Genome-wide Survey of the *GATA* Gene Family in Camptothecin-producing Plant *Ophiorrhiza Pumila*

Qikai Huang
Zhejiang Chinese Medical University
Min Shi
Zhejiang Chinese Medical University
Can Wang
Zhejiang Chinese Medical University
Jiangning Hu
Zhejiang Conba Pharmaceutical Limited Company
Guoyin Kai (guoyinkai@yahoo.com)
Zhejiang Chinese Medical University

Research Article

Keywords: Ophiorrhiza pumila, GATA, genome-wide, camptothecin, expression pattern

Posted Date: April 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-385244/v1

Abstract

Background: *Ophiorrhiza pumila* (Rubiaceae) is capable of producing camptothecin (CPT), one monoterpene indole alkaloid extensively employed in the treatment of multiple cancers. Transcription factors (TF) *GATA* are a group of transcription regulators involved in plant development and metabolism, and show the feature of binding to the GATA motif within the promoters of target genes. However, *GATA* TFs have not been characterized in *O. pumila*.

Result: In this study, a total of 18 *GATA* genes classified into four subfamilies were identified, which randomly distributed on 11 chromosomes of *O. pumila*. Gene replication and homology between *O. pumila* and other plant species such as *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, *Solanum lycopersicum*, and *Vitis vinifera* genomes were analyzed. Tissue expression pattern revealed that *OpGATA7*, *OpGATA12* and *OpGATA13* with higher transcript in leaves, which was correlated with *ASA*, *MK*, *DXS*, *CMS*, and *MECS*. *OpGATA7*, *OpGATA14* and *OpGATA15* showed high expression in roots as most of the CPT biosynthetic pathway genes did, suggesting that these *OpGATAs* may be potential candidates regulating CPT biosynthesis in *O. pumila*.

Conclusions: Genome-wide survey of the GATA gene family from *Ophiorrhiza pumila* provided insights into the involvement of GATA transcription factors in CPT biosynthesis

Background

Ophiorrhiza pumila is a dicotyledonous plant classified into Rubiaceae family and remains as a sustainable source of camptothecin (CPT). CPT is a type of monoterpene indole alkaloids (MIAs) commonly used in treatment of cancers and was initially isolated from *Camptotheca acuminata* [1], and subsequently detected in *Nothapodytes nimmoniana* and other plants [2–4]. CPT inhibits tumor growth by blocking DNA topoisomerase I [5, 6]. Topotecan and irinotecan, two drugs developed by CPT derivatives, have been extensively employed in various cancers including lung, colorectal, cervical, and ovarian cancers [7]. The biosynthesis pathway of CPT is complex and remains not fully resolved [8]. Briefly, the terpene section of CPT is derived from the 2-C-methyl-d-erythritol 4-phosphate (MEP) and mevalonate (MVA) pathways. The produced geraniol is hydroxylated to 10-hydroxygeraniol under the catalysis of geraniol 10-hydroxy (G10H) [9, 10], and then oxidized to 10-oxogeranial by 10-hydroxy geraniol oxidoreductase (10-HGO). Next, 10-oxogeraniol is converted to iridodial under the action of iridodial synthase, followed by conversion to iriotrial under the action of iridodial oxidoreductase (IO). Iriotrial is then converted to 7-deoxyloganetic acid under the action of IO. 7-deoxyloganetic acid is converted to 7-deoxyloganic acid by glucosyltransferase (7-DLGT) [6], and the product is then converted to loganic acid by 7-deoxyloganic acid hydroxylase (7-DLH) [6]. Lastly, secologanin, a precursor of CPT, is synthesized by secologanin synthase (SLS) [11, 12]. Another precursor tryptamine was produced from tryptophan under the catalyzation of tryptophan decarboxylase (TDC) [13]. Strictosidine is synthesized by condensation of tryptamine and secologanin catalyzed by strictosidine synthase (STR) [9, 14], and CPT is

then formed via a series of catalytic reactions that have not yet been elucidated. Recently, *O. pumila* has been regarded as a model plant for MIA biosynthesis [10].

To adapt to changes in the external environment and resist various biotic and abiotic stresses, plants have formed a series of complex and efficient regulatory networks causing changes in gene expression response to such stresses at multiple levels, and transcription factors (TFs) are critical regulators of these processes [15]. The GATA TF is one of the ubiquitous TF families in eukaryotes and is essential for many aspects of plant development, metabolism and signal conduction [16]. GATA proteins share a common feature of binding to the specific sequence (T/A) GATA (A/G) [17]. The DNA-binding domain of GATA contains a class IV zinc finger structure (C-X2-C-X17-20-C-X2-C), followed by a basal region [18]. Most *GATA* TFs in plants include a single C-X2-C-X18-C-X2-C motif and several contain C-X2-C-X20-C-X2-C. The first plant *GATA* TF was identified from tobacco and termed as *NTL1* harboring C-X2-C-X18-C-X2-C motif [19]. In *Arabidopsis thaliana*, 30 GATA TFs have been identified and characterized. In addition, 28 GATA genes have been identified in *Oryza sativa* [16]. Based on phylogenetic relationships, DNA binding regions, and intron-exon structures [20], *Arabidopsis* and rice GATA family genes can be categorized into four families, i.e., I, II, III, and IV.

GATA TFs have been reported involved in plant metabolism. For example, *A. thaliana* GATA nitrate-inducible carbon-metabolism-involved (GNC) and cytokinin-responsive GATA1 (CGA1) regulated chlorophyll levels, chloroplast size, photosynthetic efficiency, and carbon and nitrogen metabolism [21]. Moreover, *GNC* and *CGA1* show high expression in green tissues and are capable of mediating cytokinin to regulate plastid development [22]. *GATA8* mediated the biomass accumulation and photosynthetic efficiency in *O. sativa* seedlings [23]. Transient overexpression of *CrGATA1* in *Catharanthus roseus* seedlings increased vindoline production [24]. Under low nitrogen deposition, *GATA44* and *GATA58* genes exhibit low expression in soybean seedlings [25]. Additionally, in higher plants, the assimilation pathway of nitrate is tightly regulated. Nitrate is reduced from nitrate reductase to nitrite and then to ammonium nitrogen ($NH4^+$) by nitrite reductase to participate in the synthesis of amino acids and proteins [26]. The promoter of nitrate reductase (NIA) in tomatoes covers the required cis-acting regulatory elements capable of specifically recognizing and binding to GATA protein and then regulating nitrogen metabolism [27]. Moreover, ammonia is required for alkaloid biosynthesis. Nitrogen is an important nutritional factor affecting plant alkaloid biosynthesis and accumulation [28], and CPT contains two nitrogen atoms in its molecular structure owing to its origin as an amino acid-derived alkaloid. Thus, nitrogen metabolism may critically affect the regulation of CPT.

GATA TFs from *O. pumila* (*OpGATA*s) have not been extensively studied. In this study, GATA family TFs distributed in *O. pumila* were we systematically characterized [29]. Based on phylogenetic relationship and expression pattern combined with co-expression analysis, candidate *GATA* genes regulating CPT biosynthesis were predicted. The results provide a comprehensive analysis of *OpGATA* family genes, which shed new lights on CPT biosynthesis in *O. pumila*.

Results

Identification and phylogenetic analysis of GATA proteins

In this study, a total of 18 *OpGATA* genes were identified from the genome of *O. pumila* according to HMM search results, and renamed as *OpGATA1–OpGATA18* according to their chromosome position. Fundamental characteristics of *OpGATA1–OpGATA18* including coding sequence length, protein molecular weight, point isoelectric (pI) and subcellular location were analyzed (Table S1). The complete open reading frame (ORF) of OpGATAs varied from 160 bp (OpGATA4) to 543 bp (OpGATA15), and the molecular weights ranged from 17.73 kDa to 59.86 kDa. The pI values were predicted ranging from 5.05 (OpGATA4) to 10.11 (OpGATA6). All OpGATA proteins were predicted to localize in the nucleus.

Phylogenetic analysis and classification of OpGATA

To determine the phylogenetic relationships of OpGATA proteins, a neighbor-joining tree was constructed by complying with the full-length GATA proteins from *O. pumila*, *O. sativa*, and *A. thaliana* (Fig. 1). OpGATA proteins were classified into four distinct subfamilies (I, II, III, and IV). Seven OpGATAs (i.e., OpGATA1, OpGATA2, OpGATA7, OpGATA9, OpGATA11, OpGATA14, and OpGATA16) were classified as subfamily I; five OpGATAs (i.e., OpGATA3, OpGATA4, OpGATA8, OpGATA12, and OpGATA13) were assigned into subfamily II; five OpGATAs (i.e., OpGATA5, OpGATA6, OpGATA10, OpGATA17, and OpGATA18) were grouped into subfamily III and subfamily IV having only one OpGATA namely OpGATA15.

