



OPEN ACCESS

EDITED BY

Xiaoxu Li,
Chinese Academy of Agricultural Sciences
(CAAS), China

REVIEWED BY

Zhiyuan Li,
Chinese Academy of Agricultural Sciences
(CAAS), China
Jinhao Sun,
Chinese Academy of Agricultural Sciences,
China
Cun Guo,
Tarim University, China
Guoxin Chen,
Beijing Life Science Academy, China

*CORRESPONDENCE

Wei Ma

✉ mawei@hljucm.edu.cn

Weichao Ren

✉ lzyrenweichao@126.com

Xiubo Liu

✉ zyyxb@foxmail.com

†These authors share first authorship

RECEIVED 20 March 2025

ACCEPTED 12 May 2025

PUBLISHED 05 June 2025

CITATION

Pan M, Yue C, Jiang S, Ma J, Ma L, Kong L,
Ling Y, Ma W, Ren W and Liu X (2025)
Genome-wide identification, phylogenetic
and expression pattern analysis of GATA
gene family in *Cerasus humilis*.
Front. Plant Sci. 16:1596930.
doi: 10.3389/fpls.2025.1596930

COPYRIGHT

© 2025 Pan, Yue, Jiang, Ma, Ma, Kong, Ling,
Ma, Ren and Liu. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Genome-wide identification, phylogenetic and expression pattern analysis of GATA gene family in *Cerasus humilis*

Meitong Pan^{1†}, Chenzhuo Yue^{1†}, Shan Jiang^{1†}, Junbai Ma¹,
Lengleng Ma¹, Lingyang Kong¹, Yang Ling², Wei Ma^{1*},
Weichao Ren^{1*} and Xiubo Liu^{3*}

¹College of Pharmacy, Heilongjiang University of Chinese Medicine, Harbin, China, ²Functional Natural Products Research Center, Yichun Branch of Heilongjiang Academy of Forestry, Yichun, China,

³College of Jiamusi, Heilongjiang University of Chinese Medicine, Jiamusi, China

Introduction: The transcription factor GATA plays a pivotal role in plant growth, physical and metabolic functions, and responses to changes in the environment. After the completion of the *Cerasus humilis* genome, investigations into its GATA gene family were not pursued.

Methods: Our research team identified the GATA gene's bioinformatics techniques, analyzed its structural characteristics and evolutionary trajectory, and investigated its expression patterns among various subfamilies.

Results: In the *C. humilis* genome, 20 *ChGATAs* are divided into four unique subgroups distributed over 7 chromosomes. Collinearity analysis showed 7 pairs of segmental duplications and 2 pairs of tandem repeats. The duplication of these fragments is vital for the development of the *ChGATA* family. The evolutionary connection between *C. humilis* and the *Malus pumila* GATA gene family is more evident than with *Oryza sativa*. Moreover, the promoter region of the *ChGATA* gene family contained cis-acting elements linked to stress, hormones, and plant growth. The transcriptome heatmap showed that the expression of the *ChGATA* was specific under alkali stress, and *ChGATAs* in the same subpopulation would also show different expression patterns. qPCR analysis showed that most of the screened *ChGATAs* first decreased and then increased with time. In addition, the dual luciferase assay and protein interaction prediction showed that *ChWRKY29* could activate the expression of the differentially expressed gene *ChGATA14* in response to alkali stress, and *ChGATA14* was at the core of the protein interaction network and had a strong interaction with *ChGATA2* and *ChGATA16*.

Discussion: This study laid a theoretical and scientific foundation for further studies on the biological function of *ChGATA*.

KEYWORDS

Cerasus humilis, GATA, gene expression, gene function, abiotic stress

1 Introduction

Transcription factors (TF), often known as trans activators, are proteins capable of engaging with cis-acting components within a gene's promoter. Transcriptional control in eukaryotes is frequently linked to cis-acting components, including enhancers, promoters, and silencers (Riechmann et al., 2000; Fan et al., 2021). TFs are crucial in the development of plants, hormone signaling routes, and disease management. There are several well-known families of transcription factors in plants, including WRKY, MYB, DREB, bZIP, and GATA. GATA transcription factor possesses the type IV zinc finger motifs, which are composed of one or two highly conserved zinc-finger domains (C-X₂-CX₁₇₋₂₀-C-X₂-C; Type IVb) and DNA-binding domains (Lowry et al., 2000). The C-X₁₇₋₂₀ domain of the type IV zinc finger structure differs among different eukaryotes. In fungi, most GATA transcription factors' known type IV zinc finger structure is C-X₁₇ (Teakle et al., 2002; Behringer and Schwechheimer, 2015). Typically, the GATA factors in animals contain two conserved zinc finger areas, C-X₂-C-X₁₇-C-X₂-C. However, solely the C-terminal fingers play a role in DNA attachment, which is crucial for cellular division, growth, and development. Most plant GATA factors are limited to a single C-X₂-C-X₁₈-C-X₂-C or CX₂-C-X₂₀-C-X₂-C zinc finger domain. The *Arabidopsis thaliana* GATA family genes are generally divided into subfamilies I, II, III, and IV based on phylogenetic relationships, DNA domains, and intron-exon structures (Reyes et al., 2004; Yuan et al., 2018; Zhang et al., 2018, 2019, 2019).

A multitude of studies has identified a link between GATA TFs and phenomena such as plant development, flowering, chlorophyll generation, greening, and ageing. Furthermore, transcription factors linked to GNC (GATA21) and GNL (GATA22), facilitated by Phytochrome-interacting factors (PIF), are instrumental in regulating seed sprouting (Naito et al., 2007). Elevated levels of GNC and GNL may slow down the sprouting of *Arabidopsis thaliana* seeds, whereas deactivating GNC and GNL may enhance their germination, indicating a detrimental regulatory function of GNC and GNL in seed sprouting. The interaction of GNC/GNL with SOC1, a MADS-box transcription factor that inhibits Constans1 overexpression, affects the timing of flowering (Richter et al., 2013a; Zubo et al., 2018). Additionally, GNC and GNL markedly impede gibberellin signaling through regulation by DELLA and PIF regulators (Naito et al., 2007; Richter et al., 2010). GATA TFs are crucial in plants' reactions to non-living stressors. As an example, under saline stress conditions, the *Oryza sativa* OsGATA8 overexpression variants showed enhanced light efficiency and biomass relative to both standard and altered plants (Nutan et al., 2020). A significant decrease in the levels of GATA44 and GATA58 was observed in *Glycine max* seedlings under nitrogen scarcity (Zhang et al., 2015).

The little perennial deciduous shrub *Cerasus humilis* (*C. humilis*) is a member of the Rosaceae family. *C. humilis* fruit, rich in minerals advantageous for human health, has long been used in traditional Chinese medicine (Song et al., 2011; Yin et al., 2012). *C. humilis* is now recognized as a versatile fruit tree, offering extensive potential for various uses. Land is vital for the productivity and sustainable

development of human societies. However, the escalation in population and significant greenhouse gas emissions have escalated global temperatures and climate change, exacerbating soil salinization and hindering plant growth and development (Mahajan and Tuteja, 2005; Fedoroff et al., 2010). Soil salinization is a major environmental factor that adversely affects agricultural output and the stability of food sources (Munns et al., 2008). As a reaction to shifts in the environment, plants have developed intricate communication routes, typically involving receptors, secondary signals, plant hormones, and signal transduction mechanisms (Liu, 2014). However, there are few studies on the mechanism of salt-alkali tolerance of *C. humilis*. Consequently, this research employed bioinformatics techniques to pinpoint GATA gene family members in *C. humilis*, examining their physical and chemical traits, gene architecture, phylogenetic links, promoter cis-acting components, and expression trends in alkali stress. In addition, the differentially expressed gene *ChGATA14* was screened based on the previous transcriptome sequencing results, and the transcriptional activation of *ChWRKY29* was determined by dual luciferase assay. This study provided the theoretical basis and technical support for genetic improvement and alkali-tolerant variety breeding of *C. humilis*, accelerated the cultivation of excellent varieties that adapt to different saline-alkali environments, and promoted the development of *C. humilis* industry in saline-alkali areas.

2 Materials and methods

2.1 Genome-wide identification of GATA gene family in *C. humilis*

C. humilis genome and gene annotation file downloaded from the national genome database (<https://ngdc.cncb.ac.cn/databases/>) (Zhao et al., 2022). From the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>), download *Arabidopsis thaliana*, *Malus pumila*, *Vitis vinifera*, and the *Oryza sativa* genome file. By employing the TBtools BLAST for analyzing homologous comparisons, 30 *Arabidopsis thaliana* GATA protein sequence members were chosen from the TAIR10 database to pinpoint the *C. humilis* GATA gene families. (score ≥ 100 , $e \leq 1e-10$) (Altschul et al., 1997). HM-MER3.0 was used to search for the unique proteins containing PF00032 protein structure, and the redundant sequences between HMMsearch results and BLASTP results were removed (Finn et al., 2011). Within the GATA gene family, preservation of only the sequence with the longest amino acid occurs when a single gene generates multiple distinct protein sequences. Following this, all protein sequences are gathered using the NCBI-CDD and SMART online tools, which artificially eliminate incorrect sequences. Finally, 20 members of the ChGATA gene family were obtained, and they were named according to their order of arrangement on the *C. humilis* genome chromosome. (Table 1). The ExPASy-ProtParam tool (<http://web.expasy.org/protparam/>) is employed to ascertain the isoelectric point (pI) and molecular weight (kDa) of the ChGATA protein. Subsequently, the WoLF PSORT online tool (<https://wolfpsort.hgc.jp/>) was utilized to forecast its subcellular positioning; the biotype is plant.

TABLE 1 List of the *ChGATAs* identified in the study.

