Genome-Wide Identification, Evolution, and Expression of GDSL-Type Esterase/Lipase Gene Family in Soybean

Hong-Gang Su1,2†, Xiao-Hong Zhang†, Ting-Ting Wang3, Wen-Liang Wei3, Yan-Xia Wang4, Jun Chen2, Yong-Bin Zhou2, Ming Chen2, You-Zhi Ma2, Zhao-Shi Xu2* and Dong-Hong Min1*

1 College of Life Sciences, College of Agronomy; Northwest A&F University/State Key Laboratory of Crop Stress Biology for Arid Areas, Yangling, China, 2 Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS)/National Key Facility for Crop Gene Resources and Genetic Improvement, Key Laboratory of Biology and Genetic Improvement of Triticeae Crops, Ministry of Agriculture, Beijing, China, 3 College of Agriculture, Yangtze University, Hubei Collaborative Innovation Center for Grain Industry, Beijing, China, 4 Shijiazhuang Academy of Agricultural and Forestry Sciences, Research Center of Wheat Engineering Technology of Hebei, Shijiazhuang, China

GDSL-type esterase/lipase proteins (GELPs) belong to the SGNH hydrolase superfamily and contain a conserved GDSL motif at their N-terminus. GELPs are widely distributed in nature, from microbes to plants, and play crucial roles in growth and development, stress responses and pathogen defense. However, the identification and functional analysis of GELP genes are hardly explored in soybean. This study describes the identification of 194 GELP genes in the soybean genome and their phylogenetic classification into 11 subfamilies (A–K). GmGELP genes are disproportionally distributed on 20 soybean chromosomes. Large-scale WGD/segmental duplication events contribute greatly to the expansion of the soybean GDSL gene family. The Ka/Ks ratios of more than 70% of duplicated gene pairs ranged from 0.1–0.3, indicating that most GmGELP genes were under purifying selection pressure. Gene structure analysis indicate that more than 74% of GmGELP genes are interrupted by 4 introns and composed of 5 exons in their coding regions, and closer homologous genes in the phylogenetic tree often have similar exon-intron organization. Further statistics revealed that approximately 56% of subfamily K members contain more than 4 introns, and about 28% of subfamily I members consist of less than 4 introns. For this reason, the two subfamilies were used to simulate intron gain and loss events, respectively. Furthermore, a new model of intron position distribution was established in current study to explore whether the evolution of multi-gene families resulted from the diversity of gene structure. Finally, RNA-seq data were used to investigate the expression profiles of GmGELP gene under different tissues and multiple abiotic stress treatments. Subsequently, 7 stress-responsive GmGELP genes were selected to verify their expression levels by RT-qPCR, the results were consistent with RNA-seq data. Among 7 GmGELP genes, GmGELP28 was selected for further study owing to clear responses to drought, salt and ABA treatments. Transgenic Arabidopsis thaliana and soybean plants showed drought and
salt tolerant phenotype. Overexpression of GmGELP28 resulted in the changes of several physiological indicators, which allowed plants to adapt adverse conditions. In all, GmGELP28 is a potential candidate gene for improving the salinity and drought tolerance of soybean.

**Keywords:** GELP, expansion, intron gain and loss, gene duplication, expression profiles

**INTRODUCTION**

GDSL-type esterase/lipase protein (GELP, esterase, EC 3.1.1, lipase, EC 3.1.1) is a variety of hydrolytic enzyme (lipolytic enzyme) with broad substrate specificity, which can hydrolyze many kinds of substrates such as thioesters, aryl esters, phospholipids and amino acids (Akoh et al., 2004). GELPs hold unique structural features, possessing a conserved GDSL motif at their N-terminus, which is different from classic lipolytic enzymes containing the conserved motif GxSxG (Upton and Buckley, 1995). GDSL lipases are also named as SGNH hydrolases because of the four invariant important catalytic residues Ser, Gly, Asn and His present in conserved blocks (I, II, III and V), respectively. Meanwhile, the active-site serine of GDSL lipolytic enzymes (block I) as well as Asp and His residues (block V) composed a catalytic triad existing in all enzymes (Upton and Buckley, 1995; Akoh et al., 2004). Since first identified and reported the conserved domain PF00657 in bacteria, several studies have been conducted to research these fascinating lipolytic enzymes, and results demonstrate that GDSL lipases are spread widely among prokaryotes and eukaryotes (Chepyshko et al., 2012). To date, comprehensive studies have been carried out to systematically investigate the GDSL gene family in various species, and it was reported that there are 105, 114, 114, 96, 126 and 130 members of GDSL family in *Arabidopsis thaliana*, *Oryza sativa*, *Brassica rapa*, *Vitis vinifera*, *Populus trichocarpa* and *Sorghum bicolor*, respectively (Volokita et al., 2011; Chepyshko et al., 2012; Lai et al., 2017).

It has been demonstrated that GDSL family members play significant roles in regulating plant growth and development (Ma et al., 2018; An et al., 2019; Ding et al., 2019a; Watkins et al., 2019), organ morphogenesis (Smyth, 2017; Yadav et al., 2017; Zhang et al., 2017), secondary metabolism (Huang et al., 2015), plant immunity (Hong et al., 2008; Kwon et al., 2009; Lee et al., 2009; Kim et al., 2013, 2014; Rajarammohan et al., 2018; Ding et al., 2019b) and biotic and abiotic stresses (Naranjo et al., 2006; Hong et al., 2008; Kim et al., 2008). Very recently, it revealed that DARX1, a new active polysaccharide acetylersterase, regulating the conformation of arabinoxylan and the cross-linking mode with cell wall polymers of cellulose by acting on deacetyllating the side chain of hemicellulose and arabinoxylan, thereby controlling the advanced structure and function of the cell wall. Not only that, DARX1 can also regulate the development of fiber cells and xylem vessels, and change the mechanical strength of rice (Zhang et al., 2019). AtCDEF1 (cuticle destructing factor 1), expressed specifically in mature pollen, was demonstrated to be involved in the pollen-stigma interaction by degrading the cutin on the surface of the stigma in *Arabidopsis*. Beyond that, AtCDEF1 functions in facilitating the emergence of the lateral roots by degrading wall components (Takahashi et al., 2010). *Arabidopsis* EXL4 (extracellular lipase 4) also may function in the pollen-stigma interface to facilitate hydration, which is required for efficient pollen hydration. exl4-1 exhibited significantly reduced competitiveness in pollination and esterase activity compared with wild type (Updegraff et al., 2009). Similarly, EXL6 genes perform a crucial role in pollen development in *Arabidopsis* and *Brassica rapa* L, and BrEXL6 may be an ortholog of *AtEXL6* (Dong et al., 2016). It was reported that an endoplasmic reticulum-localized GDSL lipase, ZmMs30, specifically expressed in maize anthers, which is required for anther cuticle and pollen exine development (An et al., 2019). It was established that overexpression of *AtGDSL1* enhanced *Sclerotinia sclerotiorum* resistance in rapeseed by modulating SA- and JA-dependent pathways, resulting in increased accumulation of phosphatidic acid (PA) and activation of downstream stress response pathways after *Sclerotinia* infection (Ding et al., 2019b). AtLTL1, a novel halotolerance gene in *Arabidopsis*, was demonstrated to be involved in the response to salt stresses in yeast and transgenic plants (Naranjo et al., 2006).

Soybean, a member of the *Leguminosae* sp. family and native to China, is one of the most important global sources of seed protein and oil and rich in various beneficial nutrients, including isoflavones and vitamins. Genetically, soybean is an ancient palaeopolyploid and an outstanding model for the elucidation of the consequences of genome duplication in higher eukaryotes (Schlueter et al., 2007; Lee et al., 2013). Gene duplication can happen in varying degrees of completeness: whole genome duplication (WGD), segmental duplication or tandem repeat (Lyons and Freeling, 2008). WGD had resulted in a striking increase in several angiosperm lineages, such as *Poaceae*, *Solanaeae* and *Fabaceae* (Soltis et al., 2009). In 2010, the drafted genome sequencing of cultivar Williams 82 was completed, which set a basis to identify the soybean gene family at genome-level and also research the evolution of soybean genes in plants (Schmutz et al., 2010). Although the GDSL gene family have been analyzed in several plant species, no systematical investigation are conducted.

**Abbreviations:** ABA, Abscisic acid; ABRE, ABA-responsive elements; CDS, Coding sequence length; FPKM, Fragments per kilobase of transcript per million mapped reads; FW, Fresh weight; GELP, GDSL-type esterase/lipase proteins; GSDS, Gene structure display server; HMM, Hidden Markov model; JA, Jasmonate; Ka, Nonsynonymous substitution rate; Ks, Synonymous substitution rate; LTR, Low-temperature responsive; MBS, MYB binding site; MDA, Malonaldehyde; MEJA, Methyl jasmonate; NCBI, National Center for Biotechnology Information; PF, Pfam; RT-qPCR, Real-time quantitative polymerase chain reaction; SA, Salicylic acid; WGD, Whole genome duplication.
in soybean. Given the significance of this gene family, an accurate genome-wide identification was performed in the current study. A comprehensive analysis of the GDSL gene family, including phylogenetic relationships, chromosomal location, gene duplication as well as expression profiles was performed to illuminate GELP’s evolutionary and functional characteristics. The insights gained from this study may be of assistance to better understand the extension and evolution patterns of GELP genes in the soybean genome, and highlight their function in regulating growth and development as well as abiotic stresses.

**MATERIALS AND METHODS**

**Plant Materials, Growth Conditions, and Stress Treatments**

Soybean cultivars Zhonghuang 39 and Williams 82 were used to gene expression patterns and soybean hairy roots, respectively. Arabidopsis Col-0 plants were used for phenotypic assays. All the plants were grown in a greenhouse at 25°C with a photoperiod of 16-h light/8-h dark. One-week-old soybean seedlings were used for drought, salt and exogenous ABA treatments as previously described (Li B. et al., 2019). The leaves of seedlings under drought, salt and ABA treatments were collected at 0, 0.5, 1, 2, 4, 8, 12 and 24 h after treatments. All collected samples were frozen immediately in liquid nitrogen and then stored at -80°C for subsequent analysis.

**Identification of GELP Genes in Soybean**

The identification of soybean GDSL gene family was performed according to the method described by Rao et al. with some revisions (Rao et al., 2010). Firstly, the soybean protein sequences were downloaded from the Phytozome database to build a local protein database (Goodstein et al., 2012). Then, the Hidden Markov model (HMM) profile of the GDSL conserved domain (PF00657) was used for the purpose of scanning the local database using the local BLASTP program (with E-value < e-20), and then all obtained sequences were aligned and used to build a soybean-specific HMM profile of GDSL domain using the hmmbuild program from the HMMER v3 (Finn et al., 2011). Then, the new soybean-specific HMM profile was used to search for GELP members from the local protein database using hmmsearch program. All candidate proteins were submitted to NCBI Batch CD-search and SMART databases for examining the presence of the GDSL conserved domain (Marchler-Bauer et al., 2015). After validation, all candidate genes encoding the GDSL domain were identified in the soybean genome (Supplementary Table S1).

**Multiple Sequence Alignment and Phylogenetic Analysis**

Multiple sequence alignment was performed using MAFFT software with the E-INS-I option version 7. Non-aligned regions were removed with Gblocks version 0.91b. LG+I+G, LG+I+G and JTT+G models with a 4-categories GAMMA distribution were used in the phylogenetic trees of Arabidopsis and soybean, subfamily I and subfamily K GELP proteins, respectively, they were identified using ProTest. Possion model and pairwise deletion was used in the phylogenetic tree of soybean GELP proteins. All the phylogenetic trees were reconstructed using MEGA version 7. A bootstrap test with 1000 replications was used to determine the statistical reliability of the phylogenetic trees. The similarities of eight reported AtGELP proteins and each corresponding homologous protein in soybean are shown in Supplementary Table S2.

**Chromosomal Location and Gene Duplication**

All GELP genes were mapped on the 20 chromosomes of soybean using positional information acquired from the phytozome database by TBtools software (Chen et al., 2018). Distribution frequency of GELP genes on 20 soybean chromosomes was shown in Supplementary Figure S1. Segmental and tandem duplication events were determined as previously described (Wang et al., 2016; Fan et al., 2019). Briefly, a pair of duplicated GmGELP genes were defined when the alignments covered > 80% of the longer gene and the aligned region had an identity > 80% at the nucleotide level. Tandem duplication event was demarcated based on the chromosomal location of each of duplicated gene. Related syntenic blocks and duplicated gene pairs in soybean were obtained and visualized using TBtools software. The nonsynonymous substitution rate (Ka), synonymous substitution rate (Ks) and the Ka/Ks ratio between paralogous gene pairs were evaluated using a comparative synteny map within the soybean genome (Glycine max Wm82.a2.v1) by TBtools software, and detailed information of duplicated gene pairs can be found in Supplementary Table S3.

**Gene Structure, Intron Pattern and Conserved Motifs**

The gene structure information and intron insertion sites of GmGELP genes were obtained from the phytozome database and visualized using the GDS (Gene Structure Display Server) program and DNAMAN software, respectively (Hu et al., 2015). The schematic of conserved motifs of all members is presented in Supplementary Figure S2 based on the analysis of online MEME program, and the maximum number of motifs was set to twenty (Bailey et al., 2009).

**Expression Pattern Detected by Transcriptome Data**

To study the expression of GmGELP genes in different tissues and organs, transcriptome data extracted from a public soybean database were used to investigate the different expression of GmGELP (Severin et al., 2010; Du et al., 2018). RNA-seq data of various abiotic stresses were extracted from our previous research to study the expression of GELP genes under salt (4 h), drought (4 h) and ABA treatments (3 h) (Shi et al., 2018). FPKM (fragments per kilobase of transcript per million mapped reads) values of soybean GELP genes in different tissues and under abiotic stress treatments (drought, NaCl and ABA) were shown in Supplementary Tables S4–S7. A heatmap showing tissue-specific expression profiles was generated using the log2-transformed (FPKM + 1) values of GmGELP genes, and the expression level...
fold change relative to the normal control was used to heatmaps for abiotic stress conditions. Moreover, differentially expressed *GmGELP* genes with significant level (fold-change $\geq 2$ and $p$-value $\leq 0.01$) were used to analyze the expression of *GmGELP* genes under three abiotic stresses. Finally, visualization of the expression levels of *GmGELP* genes was accomplished using TBtools software.

**RNA Extraction and RT-qPCR**

Total RNA was isolated from soybean leaves subjected to the multiple stress treatments using RNA plant extraction kit (Zhuangmeng, Beijing, China) according to the manufacturer’s instruction. Approximately 2 μg of purified total RNA from each sample was used in reverse transcription using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China) and stored at $-20^\circ$C. RT-qPCR analysis was accomplished using PerfectStart Green qPCR SuperMix (TransGen Biotech, Beijing, China) and an Applied Biosystems 7500 Real-Time PCR System (ThermoFisher, Beijing, China). Data analysis was conducted using the $2^{-\Delta\Delta Ct}$ method. Three technical replicates were performed for each of the three biological replicates. The primers used for RT-qPCR in this study were given in Supplementary Table S8.