Gene structure and motif composition of the OpGATA gene family

To gain insights into the characteristics of OpGATA proteins, the motifs of OpGATA proteins were analyzed by MEME. In total, 10 different conservative motifs were characterized (motifs 1–10) (Fig. 2B) (Table S2). Motifs 1 and 5 were detected in all proteins except OpGATA3. Motifs 4, 6, 7, and 8 were mainly observed in subfamily I; motif 9 was mainly present in subfamily II; motif 2 and 3 remained in subfamily III; and motifs 1, 5, and 7 were mainly contained in subfamily IV. The exon and intron structures of *OpGATA* genes were obtained by comparing the corresponding genomic DNA sequences of *O. pumila*. Notably, subfamilies III and IV had more introns, whereas subfamilies I and II had only 1–3 introns (Fig. 2C). Overall, members within a single subfamily exhibited similar gene structures, and the results in this study showed similar gene structures and conserved motifs, strongly supporting the results of phylogenetic analysis of subfamily classification. Similar to previous studies of *A. thaliana*, rice, and other plants [16, 18, 20], OpGATAs classified into subfamilies I, II, and III contain the conserved domain C-X2-C-X18-C-X2-C (except OpGATA3), while domain with the C-X2-C-X20-C-X2-C structure was existed in subfamily IV (Fig. 3).

Chromosomal distribution and synteny analysis of the OpGATA gene family

A physical location map of all *OpGATA* genes in the genome of *O. pumila* was drawn (Fig. 4). The distribution of *OpGATA* genes on chromosomes was not homogeneous. The maximum number of *OpGATA* genes was distributed on Opu_Chr02 (OpGATA2–OpGATA7), whereas Opu_chr05 and Opu_chr11 had no *OpGATA* genes and the other chromosomes harbored 1–2 *OPGATA* genes, such as OpGATA1 distributed on Opu_Chr01, OpGATA8 and OpGATA9 on Opu_Chr03, OpGATA10 and OpGATA11 on Opu_Chr04, OpGATA12 on Opu_Chr06, OpGATA13 on Opu_Chr07, OpGATA14 and OpGATA15 on Opu_Chr08, OpGATA16 on Opu_Chr09, OpGATA17 and OpGATA18 on Opu_Chr10.

Moreover, replication events of OpGATAs were analyzed. The result showed that no tandem repeats were identified among the 18 genes, while four pairs of fragment repeats were detected between eight chromosomes, i.e., Opu_chr01 (*OpGATA1*)/Opu_chr09 (*OpGATA16*), Opu_chr02 (*OpGATA4*)/Opu_chr03 (*OpGATA8*), Opu_chr02 (*OpGATA5*)/Opu_chr10 (*OpGATA18*), and Opu_chr04 (*OpGATA11*)/Opu_chr08 (*OpGATA14*). Accordingly, some *OpGATA* genes may have been generated by gene replication, thereby critically affecting the amplification of *OpGATA* genes in *O. pumila*.

To gain insights into the evolution of *OpGATA* genes, the homology (Fig. 5; Table S3) of GATA proteins from *O. pumila*, *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, *Solanum lycopersicum*, and *Vitis vinifera* was evaluated. Overall, 16, 15, 14, 7, and 12 OpGATA proteins displayed homology with proteins from *G. max*, *S. lycopersicum*, *V. vinifera*, *O. sativa*, and *A. thaliana*, respectively. Furthermore, OpGATA4, OpGATA8, OpGATA12, and OpGATA14 exhibited homologous relationship in the five plant species, demonstrating that these four proteins may critically affect evolution.

Expression profiles of OpGATA genes and key enzyme genes in different tissues

Expression profiles of *OpGATA* and vital enzyme genes were evaluated in three distinct tissues and organs (roots, stems and leaves). A heat map was built according to the QTR-PCR analysis. The results showed that most of the CPT-producing pathway genes (e.g., *PMK*, *MDC*, *SLS*, *LAMT*, *TSB*, *DXR*, *AACT*, *10-HGO*, *7-DLGT*, *G10H*, *HMGR*, *STR*, *TDC*, *IO*, *IS*, *HDS*, *HDR*, *IPPI*, *CMK*, *HMGS*, *8-HGO*, and *GES*) expressed highly in roots (Fig. 6A). Several genes including *CMS*, *MECS*, *MK*, *DXS*, *ASA*, and *GPPS* showed higher expression in leaves. As indicated in Fig. 6B, most of *OpGATA* genes were highly expressed in stems, while expression level of *OpGATA9*, *OpGATA12* and *OpGATA13* was higher in leaves, and *OpGATA7*, *OpGATA14* as well as *OpGATA15* expressed higher in roots. In addition, correlations between *OpGATA*s and pathway genes were analyzed (Fig. 7, Table S4). The results showed that *OpGATA7*, *OpGATA14* and *OpGATA15* were correlated with many vital enzyme genes ((i.e., *MDC*, *TDC*, *STR*, *HMGR*, *SLS*, *PMK*, *LAMT*, *CMK*, *8-HGO*, *HDR*, *HMGS*, *IPPI*, *HDS*, *AACT*, *DXR*, *G10H*, *7-DLGT*, *GES*, *IS*, and *IO*) that were highly expressed in roots ($p < 0.05$, $r > 0$), demonstrating the involvement of these three *OpGATA* genes in regulating CPT biosynthesis. *OpGATA9* and *OpGATA12* may also regulate the upstream pathway of CPT biosynthesis and are related to genes in the MVA and MEP pathways [30] (e.g., *ASA*, *MK*, *DXS*, *CMS*, and *MECS*), all of which are highly expressed in leaves. To identify the cis-element of *OpGATA*, the 3000 bp promoter sequences of genes encoding vital enzymes in the CPT biosynthesis pathway were analyzed

using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). As predicted, the GATA motif was present in the promoters of several key enzyme genes (e.g., *TSB*, *SLS*, *MK*, *MDC*, *IPPI*, *HMGR*, *HDR*, *GES*, *G10H*, *DXS*, *AACT*, and *8-HGO*) (Table S5). Thus, these key biosynthetic genes may be regulated by *GATA* TFs.

Discussion

Camptothecin (CPT) is a widely known monoterpene indole alkaloid with excellent anticancer activity. CPT has been isolated from different plant species. CPT- producing weedy plant *Ophiorrhiza pumila* has brought about widespread attention, and the whole genome of *O. pumila* has been sequenced [29]. Recently, metabolic engineering has been applied in *O. pumila* to elevate CPT content. For example, individual introduction of *G10H* or *SLS*, and co-expression of *G10H* and *SLS* significantly enhanced CPT content in transgenic *O. pumila* hairy roots [31]. Besides, transcription regulation of CPT biosynthesis has been studied. RNA interference of *OpERF2* suppressed expression level of genes involved in MEP and secologanin-strictosidine pathways [32]. The transcription repressor OpMYB1 reduced CPT biosynthesis by downregulating expression level of *TDC* [33]. OpWRKY2 acted as a positive regulator of CPT biosynthesis by directly targeting *TDC* [10]. OpWRKY1 inhibited CPT biosynthesis by directly downregulating CPR transcription in *O. pumila* [12]. Nevertheless, transcription regulation of CPT biosynthesis needs further study.

The GATA TF family involved in many aspects of physiology-related processes has been broadly explored in a range of plants including *Arabidopsis*, rice [16], grapes [34], Moso bamboo [35], and *Gossypium* sp. [20]. The present study was first reported GATA TFs in *O. pumila*. Totally, 18 GATA TFs were identified and named *OpGATA1–OpGATA18* according to their physical location on the chromosome. The whole OpGATA family in *O. pumila* could be classified into four groups, similar to those in *A. thaliana*. In subfamily III, the GATA domain harbored 20 residues in the zinc finger domain, making up a C-X2C-X20-C-X2-C structure, and the other three subfamilies showed that C-X2C-X18-C-X2-C structure, containing 18 residues. The CCT and TIFY domains were specifically identified within subfamily III. The CCT domain was initially found in *Arabidopsis* Constans protein, which facilitates root and hypocotyl development within *A. thaliana* and mediates flowering [36]. Previously, the family with a completely conserved TIFY domain was termed TIFY [37]. However, in recent studies, the TIFY domain has been shown to exist extensively in jasmonate ZIM domain protein family and PEAPOD proteins, which are associated with the jasmonic acid pathway [38].

