



DATABASE

WheatCENet: A Database for Comparative Co-expression Network Analysis of Allohexaploid Wheat and Its Progenitors



Zhongqiu Li¹, Yiheng Hu^{2,3}, Xuelian Ma¹, Lingling Da¹, Jiajie She¹, Yue Liu¹, Xin Yi², Yaxin Cao¹, Wenying Xu¹, Yuannian Jiao^{2,3,*}, Zhen Su^{1,*}

¹ State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China

² State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

Received 12 August 2021; revised 16 March 2022; accepted 8 May 2022

Available online 1 June 2022

Handled by Long Mao

KEYWORDS

Co-expression network;
Species comparison;
Diploid and polyploid wheat;
Functional annotation

Abstract Genetic and epigenetic changes after polyploidization events could result in variable gene expression and modified regulatory networks. Here, using large-scale transcriptome data, we constructed **co-expression networks** for diploid, tetraploid, and hexaploid wheat species, and built a platform for comparing co-expression networks of allohexaploid wheat and its progenitors, named WheatCENet. WheatCENet is a platform for searching and comparing specific functional co-expression networks, as well as identifying the related functions of the genes clustered therein. **Functional annotations** like pathways, gene families, protein–protein interactions, microRNAs (miRNAs), and several lines of epigenome data are integrated into this platform, and Gene Ontology (GO) annotation, gene set enrichment analysis (GSEA), motif identification, and other useful tools are also included. Using WheatCENet, we found that the network of *WHEAT ABERRANT PANICLE ORGANIZATION 1 (WAPO1)* has more co-expressed genes related to spike development in hexaploid wheat than its progenitors. We also found a novel motif of CCWWWWWWGG (CARG) specifically in the promoter region of *WAPO-A1*, suggesting that neofunctionalization of the *WAPO-A1* gene affects spikelet development in hexaploid wheat. WheatCENet is useful for investigating co-expression networks and conducting other analyses, and thus facilitates

* Corresponding authors.

E-mail: zhensu@cau.edu.cn (Su Z), jiaoyan@ibcas.ac.cn (Jiao Y).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2022.04.007>

1672-0229 © 2023 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

comparative and functional genomic studies in wheat. WheatCENet is freely available at <http://bioinformatics.cpolar.cn/WheatCENet> and <http://bioinformatics.cau.edu.cn/WheatCENet>.

Introduction

Wheat, in the Poaceae family, is the most widely grown food crop worldwide, providing an important source of nutrients for millions of people. Global food demand is increasing rapidly, with a 60%–70% increase in food production required by 2050 [1]. Basic research and optimization of wheat breeding are necessary to meet this demand. However, the bread wheat genome is large and complex (16 Gb) [2], and thus wheat research lags behind that of rice and maize. This lag is mainly because bread wheat is a recent allohexaploid, formed via two consecutive allopolyploidization events. The diploid wheat *Triticum urartu* (AA) and a yet unknown *Aegilops* species formed tetraploid wheat *Triticum dicoccoides* (AABB) around 0.5 million years ago (MYA). Subsequently, tetraploid wheat and *Aegilops tauschii* (DD) hybridized to form the hexaploid wheat *Triticum aestivum* (AABBDD) around 0.01 MYA [3,4]. As wheat is a primary food crop, wheat scientists have worked to sequence the genome of hexaploid wheat and its progenitors, and the completion of genome sequencing has laid a solid foundation for studying the functional, comparative, and evolutionary genomics of wheat [5,6].

Currently, public databases for wheat can be divided into four types: genome, transcriptome, proteome, and others. Genome databases, like Wheat@URGI portal [7], GrainGenes [8], CerealsDB [9], Wheat-SnpHub-Portal [10], WheatGmap [11], and Triticeae-GeneTribe [12], provide genomic or genetic data and other useful tools. Transcriptome databases, including expVIP [13] and Wheat eFP Browser (https://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi), are usually used to provide the expression patterns of homeologs. Transcriptome-based co-expression networks like WheatNet have been constructed by DNA microarray datasets with an early genome assembly version [14]. Knetminer [15] and WheatOmics [16] only contain a hexaploid wheat network, not including wheat progenitors. Wheat Proteome provides searchable organ and developmental stage proteomic data [17]. Other databases like Triticeae Toolbox (T3) have phenotype and genotype data for barley, wheat, and oat [18]. The wheat microRNA Portal has integrated the abiotic stress response microRNAs (miRNAs) in wheat [19]. Despite the accumulation of large-scale RNA-seq data and the improved and high-quality wheat genomic sequence, there are still gaps in meeting the demands for co-expression analysis within allohexaploid wheat and its progenitors.

Thus, we developed the co-expression network comparison database WheatCENet for allohexaploid wheat and its progenitors, including four global networks (*T. aestivum*, *T. dicoccoides*, *T. urartu*, and *Ae. tauschii*) and two *T. aestivum* conditional networks (tissue-specific and stress-treated). WheatCENet integrates functional and epigenome sequencing data, and includes useful tools like gene set enrichment analysis (GSEA), Gene Ontology (GO) analysis, and motif analysis, which will help bench scientists easily pick key candidate genes for functional studies. In addition, this database will provide genomic scientists with a useful source for deciphering key

molecular modules during the formation, evolution, and domestication of wheat.

Construction and content

Data sources and processing

For genome data, *T. aestivum* (AABBDD) was based on the International Wheat Genome Sequencing Consortium (IWGSC) Chinese Spring v1.0 genome assembly and v1.1 annotation (URGI) [2]; *T. dicoccoides* (AABB) was based on the Zavitan WEW_v1.0 genome assembly and annotation [20]; *T. urartu* (AA) was based on the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (IGDB) genome assembly and annotation [21]; and *Ae. tauschii* (DD) was based on the Chinese Academy of Agricultural Sciences (CAAS) genome assembly and annotation [22].

For transcriptome data, we used most of the available RNA-seq datasets in addition to the common tissues of leaf, root, and grain in the four studied species for robust constructed co-expression networks [23,24]. Finally, we collected 425 transcriptomic datasets (112 for *T. aestivum*, 153 for *T. dicoccoides*, 90 for *T. urartu*, and 70 for *Ae. tauschii*) from the NCBI Sequence Read Archive (SRA) (for detailed sample information, see Table S1). Quality control was conducted based on FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and Trimmomatic software [25] was used to remove adapter sequences and low-sequencing-quality bases. The remaining sequence data (112 for *T. aestivum*, 153 for *T. dicoccoides*, 76 for *T. urartu*, and 42 for *Ae. tauschii*) were mapped to corresponding genomes, and then the fragments per kilobase of transcript per million mapped reads (FPKM) values of all protein-coding genes were calculated from each sample with parameter settings --star and --estimate-rspd using RSEM [26]. Those samples with a mapping rate < 50% were filtered out. Then, the R package “*heatmap*” (<https://github.com/raivokolde/heatmap>) was used to perform cluster analysis on all datasets, and the outlier samples were excluded (Figures S1 and S2). According to the boxplot of the reading score distribution, the quality of the remaining samples was acceptable (Figure S3). The mapping results of these data are listed in Table S2.

For epigenome data, all 25 *T. aestivum* epigenomic datasets, including H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me2, H3K9ac, H3K27ac, DNase-seq, and CENH3, were downloaded from public platforms, including SRA. Quality control was conducted based on trim_galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) software with parameter settings $q = 25$ and stringency = 3. The sequence reads were mapped to the *T. aestivum* accession Chinese Spring v1.0 reference genome with the maximal exact matches (MEM) algorithm and default parameters by BWA software [27]. The enriched regions were called by MACS software [28] with the nomodel parameter. The details and mapping results of these data are listed in Table S3.

For collinear ortholog pairs, we first computed the top 5 Basic Local Alignment Search Tool (BLAST) results between two species using protein sequences based on the rank of bit score that met an E-value threshold of 1×10^{-5} , which was suggested by MCSanX tools [29]. Then the top 5 BLAST results with general feature format (gff) files were used to establish collinear ortholog pairs in *T. urartu*, *Ae. tauschii*, *T. dicoccoides*, and *T. aestivum* through MCSanX tools with default parameters.

