

Article

Identification and Analysis of DUF506 Gene Family in Peanut (*Arachis hypogaea*)

Qing Song¹, Gideon Asare Aboagye¹, Ming Liu¹, Ying Lan¹, Minghong Hu¹, Yanbin Hong², Renfeng Wang² and Miao Chen^{1,*}

¹ College of Agricultural Sciences, Guangdong Ocean University, Zhanjiang 524091, China; songqing12@stu.gdou.edu.cn (Q.S.); gaa@stu.gdou.edu.cn (G.A.A.); liuming01@stu.gdou.edu.cn (M.L.); lany659@stu.gdou.edu.cn (Y.L.); humh384@stu.gdou.edu.cn (M.H.)

² Guangdong Key Laboratory for Crops Genetic Improvement, South China Peanut Sub-Center of National Center of Oilseed Crops Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences (GAAS), Guangzhou 510520, China; hongyanbin@gdaas.cn (Y.H.); wangrunfeng@gdaas.cn (R.W.)

* Correspondence: miaochen@gdou.edu.cn

Abstract

The Domain of Unknown Function 506 (DUF506) family, part of the PD-(D/E)XK nuclease superfamily, has been shown to play a vital role in plant development and responses to abiotic stresses. However, the function of the DUF506 family in cultivated peanuts remains unknown. This study identified 23 *AhDUF506* genes using bioinformatics approaches; these genes are spread across 15 chromosomes and grouped into 4 subfamilies. Additionally, by analyzing gene structure, upstream cis-acting elements, and transcriptional expression changes of *AhDUF506* genes in different tissues and under various stress conditions, their expression levels and response mechanisms to abiotic stresses were examined. In mature tissues, the expression levels of seven *AhDUF506* genes in flowers were significantly higher than those in other tissues. Under abiotic stress, their expression levels were all up-regulated in the roots of peanut plant seedlings. These findings provide an important foundation for a deeper understanding of the molecular characteristics of the DUF506 family in *Arachis hypogaea* (peanut), supporting future research on the functional characterization of its genes.

Keywords: DUF506; peanut; abiotic stresses; transcription analysis

1. Introduction

The “Domain of Unknown Function” (DUFs) constitutes a broad category encompassing numerous uncharacterized domains characterized by two distinct features: a relatively conserved amino acid sequence and an as-yet uncharacterized biochemical role [1–3]. Each DUF is temporarily designated with the prefix “DUF” followed by a unique identifier (for instance, DUF1, DUF2). Once experimental data reveal a domain’s function, it is either renamed to reflect its biochemical activity or integrated into an existing domain family [4]. Research indicates that DUFs are integral to various physiological processes, including plant growth, development, and adaptive responses to biotic and abiotic stresses [5]. Recent bioinformatic analyses utilizing an enhanced transfer Meta-BASIC search against updated PFAM and PDB databases have identified five novel PD-(D/E) XK nuclease families, one of which corresponds to the DUF506 family. These findings reveal that the DUF506 protein exhibits a predicted conserved core secondary structure pattern of $\alpha\alpha\alpha\beta\alpha\beta\alpha$ [6,7], although its specific function remains undefined.



Academic Editor: Hakim Manghwar

Received: 29 December 2025

Revised: 21 January 2026

Accepted: 4 February 2026

Published: 9 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Recent studies indicate that members of the DUF506 family are associated with plant stress resistance. In *Arabidopsis thaliana*, *At1g12030* responds clearly to iron deficiency, while *At2g39650* contributes to stress resistance by interacting with salt stress transcription factors [8]. In crops, *OsDUF50602* in rice actively responds to drought, and *TaDUF1.3-D* is triggered by salinity [9,10]. Similarly, the cabbage homolog *Bra017099* shows increased expression under heat and drought stress [11]. Notably, in *Malus domestica*, *MdDUF506* boosts aluminum stress tolerance by enhancing reactive oxygen species scavenging and regulating aluminum-responsive genes. Its interacting partner, *MdCNR8*, works with *MdDUF506* to further improve aluminum tolerance [12]. Studies on their biological roles suggest that DUF506 proteins may differ across species. The low sequence similarity among DUF506 proteins reveals highly divergent homologs within this family [13]. This diversity makes it challenging to study the specific functions of the DUF506 family.

Peanut (*Arachis hypogaea*) is a globally significant crop used for the production of essential oils, food, and feed [14]. Its genome has been fully sequenced and annotated, facilitating advanced molecular research [15]. Although several gene families in peanut have been characterized, such as WRKY, PP2C, and SUAR [16–18], the domain of unknown function 506 (DUF506) family has received little attention.

This study aimed to identify all DUF506 members in peanut using bioinformatics techniques, predict their subcellular localization, analyze their physicochemical properties, and evaluate their homology with DUF506 members from other species. This approach enables a comprehensive understanding of the characteristics of AhDUF506 members. Additionally, in combination with transcriptomic datasets, quantitative PCR (qPCR) was performed on seven *AhDUF506* genes to examine their transcriptional responses to abiotic stress, providing valuable insights for future gene-function studies.

2. Materials and Methods

2.1. Identification of the DUF506 Gene Family in Peanuts and Construction of a Phylogenetic Tree

The genome assembly, annotated gene models, and protein sequences of *Arachis hypogaea* are from the peanut genome database (<http://peanutgr.fafu.edu.cn/Download.php>, accessed on 20 March 2025) [19]. Whole-genome datasets for *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, and *Medicago truncatula* were obtained from phytozome v13 (<https://phytozome-next.jgi.doe.gov>, accessed on 20 March 2025) [20]. 13 AtDUF506 protein sequences served as queries in BLASTP searches against the *A. hypogaea* proteome (score > 50, E value < 0.01). The Hidden Markov Model (HMM) file for the DUF506 conserved domain (accession number: PF04720) was downloaded from the Pfam database. Genome-wide identification in *Arachis hypogaea* was conducted using the Simple HMM Search module within TBtools II [21,22]. The structure of the AhDUF506 protein was analyzed by setting the expected value to 0.010000 and the maximum number of hits to 500 through the CDD search function on NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd>, accessed 20 March 2025) [23,24]. The amino acid sequences of 23 identified AhDUF506 proteins, 13 *Arabidopsis thaliana* DUF506 proteins, 10 *Oryza sativa* DUF506 proteins, 17 *Medicago truncatula* DUF506 proteins, and 24 *Glycine max* DUF506 proteins were aligned using the MUSCLE method in MEGA11 software (MEGA11.0.13). A Maximum Likelihood (ML) phylogenetic tree was constructed with bootstrap set to 1000 replicates and using the JTT model (Jones-Taylor-Thornton model) [25,26]. The phylogenetic tree was visualized and annotated using iTOL v6 software [27].

2.2. Chromosome Localization of AhDUF506 Family Members

The most recent peanut (*Arachis hypogaea*) genome annotation (GFF3 format) was obtained from the Peanut Genome Resource database. The chromosomal localization of

DUF506 family members was mapped utilizing the “Gene Location Visualization” module in TBtools II software [28]. Based on the physical coordinates of previously characterized *AhDUF506* genes, each DUF506 locus was renamed to accurately reflect its corresponding chromosome and genomic order, thereby ensuring consistency in gene nomenclature. The amino acid sequences of all identified *AhDUF506* proteins were imported into the “Protein Parameter Calculator” module of TBtools II to analyze their physicochemical properties.

2.3. Physical and Chemical Properties and Subcellular Localization of *AhDUF506* Protein Analysis

The Protein Parameter Calculator module of TBtools II software analyzed the physicochemical properties of *AhDUF506* protein. The primary amino acid sequence was provided in FASTA format to compute the total residue count, theoretical isoelectric point (pI), hydropathicity, and instability index. The probable subcellular compartment of *AhDUF506* was predicted using WoLF PSORT (<https://wolfpsort.hgc.jp/>, accessed on 20 March 2025) and was used to determine the subcellular localization of the *AhDUF506* protein [29].

2.4. *AhDUF506* Conserved Motif, Conserved Domain, and Gene Structure Analysis

Full-length amino acid sequences of *AhDUF506* were obtained from the cultivated peanut (*Arachis hypogaea*) genome database. Conserved motifs within the *AhDUF506* protein family were identified using the MEME website (<https://meme-suite.org/meme/tools/meme>, accessed on 20 March 2025) [30]. The conserved domain architecture of *AhDUF506* proteins was predicted via NCBI's Conserved Domain Search tool (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>, accessed on 20 March 2025). Gene structures (exon–intron organization), conserved motifs, and conserved domains of the *AhDUF506* gene family were visualized concurrently using the gene structure visualization module of TBtools II software.

2.5. Analysis of *AhDUF506* Gene Promoter

The 2000 bp nucleotide sequences upstream of 23 *AhDUF506* genes were submitted to the PlantCare website (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>, accessed on 20 March 2025) for the purpose of predicting stress-related promoter elements and visualizing them using TBtools II.

2.6. Synteny and *Ka* and *Ks* Analysis of *AhDUF506* Homologous Pair

Genomic assemblies and their associated annotation files for *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, and *Medicago truncatula* were sourced from the Ensembl Plants database. Protein sequences of *Arachis hypogaea* (*AhDUF506*) were extracted to facilitate cross-species homology analysis. The collinearity between *AhDUF506* genes and their orthologous regions in other species was assessed using MCScanX [31], and syntenic relationships were subsequently visualized using the advanced circular-plot module of TBtools II. To further evaluate the duplication events of the *AhDUF506* gene, synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates were calculated for each homologous gene pair utilizing the *Ka/Ks* calculation function within TBtools II.

2.7. Expression Analysis of the *AhDUF506* Gene

An overview of the tissue-specific expression data for *AhDUF506* was obtained from the Peanut Genome Resources Public Database (<http://peanutgr.fafu.edu.cn/Transcriptome.php>, accessed on 20 March 2025). The expression profiles of all *AhDUF506* genes were quantified as transcripts per kilobase million (TPM) using the exon model, and fold changes were calculated by dividing the TPM values of the treatment group by those of the control group. A heatmap was generated with TBtools II to show the converted logarithms.

