ID: PRJNA1048624).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Key Laboratory on Molecular Safety Assessment of Agri-GMO, MARA, Beijing 100081, China

Received: 27 December 2023 / Accepted: 26 August 2024

Published online: 04 September 2024

References

- Sun Y, Shang L, Zhu QH, Fan L, Guo L. Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci.* 2022;27(4):391–401.
- Zainal-Abidin RA, Harun S, Vengatharajuloo V, Tamizi AA, Samsulrizal NH. Gene Co-Expression Network Tools and Databases for Crop Improvement. *Plants (Basel)* 2022;11(13).
- Jones DM, Vandepoele K. Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Curr Opin Plant Biol.* 2020;54:42–8.
- Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol.* 2018;19(8):489–506.
- Nimeth BA, Riegler S, Kalyna M. Alternative splicing and DNA damage response in plants. *Front Plant Sci.* 2020;11:91.
- Jangid RK, Kelkar A, Muley VY, Galande S. Bidirectional promoters exhibit characteristic chromatin modification signature associated with transcription elongation in both sense and antisense directions. *BMC Genomics.* 2018;19(1):313.
- Ahmad SS, Samia NSN, Khan AS, Turjya RR, Khan MA. Bidirectional promoters: an enigmatic genome architecture and their roles in cancers. *Mol Biol Rep.* 2021;48(9):6637–44.
- Wei W, Pelechano V, Järvelin AI, Steinmetz LM. Functional consequences of bidirectional promoters. *Trends Genet.* 2011;27(7):267–76.
- Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otililar RP, Myers RM. An abundance of bidirectional promoters in the human genome. *Genome Res.* 2004;14(1):62–6.
- Mishra RC, Grover A. Intergenic sequence between *Arabidopsis* caseinolytic protease B-cytoplasmic/heat shock protein100 and choline kinase

- genes functions as a heat-inducible bidirectional promoter. *Plant Physiol.* 2014;166(3):1646–58.
11. Wang Q, Wan L, Li D, Zhu L, Qian M, Deng M. Searching for bidirectional promoters in *Arabidopsis thaliana*. *BMC Bioinformatics.* 2009;10(Suppl 1):S29.
 12. Mitra A, Han J, Zhang ZJ, Mitra A. The intergenic region of *Arabidopsis thaliana cab1* and *cab2* divergent genes functions as a bidirectional promoter. *Planta.* 2009;229(5):1015–22.
 13. Dhadi SR, Krom N, Ramakrishna W. Genome-wide comparative analysis of putative bidirectional promoters from rice. *Arabidopsis Populus Gene.* 2009;429(1–2):65–73.
 14. Krom N, Ramakrishna W. Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, *Arabidopsis*, and *populus*. *Plant Physiol.* 2008;147(4):1763–73.
 15. Liu X, Zhou X, Li Y, Tian J, Zhang Q, Li S, Wang L, Zhao J, Chen R, Fan Y. Identification and functional characterization of bidirectional gene pairs and their intergenic regions in maize. *BMC Genomics.* 2014;15(1):338.
 16. Liu X, Li S, Yang W, Mu B, Jiao Y, Zhou X, Zhang C, Fan Y, Chen R. Synthesis of seed-specific bidirectional promoters for metabolic engineering of anthocyanin-rich maize. *Plant Cell Physiol.* 2018;59(10):1942–55.
 17. Arnaiz A, Martinez M, Gonzalez-Melendi P, Grbic V, Diaz I, Santamaria ME. Plant defenses against pests driven by a bidirectional promoter. *Front Plant Sci.* 2019;10:930.
 18. He K, Rad S, Poudel A, McLellan AD. Compact bidirectional promoters for dual-gene expression in a sleeping beauty transposon. *Int J Mol Sci* 2020, 21(23).
 19. Bai J, Wang X, Wu H, Ling F, Zhao Y, Lin Y, Wang R. Comprehensive construction strategy of bidirectional green tissue-specific synthetic promoters. *Plant Biotechnol J.* 2020;18(3):668–78.
 20. Liu X, Ma X, Wang H, Li S, Yang W, Nugroho RD, Luo L, Zhou X, Tang C, Fan Y, et al. Metabolic engineering of astaxanthin-rich maize and its use in the production of biofortified eggs. *Plant Biotechnol J.* 2021;19(9):1812–23.
 21. Wang Z, Yang C, Chen H, Wang P, Wang P, Song C, Zhang X, Wang D. Multi-gene co-expression can improve comprehensive resistance to multiple abiotic stresses in *Brassica napus* L. *Plant Sci.* 2018;274:410–9.
 22. Zinani OQH, Keseroğlu K, Özbudak EM. Regulatory mechanisms ensuring coordinated expression of functionally related genes. *Trends Genet.* 2022;38(1):73–81.
 23. Yang J, Gao L, Liu X, Zhang X, Wang X, Wang Z. Comparative transcriptome analysis of fiber and nonfiber tissues to identify the genes preferentially expressed in fiber development in *Gossypium hirsutum*. *Sci Rep.* 2021;11(1):22833.
 24. Liu X, Yang W, Mu B, Li S, Li Y, Zhou X, Zhang C, Fan Y, Chen R. Engineering of 'Purple embryo Maize' with a multigene expression system derived from a bidirectional promoter and self-cleaving 2A peptides. *Plant Biotechnol J.* 2018;16(6):1107–9.
 25. Flavell RB. Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc Natl Acad Sci U S A.* 1994;91(9):3490–6.
 26. Yang J, Wang X, Hasi A, Wang Z. Structural and functional analysis of a bidirectional promoter from *Gossypium hirsutum* in *Arabidopsis*. *Int J Mol Sci* 2018, 19(11).
 27. Seila AC, Core LJ, Lis JT, Sharp PA. Divergent transcription: a new feature of active promoters. *Cell Cycle.* 2009;8(16):2557–64.
 28. Chen WH, de Meaux J, Lercher MJ. Co-expression of neighbouring genes in *Arabidopsis*: separating chromatin effects from direct interactions. *BMC Genomics.* 2010;11:178.
 29. Tohge T, Fernie AR. Co-regulation of clustered and neo-functionalized genes in plant-specialized metabolism. *Plants (Basel)* 2020, 9(5).
 30. Liu SJ, Yue QJ, Zhang W. Structural and functional analysis of an asymmetric bidirectional promoter in *Arabidopsis thaliana*. *J Integr Plant Biol.* 2015;57(2):162–70.
 31. Yang MQ, Elnitski LL. Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genomics.* 2008;9(Suppl 2):S3.
 32. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet.* 2020;21(2):71–87.
 33. Smirnova OG, Kochetov AV. Choice of the promoter for tissue and developmental stage-specific gene expression. *Methods Mol Biol.* 2020;2124:69–106.
 34. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet.* 2014;46(6):567–72.
 35. Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, et al. Genome analysis of the smallest free-living eukaryote *Ostreococcus Tauri* unveils many unique features. *Proc Natl Acad Sci U S A.* 2006;103(31):11647–52.
 36. Delaux PM, Hetherington AJ, Coudert Y, Delwiche C, Dunand C, Gould S, Kenrick P, Li FW, Philippe H, Rensing SA, et al. Reconstructing trait evolution in plant evo-devo studies. *Curr Biol.* 2019;29(21):R1110–8.
 37. Rieseberg LH, Willis JH. Plant speciation. *Science.* 2007;317(5840):910–4.

Publisher's note

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