Motif analysis showed that all OpGATAs contained motif 1 and 5 except OpGATA3, and specific motifs were detected in other groups. For example, motif 4 was only observed in subfamily I, motif 9 was only detected in subfamily II, and motifs 2 and 3 were only detected in subfamily III, suggesting that although some motifs of GATA family genes are highly conserved, new evolutionary motifs may have distinct functions in some plants, and the functions of these new evolutionary motifs need to be further verified. The homology of *GATA* genes from *O. pumila* with those from *Arabidopsis*, rice, soybeans, tomatoes, and grapes was also explored. Notably, the *Arabidopsis* GATA TFs AtGATA1 (AT3G24050), AtGATA2

(AT3G60530), and AtGATA4 (AT2G45050) have been reported to facilitate light-dependent regulation of gene expression and photomorphogenesis [39]. Accordingly, the homologous genes OpGATA9 (Opuchr03_g0010130-1.1) and OpGATA2 (Opuchr03_g0010130-1.1) may also affect light-dependent regulation of genes. AtGATA22 (AT4G26150), which is homologous with OpGATA12 (Opuchr06_g0009000-1.1), affects the response to cytokinins and hinders root growth in *A. thaliana* [40]. Additionally, *GNC* (AT5G56860), which is homologous to OpGATA12, adversely affects seed germination, flowering, and leaf elongation, and overexpression of GNC inhibits the germination, leaf expansion, and flowering of *A. thaliana* [41]. AtGATA12 (AT5G25830), which is homologous to OpGATA14, is involved in primary dormancy in *A. thaliana* [42].

Expression level of most genes encoding the key enzymes in the CPT biosynthesis pathway (i.e., *PMK*, *MDC*, *SLS*, *LAMT*, *TSB*, *DXR*, *AACT*, *10-HGO*, *7-DLGT*, *G10H*, *HMGR*, *STR*, *TDC*, *IO*, *IS*, *HDS*, *HDR*, *IPPI*, *CMK*, *HMGS*, *8-HGO*, and *GES*) were significantly higher in roots compared with those in other tissues, whereas genes involved in the MVA and MEP pathways (i.e., *ASA*, *MK*, *DXS*, *CMS*, and *MECS*) were mostly expressed in leaves. According to the expression patterns of *OpGATA*s within different tissues, a correlational analysis was conducted. The results showed that *OpGATA7*, *OpGATA14*, and *OpGATA15* exhibited positive associations with key enzyme genes showing high expression in roots. Moreover, *OpGATA15* showed pole-strength correlations with several key genes involved in the biosynthetic pathway (i.e., *TDC*, *HMGS*, *AACT*, *PMK*, *IPPI*, *DXR*, *HDS*, *STR*, *G10H*, *LAMT*, and *SLS*). *OpGATA14* only had a pole-strength correlation ($p < 0.05$, $r > 0.8$) with the *HMGR* gene and showed strong correlations with *G10H*, *8-HGO*, and *LAMT* genes. Furthermore, *OpGATA7* did not have a pole-strength correlation with any enzyme genes but showed strong correlations ($p < 0.05$, $0.8 > r > 0.6$) with *CMK* and *8-HGO*. *OpGATA9*, *OpGATA12*, and *OpGATA13*, which were highly expressed in leaves, were found to be correlated with *ASA*, *MK*, *DXS*, *CMS*, and *MECS*, with Pearson correlation coefficients of greater than 0.6 for *OpGATA12* and *OpGATA9*. Plant terpenoids are synthesized mainly through the MVA and MEP pathways [43]. Genes that are highly expressed in the leaves are typically involved in the MEP and MVA pathways [30], demonstrating that *OpGATA9*, *OpGATA12*, and *OpGATA13* may regulate CPT biosynthesis by participating in the upstream pathway. Among the genes mentioned above, GATA motifs were found in the promoters of key enzyme genes (i.e., *SLS*, *MDC*, *IPPI*, *HMGR*, *HDR*, *GES*, *G10H*, *AACT*, and *8-HGO*) highly expressed in roots and key enzyme genes (i.e., *MK*, *DXS*) highly expressed in leaf, demonstrating that these key enzyme genes may be directly regulated by *OpGATA* genes and then affect the biosynthesis of CPT. Of which, the genes highly expressed in root (i.e., *AACT*, *IPPI*, *G10H*, *HMGR*) which is pole-strength correlations with *OpGATA14* or *OpGATA15* is a higher possibility of being directly regulated.

Overall, this comprehensive analysis of *GATA* family genes from *O. pumila* provided insights into the characteristics of *OpGATA* genes and may improve our understanding of the mechanisms regulating CPT biosynthesis in *O. pumila*.

Conclusion

In this study, OpGATA TF family in *O. pumila* were characterized and identified. Overall, a total of 18 *OpGATA* genes showing different chromosomal distribution were classified into four subfamilies. Phylogenic and homology analysis of GATA genes were conducted within several plant species including *O. pumila*, *Arabidopsis*, grapes, tomatoes, and soybeans, and the functions of some homologous genes were predicted. OpGATA genes showed different expression patterns within a range of tissues (e.g., leaves, stems, and roots) in correlation to key pathway genes, highlighting the potential roles of some OpGATA genes such as *OpGATA7*, *OpGATA9*, *OpGATA12*, *OpGATA14* and *OpGATA15* in the regulation of CPT biosynthesis in *O. pumila*. This study provides novel OpGATA TFs involved in regulating CPT biosynthesis.

Methods

Identification of OpGATAs

The hidden Markov model (PF00320) of the GATA domain originating from the Pfam database (http://pfam.xfam.org) was used to identify the *OpGATA* family [44]. The Pfam database (http://pfam.xfam.org/search/sequence) and NCBI CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) were employed to verify the integrity of the GATA domain, with an E-value cutoff of 0.01 [45]. The ProSite ExPASy server (http://web.expasy.org/protparam/) was adopted to predict the physical and chemical properties of OpGATA proteins. Subcellular localization of GATA proteins was predicted using CELLO (http://cello.life.nctu.edu.tw/).

Multiple sequence alignment and phylogenetic analysis

The GATA proteins from *A. thaliana* Information Resources (www.arabidopsis.org/index.jsp) and the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/cgi-bin/ORF_infopage.cgi) were downloaded [18]. MAFFT software was employed for multiple sequence alignment of GATA proteins [46]. The neighbor-joining tree of GATA TF families from *A. thaliana*, *O. sativa*, and *O. pumila* were built by MEGA v5 [47], with the parameters of Poisson model, pairwise deletion, and 1000 bootstrap tests.

Motifs and gene structures

The MEME was employed (http://meme.sdsc.edu/meme/itro.html) to identify the conserved motif of GATA protein in *O. pumila*, with the following parameters were adopted: 0 or 1 occurrence per sequence; maximum number of motifs = 10; and optimum motif length = 6–50 residues. Exon-intron structure of the *GATA* members was investigated by analyzing the *O. pumila* genome, and gene structure was visualized with TBtools [48].

Chromosomal distribution and gene duplication of GATA genes

The method for mapping *GATA* genes on the chromosome of *O. pumila* was identical to that of *FtAP2/ERF* genes [49]. Gene replication events were investigated using the multiple collinear scanning toolkit (MCScanX) and BLASTP method. TBtools software (https://github.com/CJ-Chen/TBtools) was adopted to build syntenic analysis maps for determining the syntenic relationships between OpGATA proteins and GATA proteins from *A. thaliana*, *O. sativa*, *G. max*, *S. lycopersicum*, and *V. vinifera*.

Expression analysis by real-time polymerase chain reaction (QRT-PCR)

Total RNA was extracted using a Plant RNAprep Pure Kit (TIANGEN, China). Corresponding sequences of *OpGATA* genes and key enzyme genes were acquired from the *O. pumila* genome sequence database (http://pumila.kazusa.or.jp/). Reverse transcription quantitative PCR (RT-qPCR) primers were designed using Primer 5 software (http://frodo.wi.mit.edu/; Table S6). The housekeeping gene *OpActin* from *O. pumila* was used as an internal control gene in QRT-PCR for normalization of all samples [10]. Relative expression levels were calculated with the $2^{-\Delta\Delta Ct}$ method. All QRT-PCR analyses were performed with three biological replicates. With the value of $2^{-\Delta\Delta Ct}$, the heatmap was constructed by TBtools software, the pearson correlation analysis by R script.