**Identification of Cis-Elements in the Promoters of 7 GmGELP Genes**

For cis-elements analysis, 2 kb sequences upstream from the start codons of 7 *GmGELP* genes were downloaded from the Phytozome database and analyzed using the PlantCARE database.

**Generation of Transgenic Arabidopsis Plants**

The *GmGELP28* coding region was amplified and inserted into the pCAMBIA1302 vector under the control of the CaMV35S promoter. The resulting construct pCAMBIA1302-*GmGELP28* was transformed into *Arabidopsis* Col-0 plants using the floral dip method. The harvested seeds were surface sterilized with sodium hypochlorite and germinated on 1/2-strength MS media. Three *GmGELP28* transgenic lines were selected for further study.

**Agrobacterium rhizogenes-Mediated Transformation of Soybean Hairy Roots**

To generate the pCAMBIA3301-*GmGELP28* overexpression vector, the coding region of *GmGELP28* was amplified from Williams 82 cDNA, and the PCR product was then ligated into the pCAMBIA3301 vector under the control of the CaMV35S promoter. The recombinant vector was transformed into soybean hairy roots by *Agrobacterium rhizogenes*-mediated transformation following the protocol described previously (Kereszt et al., 2007; Su et al., 2019). After verification, positive soybean hair roots were used to abiotic tolerance assays, and 8 plants per pot were used in three to six biological replicates.

**The Abiotic Stress Responses of Transgenic Arabidopsis Plants and Soybean Hairy Roots**

For drought tolerance assays, 3-week-old *Arabidopsis* plants in soil were subjected to drought treatment by halting watering. The plants were photographed when differences in phenotype were observed. Watering was reinitiated after 2 weeks for plant growth recovery. About a week later, the plants were photographed again and calculated the survival rates. In addition, salt stress was applied by the addition of 250 mM NaCl solution to the soil-grown *Arabidopsis* plants for 1 week, and control plants were grown under normal conditions.

Soybean drought tolerance assays were run as described above. In brief, 2-week-old soybean plants with transgenic hairy roots were subjected to dehydration for 2 weeks and rewatered for 3 days. With respect to salt treatment, 2-week-old soybean plants with transgenic hairy roots were treated with 200 mM NaCl solution for 3 days.

**The Quantification of MDA, Proline, Chlorophyll and H$_2$O$_2$**

The contents of malonaldehyde (MDA), Proline, chlorophyll and H$_2$O$_2$ were detected according to the instructions of the corresponding measurement kit (Cominbio, Suzhou, China). Three replicates were included per measurement. The leaves in each pot were mixed and sampled. The contents of MDA, Proline and chlorophyll of *Arabidopsis* samples were detected after drought treatment for 2 weeks and salt treatment for 1 week. As to soybean samples, the contents of MDA, Proline and H$_2$O$_2$ were detected after drought treatment for 10 days and salt treatment for 2 days when the leaves were slightly wilted. Each experiment was performed in triplicate.

**RESULTS**

**Identification of the GELP Genes in Soybean**

A total of 194 candidate genes encoding the GDSL domain were identified in the soybean genome (*Glycine max* Wm82.a2.v1). They were named as *GmGELP1-GmGELP194* according to their chromosomal locations. Detailed information about these predicted genes is summarized in Supplementary Table S1. The protein lengths of coding sequences range from 266 (GmGELP90) to 460 amino acids (GmGELP77), with the average sequence length of 346 amino acids. The lengths of *GmGELP* genes in genome vary from 974 to 14351 bp.

**Phylogenetic Analysis of GmGELP Genes**

As described in a previous study (Lai et al., 2017), to better elucidate the phylogenetic relationship among *Arabidopsis* and soybean GELP genes, the homology sites of 194 GmGELP and 104 AtGELP proteins (except AtGELP29,
which is absent from *Arabidopsis* (Figure 1). Combining MEME analysis and intron number statistics, 298 GELP members were further divided into 11 subfamilies containing 3 to 87 members each. Confusingly, *AtGELP34* is not clustered with any of the other members. In most of these subfamilies, the same subfamily also contains soybean GELP genes and *Arabidopsis* GELP genes, suggesting possible conservation of function within dicot species. However, all 18 members of subfamily K are from soybean, implying that *GmGELP* genes in subfamily K might have occurred very early, before the divergery of soybean and *Arabidopsis*, which may be potential candidate genes to distinguish soybean and *Arabidopsis*. Among the 8 AtGELP proteins (*AtESM1*, *AtGLIP1*, *AtGLIP2*, EXL4, CDEF1, SFAR1-5, LAE and AtLTL1) whose functions had been reported, 6 genes were found to share more than 38% similarity with their homologous proteins, and gathered together in the same subfamily of phylogenetic tree (Supplementary Table S2). However, it’s worth noting that there is no clear association between biological function and specific clades of phylogenetic tree.
Chromosomal Location, Motif Identification, and Duplication Events Analysis

As can be seen from Figure 2 and Supplementary Figure S1, 194 GmGELP genes are universally and unevenly distributed on all 20 soybean chromosomes, similarly, to the distribution characteristics of GELP genes in the rice and Arabidopsis genomes found in previous studies (Dong et al., 2016; Lai et al., 2017). Three chromosomes contain approximately 29% (56 out of 194) of GmGELP genes: chromosome 13 (19 genes, 9.8%), chromosome 15 (18 genes, 9.3%) and chromosome 19 (19 genes, 9.8%), whereas only 5, 5, 4, 2 and 3 GmGELP genes are located on chromosomes 1, 8, 9, 12 and 20, respectively. Unsurprisingly, the majority of GmGELP genes are located on chromosome ends, which is confirmed by a prior report describing that about 78% of the predicted genes are located on chromosome ends (Schmutz et al., 2010).

A total of 20 conserved motifs were discovered from 198 GmGELP proteins (with E-value below 9.4E-159) and displayed in Supplementary Figure S2. Motifs 1, 4, 5 and 8 represent the conserved blocks I, II, III and V of GDSL family, respectively, which are present in almost all 198 proteins (Figure 3). In addition, other 6 well-conserved motifs (motifs 2, 3, 6, 9, 10 and 12) were detected in more than 60% of GmGELP proteins (Supplementary Figures S2, S3). However, several motifs are specific to individual subfamilies in the phylogenetic tree (Supplementary Figure S2). For instance, subfamily F possesses specific motifs 16 and 18, while motifs 17,19 and 20 are only detected in subfamily K. These results indicate that GmGELP members of the same subfamily often have similar motif composition, which is consistent with their phylogenetic relationship.

To elucidate the amplification mechanism of GmGELPs, we conducted collinearity alignment within the soybean genome, indicating that a total of 139 GmGELP genes are located within syntenic blocks on soybean chromosomes (Figure 4 and Supplementary Table S3). Tandem duplication is defined according to the methods previously reported (Holub, 2001; Li Z. et al., 2019). Statistical results showed that approximately 11% (22 out of 194) of GmGELP genes were found to derive from tandem duplication events (Supplementary Table S3), and 9 tandemly duplicated GmGELP sets contain 2-5 GmGELP genes. Among them, 7 tandemly duplicated GmGELP sets contain 2 members (GmGELP52/53, GmGELP66/66, GmGELP96/97, GmGELP107/108, GmGELP130/131, GmGELP140/141 and GmGELP142/143), 1 tandemly duplicated GmGELP set contain 3 members (GmGELP55, 56 and 57), and 1 tandemly duplicated GmGELP set is involved in 5 members (GmGELP176, 177, 178, 179 and 180). Beside tandem duplication events, we further observed that up to 71% (137 out of 194) of GmGELP genes had participated in WGD/segmental duplication, which is much higher than the 21% in Arabidopsis genome (22 out of 105). In summary, these results suggest that WGD/segmental duplication is the main driving force for the large expansion of GELP genes in the soybean genome.

To explore the evolutionary forces acting on the 198 GmGELP genes, we evaluated the non-synonymous/synonymous substitution ratio (Ka/Ks) for each duplicated gene pair (Supplementary Table S3). The results suggest that the Ka/Ks ratios of GmGELP gene pairs are commonly less than 1 and range from 0.08 to 1.66 with an average of 0.30. Figure 5 shows an overview of the distribution frequency of the Ka/Ks ratios, results found that the Ka/Ks ratios of more than 70% of duplicated gene pairs ranged from 0.1 to 0.3, indicating that these duplicated GmGELP genes were under purifying selection pressure.

Intron Loss and Gain Events and Gene Structure

The exon-intron structure, intron size and intron position showed high conservation among soybean, which can be used to derive phylogenetic relationship (Fedorov et al., 2002; Babenko et al., 2004; Roy and Penny, 2007). A striking feature of plant GELP genes is their structure of five exons and four introns. For instance, approximately 67.6% and 49.1% of GELP members contain 4 introns in Arabidopsis and rice, respectively (Ling et al., 2006; Chepyshko et al., 2012). The intron numbers of soybean GELP genes were calculated in our study (Supplementary Figure S4). The results demonstrate that intron numbers vary from 1 to 6, and up to 74.2% (143 out of 194) of GmGELP genes harbor 4 introns. All genes contain at least 1 intron, and the highest number of introns exist in the GmGELP_112 gene (7 exons and 6 introns) (Supplementary Table S1). We further analyzed the number of introns in each subfamily (Table 1). What stands out in Table 1 is that up to 56% (10 out of 18) genes of subfamily K possess 5 introns, and about 28% (14 out of 50) genes of subfamily I contain less than 4 introns. For this reason, subfamilies K and I were used to simulate the events of intron gain and loss, respectively.

Prior studies showed that evolution of multi-gene families resulted in gene structure diversity (Luo et al., 2018). The exon-intron organization of 50 and 18 GmGELP genes from subfamilies K and I are shown in Figures 6A,B, respectively. Generally, adjacent members of the same branch have similar arrangements in terms of intron number and exon length, suggesting that exon-intron structure and phylogenetic tree are highly correlated. However, a minority of homologous gene pairs, for example, GmGELP94/112, presents slight differences in intron number and exon length. In fact, structural divergences have been widespread in duplicated genes, which could generate functionally distinct paralogs, whereas it is still unclear the mechanism that how the change of gene structure have promoted the generation of functionally distinct paralogs (Xu et al., 2012).

To further explain the reason for the variation in intron numbers of 68 genes in subfamilies K and I, 8 subfamily I and 5 subfamily K GmGELP genes were selected for analysis of intron position. Among them, three pairs of duplicated genes were included among them: GmGELP94/112, GmGELP125/164 and GmGELP16/115. The intron positions of all 13 selected gene pairs are commonly less than 1 and range from 0.08 to 1.66 with an average of 0.30. Figure 5 shows an overview of the distribution frequency of the Ka/Ks ratios, results found that the Ka/Ks ratios of more than 70% of duplicated gene pairs ranged from 0.1 to 0.3, indicating that these duplicated GmGELP genes were under purifying selection pressure.
FIGURE 2 | Chromosomal locations of the 194 GmGELP genes. The tandemly duplicated genes are labeled in purple boxes. The chromosome numbers are marked left of each chromosome, centromeric regions are indicated by ellipses. The bar located on the left side shows the size of chromosome in megabases.
the catalytic sites. However, none of the 13 GmGELP proteins contain the GDSL motif in N-terminal, and two lack the core short signature, indicating the sequence GDSL seems to be relatively conservative among GmGELP proteins, which had been confirmed by previous reports that GDSL/V motif existed in several GELP proteins of Arabidopsis and Tanacetum cinerariifolium (Gao et al., 2017). Four intron gene structure was fixated as the basic form in prior study (Volokita et al., 2011). According to the method just described, the intron positions 3, 5, 7, 11 and 4, 6, 10, 12 were defined as the basic forms of subfamilies K and I in our study. As can be seen from Figure 7A, intron positions are similar within intra-subfamily, but totally different between subfamilies K and I. Among three duplicated gene pairs, GmGELP16 and GmGELP115 show similar intron positions and contain only 2 introns, while GmGELP94 and GmGELP112 display different intron characteristics. Specifically, GmGELP94 and GmGELP112 contain 4 shared intron positions (the intron positions 4, 6, 10 and 12), while the intron positions 1 and 2 are unique to GmGELP112. Similarly, the intron position 13 exists only in GmGELP125. Based on the results of the above analysis, we speculate that the intron positions 1, 2 and 13 are produced by independent intron gain events, since extra introns exist in only a few genes. Conversely, the intron positions 4 and 10 are absent in GmGELP16 and GmGELP115, which supports the hypothesis that reduced introns are the result of later intron loss events. Moreover, we consider these events occur frequently in several genes of subfamilies K and I.

Furthermore, to reveal the ancestral gene structure of subfamilies K and I, a new model was constructed based on the 8 conserved intron positions (3, 4, 5, 6, 7, 10, 11 and 12) (Figure 7B). We presume that the differences in intron position mainly resulted from gene structure diversity, or, alternatively, from events of intron loss and gain.

**Gene Expression Patterns Analysis**

The RNA-seq data of GmGELP genes in different tissues and development stages can give reference to molecular mechanisms of plant growth and development (Severin et al., 2010). The transcription levels of a total of 168 available genes were presented in Figure 8. The expression levels of 7 genes, namely GmGELP14, 29, 81, 125, 133, 152 and 164, were not detectable, which led us to speculate that they were not expressed in the examined tissue/stage/condition, or that they were pseudogenes. A considerable number of members showed very low or no transcriptional abundance, including GmGELP4, 7, 20, 54, 61, 76, 83, 87, 140, 142, 143, 144, 151 and 181. However, a small portion of members were expressed constitutionally in soybean. For instance, GmGELP22, GmGELP149 and GmGELP186 exhibited high transcription levels in most tissues and organs throughout the growth period. Meanwhile, large amounts of genes showed tissue-specific expression, such as GmGELP10, 79, 115, 124 and 180 showed preferential expression in young leaves, flowers, pods and pod shells, while being expressed at low levels in roots and nodules, suggesting they may play distinct roles in special tissues or developmental stages. To explore the functional redundancy and differentiation of soybean GELP homologous gene pairs, the expression patterns of several duplicated genes were also investigated (Figure 8). Results indicated that some duplicated gene pairs showed similar expression patterns (for instance, GmGELP22, GmGELP149 and GmGELP186), whereas some homologous genes exhibited completely different or converse expression patterns (such as GmGELP65/66, GmGELP70/137, GmGELP87/154, GmGELP109/129 and GmGELP111/127), suggesting that they may have undergone functional difference. In addition, we examined the expression level of GmGELP20, a homologous gene of Arabidopsis GLP1, indicating GmGELP20 was expressed at extremely low levels in all tissues, while AtGLP1 showed a higher transcript levels in seedlings, stems and roots (Kwon et al., 2009). In turn, GLP2, which was highly homologous with Arabidopsis GLP1, was expressed only in seedlings, roots and stems (Lee et al., 2009), while its homologous gene GELP62 in soybean had extraordinarily weak to no expression abundance in all tissues (Lee et al., 2009). The overall expression data analysis
indicated that GmGELP genes exhibited great disparities in abundance among different tissues, which might play significant effects in accommodating different physiological processes.