Gene ID	Gene name	PI	MW	Subcellular location
<i>Chumilis00510.1</i>	<i>ChGATA1</i>	4.83	38401.46	nucleus
<i>Chumilis00511.1</i>	<i>ChGATA2</i>	5.61	31945.13	nucleus
<i>Chumilis01338.1</i>	<i>ChGATA3</i>	6.30	42288.63	nucleus
<i>Chumilis01670.1</i>	<i>ChGATA4</i>	9.08	37271.84	nucleus
<i>Chumilis04435.1</i>	<i>ChGATA5</i>	9.69	14912.09	nucleus
<i>Chumilis05700.1</i>	<i>ChGATA6</i>	9.63	20001.48	nucleus
<i>Chumilis06509.1</i>	<i>ChGATA7</i>	7.52	34515.57	nucleus
<i>Chumilis08509.1</i>	<i>ChGATA8</i>	9.50	19338.92	nucleus
<i>Chumilis10293.1</i>	<i>ChGATA9</i>	5.75	28115.91	nucleus
<i>Chumilis11028.1</i>	<i>ChGATA10</i>	6.01	37584.15	nucleus
<i>Chumilis14463.1</i>	<i>ChGATA11</i>	8.37	31533.64	chloroplast
<i>Chumilis14552.1</i>	<i>ChGATA12</i>	6.18	43512.55	chloroplast
<i>Chumilis15041.1</i>	<i>ChGATA13</i>	9.38	36454.03	nucleus
<i>Chumilis18605.1</i>	<i>ChGATA14</i>	6.05	31632.27	nucleus
<i>Chumilis19001.1</i>	<i>ChGATA15</i>	6.05	35334.07	nucleus
<i>Chumilis22132.2</i>	<i>ChGATA16</i>	5.39	31402.99	nucleus
<i>Chumilis22133.1</i>	<i>ChGATA17</i>	5.02	40285.85	nucleus
<i>Chumilis22257.1</i>	<i>ChGATA18</i>	5.49	35446.53	nucleus
<i>Chumilis24079.1</i>	<i>ChGATA19</i>	9.64	14285.56	nucleus
<i>Chumilis25938.1</i>	<i>ChGATA20</i>	6.53	60237.77	nucleus

2.2 Identifying genes in chromosomes and examining the tertiary structure of GATA family gene proteins in *C. humilis*

Utilizing the gene location visualization from the GTF/GFF feature of TBtools software, chromosome localization was executed, referencing the genome annotation file of *C. humilis*. The chromosome Gene Density was obtained using the Gene Density Profile to populate the chromosome. Employ the web-based utilities SWISS - MODEL to predict the tertiary structure of proteins using GATA protein sequences. The model with the GMQE value closest to 1 was used for modelling.

2.3 Aligning multiple sequences and conducting phylogenetic studies

Multiple amino acid sequence alignments were performed in MEGA11 using the default ClustalW parameters to eliminate sequences of different transcripts (Thompson et al., 2002). The neighborhood joining (NJ) method was used to construct the unrooted phylogenetic tree, and the parameter Settings included: The evolutionary distance was p-distance, which was based on the number of amino acid differences per site, the missing data method

was Partial deletion, the cut off was 50%, the Bootstrap was set to 1000, and the other parameters were the default parameter (Tamura et al., 2021). The evolution tree is beautified by the online software Evolvew.

2.4 Gene collinearity and evolutionary examination

The *ChGATA* location was derived from the *C. humilis* annotation GFF file, and the interspecific collinear relationship of *ChGATA* was plotted using Circos of TBtools. MCScanX facilitated the examination of *GATA* collinearity in *Arabidopsis thaliana*, *Malus pumila*, *Vitis vinifera*, *Oryza sativa*, and *C. humilis*. These findings were depicted using TBtools (v2.0697) in the Dual Synteny Plot (Jiang et al., 2023).

2.5 Investigating the genetic structure and patterns of proteins

The intron-exon configuration of the *ChGATA* in *C. humilis* was obtained from its genomic GFF annotation file. Utilizing the online MEME tool, we identified the *GATA* protein’s conserved

motif in *C. humilis*. The optimized parameters were: locus distribution: 0 or 1 for each sequence; motif width: 6-50; Number of patterns:10; Default values for other parameters (Bailey et al., 2006). TBtools software was utilized to display the exon/intron configuration and preserved motifs of the *C. humilis* GATA protein (Chen et al., 2020).

2.6 Analysis of cis-acting element

The promoter region of each gene, situated at 1500 base pairs before the transcription initiation site, was isolated from the *C. humilis* genomic sequence using TBtools software (Chen et al., 2020). Subsequently, the PlantCARE software facilitated the examination of the cis-acting segments in the promoter area (Lescot et al., 2002). Finally, the results were visualized using the Simple BioSequence Viewer of TBtools software (Chen et al., 2020).

2.7 RNA extraction and quantitative RT-PCR analysis

Two-year-old healthy *C. humilis* seedlings (variety: Red Harpo 1) were used as experimental materials. The experimental materials were divided into control and alkali-treated groups for 4 and 8 days with three biological replicates per group. The control group was irrigated with distilled water, and the experimental group was irrigated with 200 mM sodium bicarbonate solution once a day. Leaves were collected after stress and then immediately frozen in liquid nitrogen and stored at minus 80°C. Control and treated 8-d seedlings were used for transcriptome, and three biological replicates were performed for each stage. Total RNA was isolated from these specimens utilizing FreeZol Reagent R711 in compliance with the prescribed manufacturer's instructions. In essence, 50mg of the ground sample was mixed with 500 mL of FreeZol reagent for cracking. After centrifuging the lysate, the supernatant was then gathered. Subsequently, the dilution buffer was applied to the supernatant, and the mixture was precipitated using isopropyl alcohol. Post-centrifugation, the supernatant was disposed of, rinsed with 75% ethanol, and dissolved in double distilled water (ddH₂O) without RNase. Evaluation of the isolated RNA samples' integrity and concentration was conducted through 1% agarose gel electrophoresis and the use of an ultramicro nucleic acid concentration detector.

Primer 3 website (<https://primer3.org/>) was used for primer design (Supplementary Table 1). Standardized Actin primers are used as internal parameters (Livak and Schmittgen, 2001). Every test hole holds a reaction mixture of 20 µL at 95°C for one minute, succeeded by 40 repetitions at 95°C for ten seconds and 60°C for thirty seconds, forming a thermal cycle graph. After the last PCR cycle, the temperature was raised from 55°C to 99°C at a rate of 0.5°C/s to generate the melting curve of the sample. Gene expression was measured by the $2^{-\Delta\Delta C_t}$ method (Lynch and Conery, 2000). Each reaction was repeated three times, and the results were expressed as the mean of three independent biological replicates.

2.8 Analysis of gene expression

The standardized transcriptome data were used to cluster *C. humilis* leaves under different alkali treatment times. Heat maps were drawn using TBtools.

2.9 Protein interaction prediction

The GATA protein interaction network was constructed by STRING 11.0 (<https://string-db.org/cgi/input.pl>). Using *Arabidopsis thaliana* as a template, the network edge was set to confidence, the parameter was set to a medium confidence parameter (0.400), and no more than 10 interaction factors were displayed.

2.10 Transcriptional activation assay of *ChGATA14* by *ChWRKY29*

MdWRKY115 is an important regulator of osmotic stress and drought stress tolerance in apples. *MdWRKY115* can bind to the W-box elements of the promoters of downstream target genes to regulate their expression and participate in response to osmotic stress and drought stress. In this study, *ChWRKY* and *MdWRKY115* were used to construct a phylogenetic tree, and *ChWRKY29* was screened out. The *ChGATA14* gene, which was significantly up-regulated under alkaline stress, was screened by transcriptome data for gene function verification.

The full-length coding sequence (CDS) of *ChWRKY29* was inserted into the pGreenII 62-SK vector as an effector, while the w-box promoter fragment of *ChGATA14* was cloned into the pGreenII0800-LUC vector to generate a reporter gene. *Aspergillus tumefaciens* strain EHA105 was used to transfer all construct vectors. The effector and reporter were transiently co-infiltrated into 4-week-old tobacco leaves. *A. tumefaciens* cells were cultured in suspension buffer (10 mM MgCl₂, 10 mM MES, 150 mM acetosyringone, pH 5.6) until OD 600 was 0.6. After 2–3 d of injection at 23°C, tobacco leaves were sprayed with water as a control. Fluorescence in tobacco leaves was observed and photographed using an automatic chemiluminescence imager (Tanon 5200, Shanghai, China).

3 Results

3.1 Identification of the *ChGATAs*

Using the *Arabidopsis thaliana* protein sequence as a reference, we screened *ChGATA* using the BLASTP comparison method. Following HMMER, CD-Search, and SMART analysis, a total of 20 *ChGATA* genes were pinpointed and designated as *ChGATA1*–*ChGATA20*, reflecting the chromosomal arrangement of *ChGATA*. The complete coding sequence (CDSs) of the *ChGATA* gene spans 403–1657 base pairs. The physicochemical characteristics of 20 *ChGATA* proteins were then examined in more detail.

The lengths of sequences range from 131 (ChGATA19) to 542 (ChGATA20) amino acids, and their molecular weights extend from 14.286 (ChGATA19) to 60.238 (ChGATA20) kDa, indicating considerable structural differences, potentially associated with genome size and species evolution. The isoelectric point falls within the range of 4.83 (ChGATA1) to 9.69 (ChGATA5). Among these, 12 (60%) ChGATAs exhibited a pI below 7, indicating a notable prevalence of acidic amino acids in most ChGATAs. Table 1 encapsulates the essential elements of the ChGATA family. As expected, a majority of ChGATA proteins (18 out of 20) are located in the nucleus, suggesting a major function of ChGATA in regulating nuclear transcription. Two ChGATA proteins are located in chloroplasts. Considering that their functions may be involved in the physiological and biochemical regulation of chlorophyll synthesis.

3.2 Chromosomal localization of genes and tertiary structure analysis of GATA family gene proteins in *C. humilis*

To delve deeper into the chromosomal distribution of the ChGATA gene, We mapped the locations of each member of the ChGATA gene family in the *C. humilis* genome, discovering 20 ChGATAs situated on the respective Chr1-Chr8 chromosomes (Figure 1). Except for Ch-chr5, all other chromosomes contain ChGATA. Ch-chr8 contains only one gene (ChGATA20), and Ch-chr3 and Ch-chr6 both contain two genes (ChGATA9–10 and ChGATA14–15). Ch-chr1 contains the most ChGATAs, up to five (ChGATA1–5). The distribution of genes on *C. humilis* chromosomes varies, possibly indicating disparities in their size

and structure. The ChGATA gene family consists of two tandem repeat genes (ChGATA1 and ChGATA2, ChGATA16 and ChGATA17), which are located on Ch-chr1 and Ch-chr7 respectively (Figure 1), both from subfamily III. Fragment replication may be the main driving force for the expansion of members of subfamily III. The three-dimensional structure analysis of GATA protein in *C. humilis* showed that there is a significantly larger protein structure among different members (Figure 2).