2.8. Peanut Plant Treatment and RT-qPCR Analysis

Peanut (*Arachis hypogaea*) seeds were immersed in 5% hypochlorous acid (HClO) for 2 min [32], then rinsed three times with double-distilled water (ddH₂O). Sterilized seeds were placed on a moist cloth and incubated in a dark, well-ventilated environment to promote germination [33]. Once the primary roots reached 2 cm in length, seedlings were transferred to hydroponic boxes containing Hoagland nutrient solution. The seedlings were grown in a controlled-environment incubator under a 16 h photoperiod (27 °C, 60% relative humidity), followed by an 8 h dark period (25 °C, 60% relative humidity). Three-week-old seedlings, exhibiting uniform and vigorous growth, were subjected to the subsequent treatments: (i) osmotic stress: nutrient solution supplemented with 150 mM NaCl, (ii) dehydration: seedlings placed on Whatman 3 MM filter paper and dried at 25–27 °C, 60% humidity, (iii) cold stress: transferred to a growth chamber maintained at 4 °C, (iv) ABA treatment: a solution containing 100 µM ABA (KeMing Technology Co., Ltd., Zhanjiang, China) was applied to the seedlings, and (v) phosphorus deficiency stress: plants were supplemented with Hoagland nutrient solution lacking phosphorus. Harvested shoot and root tissues at 0, 3, 6, 12, and 24 h under osmotic, drought, cold stress, and ABA treatment, as well as days 0 and 8 under phosphorus deficiency stress, were immediately frozen in liquid nitrogen and stored at -80 °C for subsequent analysis. For tissue-specific expression analysis, leaves, stems, roots, pods, and dry seeds were collected from mature plants cultivated in soil.

Gene-specific primers were designed using Primer3Plus (<https://www.primer3plus.com/>, accessed on 20 March 2025) and validated for specificity using NCBI's Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome, accessed on 20 March 2025) (Table S1). Total RNA was extracted from the peanut tissue using the Magen Polysaccharide Polyphenol Plant RNA Extraction Kit. RNA was reverse-transcribed into cDNA with the Takara Reverse Transcription Kit (HiSrip III First-Strand cDNA Synthesis Kit, Shiga, Japan). Quantitative Real-Time Polymerase Chain Reaction (QRT-PCR) was performed using Yi Sheng-related products/platforms. The expression level of the peanut *ACT11* gene is used as an internal reference. Each biological sample has three biological replicates, and the relative expression of each gene was calculated using the $2^{-\Delta\Delta CT}$ method.

3. Results

3.1. Identification of AhDUF506 Members and Chromosome Localization of AhDUF506 Gene

Homologous sequence alignment was performed using BLAST, as previously described, which identified 26 potential members of the *AhDUF506* gene family. Subsequently, 23 candidate *AhDUF506* members were determined using the HMM Search module of TBtools II. Verification via CDD search identified 23 members of the *AhDUF506* gene family. In total, these 23 *AhDUF506* genes are located on 15 chromosomes (Figure 1). Chromosomal mapping indicated that these genes were spread across 14 chromosomes. Most *DUF506* genes in peanut were situated near the edges of chromosomes, with chromosome 14 containing the highest number, a total of four genes.

3.2. Physical and Chemical Properties Analysis and Subcellular Localization Prediction of AhDUF506 Protein

The physicochemical properties of 23 *AhDUF506* protein sequences were analyzed using the Protein Parameter Calc tool of TBtools II. Sequence lengths ranged from 262 to 401 amino acids. Predicted molecular weights ranged from 29,729.05 to 43,831.97, and theoretical isoelectric points (pI) varied between 5.29 and 9.06. Instability indices ranged from 42.03 to 62.45, indicating that all proteins are predicted to be unstable. The aliphatic

index ranged from 63.38 to 80.05, suggesting moderate thermostability. Hydrophilicity indices (grand average of hydropathicity) ranged from -0.607 to -0.388 , consistent with an overall hydrophilic character (Table S2).

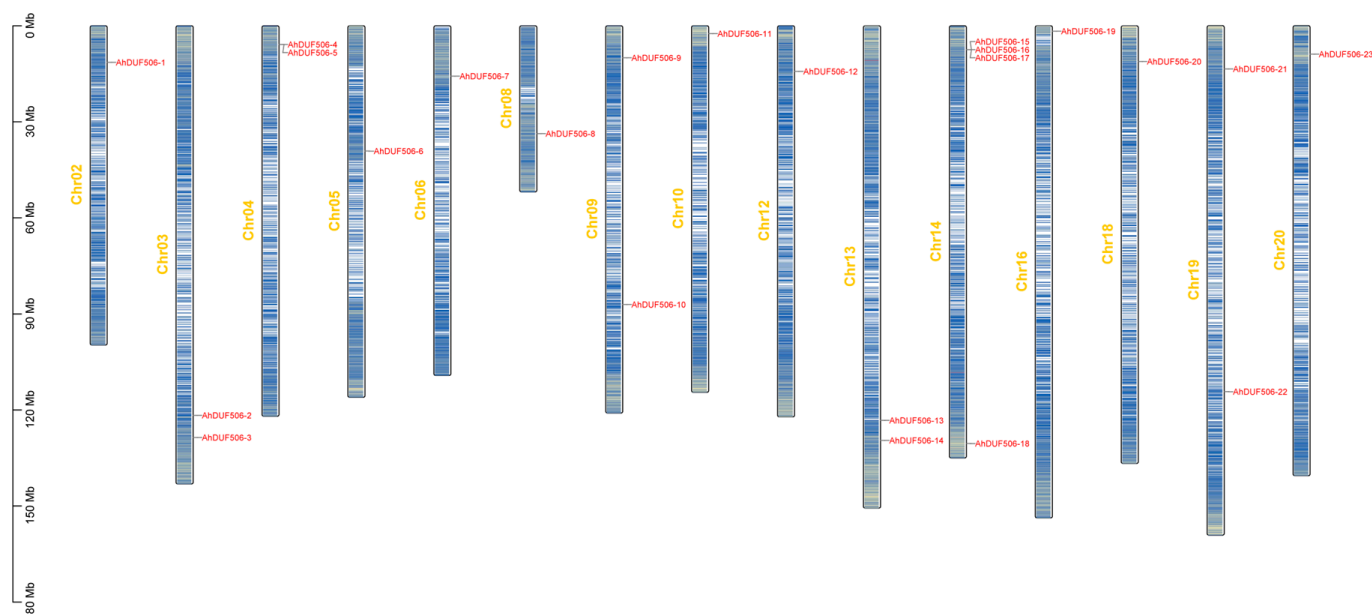


Figure 1. Chromosome density and chromosomal distribution of the *AhDUF506* gene. The density of genes on chromosomes is shown by blue-white-spaced stripes, with denser blue indicating denser genes.

3.3. Phylogenetic Analysis of DUF506 Gene in Peanuts and Four Other Plants

A total of 13, 10, 24, and 18 *DUF506* genes were identified in *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, and *Medicago truncatula*. To explain their evolutionary relationships, a Maximum Likelihood (ML) phylogenetic tree was constructed using these sequences, together with 23 *Arachis hypogaea* *DUF506* proteins, yielding a total of 88 *DUF506* members (Figure 2) [34]. The resulting tree resolved the family into 4 well-supported subfamilies (I–IV) (Table S3), in agreement with previous classifications. Notably, subfamily IIIa was the smallest, comprising only a few sequences, which suggests that this subgroup has undergone lineage-specific retention and potential functional specialization.

3.4. Analysis of Protein Structure and Gene Structure of *AhDUF506* Family Members

Members of a protein group are expected to share conserved motifs [35]. To characterize the motif and domain architecture in 23 *AhDUF506* proteins, their sequences were submitted to the MEME suite and NCBI's Conserved Domain Search (Figure 3). This analysis identified 4 conserved motifs and 2 structural domains, the PDDEXK6 superfamily and A_thal-3542. Mapping the positions of these features along each sequence revealed that the A_thal-3542 domain is the most consistently conserved across all members of *AhDUF506*.

3.5. Analysis of *AhDUF506* Cis-Acting Components

The study of promoters, primarily responsible for regulating gene expression at the transcriptional level, is vital for enhancing our fundamental comprehension of gene regulation (Figure 4) [36]. A total of 24 abiotic stress-responsive cis-acting elements were identified and categorized into 5 functional groups: (i) common elements, (ii) plant hormone-responsive elements, (iii) environmental-responsive elements, (iv) transcription factor (TF)-related elements, and (v) tissue-specific elements (Table S4).

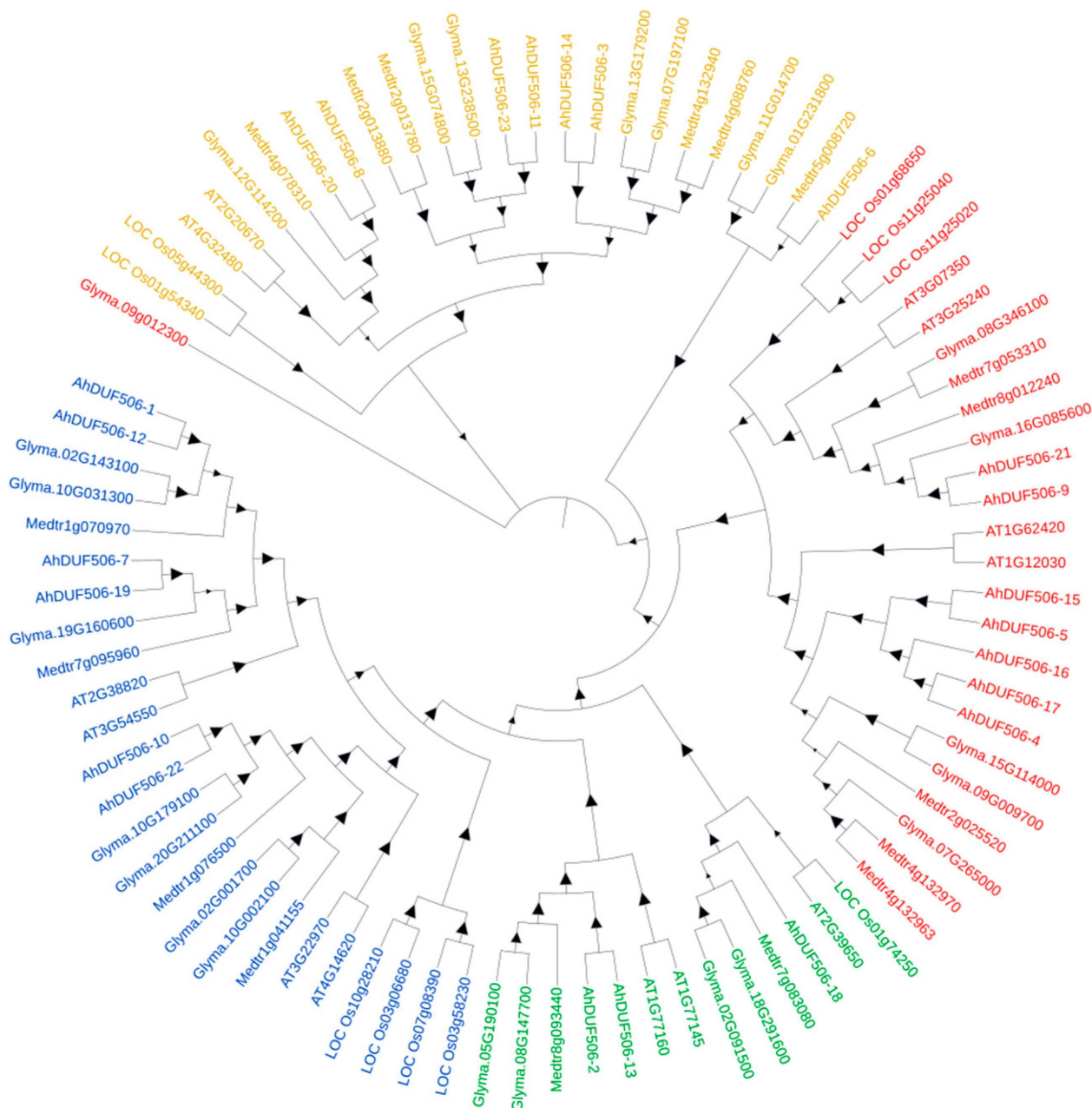


Figure 2. Phylogenetic tree of DUF506 members identified in five plant species. Members of subfamilies I, II, IIIa, and IIIb are marked in red, purple, green, and cyan, respectively. Subfamilies I, II, IIIa, and IIIb are represented by red, yellow, green, and blue, respectively. The higher the bootstrap value for a particular branch, the larger the size of the black triangles.