Extensive studies have shown that the expression of multiple GDSL family members can be induced by various abiotic stresses in some species (Naranjo et al., 2006; Hong et al., 2008; Huang et al., 2015; Li et al., 2017). To reveal the expression of GmGELP genes in response to abiotic stress, the relative expression abundances of all GmGELP genes in young leaves were investigated under three abiotic stress conditions (drought, salt and ABA). RNA-seq data indicated a total of 66, 30 and 28 GmGELP genes responded to drought, NaCl and ABA treatments, respectively (Figure 9). Overall, 89.4% (59 out of 66), 86.7% (26 out of 30) and 71.4% (20 out of 28) GmGELP genes were down-regulated under the conditions of drought, NaCl and exogenous ABA treatments, respectively. Meanwhile only 10.6% (7 out of 66), 13.3% (4 out of 30) and 28.6% (8 out of 28) GmGELP genes were up-regulated under the same conditions. Among them, several genes, such as GmGELP1, GmGELP75, GmGELP11, GmGELP43 and GmGELP8 were
specifically expressed under drought stress, GmGELP17, GmGELP102 and GmGELP110 were specifically expressed under NaCl stress, as well as GmGELP123 and GmGELP136 were significantly up-regulated under exogenous ABA treatment (Supplementary Tables S5–S7). Another important finding is that 10 GmGELP genes were down-regulated under all three types of abiotic stresses, including GmGELP25, GmGELP33, GmGELP79, GmGELP121, GmGELP148, GmGELP163, GmGELP172, GmGELP179, GmGELP187 and GmGELP188, whereas GmGELP28 and GmGELP74 were up-regulated under the same conditions. Interestingly, most homologous gene pairs showed a similar expression pattern under NaCl and drought stresses such as GmGELP33/188, GmGELP78/172 and GmGELP97/179, indicating that they might perform similar physiological functions. Nevertheless, several duplicated gene pairs displayed different expression patterns. For example, GmGELP178 was significantly up-regulated after ABA treatment, while GmGELP179 was down-regulated, suggesting there may have arisen functional difference in these genes.

Based on the results of RNA-seq, we observed that several genes were significantly induced by abiotic stresses. For this reason, 7 GmGELP genes were selected to confirm their expression patterns by RT-qPCR (Figure 10). Unsurprisingly, RT-qPCR results were consistent with RNA-seq data. For instance, GmGELP28 showed a similar expression pattern under three abiotic stress treatments. Under drought stress, GmGELP163 showed a slightly down-regulated expression (Figure 10A). GmGELP33 and GmGELP163 showed remarkable down-regulation after NaCl treatment (Figure 10B). Both GmGELP28 and GmGELP123 were significantly up-regulated at 1 h after ABA treatment (Figure 10C). The expression profiles of these stress-induced GmGELP genes provide valuable information for further revealing their roles under different abiotic stresses.

The cis-elements in promoter regions play important roles in regulating gene transcription and abiotic stress responses. Therefore, 2 kb sequences upstream from the start codons of 7 GmGELP genes were downloaded and analyzed using the PlantCARE database. The number of 6 abiotic stress response elements were counted and displayed in Table 2, including ABA-responsive elements (ABRE), drought-inducible elements (MBS, MYB binding site), low-temperature responsive elements (LTR), MEJA-responsive elements (CGTCA-motif), SA-responsive elements (TCA-element) and defense and stress responsive elements (TC-rich repeats). Among them, ABRE, MBS and CGTCA-motif were detected in almost every promoter region of the GmGELP genes. Note, the correlation between cis-elements and responses to abiotic stresses of genes needs further experimental validation. Nevertheless, cis-element analysis indicated that GmGELP genes might respond abiotic stresses.

Arabidopsis Plants Expressing GmGELP28 Were More Tolerant of Drought and Salt Stresses Than Wild Type

Among 7 GmGELP genes, GmGELP28 was selected to validate the role in abiotic stresses owing to significantly up-regulated

---

**TABLE 1** | Intron numbers in each subfamily of GmGELP gene family.

| Subfamily | 1 Intron | 2 Intron | 3 Intron | 4 Intron | 5 Intron | 6 Intron | Total |
|-----------|----------|----------|----------|----------|----------|----------|-------|
| A         | 2        | 3        | 1        | 29       | 3        | 0        | 38    |
| B         | 0        | 0        | 0        | 11       | 1        | 0        | 12    |
| C         | 0        | 0        | 0        | 3        | 0        | 0        | 3     |
| D         | 0        | 0        | 0        | 18       | 2        | 0        | 20    |
| E         | 0        | 0        | 0        | 1        | 1        | 0        | 2     |
| F         | 0        | 0        | 1        | 27       | 0        | 0        | 28    |
| G         | 0        | 0        | 0        | 11       | 0        | 0        | 11    |
| H         | 0        | 0        | 0        | 2        | 0        | 0        | 2     |
| I         | 0        | 12       | 2        | 33       | 2        | 1        | 50    |
| J         | 4        | 0        | 4        | 2        | 0        | 0        | 10    |
| K         | 0        | 1        | 0        | 7        | 10       | 0        | 18    |
| Total     | 6        | 16       | 8        | 144      | 19       | 1        | 194   |
FIGURE 6 | (A, B) Phylogenetic tree and gene structure analysis of GmGELPs from subfamilies I and K, respectively. Exon and intron are indicated by azure round-cornered rectangle and purple rectangle, respectively. Exon lengths are displayed proportionally and introns are scaled to the same length. 8 and 5 subfamily I and K genes for further study are marked with blue and red, respectively.

expression levels by all tested treatments. Three independent GmGELP28-overexpression Arabidopsis lines (OE lines) were used to investigate the performance in drought and salt stresses. The results revealed that no visible differences were observed between WT lines and OE lines under normal growth conditions (Figure 11A). But after withholding water for 2 weeks, WT plants...
exhibited seriously hypersensitive symptoms to drought and NaCl treatments, and only approximately 67% and 38% survival rates were observed upon rewatering after drought and NaCl treatments, respectively (Figure 11B). The contents of MDA, proline and chlorophyll were measured under both normal and abiotic stress conditions. Compared with WT plants, OE lines showed significantly higher contents of chlorophyll and proline in their leaves (Figures 11C,D), and featured a significantly lower MDA content (Figure 11E).

**GmGELP28 Improves Drought and Salt Stress Tolerance in Transgenic Soybean Hairy Roots**

Similarly, the drought and salt tolerance tests were performed in *Agrobacterium rhizogenes*-mediated soybean hair roots. RT-qPCR analysis showed that the expression level of *GmGELP28* in transgenic plants was significantly higher than in empty vector control plants. For drought tolerance assays, 2-weeks
soybean plants were subjected to drought stress. After 2 weeks of water deprivation, empty vector control plants exhibited severe water loss and significant withering relative to transgenic plants (Figure 12A). In addition, significantly more dried leaves were observed in empty vector control plants compared with transgenic plants after rewatering (Figure 12A). Under salt treatment, all empty vector control plants were died. In contrast, a large proportion of leaves in transgenic plants were green at this time point (Figure 12B). The contents of MDA, proline and H$_2$O$_2$ were important parameters involved with the tolerance to abiotic stresses. Under normal growth conditions, no significant differences in the contents of MDA, proline and H$_2$O$_2$ were observed. However, transgenic plants showed higher proline content and lower MDA and H$_2$O$_2$ contents than empty vector control plants when subjected to drought and salt treatments (Figures 12C–E).

**DISCUSSION**

GDSL esterase/lipase proteins (GELP) are a subfamily of lipolytic enzymes that have been discovered relatively recently (Chepyshko et al., 2012; Dong et al., 2016), and descriptions of the GDSL gene family and function have been identified in only a few plants (Dong et al., 2016; Cao et al., 2018). In fact, to the best of our knowledge, this is the first study to identify and characterize the soybean GDSL gene family. This paper set out to identify GELP members, and a search for GELP genes in soybean genome resulted in the identification of 194 members. The study aims to explore the phylogeny and evolution of the GDSL gene family, and to generate fresh insight into the different gene structures of GmGELPs. On the other hand, the study also systematically investigated the expression patterns of GmGELP members in different tissues and under drought, NaCl and ABA stress treatments.

A comparison of the number of GELP genes in soybean with several sequenced genomes displays that the GDSL gene family have a relatively large number (Chepyshko et al., 2012; Dong et al., 2016; Lai et al., 2017). Our study suggests the members of GDSL gene family are 1.86-times (194/104) more abundant in soybean than Arabidopsis. In accordance with the present results, a previous study demonstrated that the number of predicted protein-coding genes in the soybean genome was 70% higher than in Arabidopsis (Schmutz et al., 2010). Another significant feature of some GDSL lipases is their clustered distribution on chromosomes. In Arabidopsis, there were 16 cases of 2-9 genes arranged in tandem, in which one tandemly duplicated AtGELP gene sets contained 9 members (AtGELP6 to AtGELP14) on chromosome 1 (Lai et al., 2017). In rice, approximately 47% GELP genes (54 out of 114) were closely arranged on chromosomes, comprising 17 clusters, in which closely linked genes were adjacent or isolated by 1-4 genes (Chepyshko et al., 2012). Genome-wide duplication is a large-scale process of gene multiplication at the chromosomal level, which has developed the architecture and function of many higher eukaryotic genomes, and leading to highly duplicated genomes with about 75% of genes existing in multiple copies (Schmutz et al., 2010; Lee et al., 2013). Genome-wide duplication also plays a vital role in the expansion of some families, which generates the largest number of duplicated genes. For example, the plant MAPK/MAPKK gene family was more likely to be expanded through whole genome or large segmental
duplication events (Hamel et al., 2006). Consistent with this, 71% (137 out of 194) of GDSL family members occurred as WGD/segmental duplication within the soybean genome, which is the main driving force for the expansion of the soybean GDSL gene family.

Extensive loss and insertion of introns occurred during the evolution of eukaryotes, and gene duplication accelerated this process (Roy and Penny, 2007). However, the evolution of intron in duplicated genes seems to be of particular interest. For this purpose, a parsimonious reconstruction of the evolution for introns in subfamilies K and I was conducted to investigate the intron dynamics in duplicated genes on genome scale. Our results show a strong non-uniform distribution between subfamilies K and I. In addition, most intron positions are shared in duplicated genes, which reflects evolutionary conservation (Figure 7). In agreement

**FIGURE 9** | (A–C) Hierarchical clustering of expression profiles of 66, 30, and 28 drought-, NaCl- as well as ABA-responsive genes, respectively. Red and blue represent up-regulated and down-regulated expression, respectively. The color intensity indicates the expression level fold change compared to the normal control.
FIGURE 10 | (A–C) Hierarchical clustering of expression profiles of 66, 30, and 28 (A–C) RT-qPCR analysis of 7 soybean GmGELP genes under abiotic stress treatments (drought, NaCl, and ABA, respectively). The actin gene was used as an internal control. The mean expression value was calculated from three biological replicates.

with previous studies suggesting that the high conservation of intron position and phase was found more widely across angiosperms (Roy and Penny, 2007; Schmutz et al., 2010). The gene structure of more than 600 GDSL lipases from different plants had been analyzed, and the relative location of 6 conservative introns in the three subfamilies were found to be significantly different: intron positions 1 and 6 were present in all three subfamilies, intron position 5 was conserved in subfamilies A and B, while intron positions 2, 3 and 4 were specific to genes of subfamilies A, B and C, respectively (Volokita et al., 2011). However, the results of the current study do not support previous research. No conserved intron positions were found between subfamilies K and I, suggesting that the relative locations of the 13 introns in two subfamilies are significantly different, and further analysis reveal that the extra introns exist in only one or a few genes, which maybe generate by later events of intron gain (Figures 6, 7). Intron gain and loss have continually occurred in genome evolution at rather low rate, whereas the influences to structure divergence and functional differentiation are significant (Xu et al., 2012). Previous study show an excess of intron loss over intron gain in rice and Arabidopsis at low rate (Roy and Penny, 2007). A limitation of this study is that only 13 genes were used to analyze gene structure, from which the obtained conclusions cannot fully elucidate the evolutionary characteristics of introns. Therefore, there is abundant room for further progress in determining the mechanism of intron evolution.

It is well recognized that the expression of genes usually reflects their potential functions. The expression patterns of GELP genes in different tissues and abiotic stress conditions have been studied in several species, indicating the expression of a large number of GELP members can be induced by hormone, chemical, environmental stress as well as pathogen infection (Naranjo et al., 2006; Hong et al., 2008; Kim et al., 2008, 2014; Kwon et al., 2009; Chepyshko et al., 2012; Huang et al., 2015). For example, in rice, some GELP genes showed constitutive expression or nearly constitutive expression, with relatively high
transcription levels in most tissues and organs during vegetative growth phase, while very weak expression abundances were detected in some reproductive organs, such as anthers and seeds (Chepyshko et al., 2012). Our results matched those observed in earlier studies, some GmGELP genes had only background expression under normal conditions, yet they were notably induced by multiple abiotic stress treatments. For instance, GmGELP8, 17, 28 and 74 were only slightly expressed in normal conditions, but significantly higher expression levels were observed after drought treatment. In addition, it was noted that GmGELP28 was found to show a relatively high expression in three abiotic stress treatments. It is well-known that gene duplication is a main source of functional differentiation, which greatly increases functional diversity and improves the adaptability of species to the environment. Duplication events are followed by gene diversity and loss, and gene loss is the common fate of most duplicated genes (Schmutz et al., 2010), while functionally different duplicated genes are more likely to be preserved during evolution (Schlueter et al., 2007). Consistent with the find, our results suggest that duplicated genes exhibit diverse expression patterns (Figures 8, 9), which may account for the large number of duplicated genes in soybean GDSL gene family. Furthermore, previous study show that segmentally duplicated genes tend to maintain the same expression pattern, while tandemly duplicated genes can produce rapid expression differentiation (Ganko et al., 2007).

A new candidate, named as GmGELP28, was isolated from soybean and overexpressed in Arabidopsis and soybean, and a series of experiments were performed to explore its role in abiotic stress tolerance. Under drought stress, transgenic Arabidopsis plants showed higher survival rates and chlorophyll content with respect to WT plants (Figures 11B,C). Proline is implicated as the most common osmoprotectant in plants, it is involved in maintaining plant homeostasis in responses to stress conditions (Szekely et al., 2008). The higher accumulation of proline was found in GmGELP28 transgenic Arabidopsis and soybean plants under drought and salt stress conditions, suggesting that proline may contribute to the tolerance of GmGELP28 transgenic plants to drought and salt stresses (Figures 11D, 12D). High salt level can cause oxidative damage and hypertonic stress (Zhu, 2016), and the contents of MDA...
and H₂O₂ can reflect the degree of damage to plants. A great accumulation of MDA and H₂O₂ was detected in *Arabidopsis* WT plants and empty vector control soybean plants grown under drought and salt stress conditions (Figures 11E, 12C,E). In conclusion, the changes in the physiological parameters suggest the positive role of *GmGELP28* in stresses.

**CONCLUSION**

In conclusion, we found 194 *GELP* genes in the soybean genome, and a comprehensive analysis was performed, including phylogenetic analysis, evolutionary features and expression profiles. Finally, we identified a candidate *GELP* gene *GmGELP28* involved in drought and salt tolerance. These results provide insights for better understanding the evolutionary mechanism of soybean *GELP* genes and important reference for genetic improvement of soybean.

**DATA AVAILABILITY STATEMENT**

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

**AUTHOR CONTRIBUTIONS**

H-GS contributed to the conception of the study. H-GS, Z-SX, and D-HM drafted and revised the manuscript. JC, Y-BZ, Y-XW, and MC contributed to data analysis. X-HZ, T-TW, W-LW, and Y-ZM conceived of and designed the experiments. All authors reviewed and approved the final manuscript.
FUNDING

This research was financially supported by the National Transgenic Key Project of the Ministry of Agriculture of China (2016ZX08002002-010) and the National Natural Science Foundation of China (31871624).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00726/full#supplementary-material

FIGURE S1 | Distribution frequency of GELP genes on 20 soybean chromosomes.

REFERENCES

Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004). GDSL family of serine esterases/lipases. Prog. Lipid Res. 43, 534–552. doi: 10.1016/j.plipres.2004.09.002

An, X., Dong, Z., Tian, Y., Xie, K., Wu, S., Zhu, T., et al. (2019). ZmMs30 encoding a novel GDSL lipase is essential for male fertility and valuable for hybrid breeding in maize. Mol. Plant 12, 343–359. doi: 10.1016/j.molp.2019.01.011

Babenko, V. N., Rogozin, I. B., Mekhedov, S. L., and Koonin, E. V. (2004). Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Res. 32, 3724–3733. doi: 10.1093/nar/gkh686

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208. doi: 10.1093/nar/gkp335

Cao, Y., Han, Y., Meng, D., Abdullah, M., Yu, J., Li, D., et al. (2018). Expansion and evolutionary patterns of GDSL-type esterases/lipases in Rosaceae genomes. Funct. Integr. Genomics 18, 673–684. doi: 10.1007/s10292-018-0620-1

Chen, C., Xia, R., Chen, H., and He, Y. (2018). TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. bioRxiv [Preprint] doi: 10.1101/289660

Chepshko, H., Lai, C. P., Huang, L. M., Liu, J. H., and Shaw, J. F. (2012). Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (Oryza sativa L. japonica) genome: new insights from bioinformatics analysis. BMC Genomics 13:309. doi: 10.1186/1471-2164-13-309

Ding, L. N., Guo, X. J., Li, M., Fu, Z. L., Yan, S. Z., Zhu, K. M., et al. (2019a). Improving seed germination and oil contents by regulating the GDSL transcriptional level in Brassica napus. Plant Cell Rep. 38, 243–253. doi: 10.1007/s00299-018-2365-7

Ding, L. N., Li, M., Guo, X. J., Tang, M. Q., Cao, J., Wang, Z., et al. (2019b). Improving seed germination and oil contents by regulating the GDSL transcriptional level in soybean. Plant Mol. Biol. 99, 16128–16133. doi: 10.1007/pns.242624899

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HHMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37. doi: 10.1093/nar/gkr367

FIGURE S2 | Putative motifs of each soybean GELP protein by MEME.

FIGURE S3 | Motif logos detected in GmGELP proteins via MEME analysis.

FIGURE S4 | Intronic numbers of GELP genes in soybean.

TABLE S1 | Description of soybean GELP genes.

TABLE S2 | Multiple sequence alignment among 8 reported AtGELP proteins and their homologous proteins in soybean.

TABLE S3 | The Ka/Ks ratios of duplicated GELP gene pairs.

TABLE S4 | FPKM values of different tissues of soybean GELP genes.

TABLE S5 | FPKM values under drought treatment of soybean GELP genes.

TABLE S6 | FPKM values under NaCl treatment of soybean GELP genes.

TABLE S7 | FPKM values under ABA treatment of soybean GELP genes.

TABLE S8 | Primers and sequences used in this study.

Gankov, E. W., Meyers, B. C., and Vision, T. J. (2007). Divergence in expression between duplicated genes in Arabidopsis. Mol. Biol. Evol. 24, 2298–2309. doi: 10.1093/molbev/msm158

Gao, M., Yin, X., Yang, W., Lam, S. M., Tong, X., Liu, J., et al. (2017). GDSL lipases modulate immunity through lipid homeostasis in rice. PLoS Pathog. 13:e1006724. doi: 10.1371/journal.ppat.1006724

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40, D1178–D1186. doi: 10.1093/nar/gkr944

Hamel, I. P., Nicole, M. C., Sritubtim, S., Morency, M. J., Ellis, M., Ehlting, J., et al. (2006). Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. Trends Plant Sci. 11, 192–198. doi: 10.1016/j.tplants.2006.02.007

Holub, E. B. (2001). The arms race is ancient history in Arabidopsis, the wildflower. Nat. Rev. Genet. 2:516. doi: 10.1038/sj/nrg500850

Hong, J. K., Choi, H. W., Hwang, I. S., Kim, D. S., Kim, N. H., Choi, D. S., et al. (2008). Function of a novel GDSL-type pepper lipase gene, CaGELP1, in disease susceptibility and abiotic stress tolerance. Planta 227, 539–558. doi: 10.1007/s00425-007-0637-5

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GDS2.0: an upgraded gene feature visualization server. Bioinformatics 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Huang, L. M., Lai, C. P., Chen, L. O., Chan, M. T., and Shaw, J. F. (2015). Arabidopsis SEAR4 is a novel GDSL-type esterase involved in fatty acid degradation and glycol tolerance. Bot. Stud. 56:33. doi: 10.1007/s40529-015-0116-4

Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C. D., Nontachaiyapoom, S., Kinkema, M., et al. (2007). Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. Nat. Protoc. 2, 948–952. doi: 10.1038/nprot.2007.141

Kim, H. G., Kwon, S. J., Jang, Y. J., Chung, J. H., Nam, M. H., and Park, O. K. (2014). GDSL lipase 1 regulates ethylene signaling and ethylene-associated systemic immunity in Arabidopsis. FEBS Lett. 588, 1652–1658. doi: 10.1016/j.febslet.2014.02.062

Kim, H. G., Kwon, S. J., Jang, Y. J., Nam, M. H., Chung, J. H., Na, Y. C., et al. (2013). GDSL LIPASE1 modulates plant immunity through feedback regulation of ethylene signaling. Plant Physiol. 163, 1776–1791. doi: 10.1104/pp.113.225649

Kim, K. J., Lim, J. H., Kim, M. J., Kim, T., Chung, H. M., and Paek, K. H. (2008). GDSL-lipase1 (CaGELP1) contributes to wound stress resistance by modulation of CaPK-4 expression in hot pepper. Biochem. Biophys. Res. Commun. 374, 693–698. doi: 10.1016/j.bbrc.2008.07.120

Kwon, S. J., Jin, H. C., Lee, S., Nam, M. H., Chung, J. H., Kwon, S. I., et al. (2009). GDSL lipase-like 1 regulates systemic resistance associated with ethylene signaling in Arabidopsis. Plant J. 58, 235–245. doi: 10.1111/j.1365-313X.2008.03772.x

Lai, C. P., Huang, L. M., Chen, L. O., Chan, M. T., and Shaw, J. F. (2017). Genome-wide analysis of GDSL-type esterases/lipases in Arabidopsis. Plant Mol. Biol. 95, 181–197. doi: 10.1007/s11103-017-0648-y
Lee, D. S., Kim, B. K., Kwon, S. J., Jin, H. C., and Park, O. K. (2009). Arabidopsis GDSL lipase 2 plays a role in pathogen defense via negative regulation of auxin signaling. Biochem. Biophys. Res. Commun. 379, 1038–1042. doi: 10.1016/j.bbrc.2009.01.006

Lee, T. H., Tang, H., Wang, X., and Paterson, A. H. (2013). PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res. 41, D1152–D1158. doi: 10.1093/nar/gks1104

Li, B., Liu, Y., Cui, X. Y., Fu, J. D., Zhou, Y. B., Zheng, W. J., et al. (2019). Genome-wide characterization and expression analysis of soybean TGA transcription factors identified a novel TGA gene involved in drought and salt tolerance. Front. Plant Sci. 10:549. doi: 10.3389/fpls.2019.00549

Li, C., Chen, G., Mishina, K., Yamaji, N., Ma, J. F., Yamaji, N., et al. (2017). GDSL lipase 2 plays a role in pathogen defense via negative regulation of auxin signaling. Biochem. Biophys. Res. Commun. 379, 1038–1042. doi: 10.1016/j.bbrc.2009.01.006

Li, Z., Hua, X., Zhong, W., Yuan, Y., Wang, Y., Wang, Z., et al. (2019). Genome-wide identification and expression profile analysis of WRKY family genes in the autopolyploid saccharum spontaneum. Plant Cell Physiol. 61, 616–630. doi: 10.1093/pcp/pcz227

Ling, H., Zhao, J., Zuo, K., Qu, C., Yao, H., Qin, J., et al. (2006). Isolation and expression analysis of a GDSL-like lipase gene from Brassica napus L. J. Biochem. Mol. Biol. 39, 297–303. doi: 10.5483/bmbrep.2006.39.2.297

Luo, S., Hu, W., Yang, Y., Liu, B., Yan, H., and Xiang, Y. (2018). Genome-wide identification, classification, and expression of phytoestrogens in Populus trichocarpa. Planta 247, 1133–1148. doi: 10.1007/s00425-018-2849-2

Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant intron loss-dominated evolution and genome-wide comparison of exon-intron structure. BMC Plant Biol. 8:330. doi: 10.1186/1471-2229-8-330

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI’s conserved domain database. Nucleic Acids Res. 43, D222–D226. doi: 10.1093/nar/gku1221

Narjano, M. A., Forment, J., Roldan, M., Serrano, R., and Vicente, O. (2006). Overexpression of Arabidopsis thaliana LTI1, a salt-induced gene encoding a GDSL-motif lipase, increases salt tolerance in yeast and transgenic plants. Plant Cell Environ. 29, 1890–1900. doi: 10.1111/j.1365-3040.2006.01565.x

Rajarammohan, S., Pradhan, A. K., Pental, D., and Kaur, J. (2018). Genome-wide association mapping in Arabidopsis identifies novel genes underlying quantitative disease resistance to Alternaria brassicaceae. Mol. Plant Pathol. 19, 1719–1732. doi: 10.1111/mpp.12654

Rao, K. P., Richa, T., Kumar, K., Raghuram, B., and Sinha, A. K. (2010). In silico analysis reveals 75 members of mitogen-activated protein kinase kinase kinase (MAPKKK) gene family in bread wheat (Triticum aestivum L.). BMC Genomics 17:468. doi: 10.1186/1471-2164-15-468

Roy, S. W., and Penny, D. (2007). Patterns of intron loss and gain in plants. Mol. Biol. Evol. 24, 1133–1148. doi: 10.1093/molbev/msq226

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Min, D. H., Jin, L. G., Chen, J., et al. (2010). The WRKY transcription factor GmWRKY12 confers drought and salt tolerance in soybean. Int. J. Mol. Sci. 19:4087. doi: 10.3390/ijms19124087

Smyth, D. R. (2017). Wrinkles on sepalic cuticular ridges form when cuticle production outpaces epidermal cell expansion. Mol. Plant 10, 540–541. doi: 10.1016/molp.2017.02.008

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyplody and angiosperm diversification. Am. J. Bot. 96, 336–348. doi: 10.3732/ajb.0800079

Su, H. G., Li, B., Song, X. Y., Ma, J., Chen, J., Zhou, Y. B., et al. (2019). Genome-wide analysis of the DYW subgroup PPR gene family and identification of GmPPR4 responses to drought stress. Int. J. Mol. Sci. 20:5667. doi: 10.3977/j.ijms20225667

Szekely, G., Abraham, E., Cseplo, A., Rigo, G., Zaigmund, L., Csizsar, J., et al. (2008). Duplicated P5CS genes of Arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. Plant J. 53, 11–28. doi: 10.1111/j.1365-313X.2007.03318.x

Takahashi, K., Shimada, T., Kondo, M., Tamai, A., Morii, M., Nishimura, M., et al. (2010). Ectopic expression of an exterase, which is a candidate for the unidentified plant cutinase, causes cuticular defects in Arabidopsis thaliana. Plant Cell Physiol. 51, 123–131. doi: 10.1093/pcp/pcp173

Updegraff, E. P., Zhao, F., and Preuss, D. (2009). The extracellular lipase EL1X4 is required for efficient hydration of Arabidopsis pollen. Sex Plant Reprod. 22, 197–204. doi: 10.1007/s00497-009-0104-5

Upton, C., and Buckley, J. T. (1995). A new family of lipolytic enzymes? Trends Biochem. Sci. 20, 178–179. doi: 10.1016/0968-0004(95)80902-J

Volokita, M., Rosilio-Brami, T., Rivkin, N., and Zik, M. (2011). Combining comparative sequence and genomic data to ascertain phylogenetic relationships and explore the evolution of the large GDSL-lipase family in land plants. Mol. Biol. Evol. 28, 551–565. doi: 10.1093/molbev/msq226

Wang, M., Yue, H., Feng, K., Deng, P., Song, W., and Nie, X. (2016). Genome-wide identification, phylogeny and expression profiles of mitogen activated protein kinase kinase kinase (MAPKKK) gene family in bread wheat (Triticum aestivum L.). BMC Genomics 17:668. doi: 10.1186/s12864-016-2993-7

Watts, J. L., Li, M., McQuinn, R. P., Chan, K. X., McFarlane, H. E., Ermakova, M., et al. (2019). A GDSL esterase/lipase catalyzes the esterification of lutein in bread wheat. Plant Cell 31, 3092–3112. doi: 10.1105/tpc.19.00272

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. Proc. Natl. Acad. Sci. U.S.A. 109, 1187–1192. doi: 10.1073/pnas.1109047109

Yadav, V. K., Yadav, V. K., Pant, P., Singh, S. P., Maurya, R., Sable, A., et al. (2017). GhMYB1 regulates SCW stage-specific expression of the GhGDSL promoter in the fibres of Gossypium hirsutum L. Plant Biotechnol. J. 15, 1163–1174. doi: 10.1111/pbi.12706

Zhang, B., Zhang, L., Li, F., Zhang, D., Liu, X., Wang, H., et al. (2017). Control of secondary cell wall patterning involves xylan deacetylation by a GDSL esterase. Nat. Plants 3:17017. doi: 10.1038/nplants.2017.17

Zhang, L., Gao, C., Mentink-Vigier, F., Tang, L., Zhang, D., Wang, S., et al. (2019). Arabinosyl deacetylase modulates the arabinoxylan acetylation profile and secondary wall formation. Plant Cell 31, 1113–1126. doi: 10.1105/tpc.18.00894

Zhao, J., et al. (2007). Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. BMC Genomics 8:330. doi: 10.1186/1471-2164-8-330

Su et al.  Soybean GDSL Gene Family

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.