Co-expression network construction and network comparison

We used the calculated FPKM values for constructing the four global networks and two *T. aestivum* conditional (tissue-specific and stress-treated) networks based on the Pearson correlation coefficient (PCC) and mutual rank (MR) algorithm [30]; every network covered at least 84.70% of genes (Table 1). We used the PCC and MR to measure the co-expression relationship between genes. First, based on the PCC distribution diagram of all gene pairs, the thresholds for negative and positive correlation were the values of the lowest 5% PCC values (−0.25) and highest 5% PCC values (0.5) in the *T. aestivum* global network, respectively (Figure S4). Second, we used the MR method to exclude poor co-expression gene pairs, because MR has been used successfully for this purpose in several plants such as *Arabidopsis*, maize, and bamboo [31]. Furthermore, the GO terms of biological processes related to multiple genes in the interval [4,20] were used to evaluate the accuracy of the co-expression network through the receiver operating characteristic (ROC) curve [32]. Finally, to compare the *T. aestivum* global networks and conditional networks or other species' networks while considering the coverage and connectivity of each network, we set the MR of all networks to 30.

Further, in order to verify whether the networks we built are comparable among allohexaploid wheat and its progenitors, we collected genes from literature, such as the light-dependent chlorophyll accumulation gene *TaCHLH* [33], the peroxidase gene *TaPOD1* [34], the storage protein activator gene *TaSPA*, and the genes involved in the photosynthesis pathway (*LHCA2*, *LHCA3*, *LHCB7*, and *psaK*) (Table S4). We queried the expression values of these genes in the leaf, root, and grain of the four species. The results suggest that the expression trends of these genes in samples with different developmental stages or stress treatments are still consistent and comparative among four species for analyzing the network (Figure S5).

Functional module identification

We used the CFinder software to identify modules containing more densely connected genes by combining positive and negative gene pairs together [35] (Table 1). Then we used the gene set annotations, like gene families, GO terms, and metabolic pathways (detailed information is in the Functional annotation and gene annotation section), to predict the functions of modules. Non-significant entries were then filtered out using Fisher's tests and the multiple test correction method "Benjamini-Yekutieli" [false discovery rate (FDR), as referred to in the PlantGSAD [36]]. As a result, 1867 functional modules in AABDD, 625 functional modules in AAB, 524 functional modules in AA, and 851 functional modules in DD were

Table 1 Information about the networks for four wheat species

Co-expression network	AABDD (global)	AABDD (tissue-specific)	AABBDD (stress-treated)	AABB	AA	DD
No. of nodes (nodes/all genes, %)	103,512 (95.9)	102,959 (95.4)	94,849 (87.9)	56,858 (87.5)	36,613 (88.2)	36,286 (84.70)
No. of positive edges	985,277	1,004,338	892,249	564,671	366,625	392,654
No. of negative edges	418,096	459,775	388,565	274,275	160,254	164,741
Sample count (No. of excluded duplicate samples)	112 (49)	61 (32)	51 (17)	153 (20)	76 (41)	42 (20)
Tissue	Spike, rachis, ovary, root*, grain*, leaf*, crown, seedling, seed, stem, flag leaf	Spike, rachis, ovary, root*, grain*, leaf*, crown, seedling, seed, stem, flag leaf	Leaf*, grain*, root*, crown	Spike, leaf*, root*, grain*, flower, glume, lemma	Shoots, root*, leaf*, endosperm	Spikes, seedling, grain*, endosperm, leaf, root*, shoots, immature seeds
Treatment	Heat, drought, chitin, Flg22	Heat, drought, chitin, Flg22	Heat, drought, chitin, Flg22	–	Bgt, Cd, cold	–
AUC (cutoff)	0.8599 (MR30)	0.8386 (MR30)	0.8081 (MR30)	0.7486 (MR30)	0.711 (MR30)	0.579 (MR30)
Module (number)	K = 9 (1867)	K = 7 (3913)	K = 10 (1118)	K = 10 (625)	K = 10 (524)	K = 8 (851)

Note: “*” indicates that the tissue was found in all four studied species in the RNA-seq analysis. “–” indicates that the data were not available. AUC, area under curve.

obtained. These modules may be related to important agronomic traits.

Usage of co-expression network tools

In WheatCENet, a network search tool for one gene or a gene list was provided for the four global and two conditional

(tissue-specific and stress-treated) co-expression networks, which were visualized using Cytoscape [37] (**Figure 1A**). Furthermore, we established collinear ortholog pairs in AA, DD, AABB, and AABBDD using MCScanX [29] tools. We used collinear gene pairs to determine the correlations between diploid and polyploid wheat. With the relationship of homologous genes, three types of co-expression networks between