3.6. Synteny Analysis of *AhDUF506* Gene

Performing a collinearity analysis of *AhDUF506* can reveal evolutionary relationships and genome rearrangement events among different DUF506 species. Therefore, using the MCScanX program to study the genes involved in peanut replication events. A total of 18 pairs of segmental repetitive genes were identified, and no tandem gene pairs were found (Figure 5A) (Table S5). Assessing the proportions of tandem and segmental genome duplications within gene families is essential for functional analysis [37]. The K_a/K_s of the *AhDUF506* segmental repeat gene pairs were below 1 (ranging from 0.15 to 0.5), indicating that the gene functions of the peanut gene family are conserved during evolution. This discovery demonstrates that *AhDUF506* family genes undergo purifying selection after gene duplication. Notably, 10 pairs of *AhDUF506* proteins share high sequence

similarity, which likely reflects the tetraploid genome of *Arachis hypogaea* and the concurrent duplication of these genes during chromosomal doubling events (Table S6).

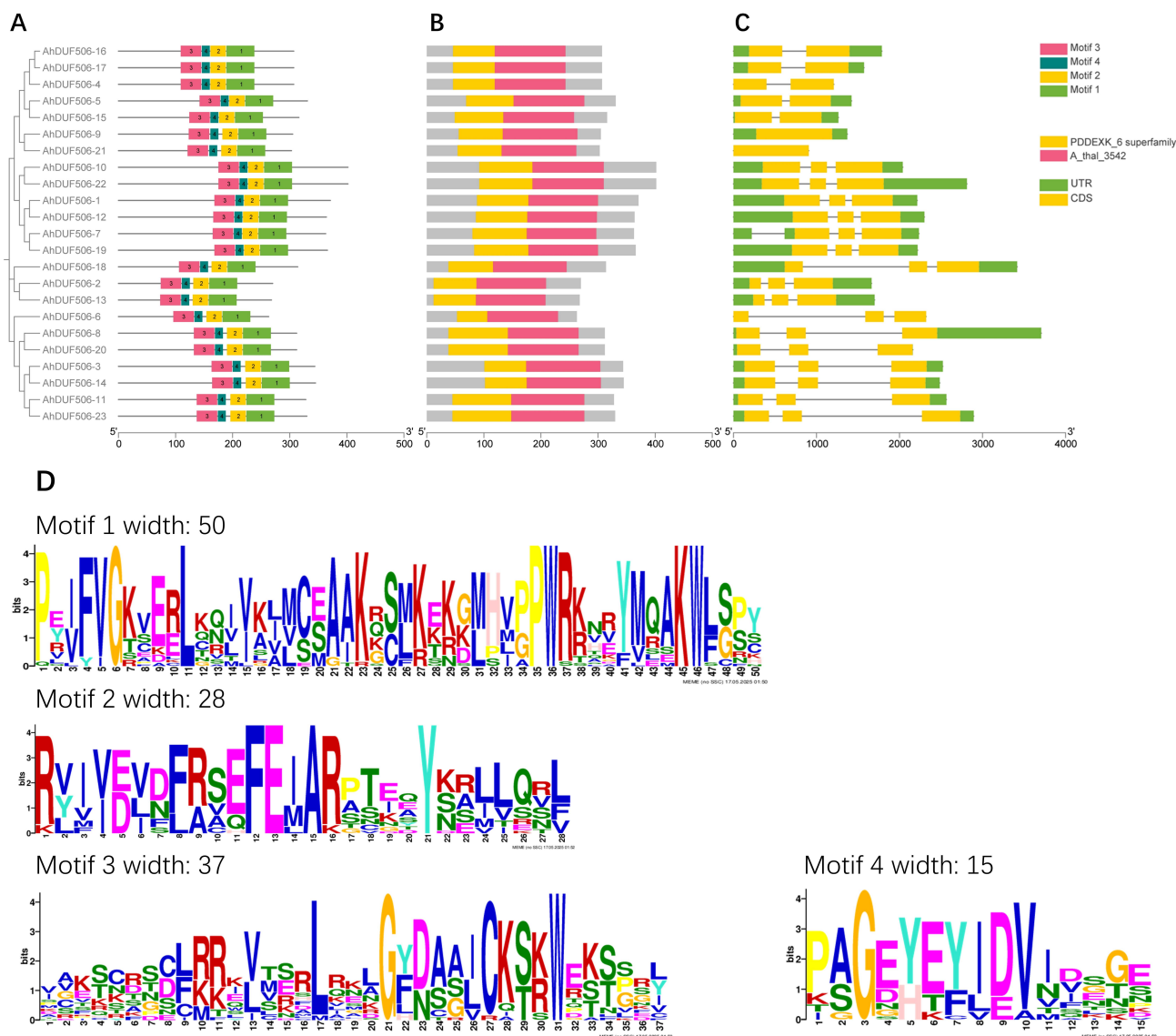


Figure 3. Distribution of motif, domain, and gene structure of AhDUF506 Family. (A) Conserved motifs of the AhDUF506 family. (B) The conserved domain of the AhDUF506 family. (C) Analysis of gene structure of the AhDUF506 family. (D) Protein sequences of conserved motifs in the AhDUF506 family.

To evaluate the genetic homology and evolutionary relationships among species, a collinearity analysis was performed between *AhDUF506* genes and the *DUF506* family genes from *Oryza sativa*, *Arabidopsis thaliana*, *Glycine max* and *Medicago truncatula*, which identified 7, 8, 52 and 32 pairs of homologous DUF506 gene pairs, respectively (Figure 5B) (Table S7). As expected, peanut and soybean share the most significant number of DUF506 orthologous pairs due to their closest genetic relationship. The Ka/Ks analysis of orthologous DUF506 gene pairs between peanut and the other four species revealed that the synonymous substitution rate in *Arabidopsis thaliana* was only slightly lower than that in *Glycine max* and *Medicago truncatula*.

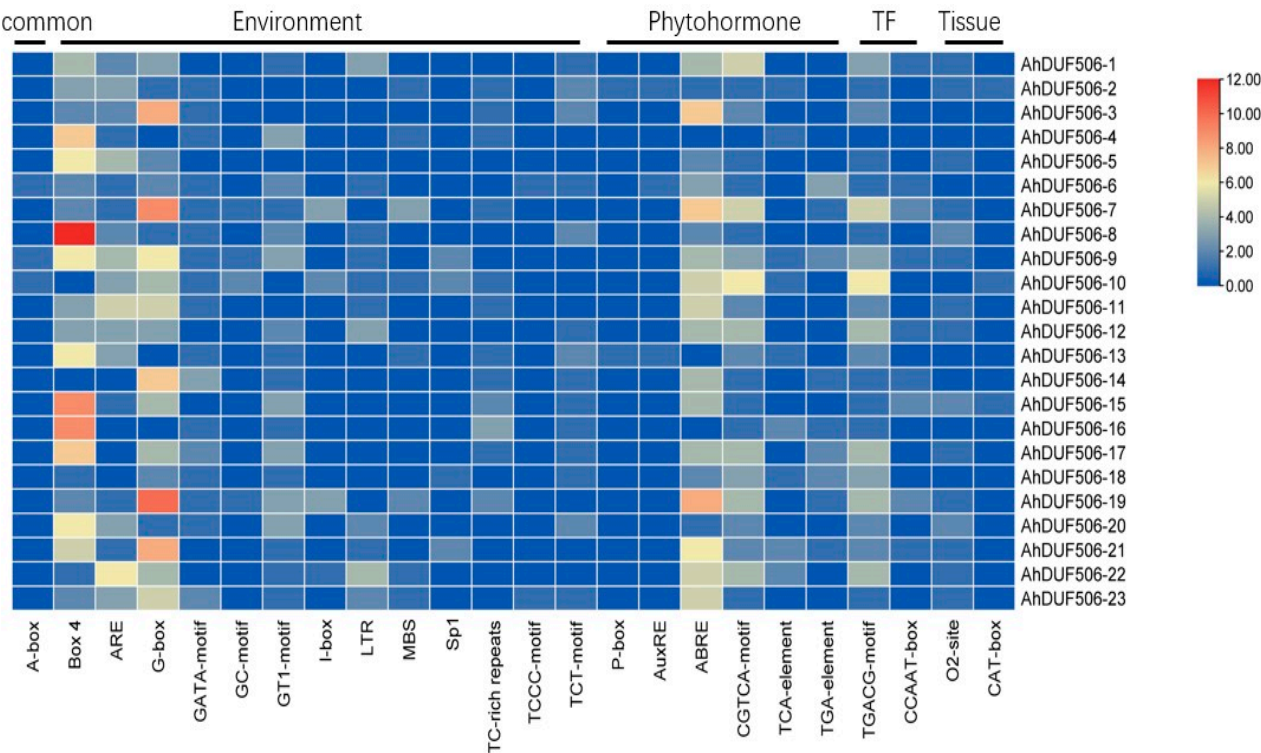


Figure 4. Heatmap of Cis-Regulatory Elements in the AhDUF506 Family. The heatmap shows the distribution of hormone-responsive, environment-responsive, and tissue-specific cis-regulatory elements across AhDUF506 genes.

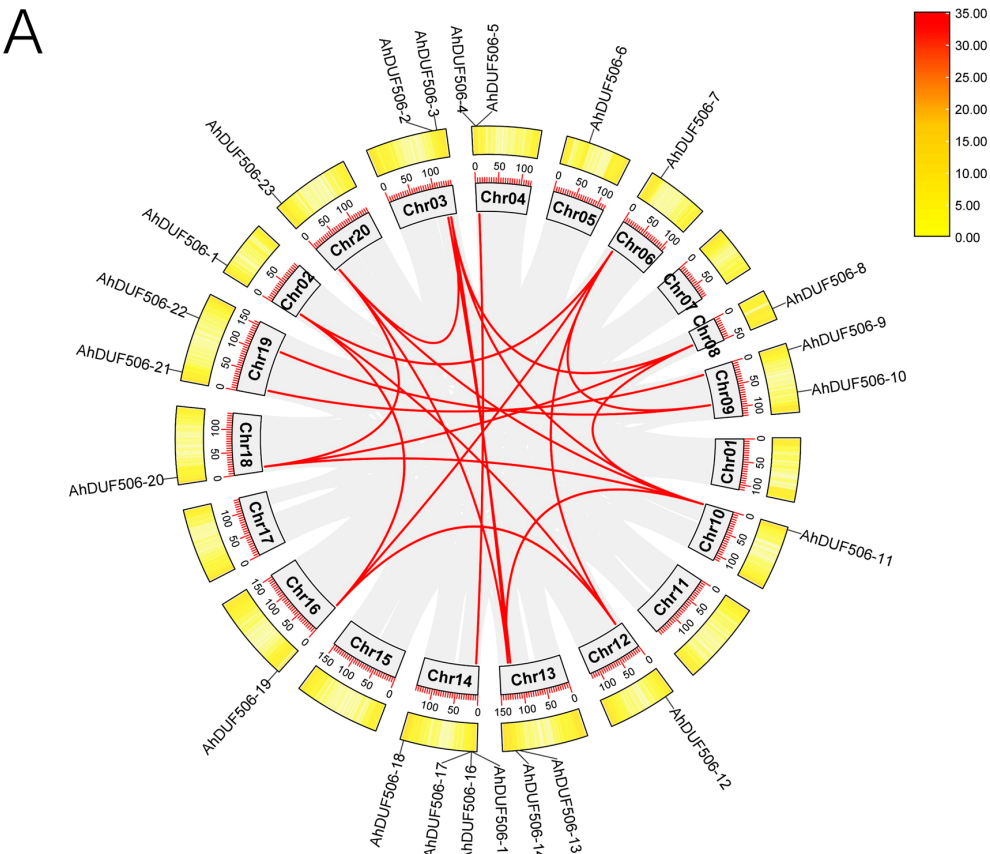


Figure 5. Cont.

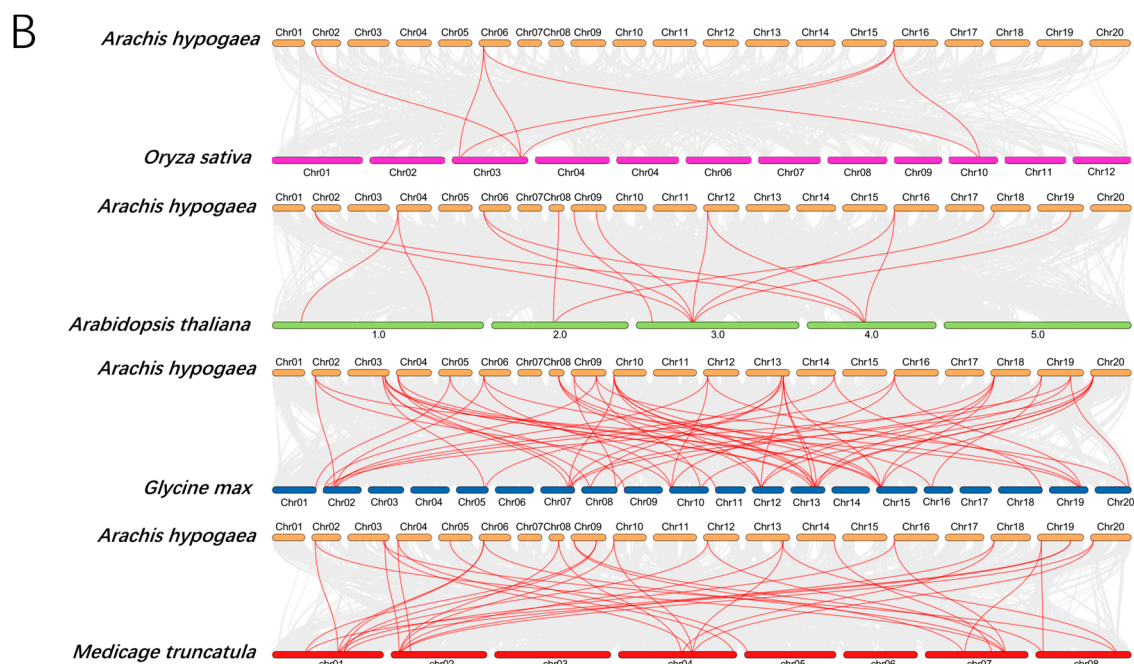


Figure 5. Gene Duplication Events of the AhDUF506 Family and Their Orthologous Relationships with DUF506 in Other Plants: (A) Gene density and the chromosomal localization of AhDUF506 members are displayed from the inside out. The gene density of chromosomes is represented by a color gradient from yellow to red, with red indicating higher density. (B) Orthologous relationships of DUF506s among *Oryza sativa*, *Arabidopsis thaliana*, *Glycine max*, and *Medicago truncatula*.

3.7. Transcriptome Analysis of AhDUF506 Gene Tissue

Analysis of publicly available transcriptomic databases revealed distinct tissue-specific expression profiles among AhDUF506 family members (Figure 6) (Table S8). *AhDUF506-2*, *AhDUF506-13* and *AhDUF506-11* showed higher transcript levels in multiple tissues. In contrast, the expression of *AhDUF506-22* was only relatively high in Testa, and the expression level of *AhDUF506-23* was also significantly higher in root tips than in other tissues. Notably, members of subfamily I demonstrated minimal or undetectable expression in all sampled tissues, whereas genes in subfamilies II and IIIa were expressed at relatively elevated levels.

3.8. Quantitative RT-PCR (RT-qPCR) Analysis

Based on the tissue-specific transcriptome data of AhDUF506 and literature references, we finally selected 7 genes from the 23 AhDUF506 genes (including *AhDUF506-15* and *AhDUF506-5* from subfamily I, *AhDUF506-23* from subfamily II, *AhDUF506-2* and *AhDUF506-13* from subfamily IIIa, as well as *AhDUF506-10* and *AhDUF506-22* from subfamily IIIb). Their expression profiles across peanut organs and developmental stages, and in response to diverse abiotic stresses, were quantified by qPCR [38].

Tissues from different parts of peanuts at the mature stage were collected, and the expression levels of 7 AhDUF506 genes in these tissues were analyzed (Figure 7). It was found that the expression levels of 6 of these genes in flowers were significantly higher than in other tissues, among which *AhDUF506-2*, *AhDUF506-5*, and *AhDUF506-22* showed the highest expression levels. This indicates that the AhDUF506 gene family plays a vital role in the biological processes of flowers. In addition, the transcription levels of 6 AhDUF506 genes were the lowest in seeds, except *AhDUF506-22*.

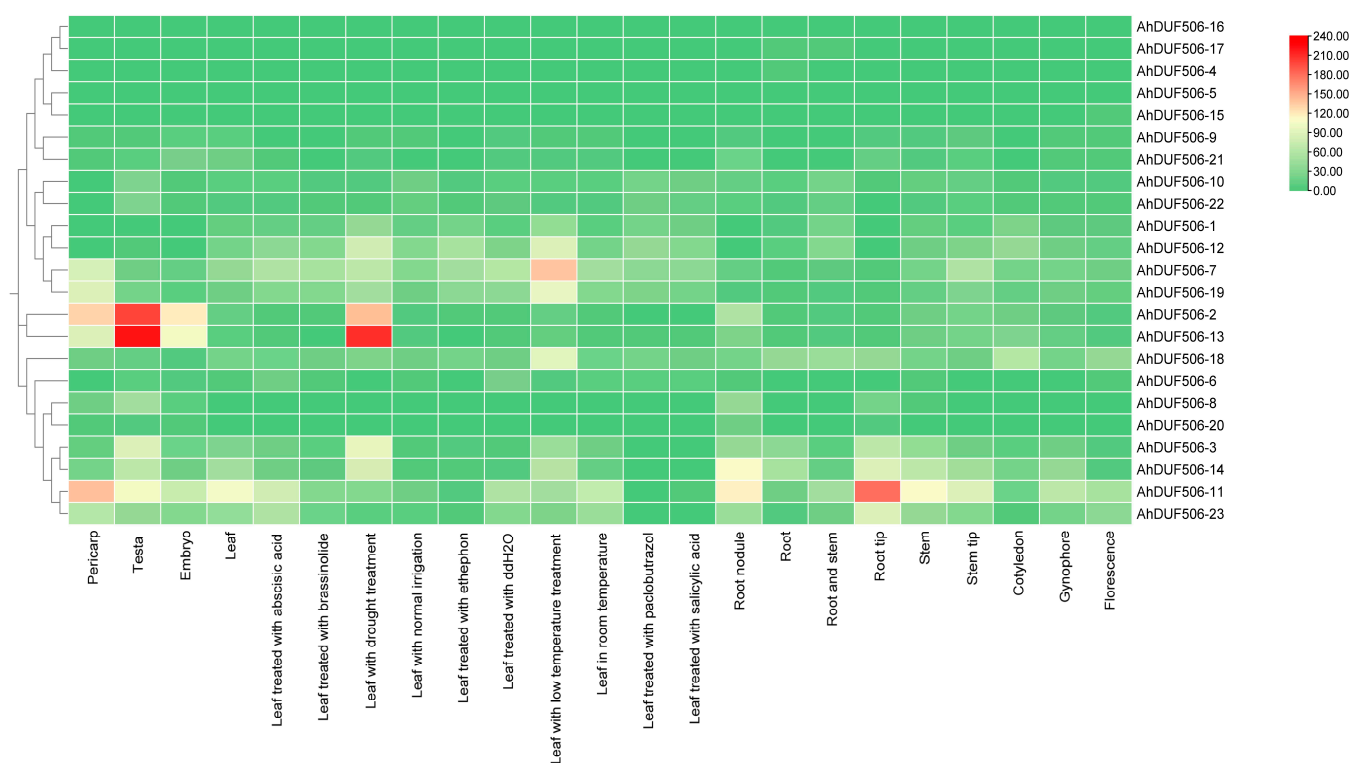


Figure 6. The heatmap shows the expression profiles of the *AhDUF506* gene family across various peanut tissues, based on transcriptome data from the Peanut Genome Resource (accessed 20 March 2025), with the adjacent green–yellow–red color bar indicating relative fold changes.

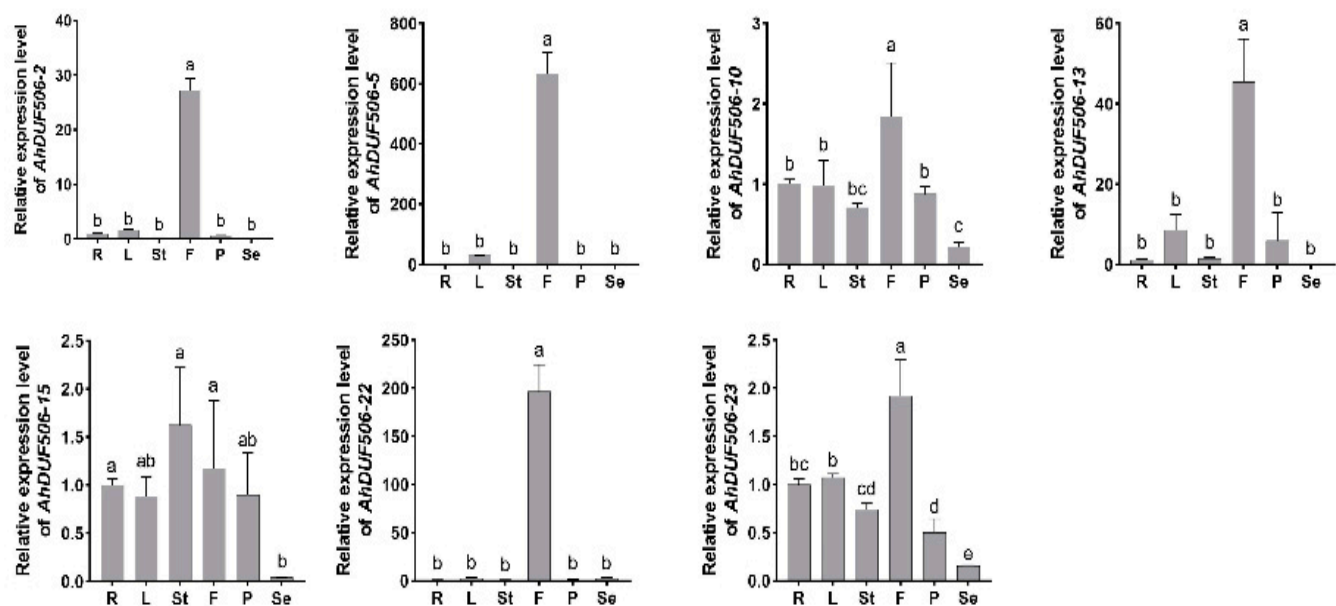


Figure 7. Transcriptional expression analysis of 7 *AhDUF506* genes in different tissues. The transcriptional level of each *AhDUF506* gene was normalized against the root as the control. L, R, and St represent leaves, roots, and stems at the mature stage; F represents flowers; P represents pods; Se represents seeds. The data represent the mean values of three biological replicates $\pm$ SD. Statistical significance of differences was tested by one-way ANOVA analysis ($p < 0.05$) and is indicated by lower case letters.

3.9. Expression Patterns of *AhDUF506* Genes Under Abiotic Stress

Under 150 mM NaCl exposure, transcript levels of all 7 *AhDUF506* genes in peanut roots were significantly upregulated, reaching their highest expression between 12 and 24 h of treatment (Figure 8). In contrast, the expression of five *AhDUF506* genes showed no significant changes in leaf tissues, while that of *AhDUF506-13* and *AhDUF506-23* was significantly downregulated. During drought stress, the 7 *AhDUF506* genes exhibited different expression patterns in leaves, while in roots, the expression trends were similar to those under osmotic stress. Interestingly, *AhDUF506-13* had a distinct expression pattern from the other six *AhDUF506* genes, being downregulated in both leaves and roots, while the other genes were upregulated in roots (Figure 9). Exposure to 4 °C elicited a uniform upregulation of all 7 *AhDUF506* genes in both roots and leaves (Figure 10). This coordinated transcriptional response indicates that members of the *AhDUF506* family actively respond when peanuts are exposed to cold stress.

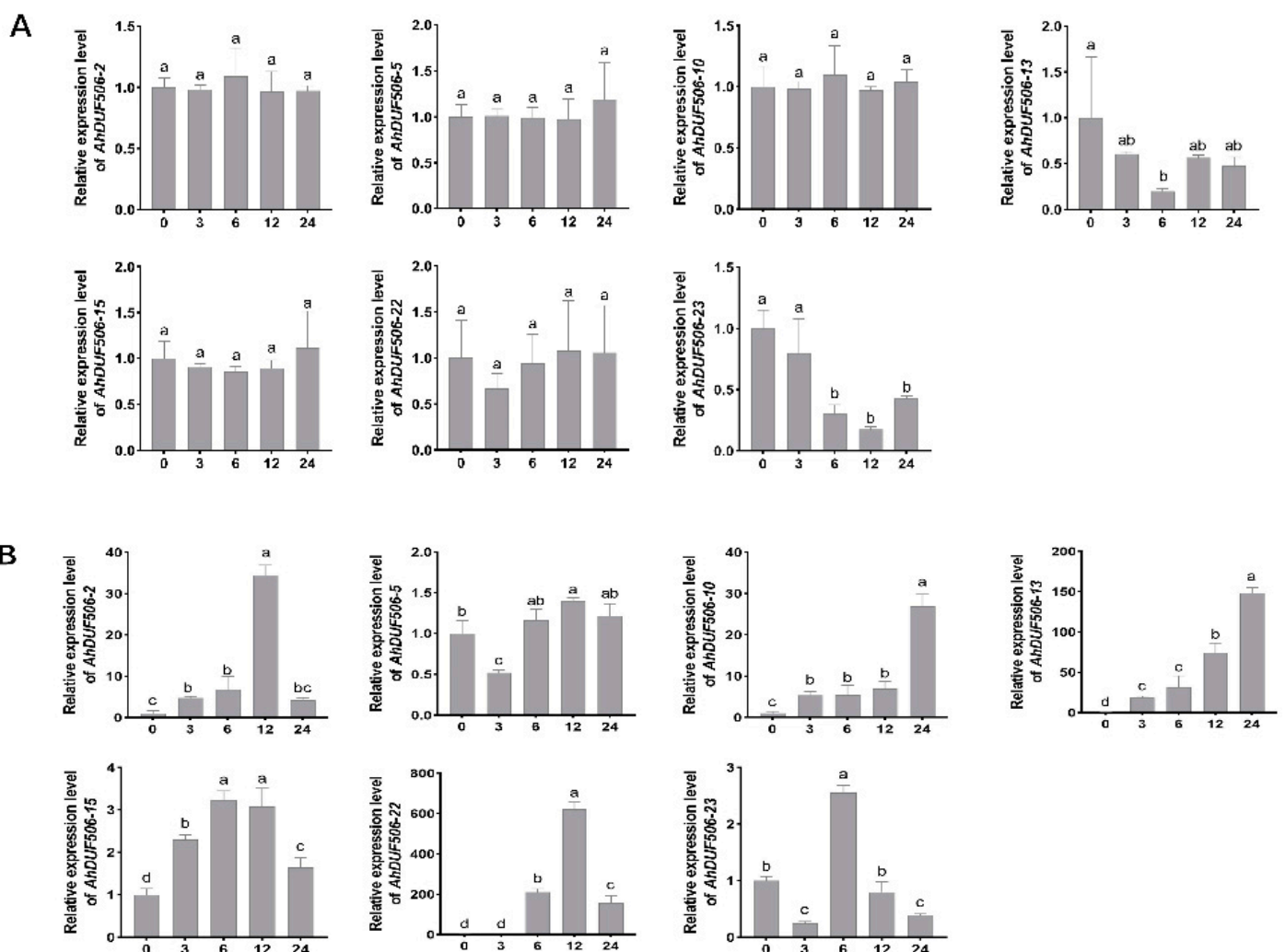


Figure 8. Transcriptional expression analysis of 7 *AhDUF506* genes at different time points under NaCl stress (150 mM NaCl). (A,B) represent peanut sprouts and roots that are three weeks old, respectively. All gene transcription level changes within 1 day are compared with the 0 h level as a control. The data represent the mean $\pm$ SD of three biological replicates. Statistical significance of differences was tested using one-way ANOVA ($p < 0.05$), and results are indicated by lowercase letters.

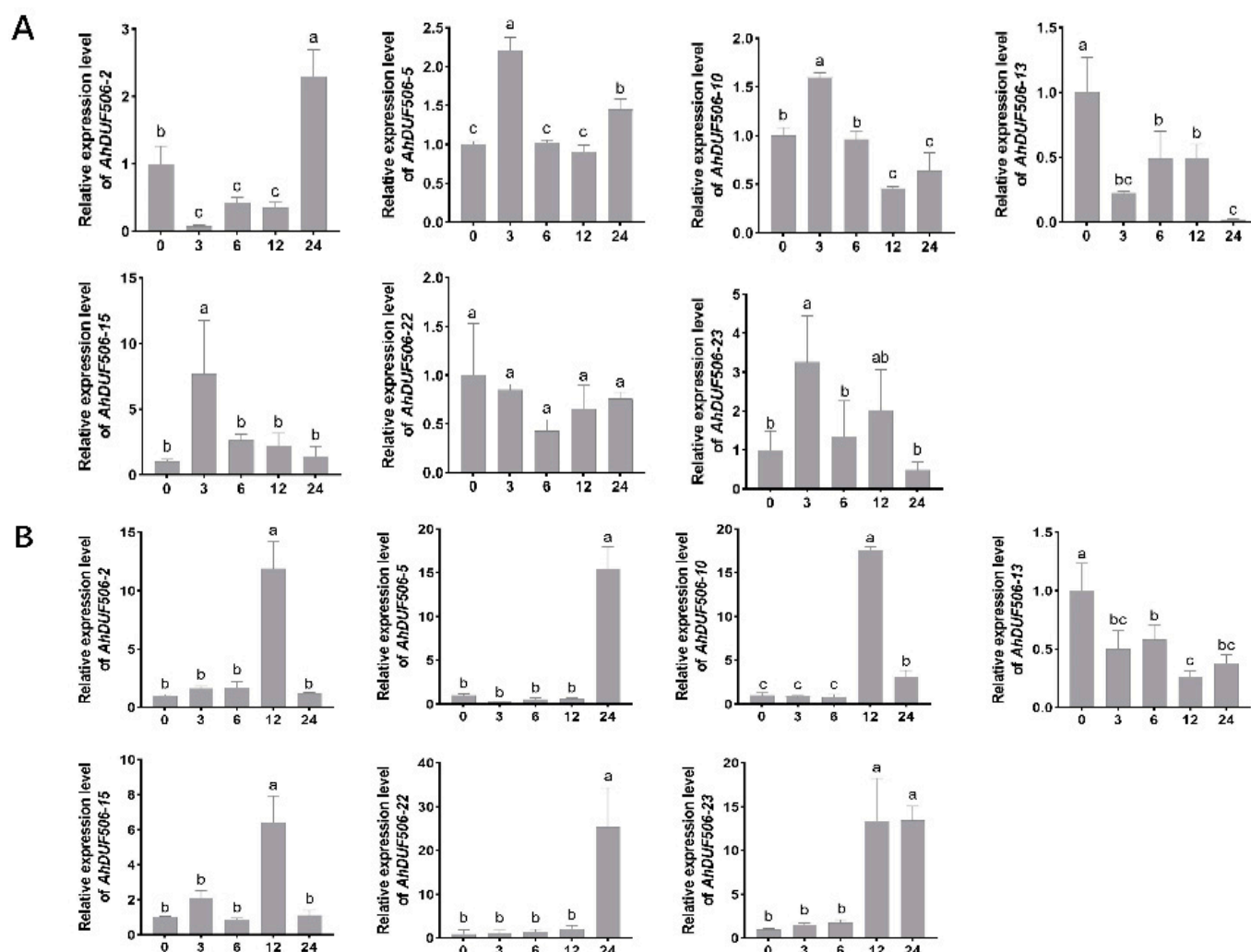


Figure 9. Transcriptional expression analysis of 7 *AhDUF506* genes at different time points under drought stress. (A,B) represent peanut sprouts and roots that are three weeks old, respectively. All gene transcription level changes within 1 day are compared with the 0 h level as a control. The data represent the mean $\pm$ SD of three biological replicates. Statistical significance of differences was tested using one-way ANOVA ($p < 0.05$), and results are indicated by lowercase letters.

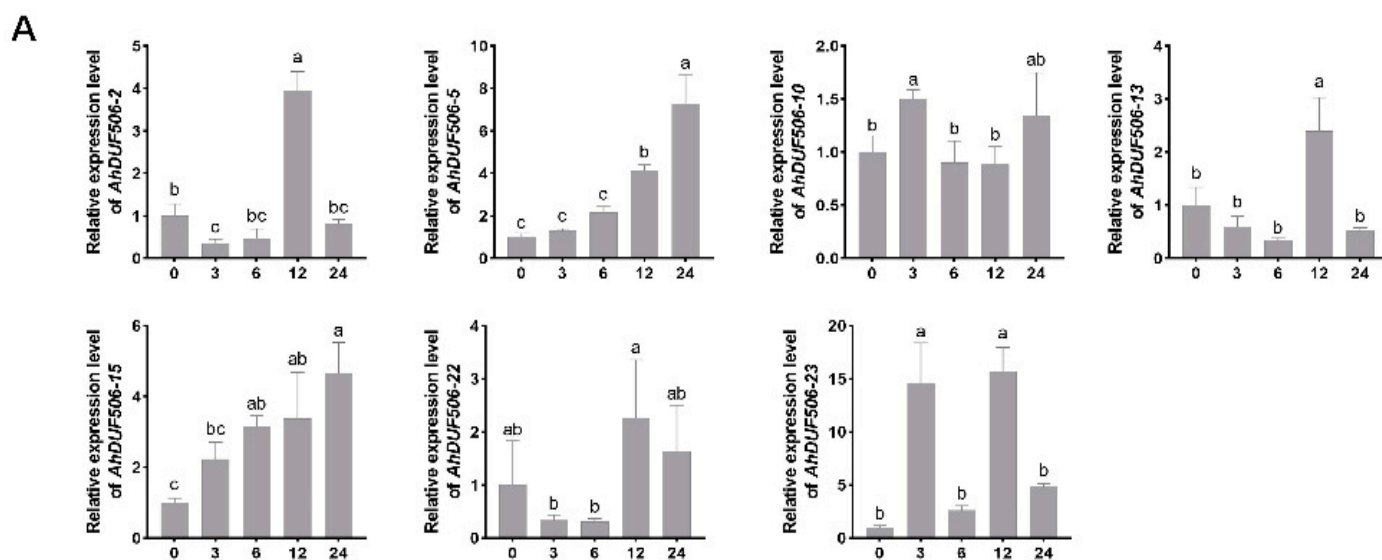


Figure 10. Cont.

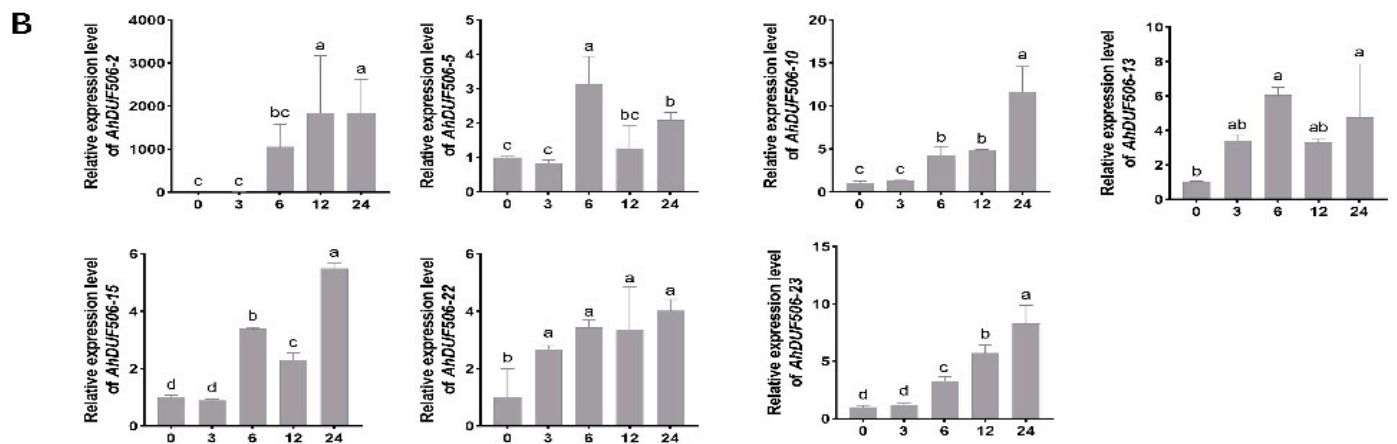


Figure 10. Transcriptional expression analysis of 7 *AhDUF506* genes at different time points under cold stress. (A,B) represent peanut sprouts and roots that are three weeks old, respectively. All gene transcription level changes within 1 day are compared with the 0 h level as a control. The data represent the mean $\pm$ SD of three biological replicates. Statistical significance of differences was tested using one-way ANOVA ($p < 0.05$), and results are indicated by lowercase letters.

Under treatment with 100 μ M abscisic acid (ABA), the expression pattern of *AhDUF506* exhibited a completely distinct profile compared with that under the three aforementioned stress conditions: the expression of seven *AhDUF506* genes in leaves all showed a pattern of initial upregulation followed by a decline, while the corresponding transcripts in roots peaked at the early stage of treatment and then decreased subsequently (Figure 11). In the phosphorus-deficiency test, most *AhDUF506* genes were upregulated in both organs (Figure 12). Exceptions included *AhDUF506-2*, which was downregulated in roots and leaves, and *AhDUF506-13* and *AhDUF506-22*, which were specifically downregulated in leaves.

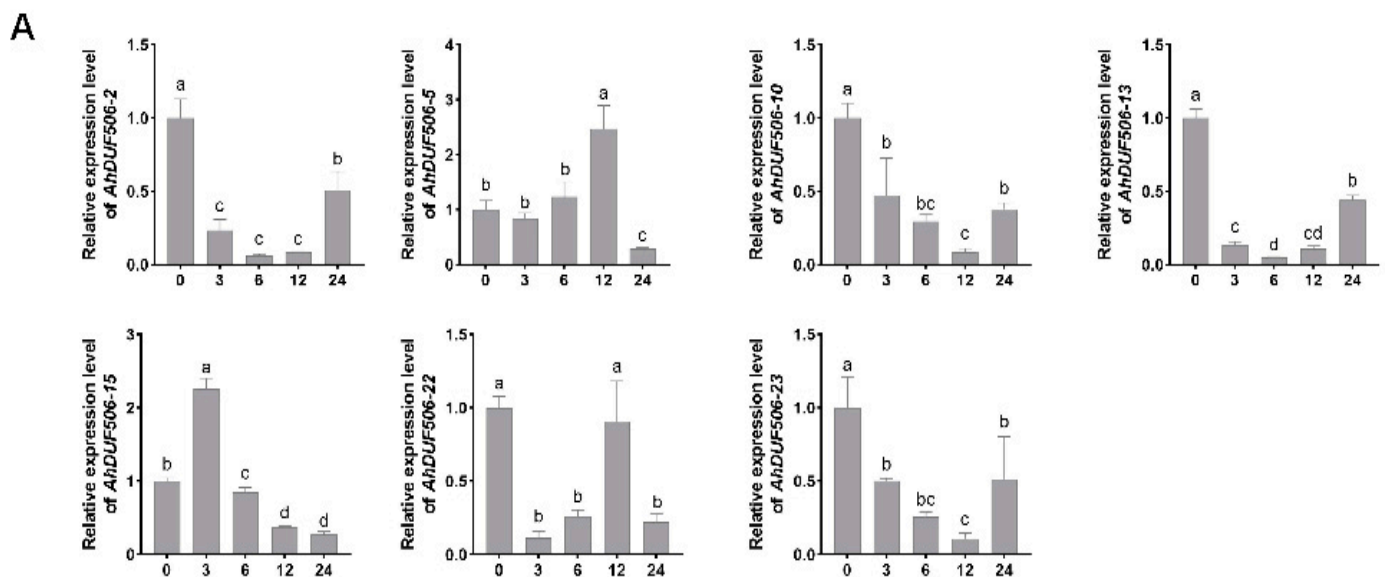


Figure 11. Cont.

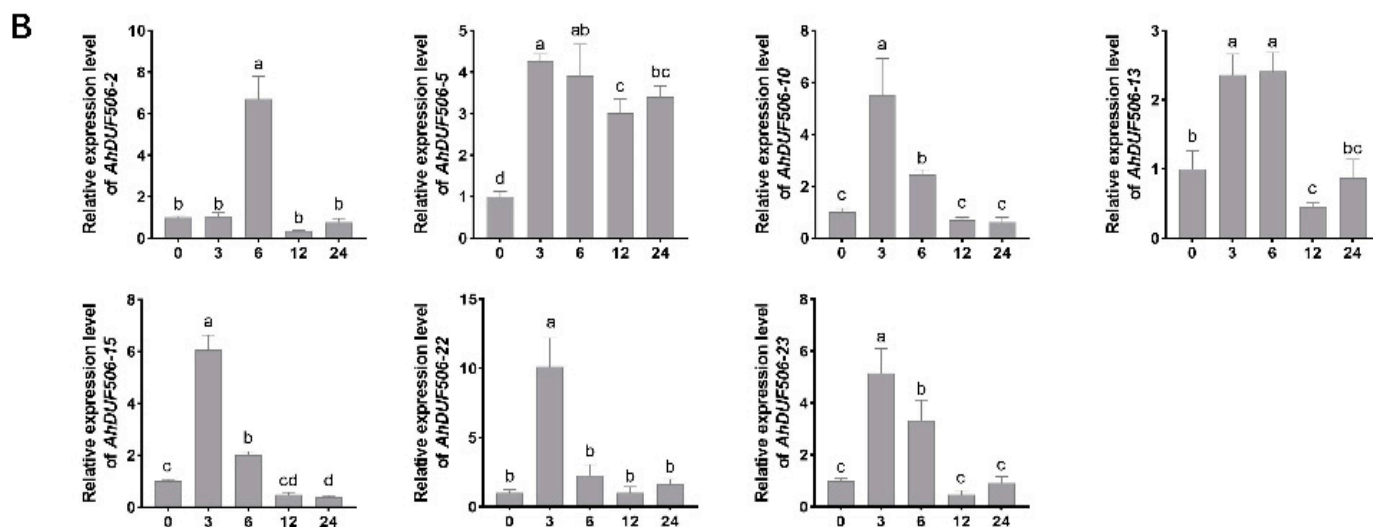


Figure 11. Transcriptional expression analysis of 7 *AhDUF506* genes under ABA treatment at different time points. (A,B) represent peanut sprouts and roots that are three weeks old, respectively. All gene transcription level changes within 1 day are compared with the 0 h level, which served as a control. The data represent the mean $\pm$ SD of three biological replicates. Statistical significance of differences was tested using one-way ANOVA ($p < 0.05$), and results are indicated by lowercase letters.

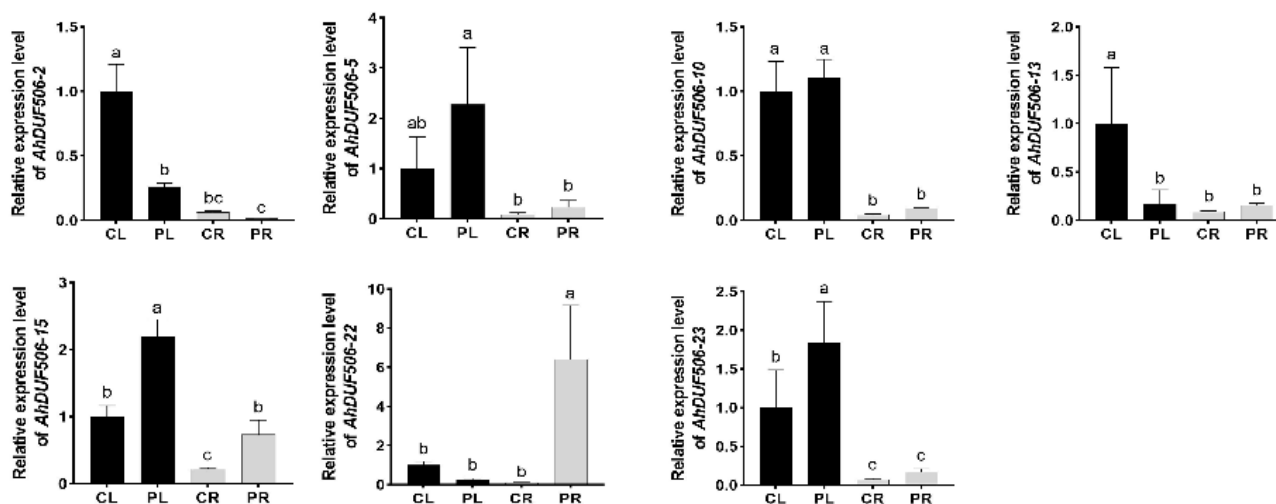


Figure 12. Transcriptional expression analysis of 7 *AhDUF506* genes in roots (white) and leaves (black) under phosphorus deficiency treatment. C, P, L, and R represent the control, phosphorus deficiency treatment, leaf, and root, respectively. The expression levels of all genes are compared with those in normally growing leaves, which serve as the control. The data represent the mean $\pm$ SD of three biological replicates. Statistical significance of differences was tested using one-way ANOVA ($p < 0.05$), and results are indicated by lowercase letters.

4. Discussion

In the protein family database (Pfam database) [39], hundreds of protein families remain uncharacterized and poorly annotated, collectively known as DUF [40]. In the updated Pfam database, 5 new PD-(D/E)XK nuclease families have been identified; 4 of these belong to DUF families (DUF506, DUF524, DUF1626, and DUF1703) and are involved in numerous nucleic acid cleavage events that are crucial for various cellular processes [6]. Previous studies have shown that DUF506 is associated with stress resistance across different plants [8–12]. However, a comprehensive analysis of the *DUF506* gene family in peanuts remains lacking. Consequently, this study identified 23 *AhDUF506* genes, comparable to the 24 reported in *Glycine max*. Homology alignment analysis further

revealed that *Arachis hypogaea* shares higher homology with soybean than with other species, confirming their evolutionary relationship [41]. Additionally, we found that the 23 *AhDUF506* genes in peanut include 10 pairs of highly homologous genes with sequence similarities exceeding 96%. *Arachis hypogaea* is a tetraploid plant with an AABB genome [42]. These 10 gene pairs display a symmetrical distribution on the chromosomes, supporting Gregory's view that *Arachis duranensis* and *Arachis cardenasii* may be ancestors of cultivated peanuts and aligning with Smartt's hypothesis that cultivated peanuts originated from the hybridization of two diploid wild species with different chromosome sets, followed by a single polyploidization event [43,44]. The high similarity among these gene pairs may result from gene redundancy caused by tetraploidization. Since peanut is a tetraploid plant, it is hypothesized that this phenomenon results from chromosome doubling events [45,46].

The correlation between the content of protein domains and their functions holds significant importance for the study and comprehension of how domain combinations encode complex functionalities [47,48]. Genetic structure analysis has demonstrated that the sequence and number of conserved motifs in members of the *AhDUF506* family differ from the two conserved motifs and one conserved domain identified in *Arabidopsis thaliana* [8,49]. The family members encompass not only the PDDEXK_6 superfamily domain but also A_thal_3542, which has an unidentified function. Gene duplication constitutes a critical mechanism for the acquisition of new genes and the generation of genetic novelty in plants [50,51]. Through collinearity analysis, we have identified 18 pairs of duplicated genes. Examination of the K_a and K_s ratio indicates significant purifying selection within the *DUF506* gene family ($K_a/K_s < 1$; Table S4), implying a central role for this gene family in the adaptive evolution of *Arachis hypogaea*. This may suggest the presence of functional redundancy [52].

Most studies indicate that the *DUF506* family is associated with plant stress resistance; however, few have examined its expression levels across different tissues [8–12]. To investigate the expression of peanut *DUF506* in various tissues and under stress conditions, we selected 7 members based on their transcriptional data and subfamily classification for further analysis. In peanut tissues, the expression level in flowers was observed to be significantly higher than that in other tissues, a pattern distinct from that in *Arabidopsis thaliana* [8]. This suggests that *AhDUF506* may be involved in floral metabolism. Under abiotic stress conditions, transcriptional alterations among the selected *AhDUF506* genes are more pronounced in roots than in leaves. The tandemly duplicated gene pair, *AhDUF506-2* and *AhDUF506-13*, is induced by dehydration or drought stress (Figures 7 and 8). Nonetheless, their expression patterns differ, likely due to variations in their cis-regulatory elements (CREs). Under phosphorus deficiency, *AhDUF506-22* is strongly induced in roots, consistent with the response observed in its homolog, *OsDUF50605*.

5. Conclusions

In this study, a comprehensive genome-wide identification and analysis were performed, resulting in the identification of 23 high-confidence *AhDUF506* genes in peanut (*Arachis hypogaea*). These 23 *AhDUF506* genes were mapped to 15 chromosomes and categorized into 4 subfamilies based on previous studies. The *AhDUF506* family is relatively conserved, with all members containing 4 conserved motifs and 2 identical structural domains. Additionally, analyses of their physicochemical properties, cis-acting elements, and genomic collinearity were conducted to characterize their features. In addition, the transcriptional expression patterns of seven representative *AhDUF506* genes were investigated across different tissues and under various environmental stress conditions, and an in-depth analysis of the expression pattern of *AhDUF506-22* was performed. Overall, these findings improve our understanding of the role of the *AhDUF506* gene family in

plant adaptation to environmental stresses and identify potential target genes for breeding stress-tolerant peanut cultivars using genome editing technologies.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biom16020270/s1>, Table S1: The primers used in RT-qPCR analysis; Table S2: Physical and chemical properties and subcellular localization of AhDUF506 protein; Table S3: The ID information of DUF506 genes in 7plant; Table S4: Promoter analysis of AhDUF506 genes; Table S5: The Ka/Ks ratios in duplicated gene pairs identified in AhDUF506s; Table S6: The Similarity of DUF506 Gene Family Members; Table S7: The Ka_Ks analysis for DUF506 genes between peanut and other model species species; Table S8: Raw data for transcriptomic analysis of AhDUF506 family in different tissues; File S1: The AhDUF506 protein sequence and rename; File S2: CDS sequence of AhDUF506 family members; File S3: AhDUF506 family member promoter sequence.

Author Contributions: This article is a collaborative effort of all the authors. Conceptualization, M.C.; methodology, Q.S.; validation, Q.S., G.A.A., and M.L.; formal analysis, Q.S.; investigation, Y.L. and M.H.; resources, Y.H. and R.W.; writing—original draft preparation, Q.S.; writing—review and editing, G.A.A., M.C.; visualization, Q.S.; supervision, M.C.; project administration, M.C.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript. This article is a collaborative effort of all the authors.

Funding: This research was funded by the Natural Science Foundation of Guangdong Province (K22082), the Scientific Research Start-up Fund of Guangdong Ocean University (R19048), the Foundation of Department of Education of Guangdong Province (230420045), the Modern Seed Industry Project (2022B0202060004), and the Science and Technology Project of Zhanjiang City (A22494).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The DNA, cDNA, and protein sequences of the peanut DUF506 family in this study can be retrieved from the peanut genome database (<http://peanutgr.fafu.edu.cn/Download.php>, accessed on 20 March 2025) and are included in Supplementary Materials File S1–S3. Plant material is available from the corresponding author upon reasonable request.

Acknowledgments: We thank Weijian Zhuang (Department of Plant Molecular and Cellular Biology, Fujian Agriculture and Forestry University) and colleagues for their support with transcriptome data analysis.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bateman, A.; Coghill, P.; Finn, R.D. DUFs: Families in search of function. *Struct. Biol. Cryst. Commun.* **2010**, *66*, 1148–1152. [CrossRef]
2. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.; Tosatto, S.C.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [CrossRef]
3. Jaroszewski, L.; Li, Z.; Krishna, S.S.; Bakolitsa, C.; Wooley, J.; Deacon, A.M.; Wilson, I.A.; Godzik, A. Exploration of uncharted regions of the protein universe. *PLoS Biol.* **2009**, *7*, e1000205. [CrossRef]
4. Luo, C.; Akhtar, M.; Min, W.; Bai, X.; Ma, T.; Liu, C. Domain of unknown function (DUF) proteins in plants: Function and perspective. *Protoplasma* **2024**, *261*, 397–410. [CrossRef]
5. Lv, P.; Wan, J.; Zhang, C.; Hina, A.; Al Amin, G.M.; Begum, N.; Zhao, T. Unraveling the diverse roles of neglected genes containing domains of unknown function (DUFs): Progress and perspective. *Int. J. Mol. Sci.* **2023**, *24*, 4187. [CrossRef] [PubMed]
6. Knizewski, L.; Kinch, L.N.; Grishin, N.V.; Rychlewski, L.; Ginalski, K. Realm of PD-(D/E) XK nuclease superfamily revisited: Detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.* **2007**, *7*, 40. [CrossRef] [PubMed]
7. Schultz, J.; Milpetz, F.; Bork, P.; Ponting, C.P. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5857–5864. [CrossRef] [PubMed]
8. Ying, S. Genome-wide identification and transcriptional analysis of Arabidopsis DUF506 gene family. *Int. J. Mol. Sci.* **2021**, *22*, 11442. [CrossRef]

9. Dong, W.; Tu, J.; Deng, W.; Zhang, J.; Xu, Y.; Gu, A.; An, H.; Fan, K.; Wang, R.; Zhang, J.; et al. Genome-wide identification of DUF506 gene family in *Oryza sativa* and expression profiling under abiotic stresses. *PeerJ* **2023**, *11*, e16168. [\[CrossRef\]](#)
10. Wang, Y.; Liu, S.; Lin, T.; Kan, S.; Yu, P. The DUF506 Gene Family in *Triticum aestivum*: Genome-Wide Identification and Expression Profiling Under Salt Stress. *Agronomy* **2025**, *15*, 281. [\[CrossRef\]](#)
11. Zhu, G.; Wang, J.; He, S.; Liang, K.; Zhang, R.; Huang, J.; Yang, X.; Zhang, X. Comprehensive analysis of BrDUF506 genes across the *Brassica rapa* genome uncovers potential functions in sexual reproduction and abiotic stress tolerance. *Int. J. Mol. Sci.* **2024**, *25*, 11087. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Wang, D.R.; Xu, M.H.; Wang, X.; Chen, G.L.; Wang, Q.; Wang, A.S.; Cao, C.; Liu, D.; Chen, Y.; Qu, M.; et al. MdDUF506 Enhances Aluminium Tolerance by Interacting With MdCNR8 in Apple. *Plant Cell Environ.* **2025**, *48*, 7041–7054. [\[CrossRef\]](#)
13. Kosinski, J.; Feder, M.; Bujnicki, J.M. The PD-(D/E) XK superfamily revisited: Identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinform.* **2005**, *6*, 172. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Variath, M.T.; Janila, P. Economic and academic importance of peanut. In *The Peanut Genome*; Springer International Publishing: Cham, Switzerland, 2017; pp. 7–26.
15. Bertoli, D.J.; Jenkins, J.; Clevenger, J.; Dudchenko, O.; Gao, D.; Seijo, G.; Leal-Bertoli, S.C.M.; Ren, L.; Farmer, A.D.; Pandey, M.K.; et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* **2019**, *51*, 877–884. [\[CrossRef\]](#)
16. Song, H.; Wang, P.; Lin, J.Y.; Zhao, C.; Bi, Y.; Wang, X. Genome-wide identification and characterization of WRKY gene family in peanut. *Front. Plant Sci.* **2016**, *7*, 534. [\[CrossRef\]](#)
17. Wu, Z.; Luo, L.; Wan, Y.; Liu, F. Genome-wide characterization of the PP2C gene family in peanut (*Arachis hypogaea* L.) and the identification of candidate genes involved in salinity-stress response. *Front. Plant Sci.* **2023**, *14*, 1093913.
18. Liu, Y.; Xiao, L.; Chi, J.; Li, R.; Han, Y.; Cui, F.; Peng, Z.; Wan, S.; Li, G. Genome-wide identification and expression of SAUR gene family in peanut (*Arachis hypogaea* L.) and functional identification of AhSAUR3 in drought tolerance. *BMC Plant Biol.* **2022**, *22*, 178. [\[CrossRef\]](#)
19. Pandey, M.K.; Monyo, E.; Ozias-Akins, P.; Liang, X.; Guimarães, P.; Nigam, S.N.; Upadhyaya, H.D.; Janila, P.; Zhang, X.; Guo, B.; et al. Advances in *Arachis* genomics for peanut improvement. *Biotechnol. Adv.* **2012**, *30*, 639–651. [\[CrossRef\]](#)
20. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [\[PubMed\]](#)
21. Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Bork, P.; Bridge, A.; Colwell, L.; et al. InterPro in 2022. *Nucleic Acids Res.* **2023**, *51*, D418–D427. [\[CrossRef\]](#)
22. Eddy, S.R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **2011**, *7*, e1002195. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Marchler-Bauer, A.; Bryant, S.H. CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* **2004**, *32*, W327–W331. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [\[CrossRef\]](#)
25. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [\[CrossRef\]](#)
26. Hall, B.G. Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **2013**, *30*, 1229–1235. [\[CrossRef\]](#)
27. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [\[CrossRef\]](#)
28. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [\[CrossRef\]](#)
30. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [\[CrossRef\]](#)
31. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.-H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [\[CrossRef\]](#)
32. Sauer, D.B.; Burroughs, R. Disinfection of seed surfaces with sodium hypochlorite. *Phytopathology* **1986**, *76*, 745–749. [\[CrossRef\]](#)
33. Navarro, S.; Donahaye, E.; Kleinerman, R.; Haham, H. The influence of temperature and moisture content on the germination of peanut seeds. *Peanut Sci.* **1989**, *16*, 6–9. [\[CrossRef\]](#)
34. Rzhetsky, A.; Nei, M. A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Evol.* **1992**, *9*, 945–967. [\[CrossRef\]](#)
35. Nevill-Manning, C.G.; Wu, T.D.; Brutlag, D.L. Highly specific protein sequence motifs for genome analysis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5865–5871. [\[CrossRef\]](#) [\[PubMed\]](#)

36. Ayoubi, T.A.; Van De Yen, W.J. Regulation of gene expression by alternative promoters. *FASEB J.* **1996**, *10*, 453–460. [[CrossRef](#)] [[PubMed](#)]
37. Cannon, S.B.; Mitra, A.; Baumgarten, A.; Young, N.D.; May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* **2004**, *4*, 10. [[CrossRef](#)]
38. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)]
39. Bateman, A.; Coin, L.; Durbin, R.; Finn, R.D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E.L.L.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2004**, *32*, D138–D141. [[CrossRef](#)]
40. Mudgal, R.; Sandhya, S.; Chandra, N.; Srinivasan, N. De-DUFing the DUFs: Deciphering distant evolutionary relationships of Domains of Unknown Function using sensitive homology detection methods. *Biol. Direct* **2015**, *10*, 38. [[CrossRef](#)]
41. Leppik, E.E. Assumed gene centers of peanuts and soybeans. *Econ. Bot.* **1971**, *25*, 188–194. [[CrossRef](#)]
42. Bertoli, D.J.; Seijo, G.; Freitas, F.O.; Valls, J.F.; Leal-Bertoli, S.C.; Moretzsohn, M.C. An overview of peanut and its wild relatives. *Plant Genet. Resour.* **2011**, *9*, 134–149. [[CrossRef](#)]
43. Gregory, M.P.; Gregory, W.C. Exotic germ plasm of *Arachis* L. interspecific hybrids. *J. Hered.* **1979**, *70*, 185–193. [[CrossRef](#)]
44. Smartt, J.; Gregory, W.C.; Gregory, M.P. The genomes of *Arachis hypogaea*. 1. Cytogenetic studies of putative genome donors. *Euphytica* **1978**, *27*, 665–675. [[CrossRef](#)]
45. Zhang, L.; Yang, X.; Tian, L.; Chen, L.; Yu, W. Identification of peanut (*Arachis hypogaea*) chromosomes using a fluorescence in situ hybridization system reveals multiple hybridization events during tetraploid peanut formation. *New Phytol.* **2016**, *211*, 1424–1439. [[CrossRef](#)]
46. Adams, K.L.; Wendel, J.F. Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **2005**, *8*, 135–141. [[CrossRef](#)]
47. Forslund, K.; Sonnhammer, E.L. Predicting protein function from domain content. *Bioinformatics* **2008**, *24*, 1681–1687. [[CrossRef](#)] [[PubMed](#)]
48. Vogel, C.; Bashton, M.; Kerrison, N.D.; Chothia, C.; Teichmann, S.A. Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* **2004**, *14*, 208–216. [[CrossRef](#)]
49. Ying, S.; Blancaflor, E.B.; Liao, F.; Scheible, W.R. A phosphorus-limitation induced, functionally conserved DUF506 protein is a repressor of root hair elongation in plants. *New Phytol.* **2022**, *233*, 1153–1171. [[CrossRef](#)] [[PubMed](#)]
50. Magadum, S.; Banerjee, U.; Murugan, P.; Gangapur, D.; Ravikesavan, R. Gene duplication as a major force in evolution. *J. Genet.* **2013**, *92*, 155–161. [[CrossRef](#)] [[PubMed](#)]
51. Flagel, L.E.; Wendel, J.F. Gene duplication and evolutionary novelty in plants. *N. Phytol.* **2009**, *183*, 557–564. [[CrossRef](#)]
52. Nowak, M.A.; Boerlijst, M.C.; Cooke, J.; Smith, J.M. Evolution of genetic redundancy. *Nature* **1997**, *388*, 167–171. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.