Abbreviations

Ophiorrhiza pumila, *O. pumila*; *Arabidopsis thaliana*, *A. thaliana*; *Oryza sativa*, *O. sativa*; *Glycine max*, *G. max*; *Solanum lycopersicum*, *S. lycopersicum*; *Vitis vinifera*, *V. vinifera*; camptothecin, CPT; point isoelectric, pI; transcription factors, TFs; monoterpene indole alkaloid, MIAs; 2-C-methyl-d-erythritol 4-phosphate, MEP; mevalonate, MVA; geraniol 10-hydroxy, G10H; 10-hydroxy geraniol oxidoreductase, 10-HGO; iridodial oxidoreductase, IO; 7-deoxyloganic acid by glucosyltransferase, 7-DLGT; 7-deoxyloganic acid hydroxylase, 7-DLH; secologanin synthase, SLS; tryptophan decarboxylase, TDC; strictosidine synthase, STR; anthranilate synthase, ASA; cytochrome P450 reductases, CPR; 1-deoxy-D-xylulose-5-phosphate reductoisomerase, DXR; 1-deoxy-D-xylulose-5-phosphate synthase, DXS; 1-hydroxy-2-methyl-2(E)-butenyl-4-diphosphate reductase, HDR; the beta-subunit of tryptophan synthase, TSB; real-time polymerase chain reaction, QRT-PCR; 8-hydroxy-geraniol oxidoreductase, 8-HGO; acetyl-CoA C-acetyltransferase, AACT; 4-(cytidine 5-diphospho)-2-C-methylerythritolkinase, CMK; 4-(cytidine 5-diphospho)-2-C-methylerythritol synthase, CMS; geraniol synthase, GES; geranyl diphosphate synthase/geranyl pyrophosphate synthase, GPPS; hydroxymethylbutenyl 4-diphosphate synthase, HDS; 3-hydroxy-3-methylglutaryl-CoA reductase, HMGR; 3-hydroxy-3-methylglutaryl-CoA synthase, HMGS; isopentenyl diphosphate isomerase, IPPI; iridoid synthase, IS; loganic acid O-methyltransferase, LAMT; mevalonate 5-diphosphate decarboxylase/mevalonate(diphospho)-decarboxylase, MDC; 2-C-methylerythritol-2,4-cyclodiphosphate synthase, MECS; mevalonate kinase, MK; phosphomevalonate kinase, PMK.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The genomic information of *Ophiorrhiza pumila* download from *Ophiorrhiza pumila* Genome DateBase (http://pumila.kazusa.or.jp/). All data generated or analyzed during this study are included in this published article and its supplementary information files

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by Zhejiang Provincial Ten Thousand Program for Leading Talents of Science and Technology Innovation [2018R52050]; National Natural Science Fund of China [82073963, 31571735, 81522049]; Zhejiang Provincial Program for the Cultivation of High-Level Innovative Health Talents. The Opening Project of Zhejiang Provincial Preponderant and Characteristic Subject of Key University (Traditional Chinese Pharmacology), Zhejiang Chinese Medical University [ZYAOX2018004, ZYAOXZD2019004, ZYAOXZD2019006].

Authors' contributions

Q.K. and M.S performed bioinformatics analyses, sample collection, experiments and drafted the manuscript. C.W. and J.H. helped in bioinformatics analyses and physiological experiments. G.K. designed the experiments and conceived the project, provided overall supervision of the study and revised the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

We thank all the staff in our lab for providing useful suggestions and technical assistance.

References

1. Lorence A, Nessler CL. Camptothecin, over four decades of surprising findings. Phytochemistry. 2004;65:2735-2749. https://doi.org/10.1016/j.phytochem.2004.09.001
2. Rather GA, Sharma A, Jeelani SM, Misra P, Kaul V, Lattoo SK. Metabolic and transcriptional analyses in response to potent inhibitors establish MEP pathway as major route for camptothecin

biosynthesis in Nothapodytes nimmoniana (Graham) Mabb. BMC Plant Biol. 2019;19:301. https://doi.org/10.1186/s12870-019-1912-x

3. Pu X, Zhang CR, Zhu L, Li QL, Huang QM, Zhang L, Luo YG. Possible clues for camptothecin biosynthesis from the metabolites in camptothecin-producing plants. Fitoterapia. 2019;134:113-128. https://doi.org/10.1016/j.fitote.2019.02.014
4. Raveendran VV, Vijayan FP, Padikkala J. Antitumor activities of an anthraquinone fraction isolated from in vitro cultures of Ophiorrhiza rugosa var decumbens. Integr Cancer Ther. 2012;11:120-128. https://doi.org/10.1177/1534735411403478
5. Zhang L, Ma D, Zhang Y, He W, Yang J, Li C, Jiang H. Characterization of DNA topoisomerase-1 in Spodoptera exigua for toxicity evaluation of camptothecin and hydoxy-camptothecin. PLoS One. 2013;8:e56458. https://doi.org/10.1371/journal.pone.0056458
6. Kai G, Wu C, Gen L, Zhang L, Cui L, Ni X. Biosynthesis and biotechnological production of anti-cancer drug camptothecin. Phytochem Rev. 2015;4: 525–539. https://doi.org/10.1007/s11101-015-9405-5
7. Coleman RL. Emerging role of topotecan in front-line treatment of carcinoma of the ovary. Oncologist. 2002;7:46-55. https://doi.org/10.1634/theoncologist.7-1-46
8. Yamazaki M, Saito K, Sirikantaramas S. Camptothecin: biosynthesis, biotechnological production and resistance mechanism(s). 2013;68:139-161. https://doi.org/10.1016/B978-0-12-408061-4.00005-5
9. Cui L, Ni X, Ji Q, Teng X, Yang Y, Wu C, Zekria D, Zhang D, Kai G. Co-overexpression of geraniol-10-hydroxylase and strictosidine synthase improves anti-cancer drug camptothecin accumulation in Ophiorrhiza pumila. Sci Rep. 5:8227. 2015;https://doi.org/10.1038/srep08227
10. Hao X, Xie C, Ruan Q, Zhang X, Wu C, Han B, Qian J, Zhou W, Nützmann HW, Kai G. The transcription factor OpWRKY2 positively regulates the biosynthesis of the anticancer drug camptothecin in Ophiorrhiza pumila. Hortic Res. 2021;8:7. https://doi.org/10.1038/s41438-020-00437-3.
11. Wang C, Wu C, Wang Y, Xie C, Shi M, Nile S, Zhou Z, Kai G. Transcription Factor OpWRKY3 Is Involved in the Development and Biosynthesis of Camptothecin and Its Precursors in Ophiorrhiza pumila Hairy Roots. Int J Mol Sci. 2019;20:3996. https://doi.org/10.3390/ijms20163996.
12. Xu M, Wu C, Zhao L, Wang Y, Wang C, Zhou W, Ming Y, Kai G. WRKY transcription factor OpWRKY1 acts as a negative regulator of camptothecin biosynthesis in Ophiorrhiza pumila hairy roots. Plant Cell Tiss Organ Cult. 2020;142: 69–78. https://doi.org/10.1007/s11240-020-01833-2
13. You D, Feng Y, Wang C, Sun C, Wang Y, Zhao D, Kai G. Cloning, characterization, and enzymatic identification of a new tryptophan decarboxylase from Ophiorrhiza pumila. Biotechnol Appl Biochem. 2020. https://doi.org/10.1002/bab.1935.
14. Lu Y, Wang H, Wang W, Qian Z, Li L, Wang J, Zhou G, Kai G. Molecular characterization and expression analysis of a new cDNA encoding strictosidine synthase from Ophiorrhiza japonica. Mol Biol Rep. 2009;36:1845-1852. https://doi.org/10.1007/s11033-008-9389-y.
15. Liu M, Wen Y, Sun W, Ma Z, Huang L, Wu Q, Tang Z, Bu T, Li C, Chen H. Genome-wide identification, phylogeny, evolutionary expansion and expression analyses of bZIP transcription factor family in

tartaty buckwheat. BMC Genomics. 2019;20:483. https://doi.org/10.1186/s12864-019-5882-z.