3.3 Evolutionary study of the ChGATA

To construct the phylogenetic tree, the amino acid sequences of 20 identified ChGATA proteins, 30 AtGATA proteins, and 28 OsGATA proteins were utilized. (Figure 3). Based on AtGATAs' categorization, there were 20 ChGATA proteins segmented into four distinct subfamilies (I, II, III, and IV). Among them, the most ChGATA members are Subgroup I and Subgroup II, both containing 7 members, followed by Subgroup III, containing 5 members, and finally Subgroup IV. Evolutionary analysis revealed that several ChGATA proteins were closely grouped with AtGATA, indicating that their homology was close and their gene functions might be similar.

3.4 Collinearity and evolutionary analysis of genes

Tandem and segmental duplication events can expand the number of gene family members, which is an important driving force for gene family amplification and the main driving force for species evolution

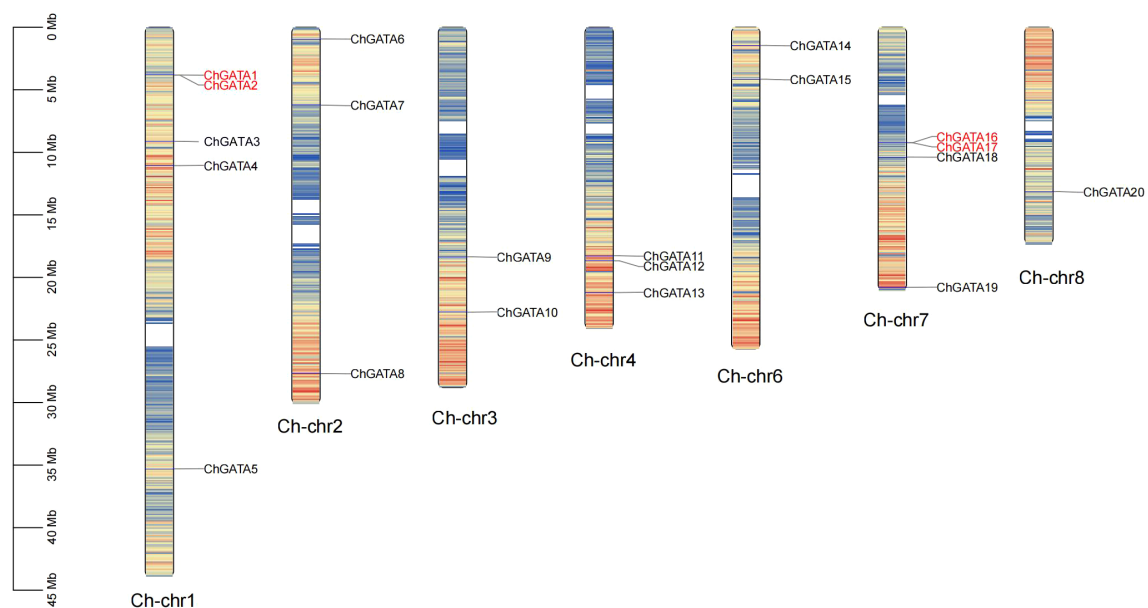


FIGURE 1

Chromosomal positioning of ChGATAs. Each of the 20 ChGATAs is displayed on the chromosomes, denoted by their names. The left scale illustrates the *C. humilis* chromosomes' length. Numbers of chromosomes are shown at each bar's base. The red lines on the chromosome indicate a dense distribution of genes. The genes marked in red font are tandemly repeated genes.

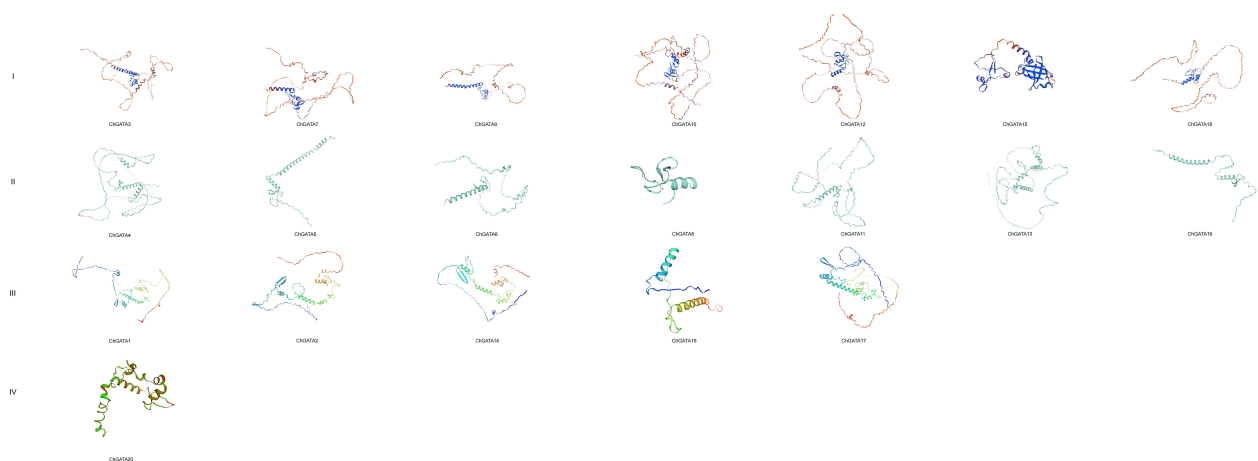


FIGURE 2
Predicted tertiary structure of ChGATA family member proteins.

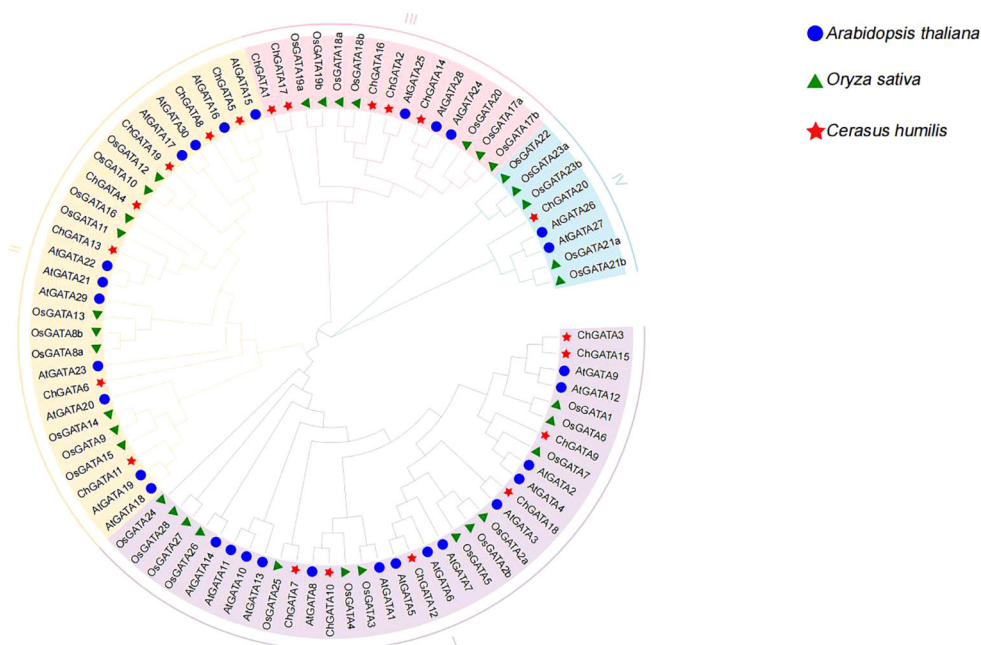
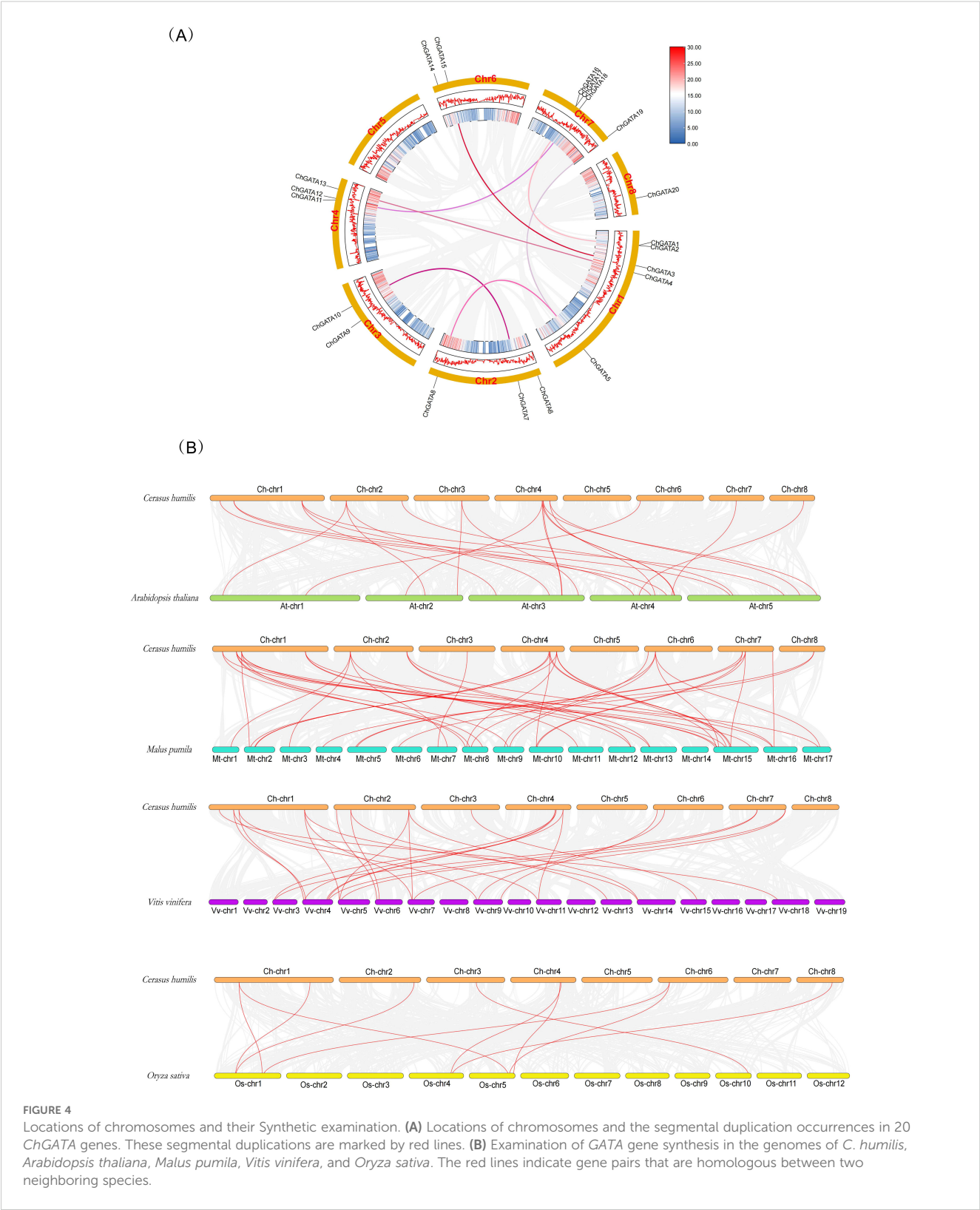


FIGURE 3
The phylogenetic analysis of GATA in *Arabidopsis thaliana*, *Oryza sativa*, and *C. humilis* by the NJ tree includes 30 GATAs from *Arabidopsis thaliana*, 28 from *Oryza sativa*, 20 from *C. humilis*, constructed using entire protein sequences.

(Altschul et al., 1997) (Figure 4). The MCSanX method was used in this study. It was found that there were 7 fragment repeat gene pairs on the repeats of *C. humilis* genome (*ChGATA1* and *ChGATA16*, *ChGATA3* and *ChGATA15*, *ChGATA4* and *ChGATA13*, *ChGATA5* and *ChGATA19*, *ChGATA5* and *ChGATA8*, *ChGATA7* and *ChGATA10*, *ChGATA12* and *ChGATA18*). Among the seven gene pairs exhibiting fragment repetition, three *ChGATA* gene pairs are part of subfamily I, another three from subfamily II, and one from subfamily III. Subfamily I and subfamily II may have mainly expanded through fragment replication during the evolutionary process.

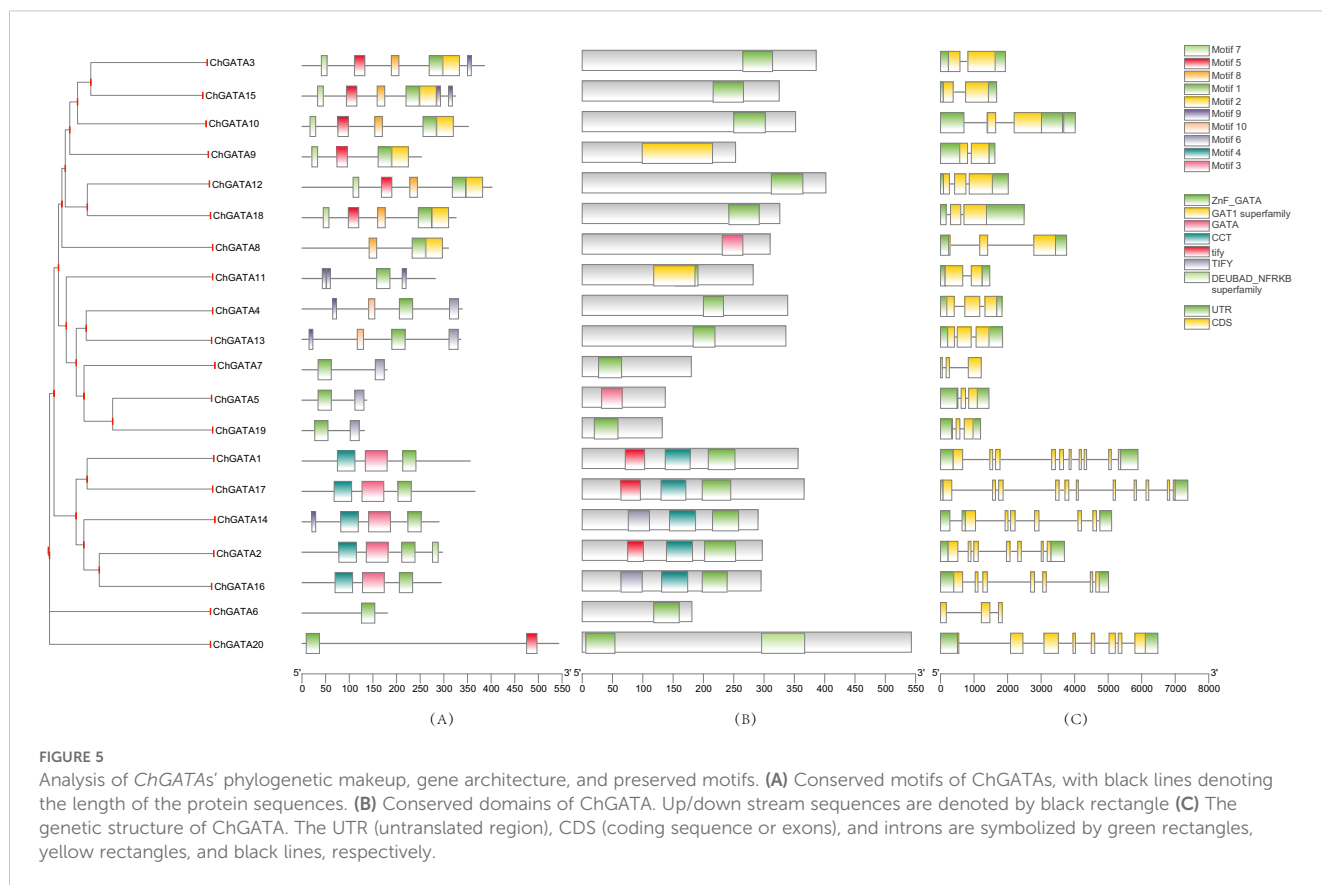
To gain a deeper insight into the genesis and development of GATA family genes, we conducted a collinear examination of various species, encompassing monocotyledonous species such as *Oryza sativa* and dicotyledonous ones like *Arabidopsis thaliana*, *Malus pumila*, and *Vitis vinifera*. A total of 23,43,26 and 10 collinear GATAs were identified. There was a higher count of collinear genes in dicotyledonous plants compared to monocotyledonous plants, with *Malus pumila* (rose family) having the most collinear gene pairs, perhaps it is because both of them originated from the Rosaceae family. These results may provide valuable reference for gene function in Rose-family cash crops.



3.5 Analysis of gene structure and protein motif

For the purpose of examining *C. humilis* *GATA*'s evolutionary diversity, the preserved patterns of 20 *ChGATA* proteins were

scrutinized via the MEME website, and a composite diagram of phylogenetic trees, motif patterns and gene structures of 20 kinds of *ChGATAs* was constructed using TBtools (Figure 5). The *ChGATA* has between one and ten introns; the majority of genes have two introns, but the *ChGATA1* and *ChGATA17*



genes have several, suggesting that the *ChGATA*'s introns may have been lost during evolution (Figure 5). Furthermore, genes grouped into subclass may display distinct structural characteristics. For instance, *ChGATA3* in subclass I has two CDS and two UTR fragments, while *ChGATA6* and *ChGATA8* in subclass II do not have UTR regions.

In the study, a total of 10 conserved motifs were identified, together referred to as Motif1-10. Most members clustered in the same subclass have similar motif structures, suggesting that members in the same subclass may have similar functions. A basic portion of the GATA domain known as motif 1 is present in every member of the *ChGATA* gene family. Members of Class II lack Motif 2, in contrast to Class I members who possess Motif 1, Motif 2, and Motif 8. The variance in the quantity and nature of preserved motifs in *C. humilis* GATA proteins mirrors their structural variety and suggests distinct biological roles.

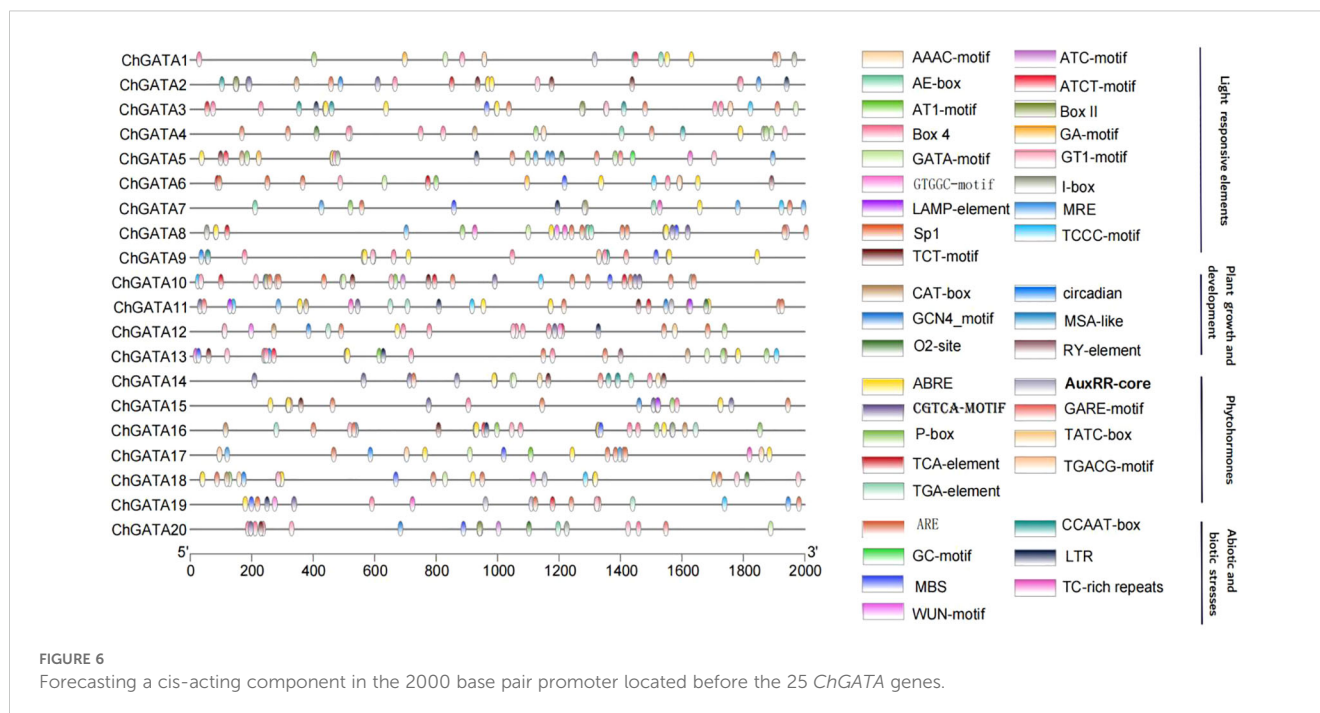
3.6 Analysis of cis-acting components

Non-coding DNA segments known as cis-acting elements, located in gene promoter areas, play a crucial role in controlling the transcription of related genes. In this instance, a specific area situated 2000 base pairs before the *ChGATA* gene's transcription initiation point was chosen as the hypothetical promoter region (Yao et al., 2022). Subsequently, the *ChGATA*'s promoter sequence

was isolated and entered into the PlantCARE database to identify cis-acting elements. Findings indicated that out of the 20 *ChGATAs* promoter sequences, the highest count of photoresponsive elements was 263, the most prevalent among all promoter sequences. This was succeeded by 75 abscisic acid response elements, prevalent in most promoter sequences (Figure 6). Within the domain of elements linked to environmental stress, the primary components of *ChGATAs* are the hypoxia-inducible element (GC-motif), low-temperature response element (LTR), and drought-inducible element (MBS). The majority of *ChGATAs* contain two hormone-related elements: ABRE, TGACG-motif, and CGTCA-motif. The majority of these genes include photoreaction-related Box 4 (ATT AAT) and Sp1 (GGG CGG), which are crucial for regulating photoresponse.

3.7 Analysis of *ChGATA* gene expression under alkali stress

An investigation into the gene expression in *C. humilis* leaves was conducted to understand the *ChGATAs*' function under alkali stress (Figure 7). Findings indicated alterations in *GATA* gene family members' expression in *C. humilis* leaves post-alkali treatment versus the control group, with distinct gene response processes. For instance, genes like *ChGATA2* and *ChGATA12* were elevated due to alkali stress, and their levels rose under alkali stress,



hinting at their potential role in resistance to alkali stress. In contrast, alkali stress led to a reduced expression of *ChGATA1*, *ChGATA4*, *ChGATA7*, and *ChGATA16*.

3.8 RNA extraction and quantitative RT-PCR analysis

To further understand the functions of *ChGATAs*, we selected 10 representative *ChGATAs* based on the heat map. The analysis was conducted using qRT-PCR technology, and the results are shown in the Figures 8, 9. Compared with the blank control group, the expression levels of *ChGATA2*, *ChGATA8*, *ChGATA10*, *ChGATA12*, *ChGATA18*, *ChGATA19*, and *ChGATA20* were significantly decreased at 4 days of alkali treatment. The expression levels of *ChGATA1*, *ChGATA14*, *ChGATA19* and *ChGATA20* significantly increased after 8 days of alkali treatment. Most *ChGATAs* show a trend of first decreasing and then increasing.

3.9 Protein interaction prediction analysis

To explore potential regulatory interactions between *ChGATA* proteins, a protein interaction network was constructed using the STRING database (Figure 10). The results showed that the 14 *ChGATA* protein interaction, among them, the *ChGATA1*, *ChGATA14*, *ChGATA17* has more protein interaction, and 10 other protein interactions. In contrast, *ChGATA8* and *ChGATA9* each have only one interacting protein. It is worth noting that *ChGATA1* *ChGATA14*, *ChGATA17*, *ChGATA2*, *ChGATA16* interaction effect is stronger, they all come from the family III,

prediction of *ChGATA* protein interaction networks could help to understand the function of the connection between these proteins.

3.10 Transcriptional activation assay of *ChGATA14* by *ChWRKY29*

In the present study, we performed a dual luciferase reporter (DLR) assay in tobacco leaves (*Nicotiana benthamiana*) (Figure 11). The relative intensity of the LUC signal was significantly increased after cotransformation of *ChWRKY29* with pGreenII0800-LUC-*ChGATA14*-w-box compared to the vector control, indicating that *ChWRKY29* activated *ChGATA14* expression. Thus, DLR analysis confirmed that *ChGATA14* may be transcriptionally activated by *ChWRKY29* to resist alkali stress.

4 Discussion

C. humilis is nutrient-rich and is known as the third generation fruit with high nutrition and high health care value together with sea buckthorn and blueberry, which has attracted more and more attention from researchers. Transcription factors in plants, including *WRKY* (Liu et al., 2019b), *MYB* (Li et al., 2019), *bHLH* (Geng and Liu, 2018), and *Bzip* (Zhang et al., 2016), are crucial in controlling genes involved in diverse biological functions related to plant growth, stress reactions, and hormone communication routes (Franco-Zorrilla et al., 2014). However, in recent years, various abiotic stresses have affected the growth and development of *C. humilis*. These environmental challenges not only hinder the physiological processes necessary for the survival of *C. humilis*, but also affect the economic vitality and conservation value of

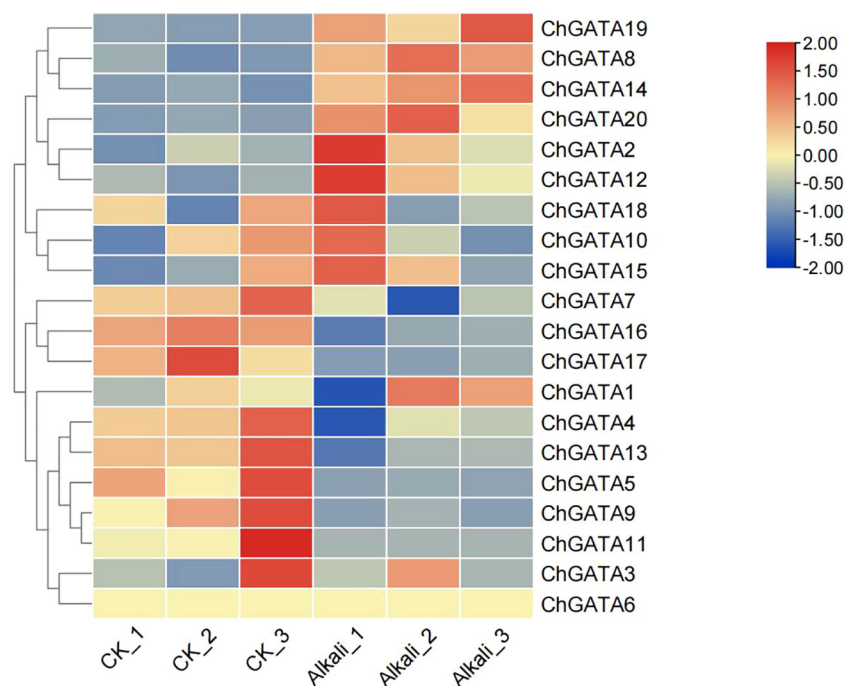


FIGURE 7

Pattern of *ChGATAs* expression under *C. humilis*' alkaline stress. This pattern, derived from FPKM plus 1 post log2 transformation, was examined using heatmap hierarchical clustering.

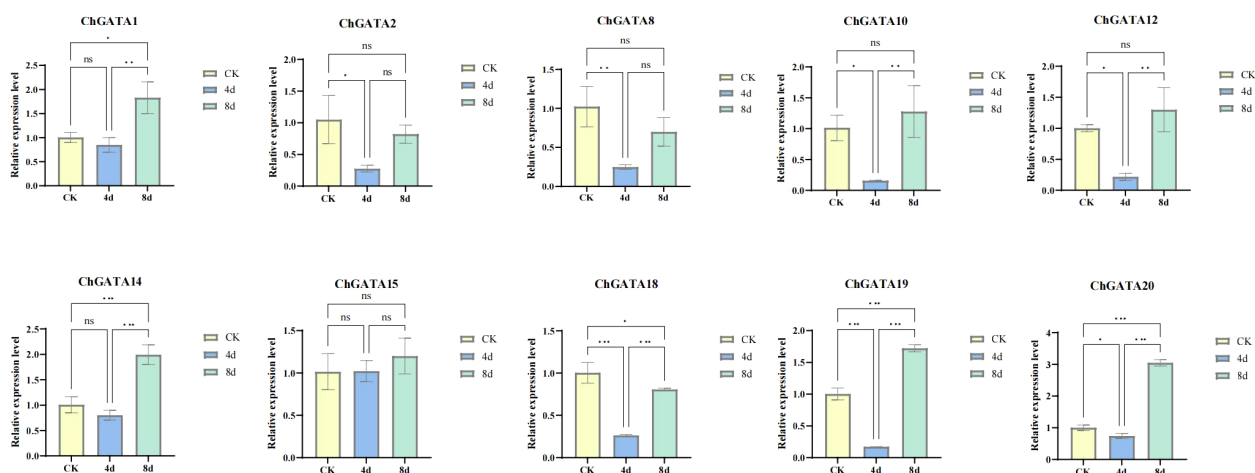


FIGURE 8

The qRT-PCR examination of *ChGATAs* under alkali stress reveals a statistically notable variance from the control group ($p^* < 0.05$; $p^{**} < 0.01$; $p^{***} < 0.001$). ns indicates no significance.