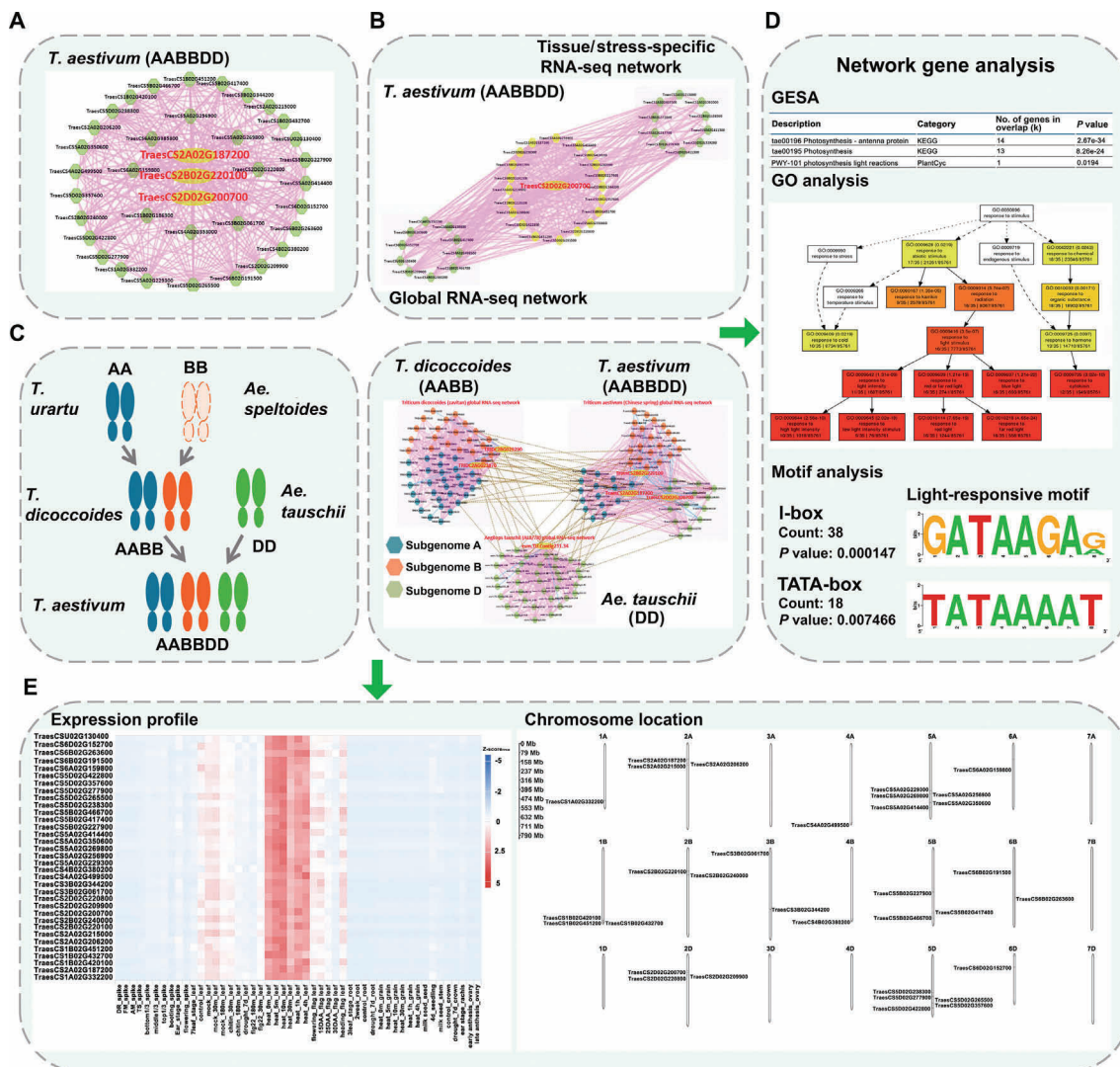


Figure 1 Description of networks in the database

A. An example of gene search results in *T. aestivum* (AABBDD). The biggest yellow node represents the gene queried; the green nodes represent the co-expressed genes of the queried gene. The pink edges link two genes that have a positive co-expression relationship. **B.** Global RNA-seq network vs. tissue/stress-specific RNA-seq network in *T. aestivum* (AABBDD). The biggest yellow node represents the gene queried; other yellow nodes represent the overlapping co-expressed genes between two networks; the green nodes in the gray box represent specific genes in their respective networks. **C.** Network comparison of *T. aestivum* (AABBDD), *T. dicoccoides* (AABB), and *Ae. tauschii* (DD). The pink edges link two genes that have a positive co-expression relationship in a species; the blue edges link two genes that have a negative co-expression relationship in a species; the brown edges link two genes with an orthologous relationship between two species; the blue nodes represent the co-expressed genes of subgenome A; the orange nodes represent the co-expressed genes of subgenome B; and the green nodes represent the co-expressed genes of subgenome D. **D.** Results of analysis tools that we provided. GSEA, GO analysis, and motif analysis can be performed directly on the network results page. **E.** Expression profiling of all genes in the network displayed by the heatmap. The chromosome positions of all genes are also displayed. *T. aestivum*, *Triticum aestivum*; *T. dicoccoides*, *Triticum dicoccoides*; *T. urartu*, *Triticum urartu*; *Ae. tauschii*, *Aegilops tauschii*; *Ae. speltoides*, *Aegilops speltoides*; DR, double-ridge stage; FM, floret meristem; AM, anther primordia stage; TS, tetrad stage; DAA, days after anthesis; GSEA, gene set enrichment analysis; GO, Gene Ontology; FPKM, fragments per kilobase of transcript per million mapped reads.

polyploid wheat and its progenitors can be compared: genes of a single species can be compared between the global network and tissue-specific network or stress-treatment network in the network comparison tool (Figure 1B); genes of two species, such as AA vs. AABB, DD vs. AABBDD, and AABB vs. AABBDD, can be compared in the network comparison tool; and genes of three species (DD, AABB, and AABBDD) can be compared in the ortholog network comparison tool. For instance, users can submit one gene of interest in our ortholog network comparison tool, and choose two or three species for the orthologous pair of this gene. Then all the orthologous gene pairs are highlighted and linked to each other with brown lines in the network to exhibit the conservation and diversification of the regulatory network during wheat evolution (Figure 1C). For all genes in the network, the annotations and relationships of genes in a network are listed in the tables. GSEA, GO analysis, and motif analysis tools are used to find the potential functions and regulations in the promoter regions of genes in the network (Figure 1D), as well as the expression profile and the distribution of genes on chromosomes of the network (Figure 1E). Moreover, the gene expression profiles and *cis*-elements of the homeologous sub-networks can be compared to find the similarities and differences of the network.

Functional annotation and gene annotation

For functional annotation (Figure 2A; Table 2), there are five types of data in WheatCENet, which can be searched by users to predict gene function. On the one hand, integrating these data is of great significance for genome-level gene annotation; on the other hand, these functional annotations can be used as background gene sets for data mining, such as enrichment analysis for network or functional modules.

Ontology has become a critical tool for enabling cross-species comparative analyses and increasing data sharing and reusability [38]. Thus, we collected 7391 GO, 444 Plant Ontology (PO), and 276 Trait Ontology (TO) annotation entries of AABBDD from URGI (<https://wheat-urgi.versailles.inra.fr/Data>), 1814 GO annotation entries of AA from MBKBASE (<https://www.mbkbase.org/Tu/>), and 7612 and 9716 GO annotation entries of DD and AABB separately annotated by Blast2GO software [39].

Metabolic pathways are responsible for the biosynthesis of complex metabolites, having an impact on the growth and development of plants or aiding plants in responding to biotic and abiotic stresses [40]. We collected Plant Reactome pathways of all species from Gramene [41]; PlantCyc pathways of AA, DD, and AABBDD were integrated from PlantCyc [42], while AABB was predicted by orthologs. For the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (<https://www.kegg.jp/kegg/>), DD came from KEGG using the ID conversion; AABBDD and AABB KEGG pathways were annotated by the KEGG tool KAAS; and AA came from MBKBASE.

Protein domains are important parts of proteins, and many domains either have specific functions or contribute to the function of their proteins in a specific way [43]. Protein domains were annotated by PfamScan tools (<https://www.ebi.ac.uk/Tools/pfa/pfamscan/>) based on a hidden Markov

model [44]. We ultimately identified 4021 functional domains with 83,690 genes in AABBDD; 4052 functional domains with 49,301 genes in AABB; 3675 functional domains with 26,636 genes in AA; and 3930 functional domains with 26,275 genes in DD. Users can submit a gene list to WheatCENet, and then the protein domains and detailed information can be extracted to proceed to the downstream analysis (like adding protein domains when constructing evolutionary trees).

Gene families, such as those of transcription factors (TFs)/transcriptional regulators (TRs), protein kinases (PKs), carbohydrate-active enzymes, and ubiquitin, play important biological roles. We used iTAK software (<http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi>) to identify TF and PK families. We obtained a hidden Markov model from UUCD (<https://uucd.biocuckoo.org/>) to identify ubiquitin families. InParanoid (<https://inparanoid.sbc.su.se/cgi-bin/index.cgi>) software and Pfam domains were used to predict the carbohydrate-active enzyme families. So we annotated 20 ubiquitin families with 6641 genes, 90 TF/TR families with 7913 genes, 125 PK families with 5217 genes, and 94 carbohydrate-active enzyme families with 4290 genes in AABBDD; 20 ubiquitin families with 2667 genes, 91 TF/TR families with 4622 genes, 126 PK families with 3126 genes, and 93 carbohydrate-active enzyme families with 2749 genes in AABB; 20 ubiquitin families with 1503 genes, 90 TF/TR families with 2325 genes, 120 PK families with 1623 genes, and 91 carbohydrate-active enzyme families with 1326 genes in AA; and 19 ubiquitin families with 2068 genes, 91 TF/TR families with 2337 genes, 122 PK families with 1183 genes, and 92 carbohydrate-active enzyme families with 1456 genes in DD. By comparing the protein sequences of hexaploid AABBDD and its progenitors with wheat protein sequences of cytochrome P450 (CYP450) genes (<https://drnelson.uthsc.edu/cytochromeP450.html>), we further identified their P450 domains using PfamScan (<https://www.ebi.ac.uk/Tools/pfa/pfamscan/>). Finally, 1564, 823, 462, and 489 full-length CYP450 genes were found in AABBDD, AABB, AA, and DD, respectively (the detailed information and related analysis about CYP450 genes are shown in File S1).

miRNAs function in plant developmental plasticity, abiotic/biotic responses, and symbiotic/parasitic interactions [45]. By integrating miRNAs data including mature miRNAs and precursor miRNAs, 9626 miRNAs in AABBDD and 173 miRNAs in DD were collected from miRbase [46] and IWGSC, while 1702 miRNAs in AA were collected from the literature [47]. Then, all miRNA sequences were mapped to the corresponding genome by GMAP (<http://research-pub.gene.com/gmap/>), and the mature targets were predicted by psRNAtarget (<https://plantgrn.noble.org/psRNATarget/> or <https://www.zhaolab.org/psRNATarget/>). We also predicted protein–protein interaction (PPI) pairs in all species, with the experimentally validated PPI pairs in *Arabidopsis* predicted by InParanoid [48]. There were 149,844 PPI pairs with 12,972 genes in AABBDD, 53,169 PPI pairs with 9452 genes in AABB, 16,726 PPI pairs with 4818 genes in AA, and 16,226 PPI pairs with 5105 genes in DD (Table 2).