16. Reyes JC, Muro-Pastor MI, Florencio FJ. The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol. 2004;134:1718-1732. https://doi.org/10.1104/pp.103.037788
17. Zhu W, Guo Y, Chen Y, Wu D, Jiang L. Genome-wide identification, phylogenetic and expression pattern analysis of GATA family genes in Brassica napus. BMC Plant Biol. 2020;20:543. https://doi.org/10.1186/s12870-020-02752-20018-009-0066-7
18. Chen H, Shao H, Li K, Zhang D, Fan S, Li Y, Han M. Genome-wide identification, evolution, and expression analysis of GATA transcription factors in apple (Malus×domestica Borkh.). Gene. 2017;627:460-472. https://doi.org/10.1016/j.gene.2017.06.049
19. Daniel-Vedele F, Caboche M. A tobacco cDNA clone encoding a GATA-1 zinc finger protein homologous to regulators of nitrogen metabolism in fungi. Mol Gen Genet. 1993;240:365-373. https://doi.org/10.1007/BF00280388
20. Zhang Z, Zou X, Huang Z, Fan S, Qun G, Liu A, Gong J, Li J, Gong W, Shi Y, Fan L, Zhang Z, Liu R, Jiang X, Lei K, Shang H, Xu A, Yuan Y. Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of Gossypium genus. Gene. 2019;680:72-83. https://doi.org/10.1016/j.gene.2018.09.039
21. Chiang YH, Zubo YO, Tapken W, Kim HJ, Lavanway AM, Howard L, Pilon M, Kieber JJ, Schaller GE. Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. Plant Physiol. 2012;160:332-348. https://doi.org/10.1104/pp.112.198705
22. Naito T, Kiba T, Koizumi N, Yamashino T, Mizuno T. Characterization of a unique GATA family gene that responds to both light and cytokinin in Arabidopsis thaliana. Biosci Biotechnol Biochem. 2007;71:1557-1560. https://doi.org/10.1271/bbb.6069.
23. Nutan KK, Singla-Pareek SL, Pareek A. The Saltol QTL-localized transcription factor OsGATA8 plays an important role in stress tolerance and seed development in Arabidopsis and rice. J Exp Bot. 2020;71:684-698. https://doi.org/10.1093/jxb/erz368
24. Liu Y, Patra B, Pattanaik S, Wang Y, Yuan L. GATA and Phytochrome Interacting Factor Transcription Factors Regulate Light-Induced Vindoline Biosynthesis in Catharanthus roseus. Plant Physiol. 2019;180:1336-1350. https://doi.org/10.1104/pp.19.00489.
25. Zhang C, Hou Y, Hao Q, Chen H, Chen L, Yuan S, Shan Z, Zhang X, Yang Z, Qiu D, Zhou X, Huang W. Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. PLoS One. 2015;10:e0125174. https://doi.org/10.1371/journal.pone.0125174.
26. Crawford NM, Glass ADM. Molecular and physiological aspects of nitrate uptake in plants. Trends in Plant Science. 1998;3:389-395. doi.org/10.1016/S1360-1385(98)01311-9
27. Dorbe MF, Caboche M, Daniel-Vedele F. The tomato nia gene complements a Nicotiana plumbaginifolia nitrate reductase-deficient mutant and is properly regulated. Plant Mol Biol. 1992;18:363-375. https://doi.org/10.1007/BF00034963

28. Bensaddek L, Gillet F, Saucedo JE, Fliniaux MA. The effect of nitrate and ammonium concentrations on growth and alkaloid accumulation of Atropa belladonna hairy roots. J Biotechnol. 2001;85:35-40. https://doi.org/10.1016/s0168-1656(00)00372-2
29. Rai A, Hirakawa H, Nakabayashi R, Kikuchi S, Hayashi K, Rai M, Tsugawa H, Nakaya T, Mori T, Nagasaki H, Fukushi R, Kusuya Y, Takahashi H, Uchiyama H, Toyoda A, Hikosaka S, Goto E, Saito K, Yamazaki M. Chromosome-level genome assembly of Ophiorrhiza pumila reveals the evolution of camptothecin biosynthesis. Nat Commun. 2021;12:405. https://doi.org/10.1038/s41467-020-20508-2
30. Vranová E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol. 2013;64:665-700. https://doi.org/10.1146/annurev-arplant-050312-120116
31. Shi M, Gong H, Cui L, Wang Q, Wang C, Wang Y, Kai G. Targeted metabolic engineering of committed steps improves anti-cancer drug camptothecin production in Ophiorrhiza pumila hairy roots. Ind Crops Prod. 2020;148:112277. https://doi.org/10.1016/j.indcrop.2020.112277
32. Udomsom N, Rai A, Suzuki H, Okuyama J, Imai R, Mori T, Nakabayashi R, Saito K, Yamazaki M. Function of AP2/ERF transcription factors involved in the regulation of specialized metabolism in Ophiorrhiza pumila revealed by transcriptomics and metabolomics. Front Plant Sci. 2016;7:1861. https://doi.org/10.3389/fpls.2016.01861
33. Rohani E R, Chiba M, Kawaharada M, Asano T, Oshima Y, Mitsuda N, Ohme-Takagi M, Fukushima A, Rai A, Saito K, Yamazaki M. An MYB transcription factor regulating specialized metabolisms in Ophiorrhiza pumila. Plant Biotechnology. 2016;33:1-9. https://doi.org/10.5511/plantbiotechnology.15.1117a
34. Zhang Z, Ren C, Zou L, Wang Y, Li S, Liang Z. Characterization of the GATA gene family in Vitis vinifera: genome-wide analysis, expression profiles, and involvement in light and phytohormone response. Genome. 2018;61:713-723. https://doi.org/10.1139/gen-2018-0042
35. Wang T, Yang Y, Lou S, Wei W, Zhao Z, Ren Y, Lin C, Ma L. Genome-Wide characterization and gene expression analyses of GATA transcription factors in Moso Bamboo (Phyllostachys edulis). Int J Mol Sci. 2019;21:14. https://doi.org/10.3390/ijms21010014
36. Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature. 2001;410:1116-1120. https://doi.org/10.1038/35074138
37. Vanholme B, Grunewald W, Bateman A, Kohchi T, Gheysen G. The tify family previously known as ZIM. Trends Plant Sci. 2007;12:239-244. https://doi.org/10.1016/j.tplants.2007.04.004
38. Bai Y, Meng Y, Huang D, Qi Y, Chen M. Origin and evolutionary analysis of the plant-specific TIFY transcription factor family. Genomics. 2011;98:128-136. https://doi.org/10.1016/j.ygeno.2011.05.002
39. Luo XM, Lin WH, Zhu S, Zhu JY, Sun Y, Fan XY, Cheng M, Hao Y, Oh E, Tian M, Liu L, Zhang M, Xie Q, Chong K, Wang ZY. Integration of light- and brassinosteroid-signaling pathways by a GATA

transcription factor in Arabidopsis. Dev Cell. 2010;19:872-883. https://doi.org/10.1016/j.devcel.2010.10.023

40. Köllmer I, Werner T, Schmülling T. Ectopic expression of different cytokinin-regulated transcription factor genes of Arabidopsis thaliana alters plant growth and development. J Plant Physiol. 2011;168:1320-1327. https://doi.org/10.1016/j.jplph.2011.02.006
41. Richter R, Behringer C, Müller IK, Schwechheimer C. The GATA-type transcription factors GNC and GNL/CGA1 repress gibberellin signaling downstream from DELLA proteins and phytochrome-interacting factors. Genes Dev. 2010;24:2093-2104. https://doi.org/10.1101/gad.594910.
42. Ravindran P, Verma V, Stamm P, Kumar PP. A Novel RGL2-DOF6 Complex Contributes to Primary Seed Dormancy in Arabidopsis thaliana by Regulating a GATA Transcription Factor. Mol Plant. 2017;10:1307-1320. doi:10.1016/j.molp.2017.09.004
43. Yu F, Utsumi R. Diversity, regulation, and genetic manipulation of plant mono- and sesquiterpenoid biosynthesis. Cell Mol Life Sci. 2009;66:3043-3052. doi:10.1007/s0
44. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29-W37. https://doi.org/10.1093/nar/gkr367
45. Chanwala J, Satpati S, Dixit A, Parida A, Giri MK, Dey N. Genome-wide identification and expression analysis of WRKY transcription factors in pearl millet (Pennisetum glaucum) under dehydration and salinity stress. BMC Genomics. 2020;21:231. https://doi.org/10.1186/s12864-020-6622-0
46. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772-780. https://doi.org/10.1093/molbev/mst010
47. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35:1547-1549. https://doi.org/10.1093/molbev/msy096
48. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13:1194-1202. https://doi.org/10.1016/j.molp.2020.06.009
49. Liu M, Sun W, Ma Z, Zheng T, Huang L, Wu Q, Zhao G, Tang Z, Bu T, Li C, Chen H. Genome-wide investigation of the AP2/ERF gene family in tartary buckwheat (Fagopyum Tataricum). BMC Plant Biol. 2019;19:84. https://doi.org/10.1186/s12870-019-1681-6

