PCP
PLANT & CELL PHYSIOLOGY

Database Paper

Peanut genome resource: a functional genomics platform for *Arachis hypogaea*

Chien-Wen Yang[1,‡], Chi-Nga Chow[2,3,‡], Hua Chen[4,5], Kuan-Chieh Tseng[3], Nai-Yun Wu[3], Yuhui Zhuang[4,5], Weijian Zhuang[4,5,*] and Wen-Chi Chang[1,3,*]

[1]Graduate Program in Translational Agricultural Sciences, National Cheng Kung University and Academia Sinica, No.1 University Road, East Dist., Tainan City 701, Taiwan, [2]School of Molecular Sciences, Arizona State University, 551 E University Dr, Physical Sciences Center PSD 104, Tempe, AZ 85287, USA, [3]Institute of Tropical Plant Sciences and Microbiology, College of Biosciences and Biotechnology, National Cheng Kung University, No.1 University Road, East Dist., Tainan City 701, Taiwan, [4]Key Laboratory of Crop Genetics and Comprehensive Utilization, Ministry of Education, No.15 Shangxiadian Road, Fuzhou 350002, China, [5]Research Center of Leguminous Oil Plant Genetics and Systems Biology, Key Laboratory of Fujian-Taiwan Crop Biological Breeding and Agriculture, Fujian Agriculture and Forestry University, No.15 Shangxiadian Road, Fuzhou 350002, China

*Corresponding authors: Weijian Zhuang, E-mail, weijianz@fafu.edu.cn; Wen-Chi Chang, E-mail, sarah321@mail.ncku.edu.tw

[‡]Chien-Wen Yang and Chi-Nga Chow contributed equally to this work.

Received 5 July 2025; Accepted 30 November 2025

Abstract

Peanut is a vital oilseed legume with considerable nutritional and economic value worldwide. Considering the global agricultural importance of this legume, researchers have sequenced the whole genomes of both wild and cultivated peanut varieties. Furthermore, databases such as PeanutBase have been established to advance peanut research and breeding. These databases compile extensive genomic resources, including reference genomes, gene annotations, and molecular markers. However, very few genes in cultivated peanut have been functionally characterized. To address this research gap, we developed an enhanced version of the Peanut Genome Resource (PGR) platform (https://pgr.itps.ncku.edu.tw), specifically focusing on providing comprehensive genomic, annotation, and phenotypic data for *Arachis hypogaea*, especially the Chinese peanut var. Shitouqi. This updated platform integrates an extensive range of genomic annotations—such as information on gene functions, protein domains, transcription factor (TF) families, gene ontology terms, and Kyoto Encyclopedia of Genes and Genomes pathways. Furthermore, PGR offers gene expression profiles across tissues and conditions as well as tools for differential gene expression and coexpression analyses. To the best of our knowledge, PGR is the first peanut-related platform to incorporate advanced bioinformatics tools for *cis*-regulatory element analyses, such as those aimed at predicting TF–binding sites; identifying CpNpG islands, tandem repeats, and single sequence repeats; and performing in silico polymerase chain reaction assays for genetic markers. With its user-friendly interface and comprehensive analytical capabilities, PGR serves as a powerful platform for advancing research on peanut genetics, breeding, and functional genomics.

Keywords: *Arachis hypogaea* • *cis*-regulatory element • coexpression • functional gene analysis • gene evaluation • molecular marker

Introduction

Cultivated peanut (*Arachis hypogaea* L.) is a major agricultural crop that originated in South America and is now cultivated worldwide (Arya et al. 2016). As an oilseed legume, peanut is the fourth-largest oil-producing crop, with global production reaching approximately 51.78 million metric tons in June 2025 (U.S. Department of Agriculture 2025a). This substantial yield highlights the importance of peanut in global agriculture and its contribution to the oilseed market. Furthermore, peanut has considerable commercial value because of its nutritional and medicinal properties (Çiftçi and Suna 2022). This legume is rich in essential nutrients and bioactive compounds, which makes it valuable for both dietary consumption and potential therapeutic applications. Given the importance of peanut in human life, the selection of superior traits and cultivars has long been a key objective for genetic breeders. Consequently, research on the peanut genome has become indispensable.

Foundational cytogenetic studies in the early 20th century established the basic genomic attributes of cultivated peanut. In the 1930s, cytogenetic investigations revealed a chromosome count of $2n = 4x = 40$ in *A. hypogaea*, confirming it as an allotetraploid species (Husted 1931, 1933, 1936). Cytological analyses indicated that the prostrate (runner) and erect (bunch) forms of cultivated peanut exhibit similar chromosome numbers, morphological traits, and meiotic behaviors,

Plant and Cell Physiology, 2026, 67(5),859–873. https://doi.org/10.1093/pcp/pcaf165, Advance Access publication on 11 December 2025, available online at https://academic.oup.com/pcp

suggesting that both varieties originated from a common ancestor (Husted 1936, Bertioli et al. 2016). These findings led to the 'single-origin hypothesis,' which suggests that modern cultivated peanut originated through hybridization between two diploid ancestors, followed by genome doubling and domestication. Because the genus *Arachis* contains more than 80 wild species with diverse genome types such as A, B, D, F, and K (Bertioli et al. 2011, Zhuang et al. 2019), identifying the diploid donors of the cultivated allotetraploid became a long-standing challenge in peanut genome research. Molecular approaches such as restriction fragment length polymorphism, genomic *in situ* hybridization, and analyses based on simple sequence repeat (SSR) markers highlighted the wild diploid species *Arachis duranensis* (A genome) and *Arachis ipaensis* (B genome) as the most likely ancestors from which cultivated peanut originated through hybridization and polyploidization (Kochert et al. 1996, Moretzsohn et al. 2012, Seijo et al. 2007).

With the advent of high-throughput sequencing technologies, substantial advances have been made in peanut genomics. In 2016, a major milestone in peanut research was reached with the whole-genome sequencing of the aforementioned ancestors (Bertioli et al. 2016). The genome assemblies of *A. duranensis* and *A. ipaensis* provided the first chromosome-scale references that clarified the evolutionary origin of cultivated peanut (Bertioli et al. 2016). This study noted a high volume of transposable elements, with 36 734 and 41 840 nontransposable element genes in *A. duranensis* and *A. ipaensis*, respectively. The researchers further identified clusters of disease resistance–related loci in the ancestral species and found evidence of structural rearrangements in the A subgenome of cultivated peanut. Building on this foundation, two different studies have revealed chromosome-scale assemblies of cultivated peanut genomes. Zhuang et al. (2019) published the reference genome of *A. hypogaea* var. Shitouqi, a widely grown Chinese cultivar. The assembly spanned 2.51 Gb, was anchored to 20 pseudomolecules, and included annotations for 83 709 protein-coding genes. Approximately 77.65% of the genome comprised repetitive elements. Analyses highlighted B-subgenome dominance in terms of gene count and expression. In addition, a long terminal repeat expansion was observed, predominantly in the A subgenome. The sequenced genome enabled the mapping of loci and candidate genes for multiple agronomically important traits, such as seed size, testa color, and foliar disease resistance. In parallel, Bertioli et al. (2019) published the genomic assembly of the Tifrunner cultivar. This assembly spanned 2.54 Gb and included sequences from 20 chromosomes, with repeats accounting for approximately 74% of the genome. It captured 99.3% of all bases and included 66 469 predicted genes, of which 31 359 originated from the A subgenome and 35 110 from the B subgenome. Comparative analyses of Tifrunner against its diploid ancestors revealed mostly one-to-one collinearity; however, large inversions and a reciprocal translocation among specific homeologous chromosome pairs were also observed. These structural variations, together with widespread homeologous exchanges, likely contributed to the phenotypic diversity of cultivated peanut. Conversely, the uniform single-nucleotide polymorphism (SNP) "fingerprints" shared by *A. hypogaea* and *Arachis monticola* indicate a single polyploid origin of cultivated peanut.

As sequencing costs decreased and analytical pipelines matured, broader access to genomic data from various *Arachis* species facilitated the integration and standardization of corresponding datasets by community resources. As part of this broader community initiative, PeanutBase—a comprehensive peanut genome database—was established with support from the International Peanut Genomics Consortium to provide researchers with the genetic and genomic data of various *Arachis* species (Dash et al. 2016). PeanutBase offers access to genome assemblies, genetic markers, quantitative trait loci (QTL), gene expression datasets, and bioinformatics tools for visualization and comparative analysis. According to the US Department of Agriculture (U.S. Department of Agriculture 2025b), China has recently become the world's leading producer of peanut, contributing to approximately 37% of the global yield. To facilitate research on cultivated varieties, researchers have sequenced the genome of the Chinese peanut cultivar *A. hypogaea* var. Shitouqi and developed a dedicated genome browser, the Peanut Genome Resource (PGR) platform, to facilitate the retrieval of genomic information (Zhuang et al. 2019). However, comprehensive genomic and phenotypic data for Chinese cultivars remain underrepresented in public repositories such as PeanutBase, limiting research and breeding initiatives focused on these varieties.