For gene annotation, gene search results included all known and predicted information. Taking *WHEAT ABRANT PANICLE ORGANIZATION 1 (WAPO-A1)* as an example, the gene detail page includes basic information: the gene locus

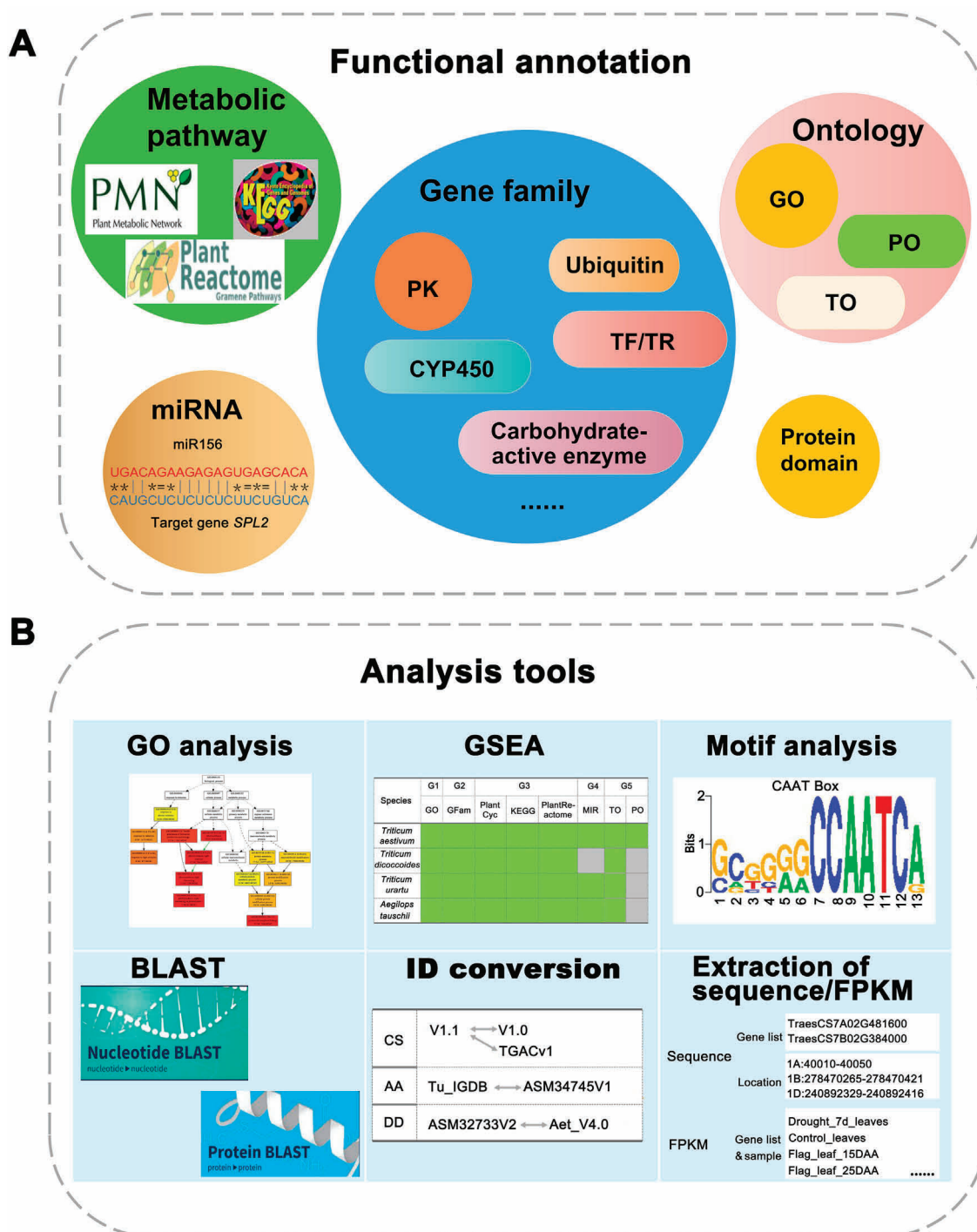


Figure 2 Description of functional annotation and supported tools in the database

A. Five types of data are included in functional annotation. Gene families include ubiquitin, TF/TR, PK, CYP450, and carbohydrate-active enzyme gene families. Ontology (such as GO, PO, and TO), protein domain, miRNA, and metabolic pathways (including PlantCyc, KEGG, and Plant Reactome), can be also browsed. **B.** Tools, like GSEA, GO analysis, motif analysis, BLAST, ID conversion, extraction of sequence/FPKM, and UCSC genome browser [including epigenome data in *T. aestivum* (AABBDD) and RNA-seq samples in *T. dicoccoides* (AABB) (not shown in the picture)], are supported in WheatCENet. miRNA, microRNA; KEGG, Kyoto Encyclopedia of Genes and Genomes; TF, transcription factor; TR, transcription regulator; PK, protein kinase; CYP450, cytochrome P450; BLAST, Basic Local Alignment Search Tool; PO, Plant Ontology; TO, Trait Ontology.

Table 2 Functional annotation collected from public databases or annotated by public software

Classification	Category (source/method)	AABBDD module count (gene count)	AABB module count (gene count)	AA module count (gene count)	DD module count (gene count)
GO	GO (IWGSC/Blast2GO/MBKBASE)	7391 (85,762)	9716 (49,242)	1814 (21,216)	7612 (24,143)
TO	TO (IWGSC/orthologous relationship)	276 (1000)	254 (528)	232 (218)	262 (308)
PO	PO (IWGSC)	444 (63,687)	–	–	–
Pfam domain	Function domain	4021 (83,690)	4052 (49,301)	3675 (26,636)	3930 (26,275)
Gene family	Ubiquitin (UUCD)	20 (6641)	20 (2667)	20 (1503)	19 (2068)
	TF/TR (iTAK)	90 (7913)	91 (4622)	90 (2325)	91 (2337)
	PK (iTAK)	125 (5217)	126 (3126)	120 (1623)	122 (1183)
	CYP450 (BLAST+Pfam)	264 (1564)	242 (823)	211 (462)	232 (489)
	Carbohydrate-active enzyme (ortholog)	94 (4290)	93 (2749)	91 (1326)	92 (1456)
Metabolic pathway	Plant Reactome (Gramene)	204 (3478)	204 (2063)	192 (654)	202 (976)
	PlantCyc (PlantCyc/ortholog)	494 (7100)	494 (4209)	426 (2057)	434 (2228)
	KEGG (KEGG/KAAS/MBKBASE)	139 (15,528)	139 (9312)	112 (2246)	140 (4692)
miRNA	miRNA target (miRbase/IWGSC/literature)	9626	–	1702	173
PPI	PPI (InParanoid)	149,844 (12,972)	53,169 (9452)	16,726 (4818)	16,226 (5105)

Note: “–” indicates that the data were not available. GO, Gene Ontology; TO, Trait Ontology; PO, Plant Ontology; miRNA, microRNA; PPI, protein–protein interaction; TF, transcription factor; TR, transcription regulator; PK, protein kinase; CYP450, cytochrome P450; KEGG, Kyoto Encyclopedia of Genes and Genomes.

is TraesCS7A02G481600; it is a 1457-bp gene with two exons located on chromosome 7A that encodes an F-box-like protein, related to spikelet number per spike (SNS); the best orthologous gene can be linked to the detailed function in *Arabidopsis thaliana* and *Oryza sativa*; and the gene sequence, coding sequence, and protein sequence can also be downloaded. In addition, the co-expression network is shown, including orthologous genes and networks (global network, tissue-specific network, and stress-specific network) linked to the corresponding functional interface in diploid and polyploid wheat. Heuristic information, such as F-box ubiquitin family and GO terms related to flower development and regulation of circadian rhythm, is displayed. Moreover, the protein domain module shows an F-box-like functional domain with alignment start and end information; the expression pattern module shows the gene expression profiling in samples (*WAPO-A1* is specifically expressed in spike); and the predicted function module shows the possible functional clues of *WAPO-A1* identified by CFinder. For histone modification, the University of California at Santa Cruz (UCSC) [49] genome browser shows obvious peaks for *WAPO-A1* in H3K4me3 and H3K27me3, which are related to spikelet or flower development (Figure 3).

Supported analyses and tools

There are three category analysis tools, including GSEA, GO analysis, and *cis*-element enrichment. GSEA was based on the data analysis processing of PlantGSAD [36]. Here, we used functional data information as gene sets. The annotation entries with FDR < 0.05 were used and displayed. GO analysis was based on agriGOv2 [50] data processing. *Cis*-element analysis (such as the functions of sequence scan, gene name scan, and custom scan) can identify significantly enriched motifs in the promoter region of one gene and thus predict possible functions. We also provided BLAST (DNA and protein), ID conversion (in different genome versions), and

Sequence (gene ID or chromosome position) and FPKM extraction tools for users to conveniently obtain the information of a gene (Figure 2B).

Case study: function analysis of the known gene *WAPO1* using WheatCENet

Polyploidy (*i.e.*, whole-genome duplication) is an important evolutionary feature in the plant kingdom, particularly in flowering plants [51], after which individual genes may experience nonfunctionalization, neofunctionalization, or subfunctionalization [52]. For example, *WAPO1* is an orthologue of rice gene *ABERRANT PANICLE ORGANIZATION 1 (APO1)* and *Arabidopsis* gene *UNUSUAL FLORAL ORGANS (UFO)*. *UFO* acts synergistically with floral meristem identity factor *LEAFY (LFY)* and restricts the expression of the class B floral organ identity genes in *Arabidopsis* [53]. The interactions between the orthologs of *LFY* and *UFO* have also been demonstrated in rice, petunia, *Antirrhinum majus*, and pea, suggesting that *LFY* and *UFO* are conserved among species [53,54]. *EVERGREEN (EVG)* encodes a WOX homologous domain protein, which is only expressed in the initial lateral IM and participates in the activation of the *UFO* homologous gene *DOUBLE TOP (DOT)* in *petunia*. The *EVG* ortholog of *Compound Inflorescence (S)* and *UFO* ortholog *ANANTHA (AN)* have a similar effect on inflorescence meristems in tomatoes and related nightshades [53].

In wheat, *WAPO-A1* showed a highly significant association with total spikelet number (TSN) [55–58]. There are three genes [TraesCS7A02G481600 (*WAPO-A1*), TraesCS7B02G384000 (*WAPO-B1*), and TraesCS7-D02G468700 (*WAPO-D1*)] in AABBDD, two genes [TRID-C7AG067090 (not expressed) and TRIDC7BG060970] in AABB, no gene in AA, and one gene (evm.TU.Contig391.11) in DD. We searched the co-expression network of *WAPO1* using WheatCENet (Figure 4), and found that spike development-related gene *wheat FLO/LFY (WFL)* [59],

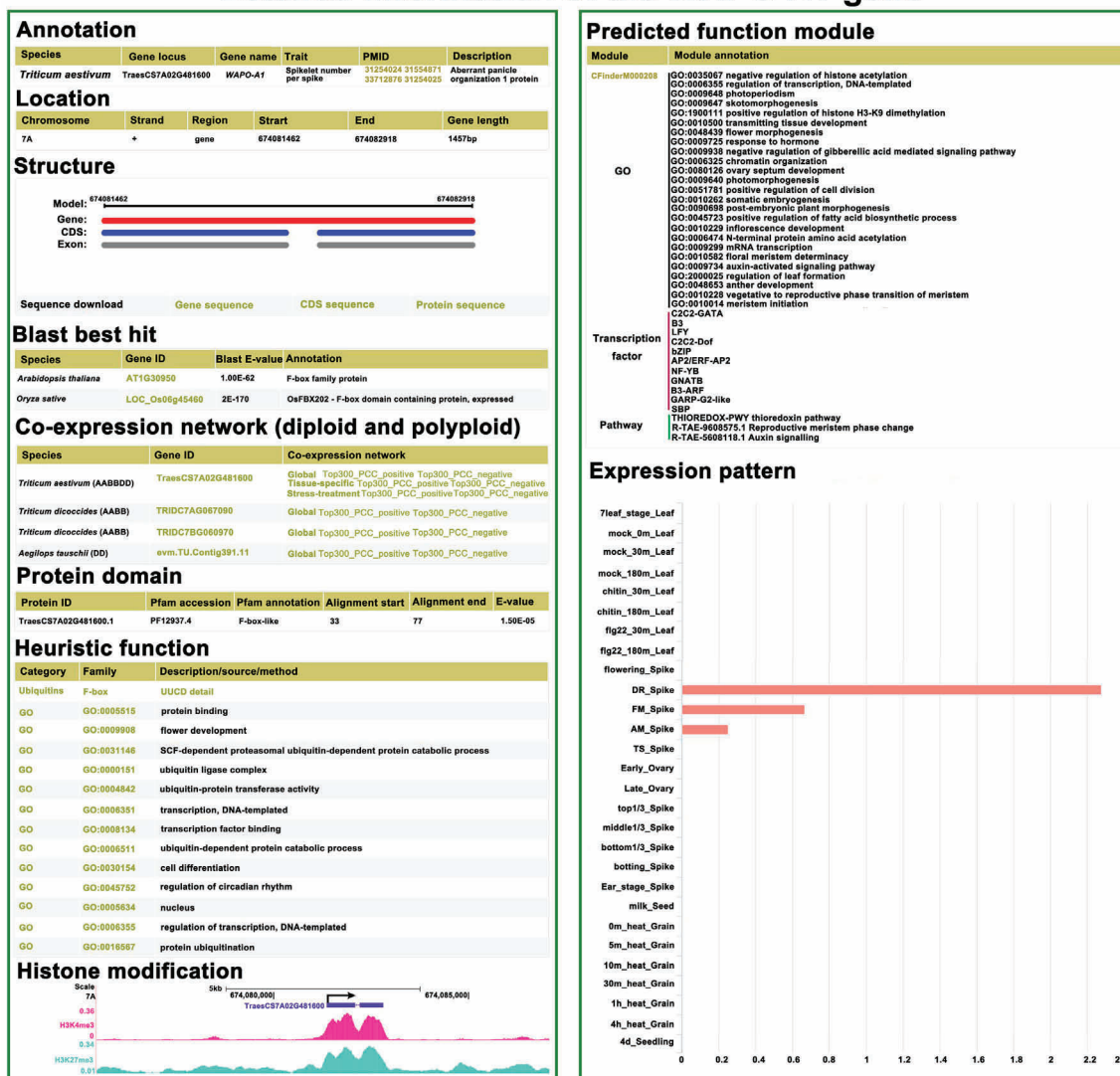
Detailed information of the *WAPO-A1* gene

Figure 3 Details of gene annotation

The interface of gene search results on the website includes gene annotation, gene location, structure, BLAST results, co-expression network, protein domain, heuristic function, histone modification, predicted function module, and expression pattern. The light yellow items link to the download page or detail page.

TaSPL14 [60], *TaSPL17* [61], and *WUSCHEL-related homeobox TaWOX4/TaWUS* [62,63], were directly co-expressed with *WAPO1* in AABBDD, but were not found in DD and AABB co-expression networks, even in the top 300 co-expressed genes. *WAPO-A1* also seems to be co-expressed with more genes related to spike development than *WAPO-B1* and *WAPO-D1*. *WAPO-A1b* in the H2 haplotype (present in the Chinese Spring IWGSC v1.0 genome) was associated with a higher SNS than *WAPO-A1c&d* in the H3 haplotype (present in *WAPO-B1* and *WAPO-D1*) and *WAPO-A1a* in the H2 haplotype [57]. *WAPO-B1* also has alleles conferring higher spikelet number at chromosome 7B [58]. Taken together, in addition to the effect of natural gene variation on gene expression, the co-expressed genes may also regulate spike development together with *WAPO1* in wheat.

GO analysis of the *WAPO1* co-expression network in AABBDD showed that these genes were related to reproductive shoot system development and post-embryonic development (Figure S6A). GSEA results for the network (Figure S6B) also indicated that the network of *WAPO1* responded to floral transition, in which some gene sets were significantly enriched with reproductive meristem phase change, miR156 miRNA_target_network (miR156 could regulate *TaSPL14* and *TaSPL17* [60,64]), and squamosal promoter binding protein (SBP) TF. After exploring the regulatory regions through motif analysis, we found a novel motif of CCWWWWWWGG (CArG) specifically identified in the promoter region of the *WAPO-A1* gene in AABBDD but not in AABB. We also found that the CArG motif appeared in the genes co-expressed with *WAPO1* (Figure 4).

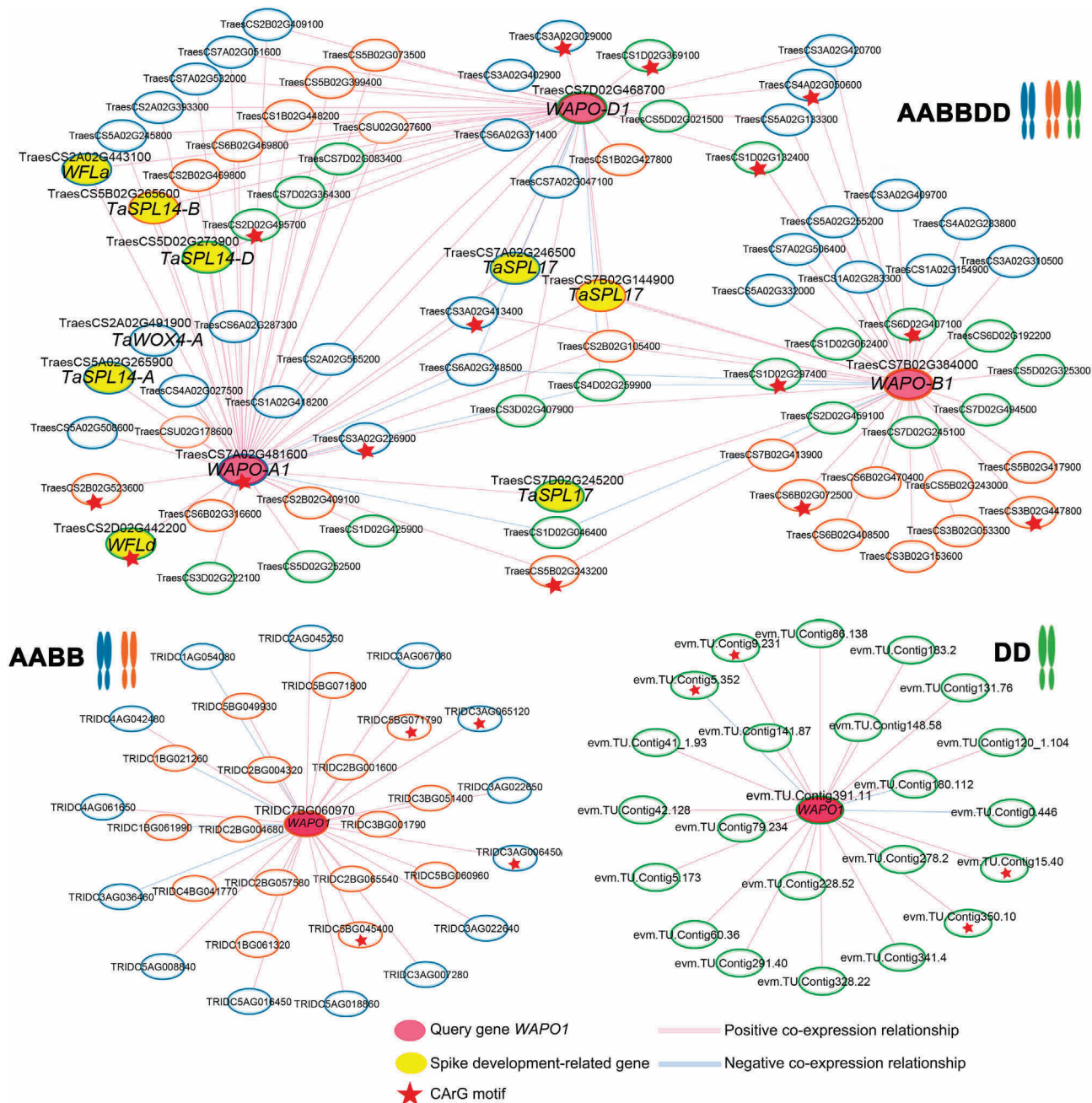


Figure 4 Global co-expression network analysis of *WAPO1*

Co-expression network of the *WAPO1* gene in DD, AABB, and AABBDD. The red node represents the query gene *WAPO1*, and the yellow node represents the spike development-related gene. The blue circle represents the subgenome A gene; the orange circle represents the subgenome B gene; the green circle represents the subgenome D gene; and the red star means the CARG motif included in the 3 kb upstream region of those genes. CARG motif, CCWWWWWWGG motif.

Previous studies revealed that the promoter of *WAPO1* ortholog contains predicted binding sites for the TFs of MADS-box and SBP-like genes [65]. The CARG motif is bound by MADS-box TFs that mainly participate in regulating flowering and floral/spikelet development [66–68]. Flower-specific TFs were confirmed to function in removing the H3K27me3 surrounding flower-specific regulatory ele-

ments in *Arabidopsis thaliana* [69,70]. In wheat, the enriched CARG-box motifs were found in the spikelet-reduced H3K27me3 peaks [71]. We also found that *WAPO1* is affected by H3K27me3 modification in different developmental stages using data from http://bioinfo.sibs.ac.cn/dynamic_epigenome/. H3K27me3 modification was lower in spikelet I at the booting stage than spikelet II at the flowering and seedling

stages, but the expression was the opposite. The newly gained CA_rG motif might have a significant role in the evolution of the *WAP01* gene for functioning in the development of spikelets in hexaploid wheat.

Discussion and conclusion

With the latest developments in sequencing technologies and assembly methods, many high-quality sequenced genomes of wheat have been produced [72]. The large amounts of data generated require a platform for experimenters to search for genes of interest. Many wheat databases have thus been developed, such as URGI and GrainGenes, which are data repositories of *T. aestivum* and its relatives, and provide tools like BLAST and JBrowse [7,8]. CerealsDB and Wheat-SnpHub-Portal mainly focus on visualizing single nucleotide polymorphism (SNP) data [9,10]; WheatGmap is committed to the analysis of whole-genome sequencing (WGS) and whole-exome sequencing (WES) data of *T. aestivum* [11]; and Triticeae-GeneTribe provides homeologous gene relationships among 12 Triticeae species and three out-groups (rice, maize, and *Arabidopsis*) [12]. None of these databases provide a comprehensive search engine for single genes, such as the gene annotation of *WAP01* (Figure 3) provided in WheatCENet. Unlike expression profiling databases such as expVIP and Wheat eFP Browser (https://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi), which only provide expression data of *T. aestivum* [13] WheatCENet integrates RNA-seq data of *T. aestivum*, *T. dicoccoides*, *T. urartu*, and *Ae. tauschii* from public platforms, which is useful for studying expression patterns of genes like *WAP01*. Existing co-expression networks like WheatNet [14], Knetminer [15], and WheatOmics [16] only provide DNA microarray datasets with an early genome assembly version or just include the co-expression networks of hexaploid wheat.