Figures

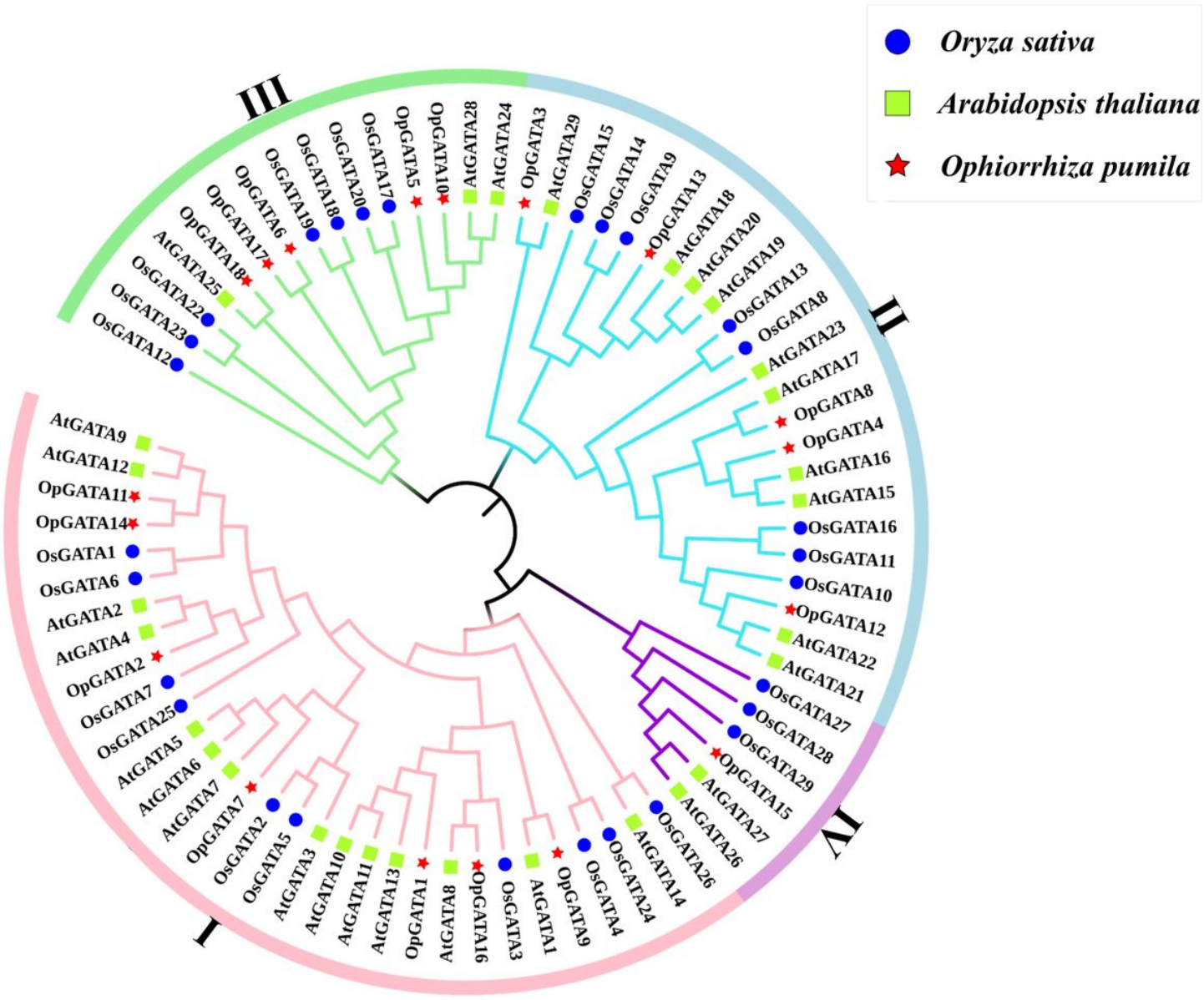

Figure 1

Neighbor-joining tree representing the relationships among GATA proteins of O. pumila, O. sativa and A. thaliana. Constructed with MEGA7 using full-length amino acid sequences and the bootstrap test replicate was set as 1000 times.

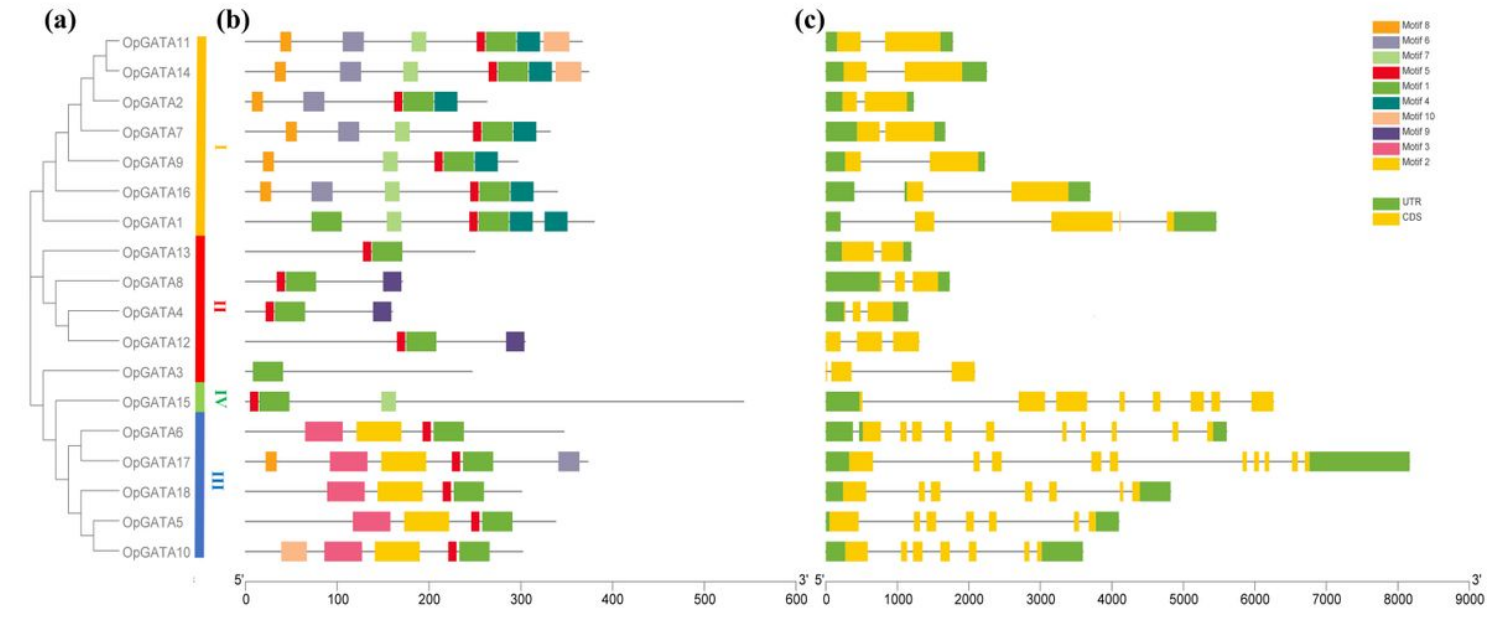

Figure 2

Schematic representation of phylogenetic relationships, conserved motifs and gene structures of the GATA genes in O. pumila. (a) a phylogenetic tree of 18 OpGATA proteins. (b) The motif composition of OpGATA proteins. The motifs are displayed in different colored boxes. (c) Exon/intron structures of OpGATA genes.

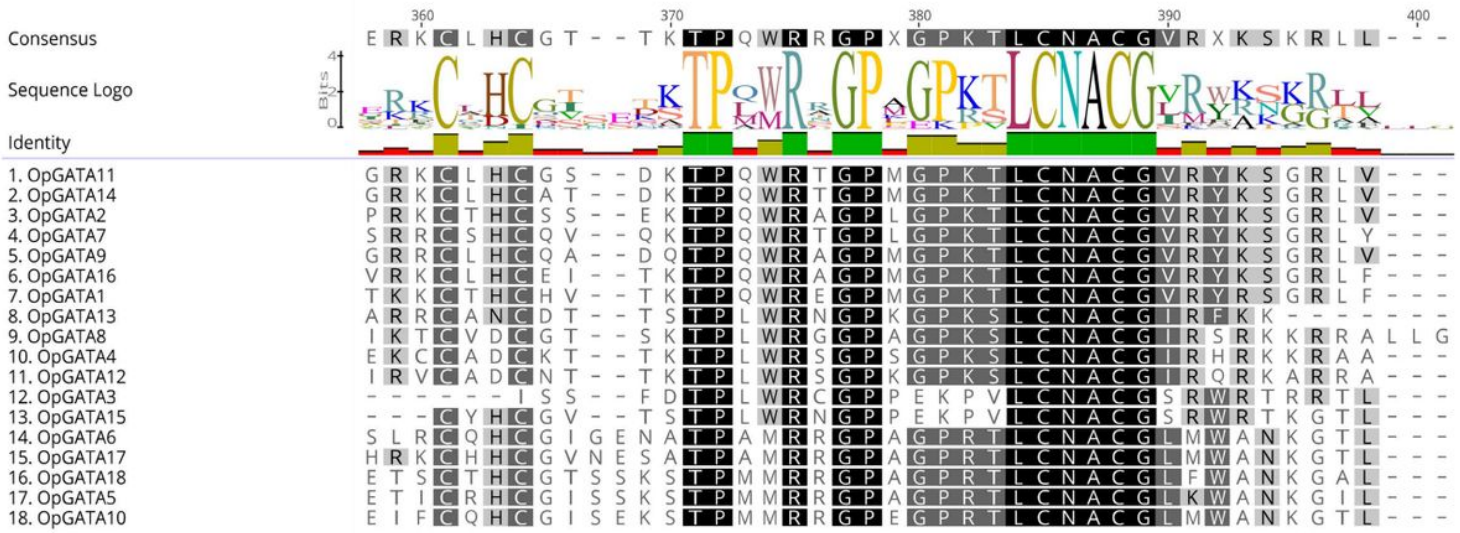

Figure 3

Alignments of GATA domain sequences of the GATA family members in O. pumila.