C. humilis. The GATA gene family plays an important role in improving plant stress tolerance and breeding, but to date, no study has been conducted on the GATA genes in *C. humilis* (Reyes et al., 2004). In this study, a total of 20 *ChGATA* family members were identified in the *C. humilis* genome, a number close to that found in four other Rosaceae species, including peach, indicating that *GATA* genes are relatively similar in number in the same genus (Manzoor et al., 2021). The striking differences in protein length, structure,

molecular weight, and theoretical isoelectric points among *ChGATA* members suggest that *GATA* would generate functional divergence to participate in different physiological processes during evolution. For example, *AtGATAs* with consensus sequences of different lengths separately regulate various physiological responses such as hypocotyl growth, root development, flowering, lateral root formation, root branching, and cell differentiation. Subcellular localization predictions suggest that *ChGATAs* function primarily

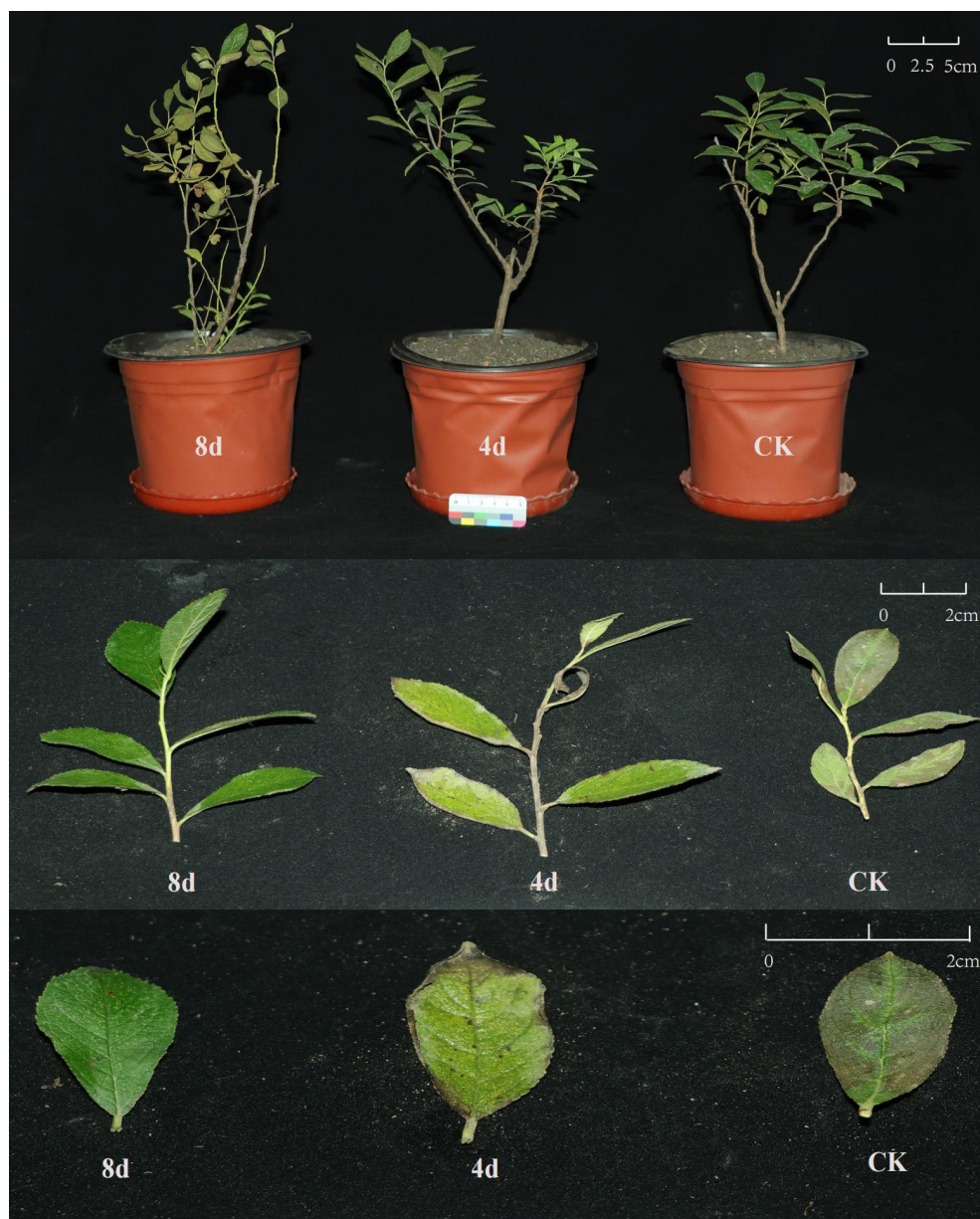


FIGURE 9
Plant diagram of *C. humilis* under alkali treatment.

in the nucleus, consistent with findings in *Cucumis melo* L, *Glycine max*, *Cucumis sativus*, and *Solanum tuberosum* L (Zheng et al., 2024).

Based on the GATA family grouping in the model plant *Arabidopsis thaliana*, the 20 *ChGATAs* were grouped into four subfamilies: subfamilies I (7), II (7), III (5), and IV (1). This is consistent with most GATA family studies, suggesting that the GATA family is relatively stable during evolution (Liu et al., 2020; Zhang et al., 2023). Analysis of the conserved motifs of the 20 *ChGATA* proteins identified motif 1 in all 20 *ChGATA* proteins, suggesting that this conserved motif is required for *ChGATA* protein function and is the zn-finger GATA motif of the family. Motif 3 corresponds to the CCT motif, and motif 7 to the TIFY

domain. CCT and TIFY domains play important roles in flowering and root development in *Arabidopsis thaliana*, suggesting that members of groups I and III may have similar regulatory functions in *C. humilis* (Zheng et al., 2024). Gene structural information provides clues to the evolution of gene family members. Consistent with most plant GATA family members, groups I and II have 2–3 exons, whereas groups III and IV have 7–11 exons. This further indicates the relative stability of GATA family members in plants. Conserved motifs/gene structures are distributed similarly across subfamily members, yet significant variances among subfamilies suggest precise subfamily classification, aligning with prior study findings (Manfield et al., 2007; Yu et al., 2019). The majority of GATA TFS proteins in plants,

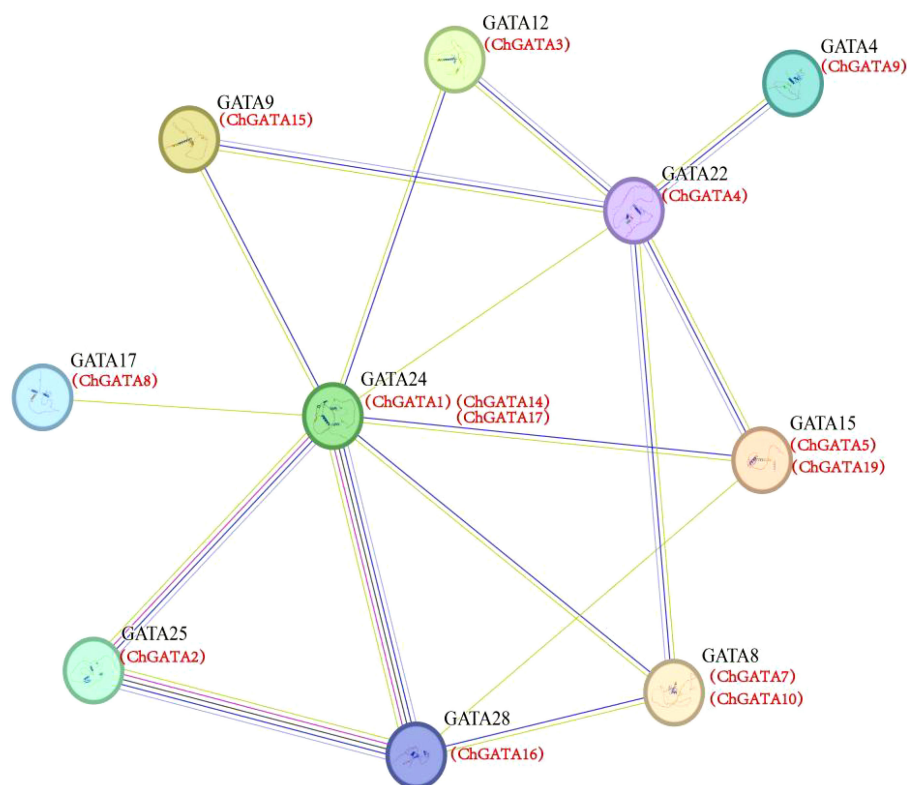


FIGURE 10
Protein interaction prediction analysis.

like those in *Arabidopsis thaliana*, *Oryza sativa*, *Triticum aestivum*, and *Vitis vinifera*, possess a solitary zinc finger domain (Reyes et al., 2004; Zhang et al., 2018; Feng et al., 2022). Nonetheless, subfamily III members include extra domains like CCT and TIFY. Earlier research has noted this and demonstrated its crucial function in various biological processes, such as regulating embryonic and flower growth, stress resistance, and diverse plant hormone signaling pathways (Richter et al., 2013b; Behringer and Schwechheimer, 2015; Gupta et al., 2017).

Gene replication fosters the creation of novel genes and their roles, significantly influencing evolutionary processes. It manifests in three primary forms: fragment replication, tandem replication, and translocation events, with segment replication and tandem replication being prevalent in the expansion of plant gene families (Zhu et al., 2014; Wang et al., 2019). Two tandem repeat gene pairs (*ChGATA1*, *ChGATA2*, *ChGATA16* and *ChGATA17*) were identified among the 20 *ChGATAs*. Tandem repeat genes may replicate during evolutionary divergence in plants and maintain their copy number at a high frequency. Analysis of the fragment repeatability of the 20 *ChGATAs* on the chromosome identified seven pairs of duplication events, mostly belonging to subfamilies I and II. This indicates that these genes are linked in their respective subfamilies and that most of the gene structural and functional information is retained after the evolution of whole genome duplication. These results indicate that genes in different subfamilies differ in structure, function, environmental adaptation, mutation, and recombination.

In addition, 20 *ChGATAs* were used to analyze interspecific collinearity with monocot and dicot plants. The results showed significant collinearity family members between calcocarps and dicots between GATA family members. Only a few GATA members are collinear with monocots. Due to chromosomal localization, fusion, and selective gene loss, specific GATA genes may fail to localize to any synteny region, making chromosomal synteny difficult to identify (Yang et al., 2008).

Acting elements within the promoter region are involved in regulating gene expression by binding to trans-acting factors to regulate the activity of target genes (Liu et al., 2019a). The presence of these elements suggests that *ChGATA* family members are involved in various functions such as growth and development, stress response, and hormonal regulation. The W-box element is a DNA cis-regulatory element that is specifically recognized and bound by the WRKY family of transcription factors. In terms of abiotic stresses, such as low temperature, drought, and salinity, WRKY transcription factors also bind to W-box elements to regulate the expression of downstream stress resistance genes and help plants resist stress. For example, the *ShWRKY33* promoter in wild hairy tomato contains W-box elements, which can induce downstream gene expression and improve the low temperature resistance of tomato under low temperature stress (Guo et al., 2022). In this study, based on *MdWRKY115*, which has been reported to play an important regulatory role in stress in apple, and combined with the transcriptome differentially expressed genes, *ChWRKY29*

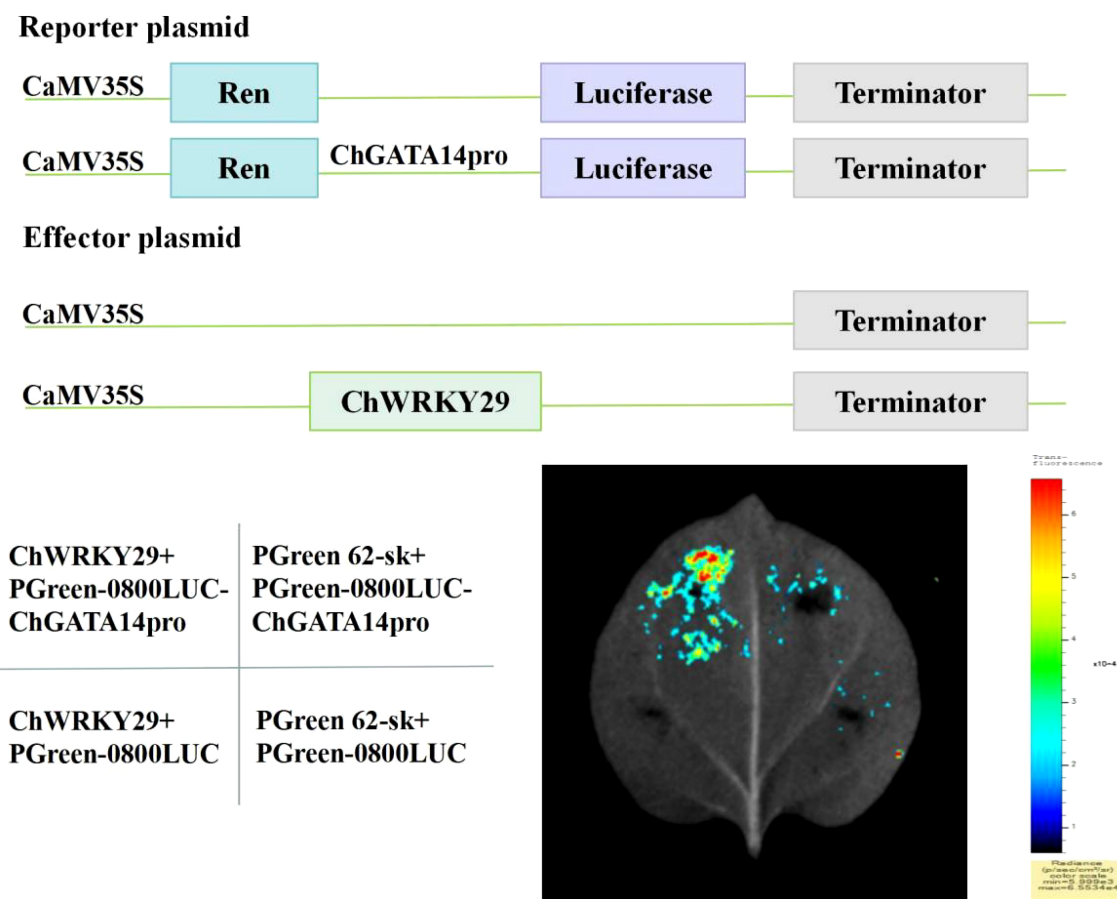


FIGURE 11

Dual luciferase imaging experiments, including plasmid construction, injection site, and imaging results.

and *ChGATA14* were selected to respond to alkali stress in *C. humilis*. *ChWRKY29* can activate *ChGATA14* expression to resist alkali stress. In addition, transcription factors usually regulate various reaction processes through protein-protein interactions. This study predicted potential interactions among 20 *ChGATA* members. The results showed that *ChGATA14* had strong interactions with *ChGATA2* and *ChGATA16*, and *ChGATA14* was a differentially expressed gene that was significantly up-regulated under alkali stress. It may play a regulatory role by forming protein complexes.

Previous studies have shown that the *GATA* gene has a distinct function in plant development. Generally speaking, gene expression patterns can offer crucial hints for predicting gene function (Zhang et al., 2019). In this study, the expression of 10 *ChGATAs* was analyzed based on RT-qPCR under alkali stress. The results showed that the expression profiles of most *ChGATAs* were first decreased and then increased with the treatment time, indicating that lye could affect the expression of *ChGATA* gene, and some upstream regulatory genes might be activated at the early stage, which inhibited the expression of *GATA* gene. As the stress continues, the expression of these upstream genes may change, or some other compensatory mechanism may be activated, which releases the inhibition of the *GATA* gene and activates the regulatory factors related to *GATA* gene

expression so that *GATA* gene expression is increased to participate in the subsequent stress response process and regulate the expression of downstream genes related to stress resistance (Gupta et al., 2017). These results can only be used as a reference, and further studies are needed to understand the specific functions of these genes. In conclusion, this study provides a new idea for the study of stress resistance of *C. humilis* and an important theoretical basis for the selection and improvement of calcaneus fruit in the future.

5 Conclusion

This research methodically pinpointed and examined the *GATA* TFs gene family in *C. humilis*, classifying twenty *ChGATAs* into four distinct categories across seven chromosomes. While class I members had Motifs 1 and 2, class II members lacked Motif 2. Fragment duplication was the main driver of the *GATA* gene family's proliferation in *C. humilis*. Transcriptome heat maps show that, the *ChGATA*'s expression is distinct under alkali stress, with the same *ChGATA* gene subgroup exhibiting varied patterns. qPCR analysis showed that most of the screened *ChGATAs* were first decreased and then increased with time, suggesting that these genes were initially repressed and subsequently turned on a compensatory mechanism.

In addition, the dual luciferase assay and the predicted protein interaction structure showed that *ChWRKY29* could activate the expression of the differentially expressed gene *ChGATA14* in response to alkali stress, and *ChGATA14* was at the core of the protein interaction network and had a strong interaction with *ChGATA2* and *ChGATA16*. These results suggest that *ChGATA14* may play a role by forming protein complexes. However, this gap in the validation of *ChGATA* function needs to be filled in future studies using knockout/overexpression experiments to confirm the role of key *ChGATAs* in alkaline stress response.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author/s.

Author contributions

MP: Data curation, Writing – review & editing. CY: Data curation, Writing – original draft. SJ: Conceptualization, Writing – original draft. JM: Methodology, Writing – original draft. LM: Software, Writing – review & editing. LK: Investigation, Writing – original draft. YL: Formal analysis, Writing – original draft. WM: Supervision, Writing – review & editing. WR: Supervision, Writing – review & editing. XL: Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The work was financially

supported by the National Key Research and Development Project, research and demonstration of collection, screening, and breeding technology of ginseng and other genuine medicinal materials (2021YFD1600901).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1596930/full#supplementary-material>

References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198
- Behringer, C., and Schwechheimer, C. (2015). B-GATA transcription factors - insights into their structure, regulation, and role in plant development. *Front. Plant Sci.* 6, 90. doi: 10.3389/fpls.2015.00090
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Fan, Y., Yan, J., Lai, D., Yang, H., Xue, G., He, A., et al. (2021). Genome-wide identification, expression analysis, and functional study of the GRAS transcription factor family and its response to abiotic stress in sorghum [*Sorghum bicolor* (L.) Moench]. *BMC Genomics* 22, 509. doi: 10.1186/s12864-021-07848-z
- Fedoroff, N. V., Battisti, D. S., Beachy, R. N., Cooper, P. J., Fischhoff, D. A., Hodges, C. N., et al. (2010). Radically rethinking agriculture for the 21st century. *Science* 327, 833–834. doi: 10.1126/science.1186834
- Feng, X., Yu, Q., Zeng, J., He, X., and Liu, W. (2022). Genome-wide identification and characterization of GATA family genes in wheat. *BMC Plant Biol.* 22, 372. doi: 10.1186/s12870-022-03733-3
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. U S A* 111, 2367–2372. doi: 10.1073/pnas.1316278111
- Geng, J., and Liu, J. H. (2018). The transcription factor CsbHLH18 of sweet orange functions in modulation of cold tolerance and homeostasis of reactive oxygen species by regulating the antioxidant gene. *J. Exp. Bot.* 69, 2677–2692. doi: 10.1093/jxb/ery065
- Guo, M., Yang, F., Liu, C., Zou, J., Qi, Z., Fotopoulos, V., et al. (2022). A single-nucleotide polymorphism in WRKY33 promoter is associated with the cold sensitivity in cultivated tomato. *New Phytol.* 236, 989–1005. doi: 10.1111/nph.v236.3
- Gupta, P., Nutan, K. K., Singla-Pareek, S. L., and Pareek, A. (2017). Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice. *Front. Plant Sci.* 8, 1944. doi: 10.3389/fpls.2017.01944
- Jiang, X., Cui, H., Wang, Z., Kang, J., Yang, Q., and Guo, C. (2023). Genome-wide analysis of the LATERAL ORGAN BOUNDARIES domain (LBD) members in alfalfa and the involvement of msLBD48 in nitrogen assimilation. *Int. J. Mol. Sci.* 24, 4644. doi: 10.3390/ijms24054644
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to

tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Li, X., Guo, C., Ahmad, S., Wang, Q., Yu, J., Liu, C., et al. (2019). Systematic analysis of MYB family genes in potato and their multiple roles in development and stress responses. *Biomolecules* 9, 317. doi: 10.3390/biom9080317

Liu, T. (2014). Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods Mol. Biol.* 1150, 81–95. doi: 10.1007/978-1-4939-0512-6_4

Liu, Y., Patra, B., Pattanaik, S., Wang, Y., and Yuan, L. (2019a). GATA and phytochrome interacting factor transcription factors regulate light-induced vindoline biosynthesis in *catharanthus roseus*. *Plant Physiol.* 180, 1336–1350. doi: 10.1104/pp.19.00489

Liu, Y., Yang, T., Lin, Z., Gu, B., Xing, C., Zhao, L., et al. (2019b). A WRKY transcription factor PbrWRKY53 from *Pyrus betulaefolia* is involved in drought tolerance and ABA accumulation. *Plant Biotechnol. J.* 17, 1770–1787. doi: 10.1111/pbi.13099

Liu, H., Zhou, X., Li, Q., Wang, L., and Xing, Y. (2020). CCT domain-containing genes in cereal crops: flowering time and beyond. *Theor. Appl. Genet.* 133, 1385–1396. doi: 10.1007/s00122-020-03554-8

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lowry, J. A., Atchley, W. R., and Atchley, W. R. (2000). Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J. Mol. Evol.* 50, 103–115. doi: 10.1007/s002399910012

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151

Mahajan, S., and Tuteja, N. (2005). Cold, salinity and drought stresses: an overview. *Arch. Biochem. Biophys.* 444, 139–158. doi: 10.1016/j.abb.2005.10.018

Manfield, I. W., Devlin, P. F., Jen, C. H., Westhead, D. R., and Gilmartin, P. M. (2007). Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family. *Plant Physiol.* 143, 941–958. doi: 10.1104/pp.106.090761

Manzoor, M. A., Sabir, I. A., Shah, I. H., Wang, H., Yu, Z., Rasool, F., et al. (2021). Comprehensive comparative analysis of the GATA transcription factors in four rosaceae species and phytohormonal response in Chinese pear (*Pyrus bretschneideri*) fruit. *Int. J. Mol. Sci.* 22, 12492. doi: 10.3390/ijms222212492

Munns, R., Tester, M., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Naito, T., Kiba, T., Koizumi, N., Yamashino, T., and Mizuno, T. (2007). Characterization of a unique GATA family gene that responds to both light and cytokinin in Arabidopsis thaliana. *Biosci. Biotech. Biochem.* 71, 1557–1560. doi: 10.1271/bbb.60692

Nutan, K. K., Singla-Pareek, S. L., and Pareek, A. (2020). The Saltol QTL-localized transcription factor OsGATA8 plays an important role in stress tolerance and seed development in Arabidopsis and rice. *J. Exp. Bot.* 71, 684–698. doi: 10.1093/jxb/erz368

Reyes, J. C., Muro-Pastor, M. I., and Florencio, F. J. (2004). The GATA family of transcription factors in Arabidopsis and rice. *Plant Physiol.* 134, 1718–1732. doi: 10.1104/pp.103.037788

Richter, R., Bastakis, E., and Schwechheimer, C. (2013a). Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in Arabidopsis. *Plant Physiol.* 162, 1992–2004. doi: 10.1104/pp.113.219238

Richter, R., Behringer, C., Müller, I. K., and Schwechheimer, C. (2010). The GATA-type transcription factors GNC and GNL/CGA1 repress gibberellin signaling downstream from DELLA proteins and PHYTOCHROME-INTERACTING FACTORS. *Genes Dev.* 24, 2093–2104. doi: 10.1101/gad.594910

Richter, R., Behringer, C., Zourelidou, M., and Schwechheimer, C. (2013b). Convergence of auxin and gibberellin signaling on the regulation of the GATA transcription factors GNC and GNL in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U S A* 110, 13192–13197. doi: 10.1073/pnas.1304250110

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., et al. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110. doi: 10.1126/science.290.5499.2105

Song, X. S., Shang, Z. W., Yin, Z. P., Ren, J., Sun, M. C., and Ma, X. L. (2011). Mechanism of xanthophyll-cycle-mediated photoprotection in *Cerasus humilis* seedlings under water stress and subsequent recovery. *Photosynthetica* 49, 523–530. doi: 10.1007/s11099-011-0065-4

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Teakle, G. R., Manfield, I. W., Graham, J. F., and Gilmartin, P. M. (2002). Arabidopsis thaliana GATA factors: organisation, expression and DNA-binding characteristics. *Plant Mol. Biol.* 50, 43–57. doi: 10.1023/A:1016062325584

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinf.* Chapter 2: Unit 2.3. doi: 10.1002/0471250953.bi0203s00

Wang, R., Ma, J., Zhang, Q., Wu, C., Zhao, H., Wu, Y., et al. (2019). Genome-wide identification and expression profiling of glutathione transferase gene family under multiple stresses and hormone treatments in wheat (*Triticum aestivum* L.). *BMC Genomics* 20, 986. doi: 10.1186/s12864-019-6374-x

Yang, S., Zhang, X., Yue, J. X., Tian, D., and Chen, J. Q. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* 280, 187–198. doi: 10.1007/s00438-008-0355-0

Yao, X., Zhou, M., Ruan, J., He, A., Ma, C., Wu, W., et al. (2022). Genome-wide identification, evolution, and expression pattern analysis of the GATA gene family in tartary buckwheat (*Fagopyrum tataricum*). *Int. J. Mol. Sci.* 23, 12434. doi: 10.3390/ijms232012434

Yin, Z. P., Shang, Z. W., Wei, C., Ren, J., and Song, X. S. (2012). Foliar sprays of photosynthetic bacteria improve the growth and anti-oxidative capability on chinese dwarf cherry seedlings. *J. Plant Nutr.* 35, 840–853. doi: 10.1080/01904167.2012.663439

Yu, M., Yu, J., Cao, H., Yong, M., and Liu, Y. (2019). Genome-wide identification and analysis of the GATA transcription factor gene family in *Ustilagoidea vires*. *Genome* 62, 807–816. doi: 10.1139/gen-2018-0190

Yuan, Q., Zhang, C., Zhao, T., Yao, M., and Xu, X. (2018). A genome-wide analysis of GATA transcription factor family in tomato and analysis of expression patterns. *Int. J. Agric. Biol.* 20 (6), 1274–1282. doi: 10.17957/IJAB/15.0626

Zhang, Y., Ali, U., Zhang, G., Yu, L., Fang, S., Iqbal, S., et al. (2019). Transcriptome analysis reveals genes commonly responding to multiple a biotic stresses in rapeseed. *Mol. Breed.* 39, 10–11. doi: 10.1007/s11032-019-1052-x

Zhang, X., He, Y., He, W., Su, H., Wang, Y., Hong, G., et al. (2019). Structural and functional insights into the LBD family involved in abiotic stress and flavonoid syntheses in *Camellia sinensis*. *Sci. Rep.* 9, 15651. doi: 10.1038/s41598-019-52027-6

Zhang, C., Hou, Y., Hao, Q., Chen, H., Chen, L., Yuan, S., et al. (2015). Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. *PLoS One* 10, e0125174. doi: 10.1371/journal.pone.0125174

Zhang, Y., Lan, H., Shao, Q., Wang, R., Chen, H., Tang, H., et al. (2016). An A20/AN1-type zinc finger protein modulates gibberellins and abscisic acid contents and increases sensitivity to abiotic stress in rice (*Oryza sativa*). *J. Exp. Bot.* 67, 315–326. doi: 10.1093/jxb/erv464

Zhang, Z., Ren, C., Zou, L., Wang, Y., Li, S., and Liang, Z. (2018). Characterization of the GATA gene family in *Vitis vinifera*: genome-wide analysis, expression profiles, and involvement in light and phytohormone response. *Genome* 61, 713–723. doi: 10.1139/gen-2018-0042

Zhang, C., Yang, R., Zhang, T., Zheng, D., Li, X., Zhang, Z. B., et al. (2023). ZmTIFY16, a novel maize TIFY transcription factor gene, promotes root growth and development and enhances drought and salt tolerance in Arabidopsis and *Zea mays*. *Plant Growth Regul.* 100, 149–160. doi: 10.1007/s10725-022-00946-2

Zhang, Z., Zou, X., Huang, Z., Fan, S., Qun, G., Liu, A., et al. (2019). Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of *Gossypium* genus. *Gene* 680, 72–83. doi: 10.1016/j.gene.2018.09.039

Zhao, Y., Fu, G., Jiang, J., Li, Y., Liu, X., Li, J., et al. (2022). Chromosome-scale *Cerasus humilis* genome assembly reveals gene family evolution and possible genomic basis of calcium accumulation in fruits. *Sci. Hortic-Amsterdam* 299, 111012. doi: 10.1016/j.scienta.2022.111012

Zheng, L., Tang, L., and Li, J. (2024). Genome-wide identification of the GATA gene family in melon (*Cucumis melo*) and analysis of their expression characteristics under biotic and abiotic stresses. *Front. Plant Sci.* 15, 1462924. doi: 10.3389/fpls.2024.1462924

Zhu, Y., Wu, N., Song, W., Yin, G., Qin, Y., Yan, Y., et al. (2014). Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. *BMC Plant Biol.* 14, 93. doi: 10.1186/1471-2229-14-93

Zubo, Y. O., Blakley, I. C., Franco-Zorrilla, J. M., Yamburenko, M. V., Solano, R., Kieber, J. J., et al. (2018). Coordination of chloroplast development through the action of the GNC and GLK transcription factor families. *Plant Physiol.* 178, 130–147. doi: 10.1104/pp.18.00414