To address the indicated gap, we developed an enhanced version of the PGR platform, aiming to provide comprehensive functional and regulatory information based on genomic and phenotypic data specific to the Chinese peanut cultivar *A. hypogaea* var. Shitouqi. PGR facilitates several aspects of peanut research. For example, it (i) offers genome annotations that integrate information on gene function, protein domains, transcription factor (TF) families, gene ontology (GO) terms, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways; (ii) offers tools for the visualization of expression profiles across 29 tissues and experimental conditions; (iii) integrates 18 genetic maps containing SNPs, SSR markers, and trait-associated QTLs derived from PeanutBase and prior studies; (iv) provides access to phenotypic information for various cultivated peanut species, selected cultivars, mutants, and accessions; (v) offers a Regulatory Element Analysis function for predicting TF-binding sites (TFBSs) within user-defined promoter regions; (vi) provides a sequence analysis module for identifying CpNpG islands, tandem repeats, and SSR regions; and (vii) offers an *in silico* polymerase chain reaction (PCR) tool (based on the Exonerate package) for predicting primer targets in user-defined genomic sequences. By integrating the aforementioned resources and tools within a user-friendly web interface, PGR provides a robust platform for peanut research, supporting genetic studies, breeding efforts, and functional genomics analyses of *A. hypogaea* var. *Shitouqi* and related cultivars.

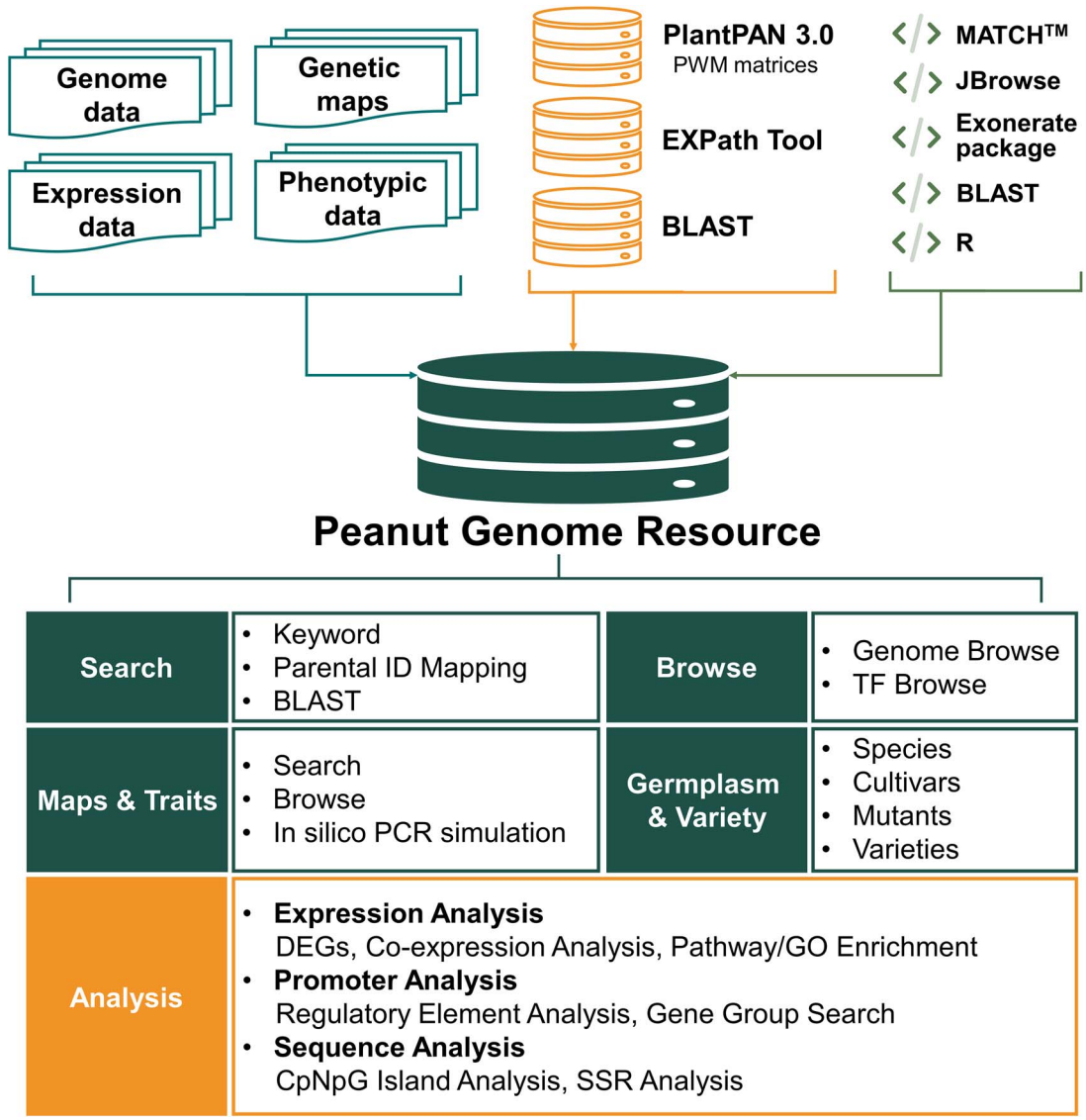


Figure 1. Overview of the PGR platform. The platform integrates genomic, genetic, expression, and phenotypic data from cultivated peanut. It offers six key functions: browse, search, germplasm & variety, maps & traits, analysis (expression, promoter, and sequence analyses), and download.

Results and Discussion

Database construction and general functions

In this study, we updated the PGR platform to provide comprehensive genomic and phenotypic information on *A. hypogaea* var. Shitouqi. The genome assembly of this cultivar comprises 20 chromosome-scale scaffolds corresponding to the diploid A and B subgenomes (chromosomes A01–A10 and B01–B10, respectively) (Zhuang et al. 2019). The updated platform indicated that the genome contains 75 810 protein-coding genes, with an average gene length of 3155 bp, an average exon count of 5.14 per gene, and an average protein length of 388 amino acids. The genes are distributed across the A subgenome (33 449 genes) and the B subgenome (41 510 genes), with 851 genes located in the unassembled region. PGR provides descriptive data for 12 *Arachis* species; the data include images, resequencing data, and genetic variation–related information for 51 peanut accessions, 59 mutants, and 6 variants or ecotypes. Regarding genetic and trait data, PGR incorporates genetic linkage maps containing 21 069 SNPs, SSR markers, and QTLs. For comparative analysis, the transcriptome component of PGR includes 29 RNA sequencing (RNA-seq) datasets and 32 microarray datasets representing various tissues and conditions (Zhuang et al. 2019). This platform operates on an Apache Web server running the Linux operating system, with data stored in a MySQL database. The Web interface was developed using PHP, HTML, and JavaScript. PGR provides six key functions: browse, search, germplasm & variety, maps & traits, analysis, and download. An overview of the PGR interface is depicted in Fig. 1.

Browse

The JBrowse tool has been integrated into the Browse section of the PGR website to facilitate genome-wide visualization

and gene exploration (Skinner et al. 2009). Users can search by gene identifiers (IDs) or specify genomic coordinates to locate genes of interest. In addition to facilitating standard genome browsing, PGR enables users to search for TFs from different families by including a dedicated TF Browser (Supplementary Fig. S1a). The display provides detailed information on the TF of interest—e.g. gene ID, transcript ID, genomic coordinates, gene description, and annotated protein domains, accompanied by graphical visualization (Supplementary Fig. S1b).

Search

To facilitate efficient retrieval of gene information, the PGR platform provides various keyword-based search functions. In addition to supporting basic searches with gene IDs, gene names, and functional descriptions, PGR facilitates advanced queries with GO terms, KEGG ortholog and pathway IDs, and InterPro domain IDs. Each search result links to a detailed gene information page that displays GO and KEGG pathway annotations and protein domain classifications (Supplementary Fig. S2a–c). The page also provides access to the coding sequence, the protein sequence, and downloadable FASTA files for downstream analyses. These integrated features enhance data accessibility and facilitate gene information retrieval.

PGR offers a Parental ID Mapping function that enables users to identify homeologous transcript IDs between the allotetraploid AABB genome of *A. hypogaea* var. Shitouqi and its parental diploid genomes, *A. duranensis* (AA) and *A. ipaensis* (BB). On the one hand, users can input a transcript ID from the AABB genome to retrieve the corresponding homeologous transcript(s) in the parental genomes; on the other hand, they can use transcript IDs from the parental genomes to identify their counterparts in the AABB genome. The output displays protein domains for both the query transcript and its mapped homeolog, along with the detailed protein sequence alignments (Fig. 2a–c). For parental diploid transcripts, PGR provides external links that direct users to corresponding gene search pages on PeanutBase.

Expression analysis

To facilitate the analysis of gene expression in *A. hypogaea* var. Shitouqi, PGR integrates 29 RNA-seq datasets and 32 microarray datasets corresponding to various tissues, developmental stages, and stress conditions. Users can explore the expression profiles of target genes under different conditions by searching using transcript IDs or microarray probe IDs. Expression levels are visualized as line charts, enabling an intuitive comparison of gene expression dynamics across experimental contexts. To facilitate the identification of key genes and regulatory networks, PGR enables users to perform comparative analyses of expression data obtained from RNA-seq and microarray data. Several analytical tools are available in PGR: (i) tissue-specific search identifies genes specifically expressed in the target tissue (Fig. 3a); (ii) differentially expressed genes (DEGs) Search lists genes with different expression profiles in leaves subjected to hormone or abiotic stress treatments (Fig. 3b); (iii) KEGG pathway and GO enrichment analyses unveil significantly enriched (P-value < 0.5) biological processes and signaling cascades associated with user-defined gene sets (Fig. 3c); and (iv) Coexpression analysis detects gene sets with similar expression patterns on the basis of either a single gene or a set of genes. PGR offers two modes of gene expression analysis: correlation of single gene and correlation network. These modes reveal genes coexpressed with the gene of interest or identify coexpression patterns within a given set of genes (Fig. 4a and b).