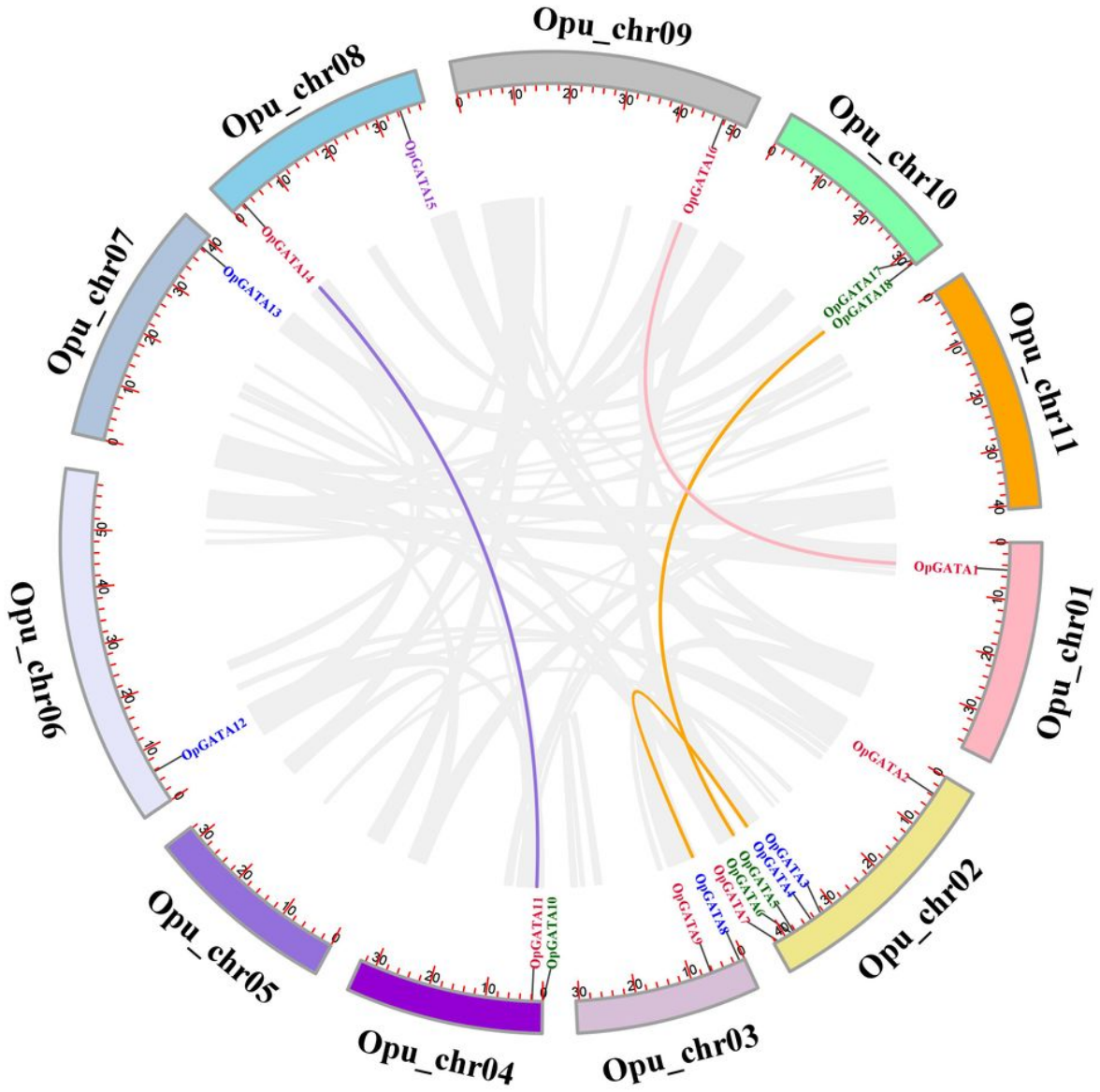

Figure 4

The chromosomal distribution and synteny analysis of OpGATA genes in O. pumila. The locations of all the OpGATA genes are depicted in the chromosomes. Red-colored genes belong to subfamily I, blue-colored genes belong to subfamily II, green-colored genes belong to subfamily III, purple-colored genes belong to subfamily IV. Background gray lines indicate all O. pumila genome synteny blocks, and the Colored lines highlight the duplicated OpGATA gene pairs. ID of the chromosomes is indicated at the bottom of each chromosome.

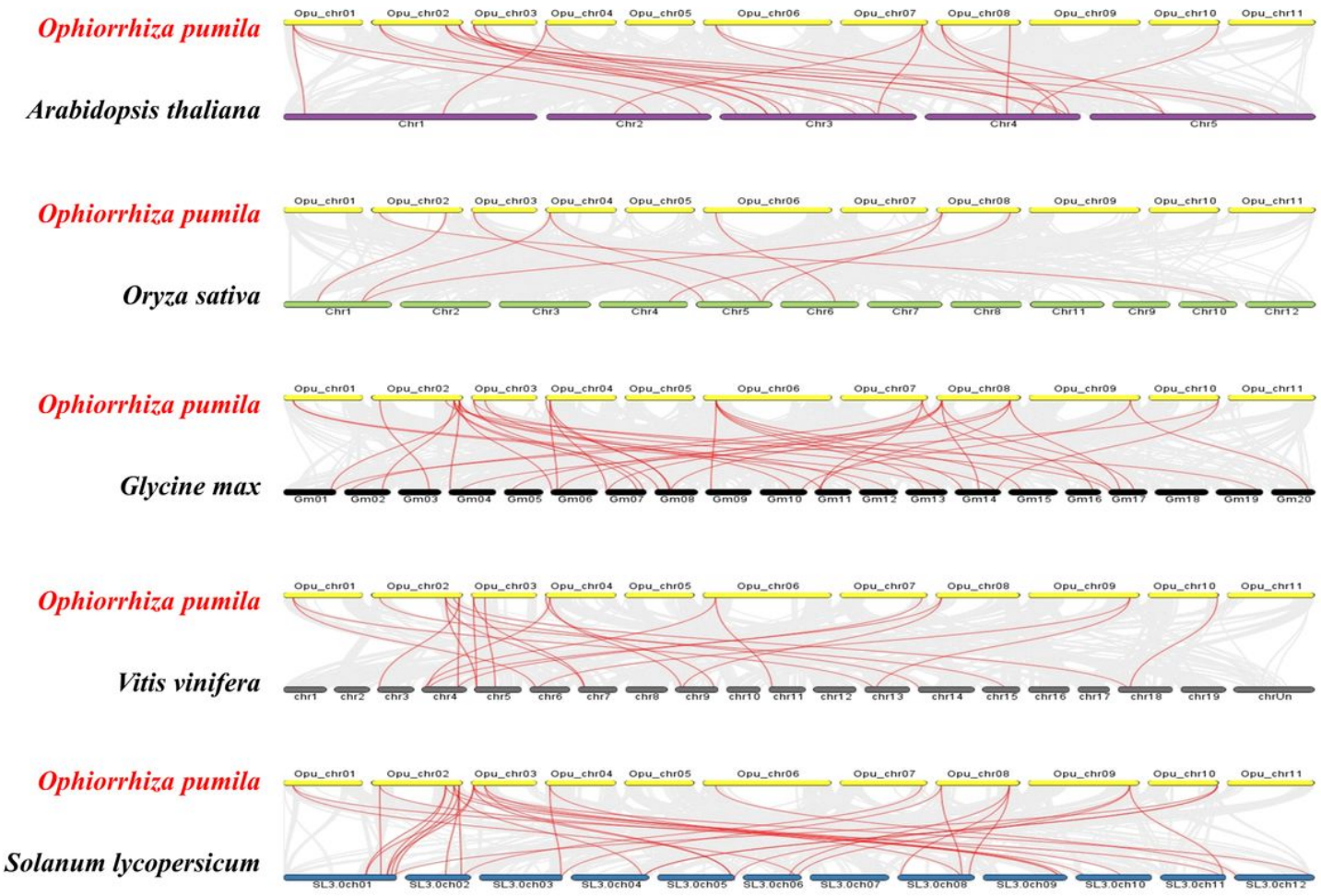

Figure 5

Synteny analysis of GATA genes between O. pumila and five representative plant species (A. thaliana, O. sativa, G. max, S. lycopersicum and V. vinifera.). Gray lines in the background indicate the collinear blocks within O. pumila and other plant genomes, while red lines highlight syntenic GATA gene pairs.