The WheatCENet database aims to provide an online service platform for comparative analysis of gene functions from a multidimensional network across diploid and polyploid wheat species. WheatCENet also includes comprehensive functional annotations (e.g., gene family, GO, miRNA, and metabolic pathways) to predict gene functions. Moreover, WheatCENet includes online tools like GSEA, GO, module, and motif analysis to determine the possible functions of gene sets, and BLAST and ID conversion allow for gene ID conversion between different genome versions. By using WheatCENet, wheat researchers can quickly find key information about the desired gene, predict biological process(es) the gene may participate in, and study the evolutionary history of the gene in wheat of different ploidies.

For genes that have been cloned but whose functional analysis is not very accurate in wheat, WheatCENet can help to further analyze the function. For example, we found that *Arabidopsis* *UFO* and wheat *WAP01* have the same 15th amino acid, phenylalanine (F) [55], so we also searched the motif of *UFO*, and identified a variant of the CA_rG motif in the promoter. Using our UCSC genome browser, we also found that H3K27me3 and H3K4me3 modifications have a peak in the gene body of *WAP01*. However, H3K27me3 modification has obvious differences in different tissues and developmental stages, while H3K4me3 does not. Spikelet-reduced H3K27me3

peaks carrying the enriched CA_rG-box motifs have been found in wheat [71]. Taken together, gaining the CA_rG motif may affect *WAP01* functions in regulating the total number of spikelets in AABBDD, though this requires further experimental verification. Gene analysis methods like that used for *WAP01* can also be applied to other cloned genes, to find the evolutionary differences in wheat of different ploidies.

WheatCENet has more possibilities for improvement, and we will continue to update it in the future. For example, RNA-seq samples from different growth stages, various stress treatments, and more tissue types can be integrated to build a more robust co-expression network. Concurrently, the 10+ Wheat Genomes Project has provided 15 assemblies for different wheat lines from global breeding programs. These wheat accessions and tetraploid durum wheat could be used to analyze co-expression networks and modules with the increase of corresponding RNA-seq samples, so as to more closely link the network with variation and evolution. Epigenomic data like DNase-seq, ChIP-seq, ATAC-seq, MNase-seq, MeDIP-seq, and BS-seq, which can be used to find peaks with gene expression, can be integrated to clarify the complex relationship between gene expression and chromatin structure. These new additions to WheatCENet will help to mine gene function and breeding in wheat.

By constructing and comparing networks in diploid and tetraploid wheat progenitors, we can dissect the origin and evolution of co-expression networks to better understand the underlying genetic basis for various agronomically important traits of bread wheat. The new WheatCENet platform could facilitate bench scientists identifying key candidate genes for functional studies, and provide genomic scientists a reliable source to decipher key molecular modules during the formation, evolution, and domestication of wheat.

Data availability

WheatCENet is freely available at <https://bioinformatics.copolar.cn/WheatCENet> and <http://bioinformatics.cau.edu.cn/WheatCENet/>.

Competing interests

The authors have declared no competing interests.

CRedit authorship contribution statement

Zhongqiu Li: Data curation, Formal analysis, Visualization, Writing – original draft. **Yiheng Hu:** Data curation, Formal analysis, Writing – review & editing. **Xuelian Ma:** Visualization, Formal analysis, Writing – review & editing. **Lingling Da:** Visualization, Formal analysis. **Jiajie She:** Software, Visualization. **Yue Liu:** Software, Formal analysis. **Xin Yi:** Visualization. **Yaxin Cao:** Software. **Wenyong Xu:** Conceptualization, Methodology, Writing – review & editing. **Yuannian Jiao:** Conceptualization, Methodology, Writing – review & editing. **Zhen Su:** Conceptualization, Methodology, Supervision, Project administration, Writing – review & editing. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (Grant Nos. 31970629 and 31771467 to ZS, and 31870209 to YJ).

Supplementary material

Supplementary data to this study can be found online at <https://doi.org/10.1016/j.gpb.2022.04.007>.

ORCID

ORCID 0000-0003-1355-7353 (Zhongqiu Li)
 ORCID 0000-0003-4404-7241 (Yiheng Hu)
 ORCID 0000-0002-0774-8439 (Xuelian Ma)
 ORCID 0000-0001-6572-0066 (Lingling Da)
 ORCID 0000-0003-4286-6411 (Jiajie She)
 ORCID 0000-0002-7926-9947 (Yue Liu)
 ORCID 0000-0003-4240-7039 (Xin Yi)
 ORCID 0000-0002-2439-3406 (Yaxin Cao)
 ORCID 0000-0002-7952-2244 (Wenying Xu)
 ORCID 0000-0002-8987-2782 (Yuannian Jiao)
 ORCID 0000-0001-6579-3847 (Zhen Su)

References

- Alexandratos N, Bruinsma J. World agriculture towards 2030/2050 - the 2012 revision. Rome: Food and Agriculture Organization; 2012.
- International Wheat Genome Sequencing Consortium, Appels R, Eversole K, Stein N, Feuillet C, Keller B, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 2018;361:eaar7191.
- Feldman M, Levy AA. Genome evolution due to allopolyploidization in wheat. *Genetics* 2012;192:763–74.
- Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, et al. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci U S A* 2002;99:8133–8.
- Guan JT, Garcia DF, Zhou Y, Appels R, Li AL, Mao L. The battle to sequence the bread wheat genome: a tale of the three kingdoms. *Genomics Proteomics Bioinformatics* 2020;18:221–9.
- Walkowiak S, Gao LL, Monat C, Haberer G, Kassa MT, Brinton J, et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* 2020;588:277–83.
- Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, et al. Linking the international wheat genome sequencing consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol* 2018;19:111.
- Blake VC, Woodhouse MR, Lazo GR, Odell SG, Wight CP, Tinker NA, et al. GrainGenes: centralized small grain resources and digital platform for geneticists and breeders. *Database (Oxford)* 2019;2019:baz065.
- Wilkinson PA, Winfield MO, Barker GL, Allen AM, BurrIDGE A, Coghill JA, et al. CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC Bioinformatics* 2012;13:219.
- Wang W, Wang Z, Li X, Ni Z, Hu Z, Xin M, et al. SnpHub: an easy-to-set-up web server framework for exploring large-scale genomic variation data in the post-genomic era with applications in wheat. *Gigascience* 2020;9:giaa060.
- Zhang L, Dong C, Chen Z, Gui L, Chen C, Li D, et al. WheatGmap: a comprehensive platform for wheat gene mapping and genomic studies. *Mol Plant* 2021;14:187–90.
- Chen Y, Song W, Xie X, Wang Z, Guan P, Peng H, et al. A collinearity-incorporating homology inference strategy for connecting emerging assemblies in the triticeae tribe as a pilot practice in the plant pangenomic era. *Mol Plant* 2020;13:1694–708.
- Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiol* 2016;170:2172–86.
- Lee T, Hwang S, Kim CY, Shim H, Kim H, Ronald PC, et al. WheatNet: a genome-scale functional network for hexaploid bread wheat, *Triticum aestivum*. *Mol Plant* 2017;10:1133–6.
- Hassani-Pak K, Castellote M, Esch M, Hindle M, Lysenko A, Taubert J, et al. Developing integrated crop knowledge networks to advance candidate gene discovery. *Appl Transl Genom* 2016;11:18–26.
- Ma SW, Wang M, Wu JH, Guo WL, Chen YM, Li GW, et al. WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat. *Mol Plant* 2021;14:1965–8.
- Duncan O, Trosch J, Fenske R, Taylor NL, Millar AH. Resource: mapping the *Triticum aestivum* proteome. *Plant J* 2017;89:601–16.
- Blake VC, Birkett C, Matthews DE, Hane DL, Bradbury P, Jannink JL. The Triticeae toolbox: combining phenotype and genotype data to advance small-grains breeding. *Plant Genome* 2016;9:e2012–4.
- Remita MA, Lord E, Agharbaoui Z, Leclercq M, Badawi MA, Sarhan F, et al. A novel comprehensive wheat miRNA database, including related bioinformatics software. *Curr Plant Biol* 2016;7:8:31–3.
- Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 2017;357:93–7.
- Ling HQ, Ma B, Shi X, Liu H, Dong L, Sun H, et al. Genome sequence of the progenitor of wheat a subgenome *Triticum urartu*. *Nature* 2018;557:424–8.
- Zhao G, Zou C, Li K, Wang K, Li T, Gao L, et al. The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat Plants* 2017;3:946–55.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol* 2014;32:915–25.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137.
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 2012;40:e49.
- Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T. ATTED-II in 2016: A plant coexpression database towards lineage-specific coexpression. *Plant Cell Physiol* 2016;57:e5.

- [31] Rao X, Dixon RA. Co-expression networks for plant biology: why and how. *Acta Biochim Biophys Sin (Shanghai)* 2019;51:981–98.
- [32] You Q, Xu W, Zhang K, Zhang L, Yi X, Yao D, et al. ccNET: database of co-expression networks with functional modules for diploid and polyploid *Gossypium*. *Nucleic Acids Res* 2017;45:D1090–9.
- [33] Guo WL, Xin MM, Wang ZH, Yao YY, Hu ZR, Song WJ, et al. Origin and adaptation to high altitude of Tibetan semi-wild wheat. *Nat Commun* 2020;11:5085.
- [34] He X, Fang JJ, Li JJ, Qu BY, Ren YZ, Ma WY, et al. A genotypic difference in primary root length is associated with the inhibitory role of transforming growth factor-beta receptor-interacting protein-1 on root meristem size in wheat. *Plant J* 2014;77:931–43.
- [35] Adamcsek B, Palla G, Farkas IJ, Derenyi I, Vicsek T. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 2006;22:1021–3.
- [36] Ma X, Yan H, Yang J, Liu Y, Li Z, Sheng M, et al. PlantGSAD: a comprehensive gene set annotation database for plant species. *Nucleic Acids Res* 2021;9:D1456–67.
- [37] Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* 2016;32:309–11.
- [38] Walls RL, Athreya B, Cooper L, Elser J, Gandolfo MA, Jaiswal P, et al. Ontologies as integrative tools for plant science. *Am J Bot* 2012;99:1263–75.
- [39] Conesa A, Gotz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* 2008;2008:619832.
- [40] Hartmann T. From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry* 2007;68:2831–46.
- [41] Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A, et al. Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Res* 2018;46:D1181–9.
- [42] Schlapfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, et al. Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol* 2017;173:2041–59.
- [43] Forslund SK, Kaduk M, Sonnhammer ELL. Evolution of protein domain architectures. *Methods Mol Biol* 2019;1910:469–504.
- [44] Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 2016;44:D279–85.
- [45] Song XW, Li Y, Cao XF, Qi YJ. MicroRNAs and their regulatory roles in plant-environment interactions. *Annu Rev Plant Biol* 2019;70:489–525.
- [46] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;42:D68–73.
- [47] Alptekin B, Akpinar BA, Budak H. A comprehensive prescription for plant miRNA identification. *Front Plant Sci* 2016;7:2058.
- [48] Sonnhammer ELL, Ostlund G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res* 2015;43:D234–9.
- [49] Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, et al. The UCSC genome browser database: 2016 update. *Nucleic Acids Res* 2016;44:D717–25.
- [50] Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* 2017;45:W122–9.
- [51] Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A* 2009;106:13875–9.
- [52] Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science* 2000;290:1151–5.
- [53] Mckim S, Hay A. Patterning and evolution of floral structures - marking time. *Curr Opin Genet Dev* 2010;20:448–53.
- [54] Liang WH, Shang F, Lin QT, Lou C, Zhang J. Tillering and panicle branching genes in rice. *Gene* 2014;537:1–5.
- [55] Muqaddasi QH, Brassac J, Koppolu R, Plieske J, Ganai MW, Roder MS. *TaAPO-A1*, an ortholog of rice *ABERRANT PANICLE ORGANIZATION 1*, is associated with total spikelet number per spike in elite European hexaploid winter wheat (*Triticum aestivum* L.) varieties. *Sci Rep* 2019;9:13853.
- [56] Voss-Fels KP, Keeble-Gagnere G, Hickey LT, Tibbits J, Nagornyy S, Hayden MJ, et al. High-resolution mapping of rachis nodes per rachis, a critical determinant of grain yield components in wheat. *Theor Appl Genet* 2019;132:2707–19.
- [57] Kuzay S, Xu YF, Zhang JL, Katz A, Pearce S, Su ZQ, et al. Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theor Appl Genet* 2019;132:2689–705.
- [58] Corsi B, Obinu L, Zanella CM, Cutrupi S, Day R, Geyer M, et al. Identification of eight QTL controlling multiple yield components in a German multi-parental wheat population, including *Rht24*, *WAPO-A1*, *WAPO-B1* and genetic loci on chromosomes 5A and 6A. *Theor Appl Genet* 2021;134:1435–54.
- [59] Shitsukawa N, Takagishi A, Ikari C, Takumi S, Murai K. *WFL*, a wheat *FLORICAULA/LEAFY* ortholog, is associated with spikelet formation as lateral branch of the inflorescence meristem. *Genes Genet Syst* 2006;81:13–20.
- [60] Cao J, Liu KY, Song WJ, Zhang JN, Yao YY, Xin MM, et al. Pleiotropic function of the *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE* gene *TaSPL14* in wheat plant architecture. *Planta* 2021;253:44.
- [61] Zhang B, Liu X, Zhao GY, Mao XG, Li A, Jing RL. Molecular characterization and expression analysis of *Triticum aestivum* squamosa-promoter binding protein-box genes involved in ear development. *J Int Plant Biol* 2014;56:571–81.
- [62] Rathour M, Sharma A, Kaur A, Upadhyay SK. Genome-wide characterization and expression and co-expression analysis suggested diverse functions of *WOX* genes in bread wheat. *Heliyon* 2020;6:e05762.
- [63] Li Z, Liu D, Xia Y, Li Z, Jing D, Du J, et al. Identification of the WUSCHEL-related homeobox (*WOX*) gene family, and interaction and functional analysis of *TaWOX9* and *TaWUS* in wheat. *Int J Mol Sci* 2020;21:1581.
- [64] Liu J, Cheng X, Liu P, Sun J. miR156-targeted SBP-box transcription factors interact with *DWARF53* to regulate *TEOSINTE BRANCHED1* and *BARREN STALK1* expression in bread wheat. *Plant Physiol* 2017;174:1931–48.
- [65] Kusters E, Della Pina S, Castel R, Souer E, Koes R. Changes in *cis*-regulatory elements of a key floral regulator are associated with divergence of inflorescence architectures. *Development* 2015;142:2822–31.
- [66] de Folter S, Angenent GC. *trans* meets *cis* in MADS science. *Trends Plant Sci* 2006;11:224–31.
- [67] Guo SY, Xu YY, Liu HH, Mao ZW, Zhang C, Ma Y, et al. The interaction between *OsMADS57* and *OsTB1* modulates rice tillering via *DWARF14*. *Nat Commun* 2013;4:1566.
- [68] Tao J, Liang W, An G, Zhang D. *OsMADS6* controls flower development by activating rice *FACTOR OF DNA METHYLATION LIKE1*. *Plant Physiol* 2018;177:713–27.
- [69] Mozgova I, Hennig L. The polycomb group protein regulatory network. *Annu Rev Plant Biol* 2015;66:269–96.
- [70] Wang H, Liu C, Cheng J, Liu J, Zhang L, He C, et al. *Arabidopsis* flower and embryo developmental genes are repressed in seedlings

- by different combinations of polycomb group proteins in association with distinct sets of *cis*-regulatory elements. *PLoS Genet* 2016;12:e1005771.
- [71] Wang M, Li Z, Zhang Y, Zhang Y, Xie Y, Ye L, et al. An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. *Plant Cell* 2021;33:865–81.
- [72] Uauy C. Wheat genomics comes of age. *Curr Opin Plant Biol* 2017;36:142–8.