Promoter and sequence analyses

To facilitate the identification of putative *cis*-regulatory elements, PGR includes a regulatory element analysis function that scans the upstream and downstream regions of target genes (Fig. 5a–c). This function reveals predicted TFBSs, CpNpG islands, and tandem repeats. Users can refine analyses by selecting species-specific TF-binding matrices, uploading custom motifs, and specifying the length of the genomic region to be analyzed. In addition to enabling the prediction of TFBSs, PGR facilitates the identification of CpNpG islands and tandem repeats, thereby serving as a comprehensive platform for investigating gene regulatory elements.

PGR also offers a gene group analysis function for the identification of shared regulatory mechanisms within a set of genes. This function identifies TFs whose binding sites are located within the regulatory regions of multiple target genes, thereby highlighting potential common regulators and coordinated transcriptional control across the gene set.

Germplasm & variety

The germplasm & variety function of PGR provides an overview of cultivated and wild peanut species, covering four key categories: species, cultivars, mutants, and varieties. The species page displays descriptive information and resource links for 6 cultivated peanut ecotypes and 11 wild peanut species. The cultivars page features 51 peanut cultivars with phenotypic data and breeding histories. The mutant's page features 59 mutant lines with phenotypic descriptions and photographs. Finally, the varieties page features six cultivated peanut varieties and presents information on their landrace origins.

Maps & traits

The maps & traits function of PGR enables users to search for genetic markers, SSRs, SNPs, and QTLs across 18 genetic linkage maps. Marker information can be accessed using two primary methods: by entering a keyword or primer sequence or by browsing the marker table. The output page provides detailed data, including information on marker name, type, motif, reference, and sequence as well as PCR primer sequence, if available. In the linkage maps section, users can locate the genetic map(s) containing the queried marker and the corresponding map position (Fig. 6a) and access relevant visual representations (Fig. 6b). PGR further enables users to perform *in*

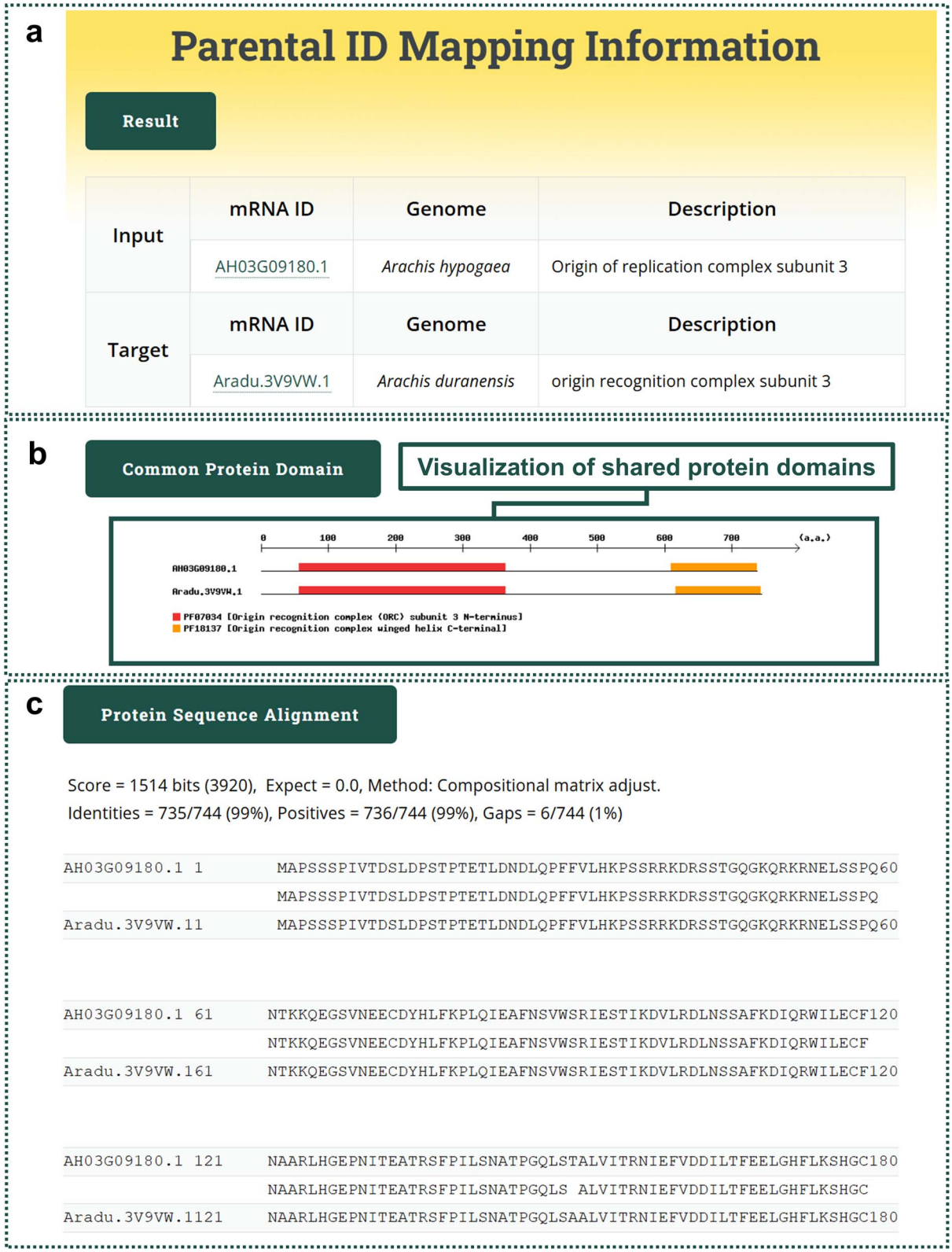
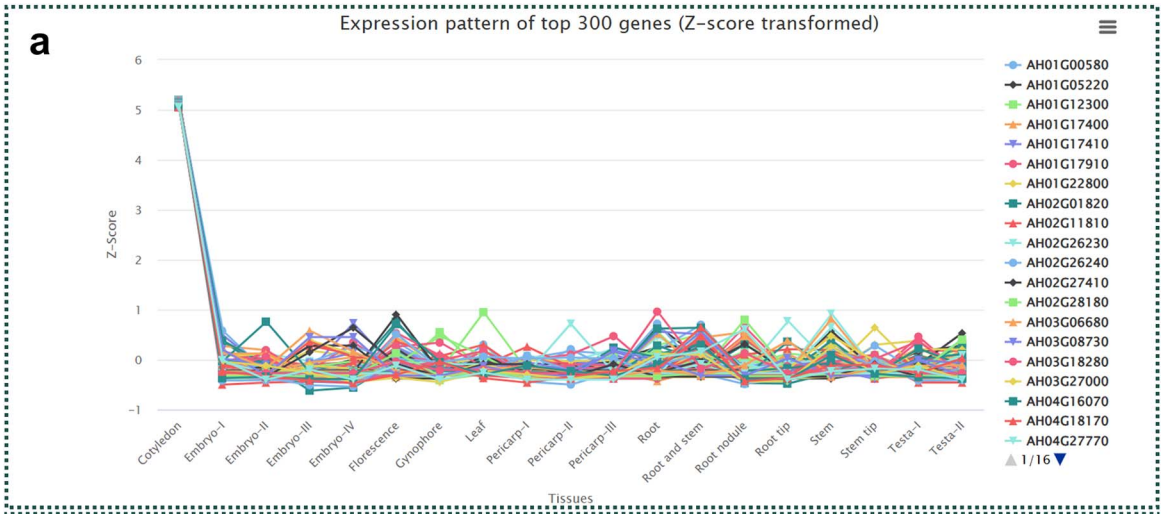
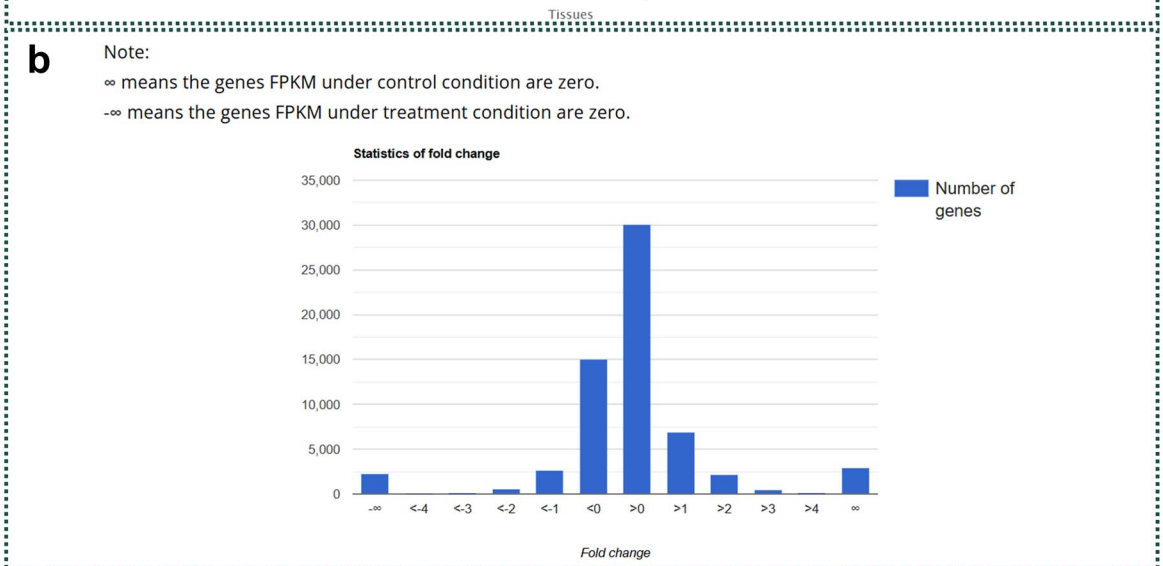


Figure 2. Screenshots depicting the output of the Parental ID Mapping function. Mapping result for AH03G09180.1 includes (a) basic information for both query and target transcripts, (b) visualization of shared protein domains, and (c) a detailed protein sequence alignment between the query and corresponding homoeologous transcript.





c

	Pathway ID	Pathway name	Hit number (Query)	Percentage in query	P-value	FDR	View pathway
+	00910	Nitrogen metabolism	7	43.75%	2.50E-12	5.00E-11	View
+	00250	Alanine, aspartate and glutamate metabolism	4	25.00%	1.66E-5	1.66E-4	View
+	00630	Glyoxylate and dicarboxylate metabolism	4	25.00%	4.44E-5	2.96E-4	View

Query ID	Gene ID	Definition of KEGG ortholog
AH02G05660	AH02G05660	malate dehydrogenase [EC:1.1.1.37]
AH03G26540	AH03G26540	glutamate synthase (ferredoxin) [EC:1.4.7.1]
AH06G02890	AH06G02890	glutamate synthase (ferredoxin) [EC:1.4.7.1]
AH16G05380	AH16G05380	glutamate synthase (ferredoxin) [EC:1.4.7.1]

Figure 3. Overview of the expression analysis function. (a) Output of the tissue-specific gene search function for cotyledon, displaying the expression profiles of the top 300 tissue-specific genes. (b) Output of the DEG search function for "leaf treated with abscisic acid," displaying the distribution of DEGs on the basis of log2 fold-change values. (c) Output of the pathway/GO enrichment analysis function, displaying significantly enriched KEGG pathways and GO terms. The example presents pathway enrichment results with enrichment scores and adjusted *P* values.

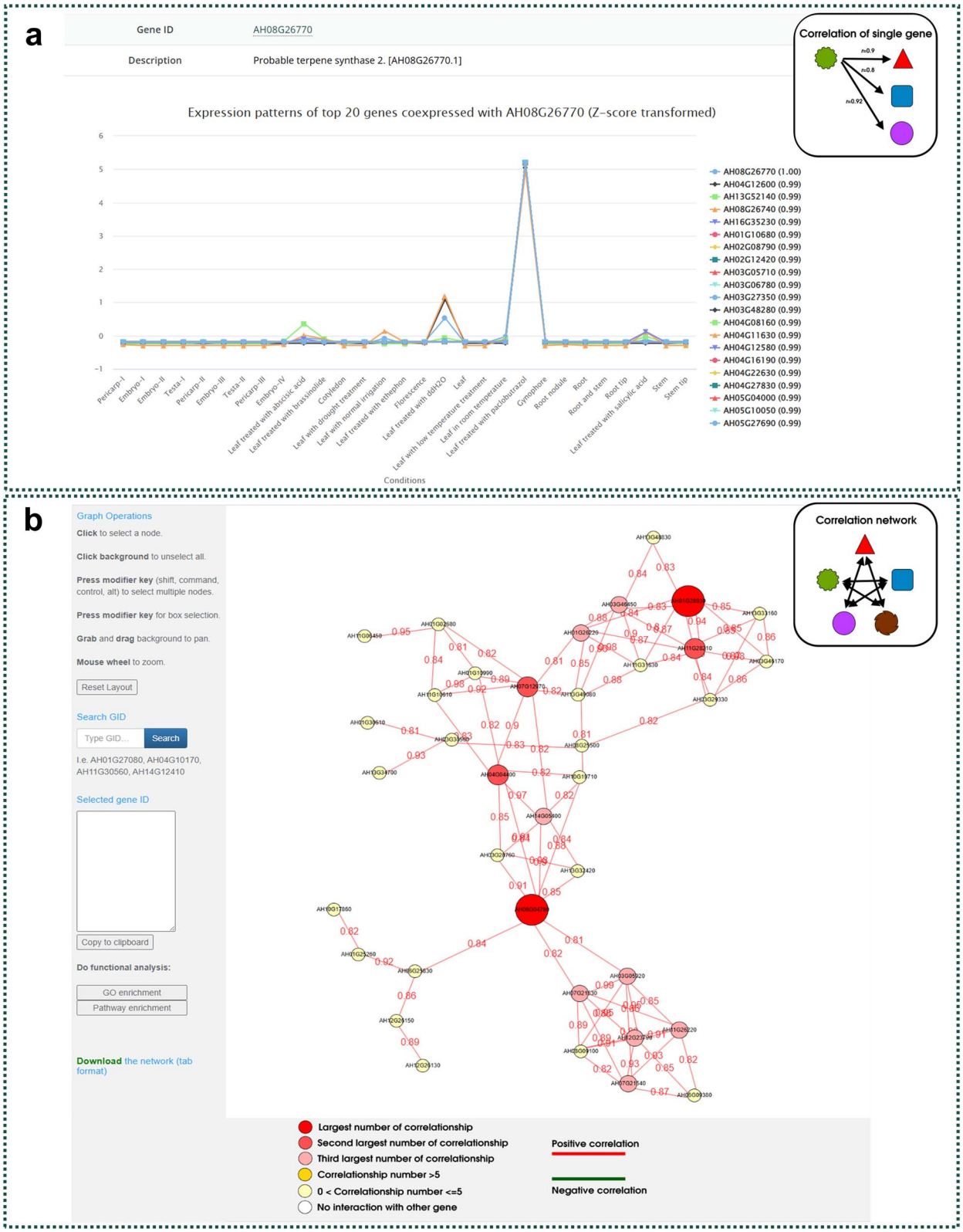


Figure 4. Output of the coexpression analysis function. (a) Coexpression profile of AH08G26770 generated using the single-gene correlation mode. (b) Coexpression network of WRKY genes generated using the correlation network mode.

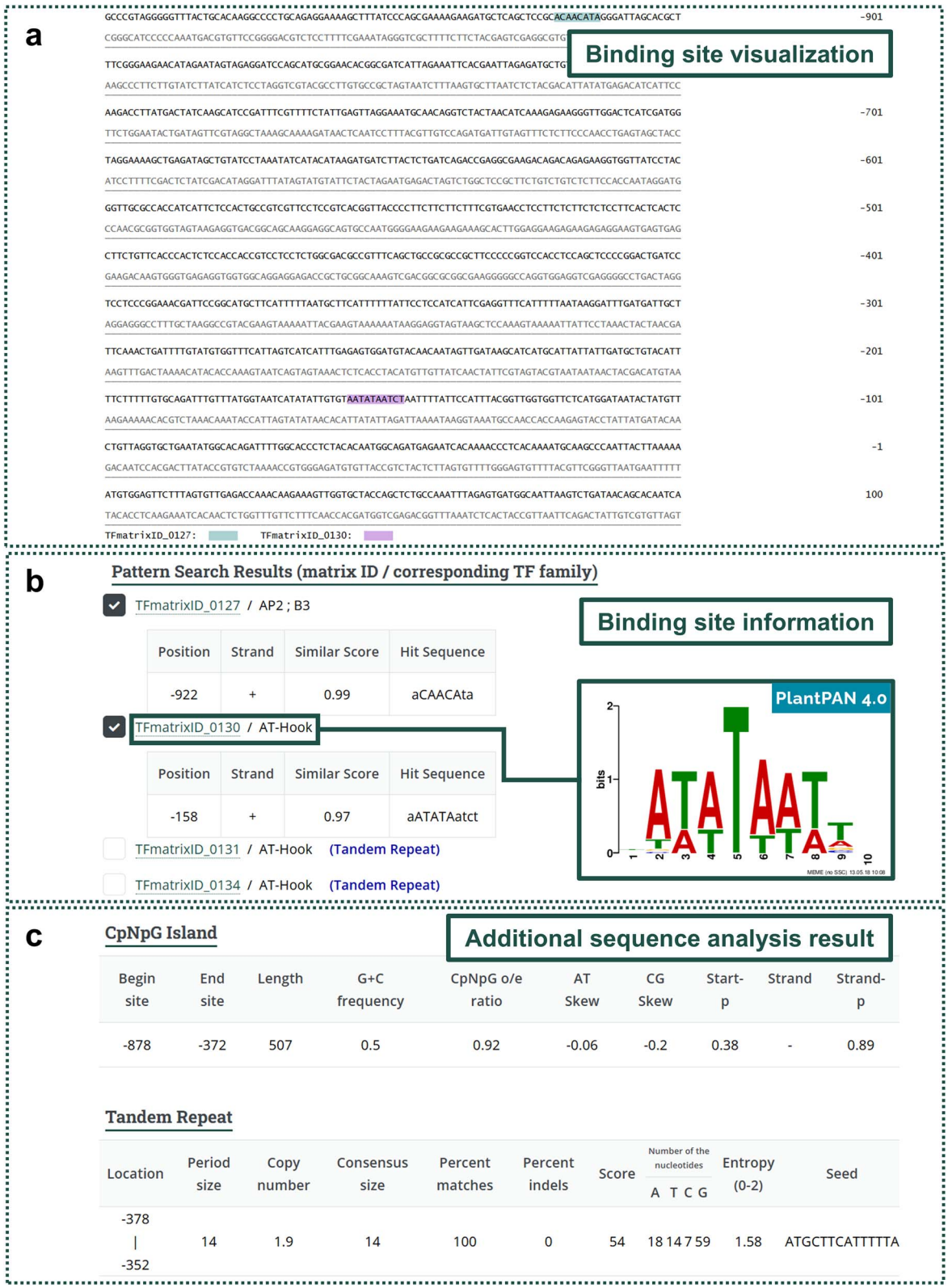


Figure 5. Screenshots depicting the output of the regulatory element analysis function. The results page showing the potential (a) TF-binding sites in the upper viewer, with (b) detailed results of binding site, each TF matrix ID provides a hyperlink to PlantPAN, allowing users to access the sequence logo of the corresponding TF binding site. (c) CpNpG island and tandem repeat analyses.

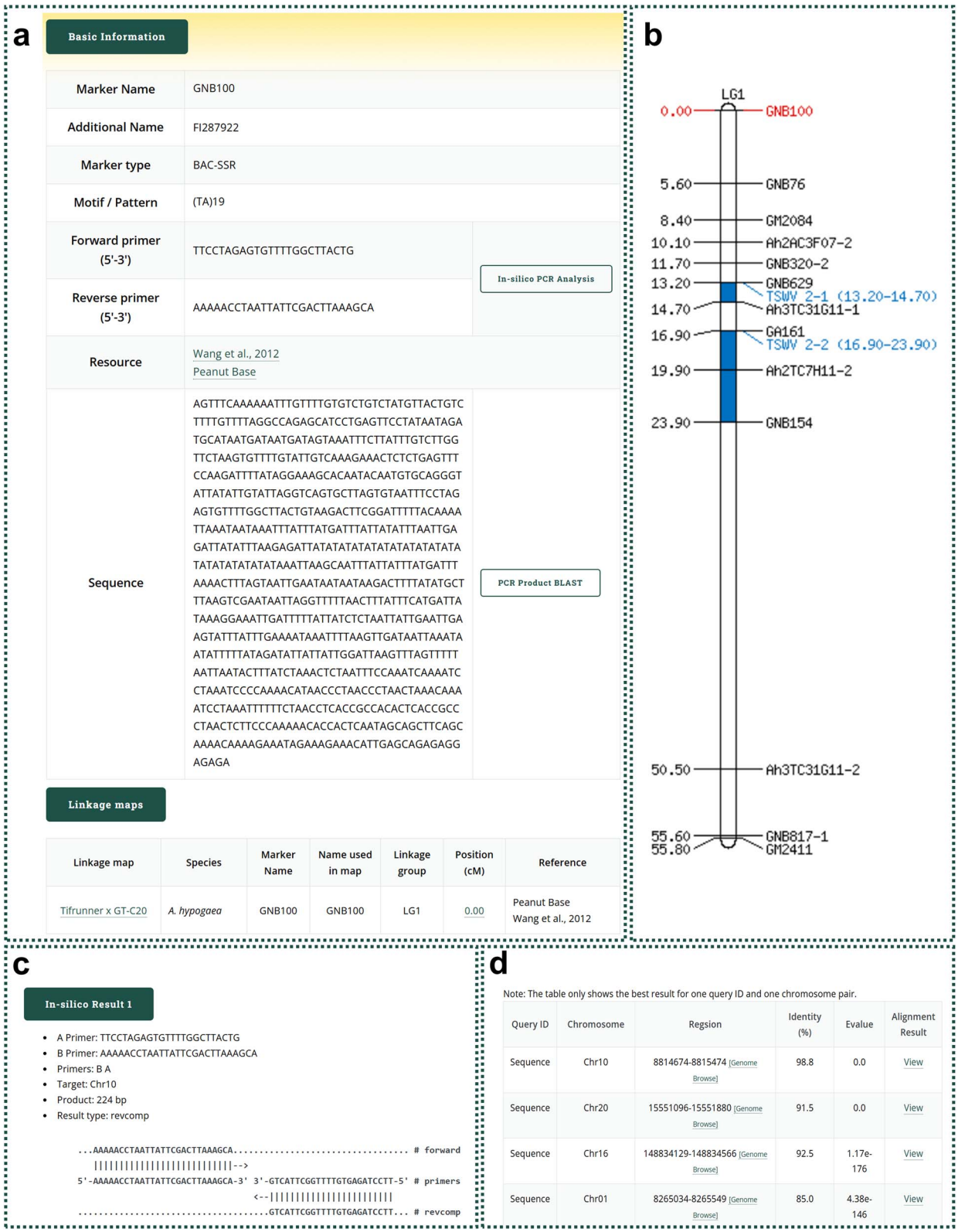


Figure 6. Screenshots depicting the output of the maps & traits function. The example output using GNB100 as query includes (a) marker information, (b) marker position visualization, (c) *in silico* PCR results, and (d) BLAST output for the marker sequence in the *Arachis hypogaea* var. Shitouqi genome.

silico PCR assays and conduct BLAST searches for PCR products to predict marker positions in the Shitouqi genome (Fig. 6c and d). BLAST results include alignment identity, *E*-value, and alignment length, helping users evaluate potential off-target amplifications. Collectively, the maps & traits function facilitates the rapid identification of candidate genes within trait-associated regions and supports marker-assisted trait selection in peanut breeding and genomics research.

Case study: Integrative cis-regulatory and coexpression network analysis reveals candidate transcriptional regulators of LEA5 genes

Late embryogenesis abundant (LEA) proteins were first identified in *Gossypium hirsutum* (cotton), where they accumulate during the later stages of seed development (Dure III et al. 1981). These proteins have since been identified in diverse plant species, such as angiosperms, gymnosperms, and algae, and even in prokaryotes (Hundertmark and Hincha 2008, Lan et al. 2013). Among the LEA proteins, LEA5—a member of the LEA type 3 family—mediates responses to abiotic stressors such as drought, salinity, and heat (Naot et al. 1995, Hundertmark and Hincha 2008, Karpinska et al. 2022). In *Arabidopsis thaliana*, LEA5 is highly expressed in leaves, roots, pollen, and flowers (Evans et al. 2024, Salleh et al. 2012). Yeast one-hybrid assays and TF mutant analysis highlighted several TFs belonging to the NAC, ERF, and WRKY families—e.g. NAC domain-containing protein 42 and NAC domain-containing protein 71, Ethylene response factor 1, and WRKY DNA-binding protein 15 and WRKY DNA-binding protein 63, respectively—as key regulators of *LEA5* (Evans et al. 2024).

A PGR-based analysis revealed two expression profiles for four *LEA5* genes (*AH01G27080*, *AH04G10170*, *AH11G30560*, and *AH14G12410*) across 29 transcriptomic datasets corresponding to various tissues and conditions. The first expression profile—characterized by consistently high expression levels in embryo, pericarp, and leaf tissues under drought stress and significantly low expression levels in leaf tissues treated chemically (abscisic acid, brassinolide, ethephon, distilled water, paclobutrazol, or salicylic acid) or maintained at room temperature—was exhibited by *AH01G27080* and *AH11G30560* (first group; Fig. 7a and b). The second expression profile—characterized by low expression levels in most tissues but relatively high expression levels in leaf tissue—was exhibited by *AH04G10170* and *AH14G12410* (second group). Notably, the expression levels of these genes were upregulated in response to chemical treatment (abscisic acid, brassinolide, ethephon, distilled water, or salicylic acid) and incubation at room temperature, exhibiting a profile opposite to that of *AH01G27080* and *AH11G30560* (Fig. 7c and d). To identify potential regulatory elements associated with *LEA5* genes in *A. hypogaea* var. Shitouqi, we analyzed the 2000-bp upstream regions and 5′ untranslated regions of the four *LEA5* genes by using the regulatory element analysis and gene group analysis functions in PGR. For gene group analysis, a 100% conservation threshold was applied to retain only TF-binding motifs shared across all *LEA5* genes. Although the predicted TFs varied in family classification, the Myb/SANT, AP2, ERF, WRKY, and NAC families were consistently conserved across the *LEA5* genes (Fig. 8a; Supplementary Figs. S3–S6). Similarly, Evans et al. (2024) highlighted TFs belonging to the ERF, WRKY, and NAC families as key regulators of *LEA5*.

To narrow down potential TFs that regulate *LEA5*, we analyzed coexpression patterns between the four *LEA5* genes and TFs belonging to the NAC, WRKY, and ERF families (Supplementary Tables S1–S4). Several NAC family members, such as NAC domain-containing protein 68, NAC domain-containing protein 78 (NAC78), NAC domain-containing protein 37, nascent polypeptide-associated complex subunit alpha-like protein 2, and nascent polypeptide-associated complex subunit beta, coexpressed with the *LEA5* genes (Fig. 8b).

Among the coexpressed TFs, NAC78 serves as a transcriptional activator in flavonoid biosynthesis and enhances anthocyanin accumulation in *A. thaliana* under high light stress conditions (Morishita et al. 2009). The expression of NAC78 is upregulated under heat and drought stress conditions, which suggests its regulatory role in stress response (Kim et al. 2006). In rice, NAC78 scavenges reactive oxygen species by directly activating *Tau class glutathione S-transferase 37* during drought stress (Yu et al. 2024). These findings may explain the observed upregulation of *AH01G27080* and *AH11G30560* under drought conditions (Figs. 7a, b and 8a).

An analysis of coexpression patterns between the four *LEA5* genes and TFs belonging to the WRKY family indicated that although WRKY TFs formed extensive coexpression networks, none of the family members exhibited a direct coexpression relationship with the *LEA5* genes (Fig. 8c; Supplementary Fig. S7). By contrast, within the ERF family, Cytokinin response factor 2, Related to APETALA2 11 (RAP2−11), and ethylene-responsive TF 61 coexpressed with the *LEA5* genes (Fig. 8d). Notably, RAP2−11, which coexpressed with *AH14G12410*, was located within a densely interconnected coexpression cluster and exhibited numerous coexpression linkages with TFs belonging to the WRKY family. This highly connected network suggests that *RAP2−11* serves as a central regulatory hub, likely orchestrating stress-responsive transcriptional programs and influencing *AH14G12410* expression dynamics under various conditions (Fig. 7d; Supplementary Fig. S8).

Coexpression network analysis with the expression profiles of the four *LEA5* genes revealed that none of the TFs belonging to the three families (NAC, ERF, or WRKY) coexpressed with the two *LEA5* groups with distinct expression profiles. This finding suggests that the expression of the two *LEA5* groups is regulated by distinct, group-specific TF networks rather than by shared regulatory factors.

Taken together, this case study indicates that the regulation of *LEA5* in cultivated peanut is modular and group-specific, involving distinct sets of TFs that do not overlap between expression groups. These findings highlight the complexity of

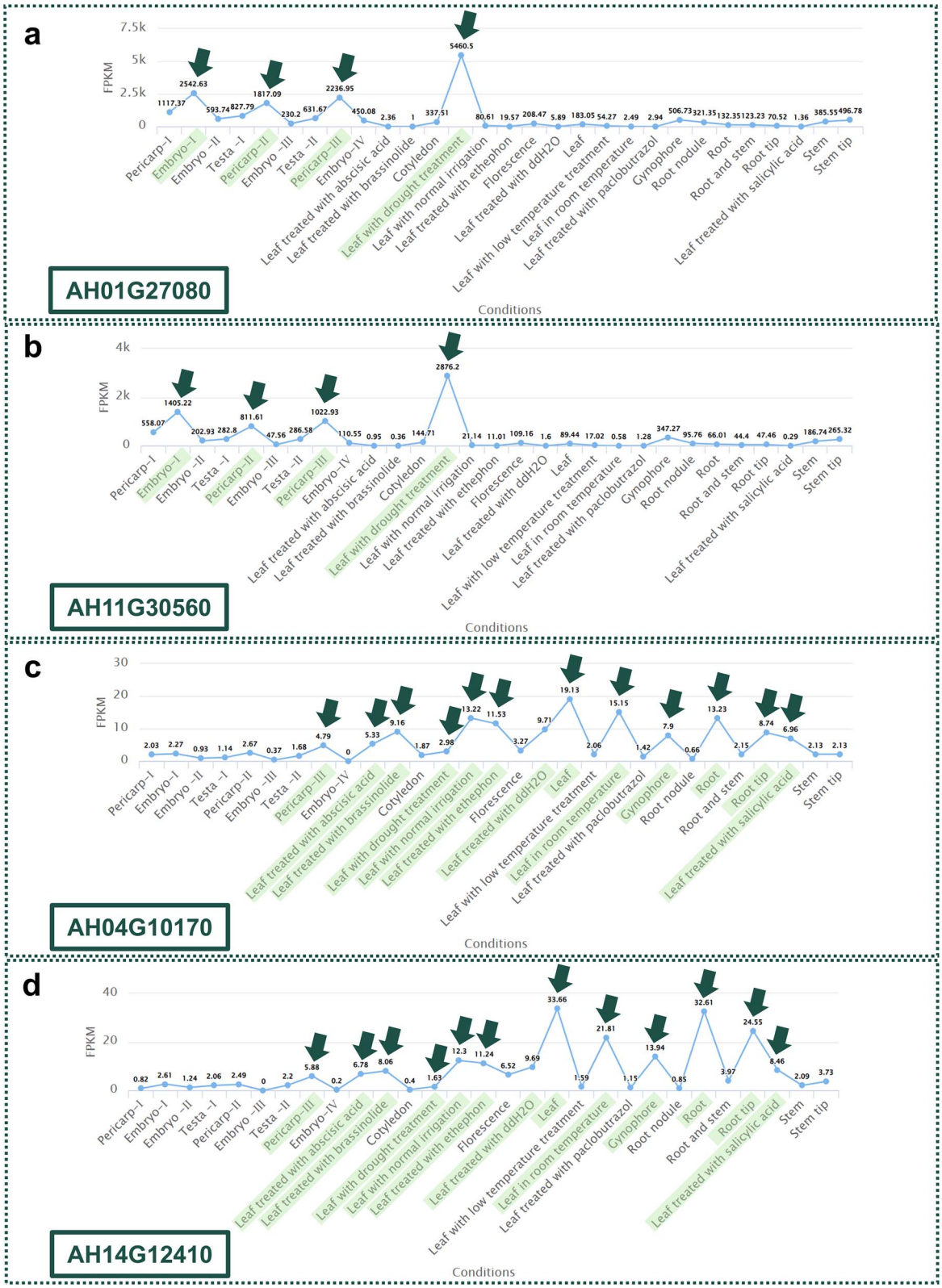


Figure 7. Expression profiles of four LEA5 genes across 29 tissues or conditions. (a) AH01G27080, (b) AH11G30560, (c) AH04G10170, and (d) AH14G12410. AH01G27080 and AH11G30560 exhibited similar expression patterns, while AH04G10170 and AH14G12410 showed comparable profiles. The four LEA5 genes can thus be divided into two expression groups.

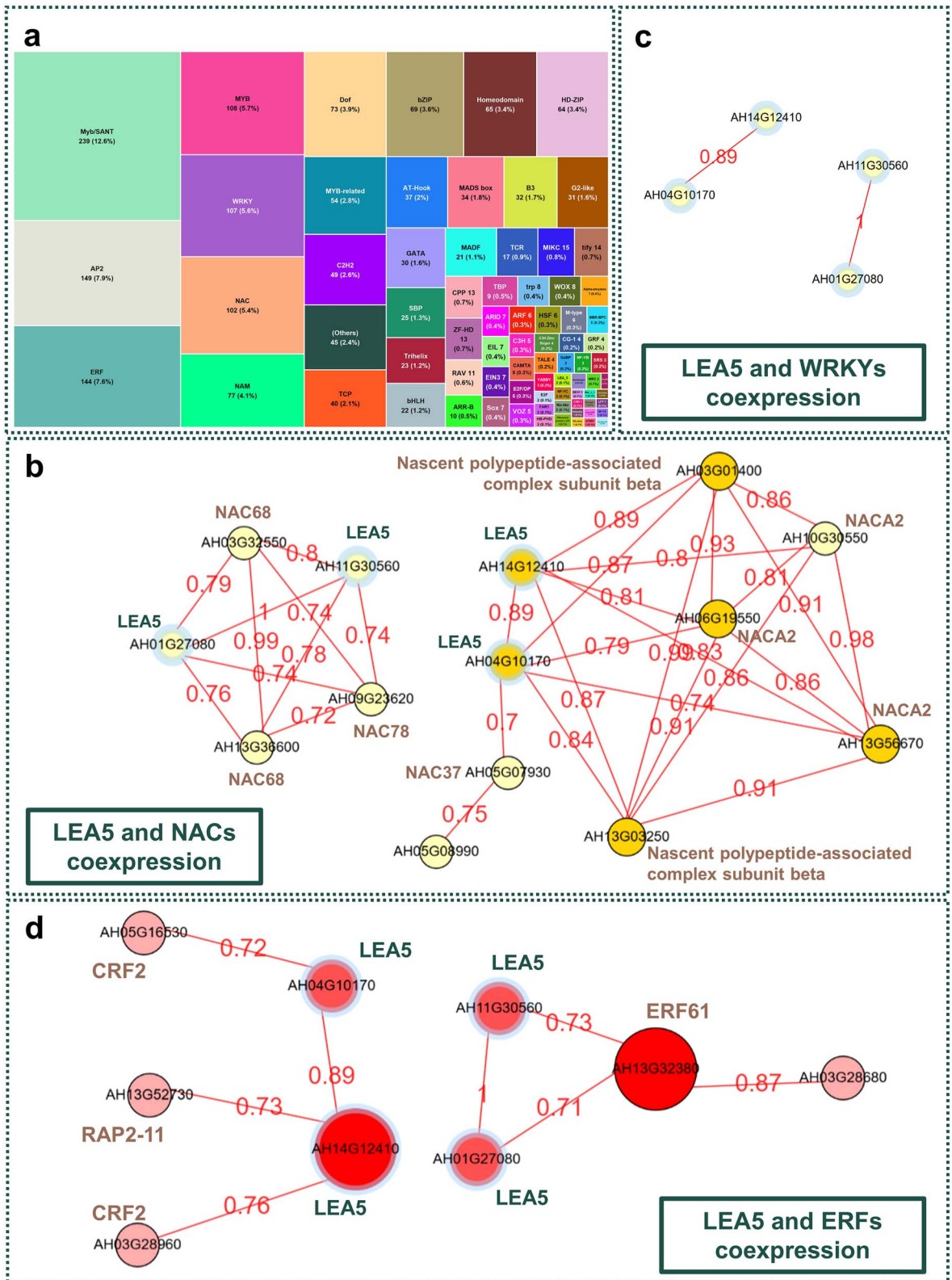


Figure 8. *Cis*-regulatory sequence and transcriptome analyses of *LEA5* genes. (a) Treemap visualization of potential TF families that bind to *LEA5* promoters yielded intersected results. Block size indicates the number of predicted TF-binding motifs per family. Different colors correspond to different TF families. (b–d) Coexpression networks between the four *LEA5* genes and TFs belonging to the NAC (b), WRKY (c), and ERF (d) families. The coexpression networks were constructed at a threshold of 0.7. Nodes represent genes, and edges indicate significant coexpression relationships.

stress-responsive gene regulation in peanut and the need for integrating transcriptomic and regulatory element analyses to elucidate gene function. Ultimately, this study may guide the development of a comprehensive peanut genome database to support functional and regulatory genomic research in cultivated peanut.

Conclusion

The updated PGR platform represents a major advancement in genomic research for *A. hypogaea* var. Shitouqi. PGR integrates genomic annotations, transcriptomic expression data, genetic linkage maps, and phenotypic information, addressing key limitations of prior peanut genomic resources. The platform supports multiple functional analyses, including DEG, coexpression, and promoter analyses. PGR's integrated bioinformatics tools for regulatory element prediction, sequence analysis, and in silico PCR increase the accessibility and utility of genomic data. Collectively, PGR provides researchers and breeders with a robust set of tools for investigating genetic traits, regulatory networks, and molecular markers, thereby supporting improved peanut breeding strategies and advancing research in peanut functional genomics.

Materials and Methods

Collection of peanut metadata

The genome assembly data for *A. hypogaea* var. Shitouqi used in PGR are available at the National Center for Biotechnology Information under the BioProject accession number PRJNA480120. Genome assemblies and annotation data for the parental species *A. duranensis* and *A. ipaensis* were obtained from PeanutBase (Dash et al. 2016). Additional resources, such as genome annotations and peptide sequences for the Shitouqi cultivar, transcriptome RNA-seq data corresponding to 29 tissues and conditions (Zhuang et al. 2019), and 32 microarray samples (Supplementary Tables S5 and S6), along with descriptive information and images of other peanut species, cultivars, and mutants, were provided by Dr Weijian Zhuang's laboratory at Fujian Agriculture and Forestry University, China. Most of these cultivars and mutants are not published, but some of them, like mutant varieties with high oleic acid contents, have been reported previously (Zhuang et al. 2019). A total of 21 069 SNPs, SSRs, and QTLs for cultivated peanut were manually curated and integrated into 18 genetic linkage maps compiled from PeanutBase and 12 published studies (Foncéka et al. 2009, Ravi et al. 2011, Qin et al. 2012, Shirasawa et al. 2012, 2013, Sujay et al. 2012, Wang et al. 2012, Gautami et al. 2012a, 2012b, Dash et al. 2016, Huang et al. 2016a, 2016b).

Classification of TFs

To classify TFs into families, we obtained a TF–Pfam correspondence table from the PlantPAN database (Chow et al. 2024), which maps Pfam accession numbers to specific TF families. All protein sequences from *A. hypogaea* var. Shitouqi were annotated using InterProScan to identify Pfam domains (Jones et al. 2014). Then, each protein's Pfam accession number was cross-referenced with the PlantPAN mapping table to assign it to the appropriate TF family.

Analysis of gene expression data

PGR's functions for identifying DEGs and conducting coexpression analysis were developed using the ExPath tool (Zheng et al. 2017). To facilitate the detection of DEGs under user-defined conditions, PGR allows for both chemical treatment (e.g. abscisic acid, brassinolide, ethephon, paclobutrazol, and salicylic acid) and physical treatment (e.g. drought and low temperature); leaves treated with distilled water serve as the control group. The t.test() function from the R package was used to identify DEGs. As mentioned, PGR offers two modes for coexpression analysis. The correlation of single gene mode identifies the top 20 coexpressed genes across all experimental conditions for a given gene. By contrast, the correlation network mode visualizes networks of coexpressed genes on the basis of user-defined inputs. Pearson's correlation and Spearman's rank correlation analyses were performed using the cor() function in R. The resulting networks were visualized using Cytoscape.js (Franz et al. 2015).

KEGG pathway and GO enrichment analyses

For KEGG pathway and GO enrichment analyses, transcript sequences of cultivated peanut were annotated using InterProScan to predict protein domains and associated functions (Jones et al. 2014). The resulting data were used to extract UniProt accession numbers and corresponding GO terms. KEGG orthology IDs were assigned on the basis of UniProt accession numbers (Kanehisa and Goto 2000). Enrichment analyses for GO and KEGG terms were performed using the hypergeometric distribution. Cumulative probabilities (*P* values) were calculated using the dhyper() and phyper() functions in R to evaluate the statistical significance of functional enrichment. The resulting *P* values were adjusted for multiple comparisons by using the Benjamini–Hochberg method.

Sequence analysis tools

The TFBS prediction tool in PGR was integrated from the PlantPAN database. Position weight matrix data and corresponding cutoff profiles from the PlantPAN database were used to identify putative TFBSs with the MATCH algorithm (version 9.0) (Kel et al. 2003). For species-specific motif detection, the position weight matrix data were classified into 8 groups (Supplementary Table S7): (i) *A. thaliana*, (ii) *Glycine max*, (iii) *Medicago truncatula*, (iv) *Oryza sativa*, (v) *Zea mays*, (vi) Fabaceae family members ($n = 7$), (vii) eudicot species ($n = 52$) and genera ($n = 2$), and (viii) monocot species ($n = 12$). For promoter structure analysis, CpG islands were detected using CpG Island Promoter Detection tool, and tandem repeats were identified using Tandem Repeats Finder (version 4.07b) (Benson 1999, Ponger and Mouchiroud 2002). SSRs were detected using the MIcroSAtellite identification tool (version 2.1), which also enables users to define the maximum allowable distance between two SSRs (Beier et al. 2017).

In silico PCR

To predict the locations of genetic markers in the peanut genome, PGR uses ipcress, a tool from the Exonerate package that simulates PCR amplification by predicting potential PCR products (Slater and Birney 2005). On the PGR website, ipcress is implemented with a default mismatch threshold of three (–mismatch 3). The minimum product length (min_product_len) was set to match the original amplicon size for each primer pair, and the maximum product length (max_product_len) was restricted to 2000 bp to avoid overly long and improbable products. Furthermore, primer pairs generating AA or BB configurations were filtered out to reduce nonspecific amplification, including potential primer-dimer formation.

Acknowledgments

We would like to thank National Core Facility for Biopharmaceuticals (NCFB, NSTC 114-2740-B-492-001) and National Center for High-performance Computing (NCHC) of National Institutes of Applied Research (NIAR) of Taiwan for providing computational and storage resources. This manuscript was edited by Wallace Academic Editing.

Supplementary Data

Supplementary Data is available at *PCP* online.

Conflict of Interest

The authors have no conflict of interest to declare.

Funding

This work was supported by the National Science and Technology Council (NSTC 113-2311-B-006-007-MY3 and NSTC 112-2311-B-006-003) and National Cheng Kung University, and grants from the NSFs of China (32272155, 32401787, U1705233) and Fujian Agriculture and Forestry University.

Data Availability

The data underlying this article are available on the PGR Download page (https://pgr.itps.ncku.edu.tw/download_genomic_data.php).

References

Arya, S.S., Salve, A.R., Chauhan, S. (2016) Peanuts as functional food: a review. *J. Food Sci. Technol.* 53:31–41. https://doi.org/10.1007/s13197-015-2007-9

Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M. (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580. https://doi.org/10.1093/nar/27.2.573

Bertioli, D.J., Seijo, G., Freitas, F.O., Valls, J.F.M., Leal-Bertioli, S.C.M., Moretzsohn, M.C. (2011) An overview of peanut and its wild relatives. *Plant Genet. Resour.* 9:134–149. https://doi.org/10.1017/S1479262110000444

Bertioli, D.J., Cannon, S.B., Froenicke, L., Huang, G., Farmer, A.D., Cannon, E.K.S., et al. (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48:438–446. https://doi.org/10.1038/ng.3517

Bertioli, D.J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., et al. (2019) The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* 51:877–884. https://doi.org/10.1038/s41588-019-0405-z

Chow, C.N., Yang, C.W., Wu, N.Y., Wang, H.T., Tseng, K.C., Chiu, Y.H., et al. (2024) PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters. *Nucleic Acids Res.* 52:D1569–D1578. https://doi.org/10.1093/nar/gkad945

Çiftçi, S. and Suna, G. (2022) Functional components of peanuts (*Arachis Hypogaea* L.) and health benefits: a review. *Future Foods* 5:100140. https://doi.org/10.1016/j.fufo.2022.100140

Dash, S., Cannon, E.K.S., Kalberer, S.R., Farmer, A.D., Cannon, S.B. (2016) Chapter 8 - PeanutBase and other bioinformatic resources for peanut. In: Stalker, HT and Wilson, FR (eds) *Peanuts*. AOCS Press, Urbana, Illinois, USA, pp 241–252. https://doi.org/10.1016/B978-1-63067-038-2.00008-3

Dure, L., III, Greenway, S.C., Galau, G.A. (1981) Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry* 20:4162–4168. https://doi.org/10.1021/bi00517a033

Evans, K.V., Ransom, E., Nayakoti, S., Wilding, B., Mohd Salleh, F., Gržina, I., et al. (2024) Expression of the *Arabidopsis* redox-related LEA protein, SAG21 is regulated by ERF, NAC and WRKY transcription factors. *Sci. Rep.* 14:7756. https://doi.org/10.1038/s41598-024-58161-0

Foncéka, D., Hodo-Abalo, T., Rivallan, R., Faye, I., Sall, M.N., Ndoye, O., et al. (2009) Genetic mapping of wild introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. *BMC Plant Biol.* 9:103. https://doi.org/10.1186/1471-2229-9-103

Franz, M., Lopes, C.T., Huck, G., Dong, Y., Sumer, O., Bader, G.D. (2015) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* 32:309–311. https://doi.org/10.1093/bioinformatics/btv557

Gautami, B., Foncéka, D., Pandey, M.K., Moretzsohn, M.C., Sujay, V., Qin, H., et al. (2012a) An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.). *PLoS One* 7:e41213. https://doi.org/10.1371/journal.pone.0041213

Gautami, B., Pandey, M.K., Vadez, V., Nigam, S.N., Ratnakumar, P., Krishnamurthy, L., et al. (2012b) Quantitative trait locus analysis and construction of consensus genetic map for drought tolerance traits based on three recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol. Breed.* 30:757–772. https://doi.org/10.1007/s11032-011-9660-0

Huang, L., Ren, X., Wu, B., Li, X., Chen, W., Zhou, X., et al. (2016a) Development and deployment of a high-density linkage map identified quantitative trait loci for plant height in peanut (*Arachis hypogaea* L.). *Sci. Rep.* 6:39478. https://doi.org/10.1038/srep39478

Huang, L., Wu, B., Zhao, J., Li, H., Chen, W., Zheng, Y., et al. (2016b) Characterization and transferable utility of microsatellite markers in the wild and cultivated *Arachis* species. *PLoS One* 11:e0156633. https://doi.org/10.1371/journal.pone.0156633

Hundertmark, M. and Hincha, D.K. (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9:118. https://doi.org/10.1186/1471-2164-9-118

Husted, L. (1931) Chromosome number in species of peanut, *Arachis*. *Am. Nat.* 65:476–477. https://doi.org/10.1086/280391

Husted, L. (1933) Cytological studies on the peanut, *Arachis* I. Chromosome number and morphology. *Cytologia* 5:109–117. https://doi.org/10.1508/cytologia.5.109

Husted, L. (1936) Cytological studies on the peanut, *Arachis*. II Chromosome number, morphology and behavior, and their application to the problem of the origin of the cultivated forms. *Cytologia* 7:396–423. https://doi.org/10.1508/cytologia.7.396

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30. https://doi.org/10.1093/nar/28.1.27

Karpinska, B., Razak, N., Shaw, D.S., Plumb, W., Van De Slijke, E., Stephens, J., et al. (2022) Late embryogenesis abundant (LEA)5 regulates translation in mitochondria and chloroplasts to enhance growth and stress tolerance. *Front. Plant Sci.* 13:875799. https://doi.org/10.3389/fpls.2022.875799

Kel, A.E., Gössling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., Wingender, E. (2003) MATCH: a tool for searching transcription

factor binding sites in DNA sequences. *Nucleic Acids Res.* 31:3576–3579. https://doi.org/10.1093/nar/gkg585

Kim, S.-Y., Kim, S.-G., Kim, Y.-S., Seo, P.J., Bae, M., Yoon, H.-K., et al. (2006) Exploring membrane-associated NAC transcription factors in *Arabidopsis*: implications for membrane biology in genome regulation. *Nucleic Acids Res.* 35:203–213. https://doi.org/10.1093/nar/gkl1068

Kochert, G., Stalker, H.T., Gimenes, M., Galgaro, L., Lopes, C.R., Moore, K. (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *Am. J. Bot.* 83:1282–1291. https://doi.org/10.1002/j.1537-2197.1996.tb13912.x

Lan, T., Gao, J., Zeng, Q.-Y. (2013) Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in *Populus trichocarpa*. *Tree Genet. Genomes* 9:253–264. https://doi.org/10.1007/s11295-012-0551-2

Moretzsohn, M.C., Gouvea, E.G., Inglis, P.W., Leal-Bertioli, S.C.M., Valls, J.F.M., Bertioli, D.J. (2012) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* 111: 113–126. https://doi.org/10.1093/aob/mcs237

Morishita, T., Kojima, Y., Maruta, T., Nishizawa-Yokoi, A., Yabuta, Y., Shigeoka, S. (2009) Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant Cell Physiol.* 50: 2210–2222. https://doi.org/10.1093/pcp/pcp159

Naot, D., Ben-Hayyim, G., Eshdat, Y., Holland, D. (1995) Drought, heat and salt stress induce the expression of a citrus homologue of an atypical late-embryogenesis Lea5 gene. *Plant Mol. Biol.* 27:619–622. https://doi.org/10.1007/BF00019327

Ponger, L.C. and Mouchiroud, D. (2002) CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. *Bioinformatics* 18:631–633. https://doi.org/10.1093/bioinformatics/18.4.631

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., et al. (2012) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. *Theor. Appl. Genet.* 124: 653–664. https://doi.org/10.1007/s00122-011-1737-y

Ravi, K., Vadez, V., Isobe, S., Mir, R.R., Guo, Y., Nigam, S.N., et al. (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (Arachishypogaea L.). *Theor. Appl. Genet.* 122:1119–1132. https://doi.org/10.1007/s00122-010-1517-0

Salleh, F.M., Evans, K., Goodall, B., Machin, H., Mowla, S.B., Mur, L.A.J., et al. (2012) A novel function for a redox-related LEA protein (SAG21/AtLEA5) in root development and biotic stress responses. *Plant Cell Environ.* 35:418–429. https://doi.org/10.1111/j.1365-3040.2011.02394.x

Seijo, G., Lavia, G.I., Fernández, A., Krapovickas, A., Ducasse, D.A., Bertioli, D.J., et al. (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. *Am. J. Bot.* 94:1963–1971. https://doi.org/10.3732/ajb.94.12.1963

Shirasawa, K., Koilkonda, P., Aoki, K., Hirakawa, H., Tabata, S., Watanabe, M., et al. (2012) In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. *BMC Plant Biol.* 12:80. https://doi.org/10.1186/1471-2229-12-80

Shirasawa, K., Bertioli, D.J., Varshney, R.K., Moretzsohn, M.C., Leal-Bertioli, S.C.M., Thudi, M., et al. (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. *DNA Res.* 20: 173–184. https://doi.org/10.1093/dnares/dss042

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., Holmes, I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.* 19:1630–1638. https://doi.org/10.1101/gr.094607.109

Slater, G.S.C. and Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. https://doi.org/10.1186/1471-2105-6-31

Sujay, V., Gowda, M.V.C., Pandey, M.K., Bhat, R.S., Khedikar, Y.P., Nadaf, H.L., et al. (2012) Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol. Breed.* 30:773–788. https://doi.org/10.1007/s11032-011-9661-z

U.S. Department of Agriculture, F.A.S (2025a) *Oilseeds: World Markets and Trade*. https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf (June 12, 2025, date last accessed).

U.S. Department of Agriculture, F.A.S (2025b) *Peanut Commodity View*. U.S. Department of Agriculture, Foreign Agricultural Service, https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2221000&sel_year=2025&rankby=Production (July 1, 2025, date last accessed).

Wang, H., Penmetsa, R.V., Yuan, M., Gong, L., Zhao, Y., Guo, B., et al. (2012) Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 12:10. https://doi.org/10.1186/1471-2229-12-10

Yu, X., Xie, Y., Wang, L., Li, L., Jiang, S., Zhu, Y., et al. (2024) Transcription factor NAC78 cooperates with NAC78 interacting protein 6 to confer drought tolerance in rice. *Plant Physiol.* 196:1642–1658. https://doi.org/10.1093/plphys/kiae395

Zheng, H.-Q., Wu, N.-Y., Chow, C.-N., Tseng, K.-C., Chien, C.-H., Hung, Y.-C., et al. (2017) EXPath tool—a system for comprehensively analyzing regulatory pathways and coexpression networks from high-throughput transcriptome data. *DNA Res.* 24:371–375. https://doi.org/10.1093/dnares/dsx009

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M.K., Zhang, C., et al. (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51: 865–876. https://doi.org/10.1038/s41588-019-0402-2

Plant and Cell Physiology, 2026, 67(5),859–873. https://doi.org/10.1093/pcp/pcaf165, Advance Access publication on 11 December 2025, available online at https://academic.oup.com/pcp
Database Paper