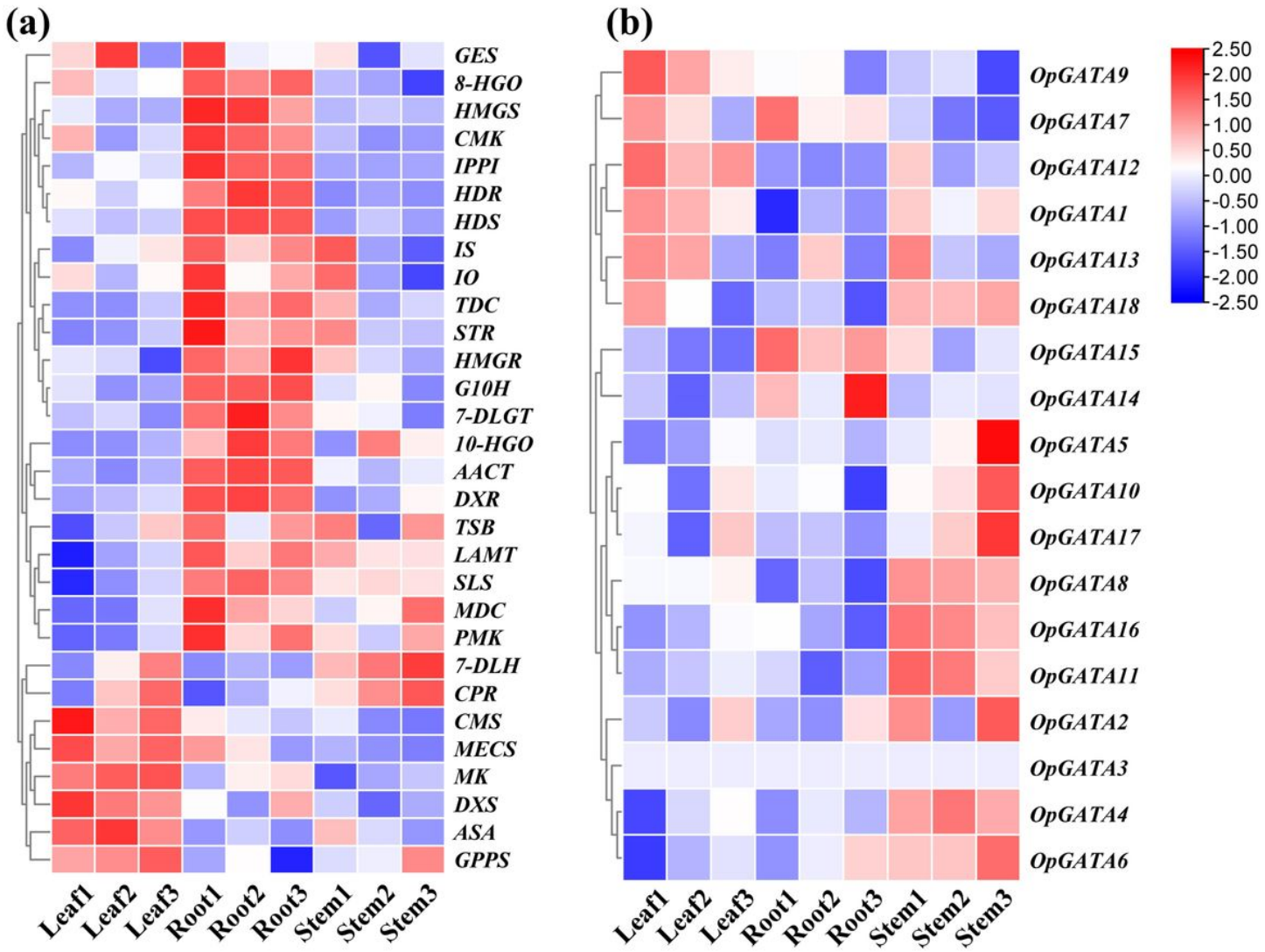

Figure 6

The expression patterns of key enzyme genes and OpGATAs in leaf, root and stem tissues examined by QRT-PCR. The color scale represents relative expression levels from high (red colored) to low (blue color).

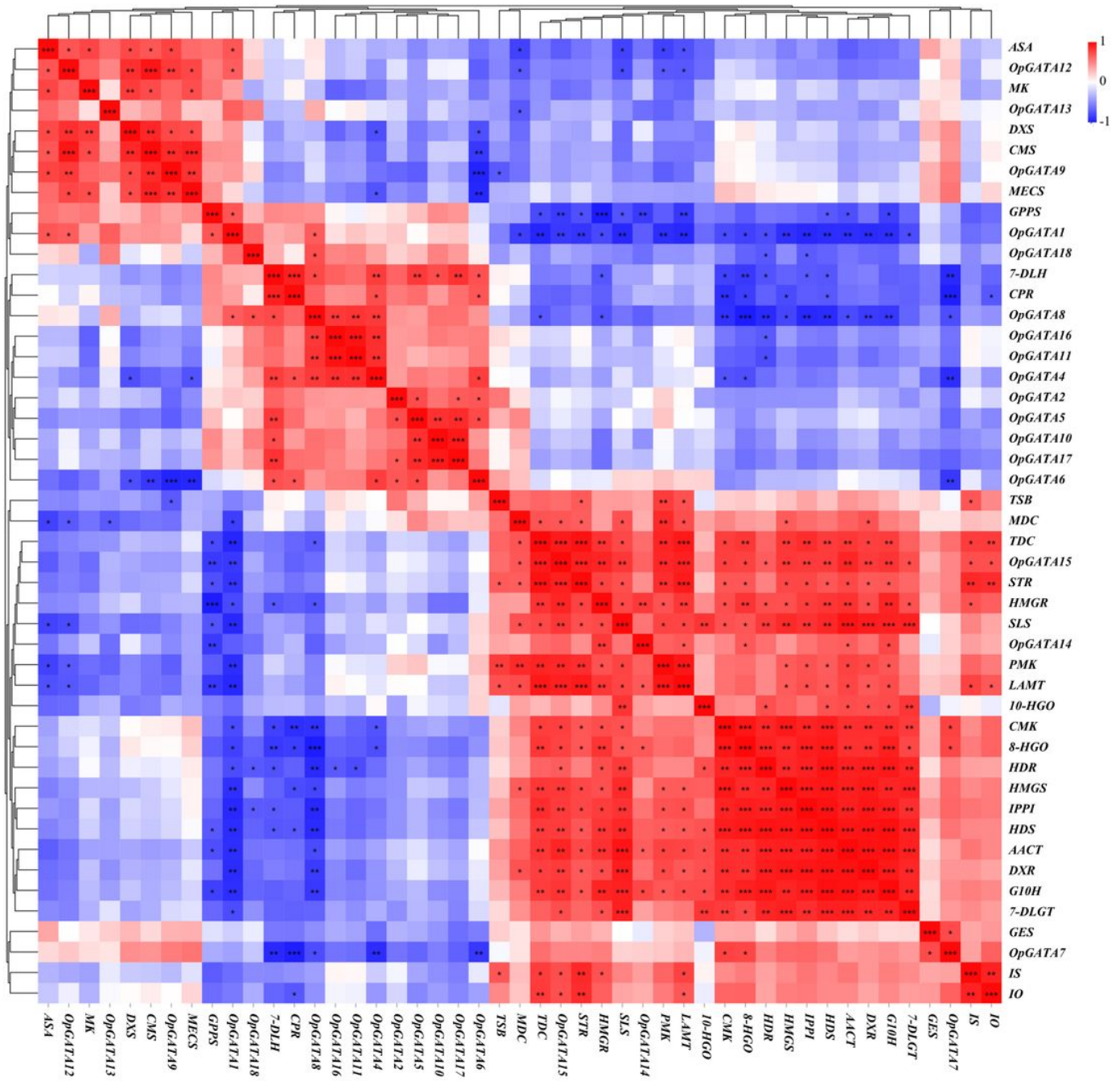

Figure 7

The correlation between the gene expression patterns of OpGATA and key enzyme genes. Red: positively correlated; blue: negatively correlated. *: significant difference (P < 0.05)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx

- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx