scientific reports

Check for updates

OPEN

Genome-wide analysis of class III peroxidase gene family in *Glycine max* and functional roles in stress response

Sushuang Liu[1,4], Yizhou Wu[1,4], Yingsheng Qiu[2], Dandan He[1], Kangqi Zhao[1], Liuyang Yang[1], Yang Li[3], Zaibao Zhang[1], Huasong Zou[1] & Yanmin Liu[1]✉

Soybeans are one of the world's most widely grown crops, renowned for their 35–40% plant protein content. However, climate change-induced high temperature and humidity stress can lead to irreversible seed deterioration. It reduced germination and field emergence rates, and diminished nutritional value, ultimately lowering market value. Class III peroxidases (PODs) are plant-specific enzymes that play crucial roles in growth, development, and environmental stress response. This research investigates the response mechanism of the soybean GmPOD gene to high temperature and humidity stress. We identified GmPOD gene family members and predicted their chromosomal positions, evolutionary relationships, protein interaction networks, promoter cis-regulatory elements, and tissue-specific expression patterns using soybean genome data. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis revealed the expression patterns of 180 relatively conserved GmPOD members, which are distributed across twenty chromosomes in six subfamilies. The promoter region of GmPOD contains cis-acting elements linked to growth, hormone regulation, and stress response, regulating the expression of related genes. This study offers insights into the evolution of the GmPOD gene family and establishes a foundation for exploring plant GmPOD gene functions.

Keywords Soybean (*Glycine max*), Class III peroxidase gene family, Gene expression, High-temperature and -humidity stress, Bioinformatics

Peroxidases, or PODs, are members of large multigene families found in many different organisms; these enzymes use hydrogen peroxide (H_2O_2) as an electron receiver at their reactive sites to aid oxidation processes, with the help of metals. PODs may be divided into two main types, based on their catalytic properties and architectural features: heme-containing PODs, and those without heme[1]. Moreover, hemoglobin-associated PODs can be further separated into animal-derived and non-animal-derived divisions[2]. There are three main subclasses within the ungulate family, Classes I, II, and III[3], among which, the plant-specific Class III peroxidases (EC 1.11.1.7) are denoted by a number of acronyms, such as POX, POD, Px, PER, and Prx[4]; we shall refer to POD as "Class III peroxidase" for the duration of this study.

Class III plant peroxidase enzymes have a highly conserved collection of amino acids that include protoporphyrin IX domains and single peptide chains; they also have eight cysteine residues that form disulfide bridges, as well as two histidine residues that interact with a single heme[5,6]. These enzymes mainly take part in the peroxide and hydroxyl radical cycles, which are intended to limit the production of reactive oxygen species (ROS) and hydrogen peroxide[7]. Most plant PODs form glycosylated proteins through their association with carbohydrate side chains, which protects them from proteolytic degradation and maintains their stability[8–10]. Additionally, POD proteins support a variety of physiological functions, including the integration of cell wall components, the modification of plant hormones, pathogen resistance, the breakdown of toxic chemicals, and enhancements in salt tolerance[11]. The function of POD proteins in plants' responses to biotic and abiotic stress conditions is supported by a number of genetic studies; for example, overexpressing several POD genes (*AtPrx22*, *AtPrx39*, and *AtPrx69*) in Arabidopsis thaliana increases its resistance to cold[7], and transgenic tobacco's tolerance to aluminum stress is increased by upregulating the expression of *AtPrx64*[12]. While pepper plants with

[1]Department of Life Sciences and Health, Huzhou College, Huzhou 313000, China. [2]College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China. [3]College of Life Science, Huzhou University, Huzhou 313000, China. [4]These authors contributed equally: Sushuang Liu and Yizhou Wu. ✉email: liuyanmin@zjhzu.edu.cn

overexpressed *CaPOD2* in Arabidopsis show resilience to bacterial assaults, those with silenced *CaPOD2* genes are more vulnerable to xanthomonas infections[13]. Furthermore, the *GhPOX1* variation in cotton is linked to an increase in the generation of reactive oxygen species[14]. Ultimately, these investigations highlight the beneficial impacts of Class III plant peroxidases on the plant's response to biological and environmental stressors.

Soybean is among the most widely grown crops in the world and is an essential source of both cooking oil and plant protein[15]. All stages of soybean growth and maturity are at risk from the stressors of high temperatures and humidity, which are worsened by climate change. A variety of harmful outcomes, such as weakened cell membrane integrity, deformed leaf and flower tissues, decreased pollen viability and germination rates, pollen abortion, reduced pod setting, abnormal seed composition, lower seed vitality, slowed or stopped germination, decreased field emergence rates, and diminished nutritional content, can result from such unfavorable conditions, seriously hindering the normal progression of plant growth; when all of these things come together, crop output and quality significantly decline[16–20]. It is estimated that worldwide soybean output would decrease by 3.1% for every degree Celsius by which the average global temperature climbs[21], as stress from both high temperatures and humidity is a major factor affecting the yield and quality of soybean[22]. The exact method by which the GmPOD gene family functions under high-temperature and -humidity stress conditions in soybean is still unclear, despite the gene family's substantial documentation in a variety of plants; therefore, determining the GmPOD gene family in soybean is crucial for understanding genes associated with tolerance to high-temperature and -humidity stress conditions, both theoretically and practically. This information may make it easier to understand how soybean responds to stress and to create cultivars that are more resistant to it. In order to provide insights into how GmPOD family members regulate soybean development and stress responses, we conducted this study with the aims of identifying the GmPOD gene family in soybean and conducting thorough analyses of its phylogenetic relationships, gene structure, domains, chromosomal positioning, collinearity, protein interaction networks, cis-regulatory elements, and tissue-specific expression patterns.

Materials and methods

Identification of GmPOD family members in soybean

The soybean genome sequence source was Soybase (https://www.soybase.org/)(the study used the genome of the version to Wm82 a4. V1, the version from the Phytozome database (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a4_v1)); protein sequences for POD proteins from pepper (Use the version for PI 159236 v1.55 genome, from Phytozome database (https://phytozome-next.jgi.doe.gov/info/Cannuum_PI159236_v1_55)) and Arabidopsis were obtained from the Phytozome Araport11 version in the database (https://phytozome-next.jgi.doe.gov/info/Athaliana_Araport11). Using known GmPOD protein sequences (Additional file 1), Hidden Markov Model (HMM) profiles were constructed, and then used to scan soybean protein databases using the HMMER program (version 3.3.2)[23]. The Pfam peroxidase (PF00141) structural domain was used to validate the identified soybean GmPOD proteins using the InterPro database (https://www.ebi.ac.uk/interpro/), SMART database (http://smart.embl-heidelberg.de/), and NCBI's Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/). Other potential proteins were eliminated, and proteins containing the entire Pfam peroxidase (PF00141) domain were identified as GmPOD members. The names of the potential soybean GmPOD genes were determined by their locations on the physical map. The Compute pI/Mw program (https://web.expasy.org/compute_pi/) was used to ascertain the isoelectric point (pI) and molecular weight (MW) of each GmPOD protein[24]. Furthermore, amino acid counts were predicted using the ProtParam program[25].

Chromosomal locations, gene duplication analysis, and synteny analyses of GmPOD genes

GFF3 files were used to collect information on the chromosomal locations of the members of the GmPOD gene family. Use the "One Step MCScanX" function of TBtools (version 1.120) to complete the lineage analysis of the GmPOD gene family, the detection of gene duplication events, and draw chromosome mapping and collinearity relationship diagrams through this software[26]. Circos (Version 0.69–9.69 is adopted) was used to display the location of gene duplication events[27]. Using MCScanX (This study used the original version released in 2012), the duplication patterns and collinearity of GmPOD genes within soybean species were investigated, and TBtools (version 1.120) was used to display the results[28]. Using the Simple Ka/Ks calculator in the TBtools program, the non-synonymous substitution rate (Ka), synonymous substitution rate, and the ratio of non-synonymous to synonymous substitution rates (Ka/Ks) were calculated from the coding sequences[26,29].

Protein–protein interaction network of GmPOD genes

A search for the full appropriate typical transcript protein sequences was conducted on the STRING database (version v11.5) using the protein sequences from the 180 discovered GmPOD genes[30]. The data on Arabidopsis's protein–protein interaction network and sample protein sequence sets were also made available via the STRING database[30]. When constructing the GmPOD protein-protein interaction network through the STRING database (version 11.5) in this study, the following key parameters were set: score critical value: Use the default medium confidence threshold of the database (combined score ≥ 0.4) to screen significant interaction relationships to exclude protein interactions with low confidence. Analysis of the protein–protein interaction (PPI) network linked to the GmPOD family was conducted using the TBtools program[26]; then, the results were visualized using Cytoscape (version 3.9.1).

Conserved motif and gene structure analysis of GmPOD genes

TBtools version 1.120 was used to study the GmPOD genomic architecture in order to clarify the exon/intron organization of soybean GmPOD genes[26]. Utilizing the MEMESuite program (https://meme-suite.org/meme/index.html), the conserved motifs found in the GmPOD gene were investigated; all other parameters were maintained at their normal values, and the analysis was limited to identifying up to 10 motifs[31]. Gene structure

and conserved motif analysis: After predicting the conserved motifs of the GmPOD gene using MEME Suite, the exon/intron structure of the gene (CDS regions marked with yellow boxes and introns marked with black lines) and the distribution of 10 conserved motifs were presented through TBtools (version 1.120).

Phylogenetic tree construction of GmPOD protein

The POD amino acid sequences were obtained and concatenated from Arabidopsis[32], pepper[33], and soybean (source: https://www.soybase.org), after which multiple sequence alignment was carried out on this composite dataset using ClustalW version 2.0.11. Each protein sequence was then aligned using MAFFT version 7. The maximum likelihood (ML) approach was used to generate a phylogenetic tree without a root, using 1000 bootstrap replicates and MEGA 11.0 software[34]. The GmPOD genes were categorized using the topological structure of the evolutionary tree and the taxonomy of their Arabidopsis counterparts[35]. A more accurate visualization of the phylogenetic tree was achieved using the online application iTol (Use the online version of iTOL v6) (http://itol.embl.de)[36].

Analysis of cis-acting elements of GmPOD gene family promoters

We obtained the promoter sequences, which extend two kilobases upstream of the transcription start site for every GmPOD gene, from the soybean genome database and performed an analysis to determine putative cis-acting regulatory elements within these sequences using the PlantCARE(Uses the core version updated in 2012) online tool (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed 27 April 2024)[37], after which we examined features linked to phytohormones and stress response[38] and used TBtools to visualize the results of the promoter analysis[26]. In this study, when extracting the promoter sequence of the target gene (the upstream 2 kb region), for the two special cases of "overlapping genes" and "short promoter region (less than 2 kb)", the following standardized processing strategies were adopted: Handling of overlapping promoter regions When the upstream 2 kb regions of two adjacent genes (such as gene A and gene B, located on the same chromosome and with the same/opposite transcription directions) overlap (i.e., the distance between the downstream end of gene A and the upstream end of gene B is less than 2 kb), we take the midpoint between the transcription start sites (TSS) of the two genes as the boundary. Extract the respective promoter sequences respectively. Handling of shorter promoter regions: When the target gene is located at the end of the chromosome or its upstream is adjacent to the chromosomal structural region (such as centromeres, telomeres), making it impossible to extract the complete 2 kb upstream sequence, we take the actual length from the gene TSS to the upstream chromosomal boundary/structural region as the promoter sequence.

Tissue-specific expression analysis of GmPOD gene family

The Yanglab website (https://yanglab.hzau.edu.cn/SoyMD/#/transcriptomics/expression?) provided information on the expression levels of GmPOD family members in soybean across a variety of tissues, including the roots, stems, young leaves, old leaves, flowers, seeds, embryos, and endosperms. Tissue-specific expression analysis After obtaining the GmPOD gene expression data from the Yanglab website, the original data were converted to log2(tpm + 1) values using TBtools (version 1.120) (genes with log2(tpm + 1) < 1 were defined as low expression or no expression), and tissue expression heat maps were drawn. The resultant data were processed as heatmaps, categorized by expression levels, using the TBtools program[26].

Prediction of MiRNA targeting sites and analysis of GmPOD protein structure

Known soybean miRNA sequences were downloaded from the miRBase database (version v22.1)(https://www.mirbase.org/download/). miRNA targeting sites were predicted using psRNATarget (Adopts version 2.0) online tool (https://www.zhaolab.org/psRNATarget/analysis?function=2)[39]. The prediction results were screened, the prediction sites with the expected value were retained, and the unreasonable prediction results were removed. Using AlphaFold2 online platform (version v2.3.1)(https://alphafold.ebi.ac.uk/) predict the three-dimensional structure of protein genes. Tertiary structure similarity among members of the GmPOD family was evaluated using the TM-scoring tool (http://zhanggroup.org/TM-score/).

Plant materials, growth conditions, and high-temperature and -humidity treatment

May White soybean seeds were soaked overnight at room temperature in an incubator, after which the seeds were spread out on damp filter paper and allowed to pre-germinate for 48 h at 25 °C in a Petri dish (The soybean variety "May White" selected in this study is a conventional medium-maturity variety traditionally cultivated in the middle and lower reaches of the Yangtze River. Germplasm registration has been completed on the national crop germplasm Resources platform (https://www.cgris.net/) (registration number: ZJHZ20220518-01). The specific agronomic traits of this variety are as follows: the plant height is 65–75 cm, with a limited podding habit. The leaves are ovate, the flowers are white, and the pods are yellowish-brown, with 2–3 seeds per pod. The seeds are round, with white seed coats and a weight of 18–20 g per 100 grains.)[40]. After pre-germination, the seeds were planted in pots with a soil and vermiculite combination (2:1) and cultivated for two weeks at 20 °C in a standard greenhouse with a 16/8-hour light/dark cycle. Thereafter, the seedlings were divided into experimental and control groups. Twelve hours before stress treatment, the potted soybeans in the experimental group were overwatered at one time (300 mL per pot) to maintain the soil relative humidity at 85%–90%; During the entire stress treatment period (0–48 h), the temperature was maintained at 40 °C (with an error of ± 0.5 °C), and the photopod was 16/8 h (light intensity 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$), forming a clear stress gradient with the control group (20 °C, air relative humidity 50%–55%). In order to evaluate the levels of GmPOD gene expression, leaves from the fourth node were taken at 0, 24, and 48 h; the samples were then subjected to real-time fluorescence quantitative PCR (qRT-PCR) analysis.

RNA extraction and qRT-PCR analysis

With the use of a HiPure Universal RNA mini-kit (MGBio, Shanghai, China), RNA was extracted from soybean leaves. Using MonadTMRTIII Super combined with dsDNase (provided by Monad Biotechnology Co., LTD., Shanghai, China), reverse transcription was carried out. Following the manufacturer's instructions, the MonAmpTMChemoHS qPCR Mix (from Monad Biotechnology Co., LTD., Shanghai, China) was used to perform quantitative PCR for the target gene on the CFX96 Real-Time PCR machine (BioRad). Relative mRNA expression levels were determined using the comparative threshold cycle (CT) technique[41], which took into account information from three different biological replicates. Supplementary File 3 contains a list of the primers used for real-time PCR. In this study, the qRT-PCR experiment selected the soy constitutive actin gene (GmACT1, GenBank registry number: NM_001250579.1) as the internal reference gene for expression level normalization.

Statistical analysis

Three independent replicates were used for the experimental protocols, and the results are shown as averages plus or minus standard deviation (SD). Duncan's-test was used to determine the statistical significance of any alterations, with a P value of less than 0.05 being considered statistically significant.

Results

Genome-wide identification of GmPOD gene family in soybean

In this study, POD protein sequences from *Arabidopsis thaliana* and *Capsicum annuum* L. were used to create a Hidden Markov Model (HMM), which allowed for the discovery of the soybean GmPOD gene family (Supplementary file 1); in total, 180 GmPOD genes were found in soybean as a result of this procedure and were given the names GmPOD-1 through GmPOD-180, based on where in the genome they were found (Supplementary file 4). These detected GmPOD have genomic mRNAs ranging in length from 605 to 21,060 base pairs; the coding sequences for the putative GmPOD proteins ranged from 315 to 1308 base pairs; the proteins themselves had between 104 and 989 amino acid residues, while their molecular weights (MW) and isoelectric points (pI) ranged from 11.48 to 47.34 kDa and 4.14 to 9.84, respectively (Supplementary file 4).

Chromosome localization, gene replication events, and synchronic analysis of GmPOD family

Based on the genetic information acquired, the chromosomal locations of these genes were mapped in order to examine the chromosomal distribution of GmPOD genes in soybean. We discovered that the 180 identified GmPOD genes were dispersed among 20 chromosomes (Fig. 1), as follows: Chr09 had 17 GmPOD genes; Chr02, Chr03, and Chr15 had 15 genes; Chr17 had 12 genes; Chr11 had 11 genes; Chr01, Chr10, and Chr13 had 9 genes; Chr14 and Chr16 had 8 genes; Chr06, Chr12, Chr18, and Chr20 had 7 genes; Chr04, Chr08 and Chr19

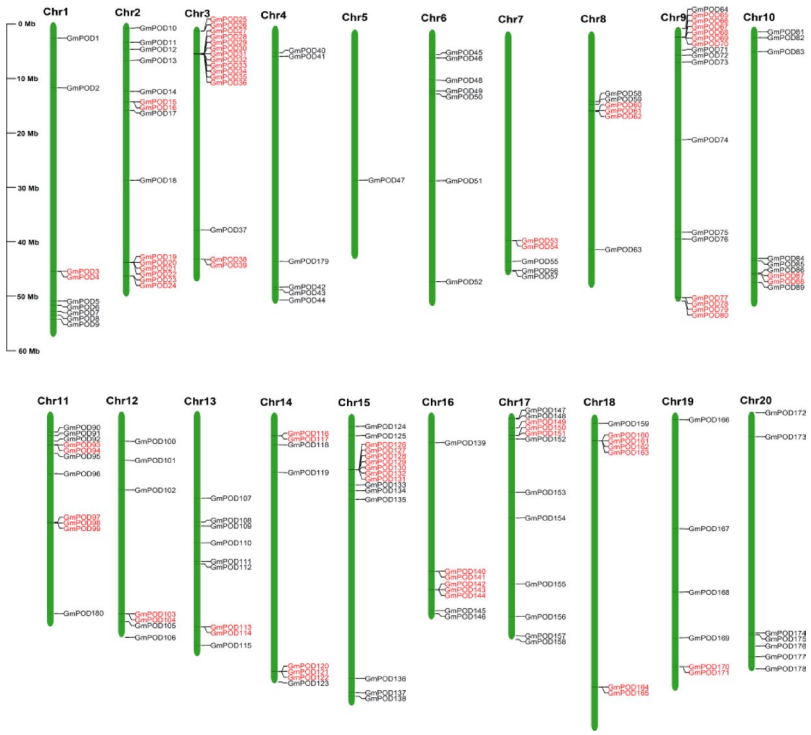

Fig. 1. The locations of GmPOD genes on chromosomes. Tandem duplication-containing genes are indicated in red. Every bar chart shows the relevant chromosomal number at the top. On the left side of the figure, the chromosomal size is expressed in millions of base pairs (Mb).

had 6 genes; Chr07 had 5 genes; and only Chr05 had 1 GmPOD gene (Figs. 1 and 2). Based on each gene's chromosomal location, homology and gene duplication events were examined to clarify the duplication history of GmPOD genes over the course of evolution, and 80 (44.4%) of the GmPOD genes were found to have tandem duplications. Red lines in Fig. 2 indicate the locations of 30 pairs of tandem repeats and 56 pairs of segmental repetitions that were found throughout our study, also presented in Table 1. When comparing the homologous fragments of the genomes of peppers, soybeans and *Arabidopsis thaliana*, many POD genes are the same (Fig. 3). It was particularly found that 82 POD genes were homologous in *Glycine max* and *Arabidopsis thaliana* (Fig. 3).

Utilizing the TBtools (v.1.120) program, the non-synonymous (Ka) and synonymous (Ks) substitution rates for these gene pairs were determined in order to obtain a better understanding of the evolutionary dynamics of the GmPOD gene family, as well as to identify the times of evolution and selective pressures operating on the duplicated genes. According to the literature, segmental and tandem repeat occurrences in plant gene pairs are subject to positive, purifying, and neutral selection when the Ka/Ks ratios are > 1, <1, and = 1, respectively[42,43]. Given that the Ka/Ks ratio was considerably smaller than 1, purifying selection appears to have had a major role in the development of GmPOD gene pairs in soybean. Moreover, the dicotyledonous formula T = Ks/2λ was employed to determine the evolutionary dates of these recurrences[44,45]. According to Table 1, these estimations suggest that the tandem duplications may have occurred between 1.38 and 55.16 million years ago, and the segmental duplications between 4.38 and 67.46 million years ago.

Construction of GmPODs family protein interaction network

Analysis of protein–protein interactions, or PPIs, is a basic method for facilitating comprehension of the interactions between proteins and how they affect biological processes. Through close examination of the PPI network's linkages, interactions between different proteins can be found[46,47]. Using the STRING database, a network comprising 29 soybean GmPOD proteins was created for this investigation (Fig. 4). *GmPOD4*, *GmPOD20*, *GmPOD30*, *GmPOD73*, *GmPOD78*, *GmPOD148*, and *GmPOD151* show very robust interactions with other members of the family, as the network illustrates; thus, these proteins could play a key role in preserving cellular homeostasis and responding to stress conditions associated with high temperatures and humidity. Moreover, *GmPOD76*, which is located in the network's center, interacts with other genes, demonstrating its crucial function in investigating the roles of the GmPOD gene in soybean under diverse stress circumstances.

Phylogenetic analysis of GmPOD proteins

We created phylogenetic trees, utilizing 180 GmPOD, 75 CaPOD, and 73 AtPOD proteins, in order to clarify the evolutionary relationships among the members of the POD gene family and to infer their analogous activities; these proteins were categorized into six groups (Groups I–VI), as shown in Fig. 5. Sets of three (Ia, Ib, and Ic), three (IIa, IIb, and IIc), two (IIIa and IIIb), and two (IVa and IVb) subgroups were further separated into Groups I, II, and III. Group I was made up of seventeen AtPrx members, twenty-three CaPOD members, and seventy-three GmPOD members. Group II consisted of ten AtPrx members, nine CaPOD members, and fifteen GmPOD members. Group III is the largest subgroup, consisting of twenty-eight AtPrx members, twenty-eight CaPOD members, and fifty-three GmPOD members. Group IV consisted of fourteen AtPrx members, ten CaPOD members, and seventeen GmPOD members. Group V consisted of four AtPrx members, five CaPOD

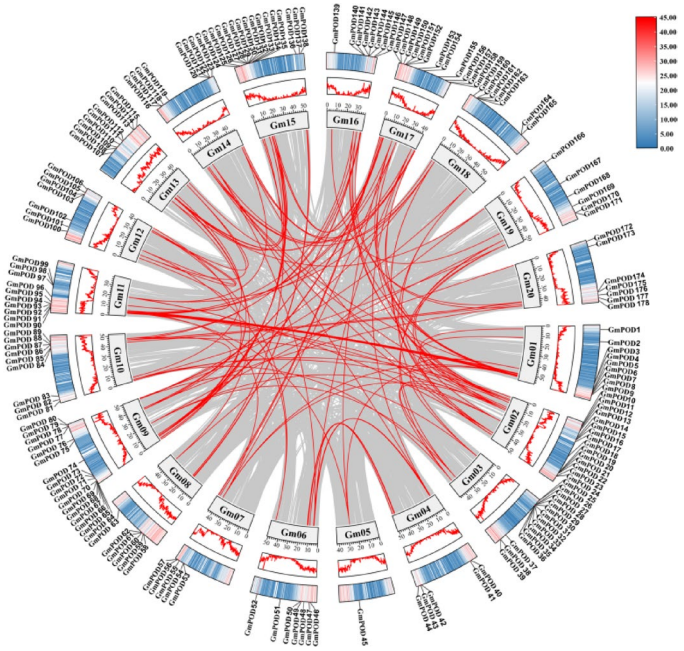

Fig. 2. GmPOD gene homology. Gray lines depict all homologous regions within the soybean genome. The red lines in the center signify gene duplication events.

Seq1	Seq2	Ka	Ks	Ka/Ks Ratio	Date (MY)	Duplication Type
GmPOD164	GmPOD165	0.139	0.461	0.302	15.38	WGD or Segmental
GmPOD65	GmPOD66	0.257	0.861	0.298	28.70	WGD or Segmental
GmPOD65	GmPOD67	0.220	0.686	0.320	22.85	WGD or Segmental
GmPOD65	GmPOD68	0.276	0.852	0.324	28.41	WGD or Segmental
GmPOD65	GmPOD69	0.256	0.904	0.283	30.12	WGD or Segmental
GmPOD65	GmPOD70	0.288	0.844	0.341	28.15	WGD or Segmental
GmPOD66	GmPOD67	0.113	0.569	0.199	18.96	WGD or Segmental
GmPOD66	GmPOD68	0.243	0.648	0.374	21.62	WGD or Segmental
GmPOD66	GmPOD69	0.225	0.693	0.325	23.11	WGD or Segmental
GmPOD66	GmPOD70	0.124	0.340	0.366	11.32	WGD or Segmental
GmPOD67	GmPOD68	0.165	0.581	0.284	19.35	WGD or Segmental
GmPOD67	GmPOD69	0.137	0.620	0.222	20.68	WGD or Segmental
GmPOD67	GmPOD70	0.169	0.529	0.320	17.63	WGD or Segmental
GmPOD68	GmPOD69	0.099	0.283	0.349	9.44	WGD or Segmental
GmPOD68	GmPOD70	0.264	0.713	0.370	23.77	WGD or Segmental
GmPOD69	GmPOD70	0.266	0.780	0.341	26.00	WGD or Segmental
GmPOD116	GmPOD117	0.353	NaN	NaN	#VALUE!	WGD or Segmental
GmPOD140	GmPOD141	0.132	0.198	0.666	6.61	WGD or Segmental
GmPOD150	GmPOD151	0.040	0.163	0.246	5.43	WGD or Segmental
GmPOD38	GmPOD39	0.056	0.232	0.241	7.72	WGD or Segmental
GmPOD27	GmPOD28	0.163	0.336	0.487	11.19	Tandem
GmPOD27	GmPOD29	0.094	0.280	0.334	9.34	Tandem
GmPOD27	GmPOD30	0.095	0.281	0.338	9.36	Tandem
GmPOD27	GmPOD31	0.095	0.294	0.323	9.79	Tandem
GmPOD27	GmPOD32	0.100	0.293	0.341	9.77	Tandem
GmPOD27	GmPOD33	0.113	0.307	0.369	10.25	Tandem
GmPOD27	GmPOD34	0.184	0.442	0.417	14.73	Tandem
GmPOD28	GmPOD29	0.174	0.298	0.585	9.92	Tandem
GmPOD28	GmPOD30	0.176	0.312	0.564	10.39	Tandem
GmPOD28	GmPOD31	0.174	0.332	0.524	11.07	Tandem
GmPOD28	GmPOD32	0.174	0.318	0.548	10.59	Tandem
GmPOD28	GmPOD33	0.160	0.281	0.568	9.36	Tandem
GmPOD28	GmPOD34	0.216	0.598	0.361	19.92	Tandem
GmPOD29	GmPOD30	0.004	0.056	0.074	1.85	Tandem
GmPOD29	GmPOD31	0.004	0.060	0.068	2.01	Tandem
GmPOD29	GmPOD32	0.008	0.055	0.148	1.85	Tandem
GmPOD29	GmPOD33	0.075	0.285	0.265	9.50	Tandem
GmPOD29	GmPOD34	0.148	0.497	0.297	16.58	Tandem
GmPOD30	GmPOD31	0.005	0.041	0.132	1.38	Tandem
GmPOD30	GmPOD32	0.010	0.065	0.147	2.18	Tandem
GmPOD30	GmPOD33	0.077	0.286	0.269	9.52	Tandem
GmPOD30	GmPOD34	0.146	0.482	0.303	16.06	Tandem
GmPOD31	GmPOD32	0.007	0.041	0.166	1.38	Tandem
GmPOD31	GmPOD33	0.078	0.292	0.269	9.74	Tandem
GmPOD31	GmPOD34	0.151	0.448	0.337	14.95	Tandem
GmPOD32	GmPOD33	0.078	0.285	0.276	9.49	Tandem
GmPOD32	GmPOD34	0.151	0.472	0.320	15.72	Tandem
GmPOD33	GmPOD34	0.153	0.460	0.333	15.33	Tandem
GmPOD103	GmPOD104	0.028	0.131	0.216	4.38	WGD or Segmental
GmPOD77	GmPOD78	0.151	0.463	0.326	15.43	WGD or Segmental
GmPOD53	GmPOD54	0.201	0.223	0.898	7.45	WGD or Segmental
GmPOD103	GmPOD104	0.028	0.131	0.216	4.38	WGD or Segmental
GmPOD113	GmPOD114	0.027	0.152	0.179	5.07	WGD or Segmental
GmPOD15	GmPOD16	0.139	0.206	0.672	6.87	WGD or Segmental
GmPOD3	GmPOD4	0.119	0.406	0.293	13.55	WGD or Segmental
GmPOD179	GmPOD180	0.075	0.297	0.253	9.91	Tandem
Continued						

Seq1	Seq2	Ka	Ks	Ka/Ks Ratio	Date (MY)	Duplication Type
GmPOD161	GmPOD162	0.125	0.785	0.159	26.15	WGD or Segmental
GmPOD60	GmPOD61	0.260	0.863	0.301	28.76	WGD or Segmental
GmPOD25	GmPOD26	0.288	1.655	0.174	55.16	Tandem
GmPOD20	GmPOD21	0.181	0.752	0.241	25.08	WGD or Segmental
GmPOD87	GmPOD88	0.366	1.013	0.361	33.78	WGD or Segmental
GmPOD142	GmPOD143	0.197	0.702	0.280	23.39	WGD or Segmental
GmPOD142	GmPOD144	0.382	1.194	0.320	39.80	WGD or Segmental
GmPOD143	GmPOD144	0.460	2.024	0.227	67.46	WGD or Segmental
GmPOD170	GmPOD171	0.791	NaN	NaN	#VALUE!	WGD or Segmental
GmPOD126	GmPOD127	0.226	0.696	0.325	23.19	WGD or Segmental
GmPOD126	GmPOD128	0.195	0.588	0.331	19.61	WGD or Segmental
GmPOD126	GmPOD129	0.280	0.806	0.348	26.85	WGD or Segmental
GmPOD126	GmPOD130	0.247	0.847	0.292	28.24	WGD or Segmental
GmPOD126	GmPOD131	0.419	1.002	0.418	33.40	WGD or Segmental
GmPOD126	GmPOD132	0.282	1.053	0.268	35.09	WGD or Segmental
GmPOD127	GmPOD128	0.128	0.474	0.270	15.80	WGD or Segmental
GmPOD127	GmPOD129	0.232	0.822	0.283	27.41	WGD or Segmental
GmPOD127	GmPOD130	0.238	0.678	0.351	22.61	WGD or Segmental
GmPOD127	GmPOD131	0.233	0.476	0.490	15.86	WGD or Segmental
GmPOD127	GmPOD132	0.200	0.831	0.241	27.70	WGD or Segmental
GmPOD128	GmPOD129	0.201	0.858	0.234	28.60	WGD or Segmental
GmPOD128	GmPOD130	0.147	0.640	0.230	21.34	WGD or Segmental
GmPOD128	GmPOD131	0.238	0.765	0.312	25.48	WGD or Segmental
GmPOD128	GmPOD132	0.130	0.709	0.183	23.64	WGD or Segmental
GmPOD129	GmPOD130	0.124	0.311	0.400	10.36	WGD or Segmental
GmPOD129	GmPOD131	0.275	0.703	0.391	23.44	WGD or Segmental
GmPOD129	GmPOD132	0.258	0.597	0.433	19.91	WGD or Segmental
GmPOD130	GmPOD131	0.381	0.861	0.443	28.70	WGD or Segmental
GmPOD130	GmPOD132	0.174	0.508	0.343	16.92	WGD or Segmental
GmPOD131	GmPOD132	0.368	0.833	0.442	27.78	WGD or Segmental

Table 1. Ka, Ks, and Ka/Ks ratio values for the duplication of gene pairs.

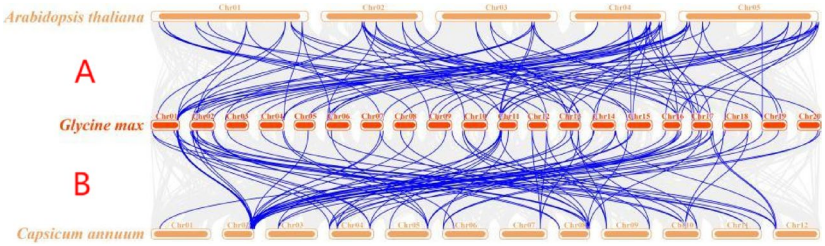

Fig. 3. Analysis of GmPOD genes in soybean alongside two other model plants (Arabidopsis and capsicum). (A) Comparison of Arabidopsis and soybean. (B) Comparison soybean and chili pepper. Gray lines indicate significant collinear regions within and between plant genomes; blue lines depict homologous GmPOD gene pairs. The chromosome number of is displayed above each chromosome.

members, and ten GmPOD members. Group VI had only seven members, *GmPOD41*, *GmPOD44*, *GmPOD46*, *GmPOD48*, *GmPOD94*, *GmPOD96*, and *GmPOD100*. These results suggest that soybean has evolved many of its PODs during its long-term domestication.

Gene structure and conserved domain analysis of GmPOD family

All members of the GmPOD gene family have numerous exons, according to an intron/exon makeup analysis of these genes. High degrees of conservation and tight evolutionary affinity were shown in the similar intron/exon architecture across the various GmPOD genes within each subgroup (Fig. 6).

Using the MEME suite (Version v5.5.3) (http://meme-suite.org/tools/meme/), we examined the conserved motifs of GmPOD proteins to obtain a thorough grasp of their functional variety. Ten conserved motifs (motifs

Fig. 4. A demonstration of the GmPOD protein interaction network. Each node represents a protein, and each connection represents an interaction. The purple lines represent protein interactions that have been verified through experimentation, the green lines represent protein interactions that are inferred from gene proximity, and the blue lines represent protein interactions that are inferred from gene co-presence.

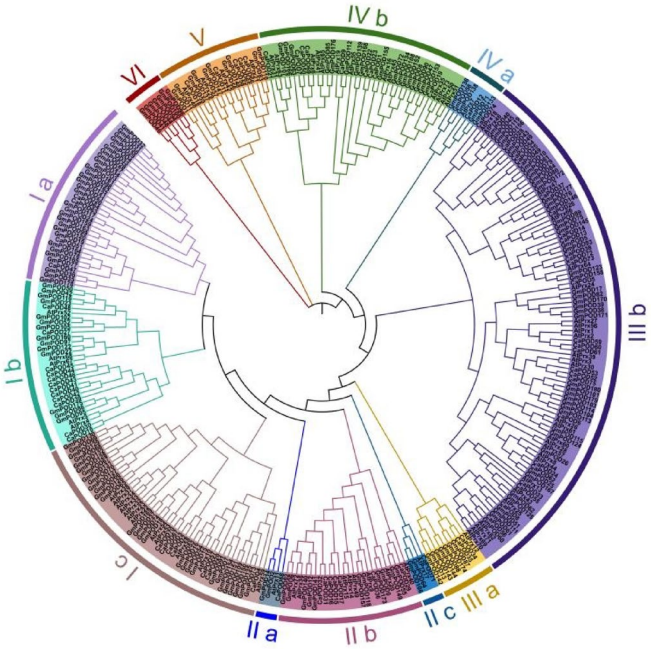

Fig. 5. Analysis of the phylogeny of proteins in soybean, Arabidopsis, and capsicum. The complete amino acid sequences of 180 GmPOD proteins, 73 AtPrx proteins, and 75 CaPOD proteins were aligned using the MUSCLE algorithm within MEGA 11.0. A phylogenetic tree was then built with 1000 bootstrap replicates through MEGA 11.0.

1–10) were found in the GmPOD protein sequences, as shown in Fig. 7. The phylogenetic tree and MEME analysis show that, although the motif compositions of GmPODs vary greatly between proteins in different phylogenetic groups, they are more consistent within the same group (Fig. 7). The motifs 1, 2, 3, 4, 5, 6, and 7 are common to many GmPODs. Subgroups II and IV are the only ones with motif 9. Overall, the structural domain compositions and configurations of proteins within the same subgroup are comparable, indicating that their activities may be similar, while variations between individual proteins can be used to deduce functional diversity.

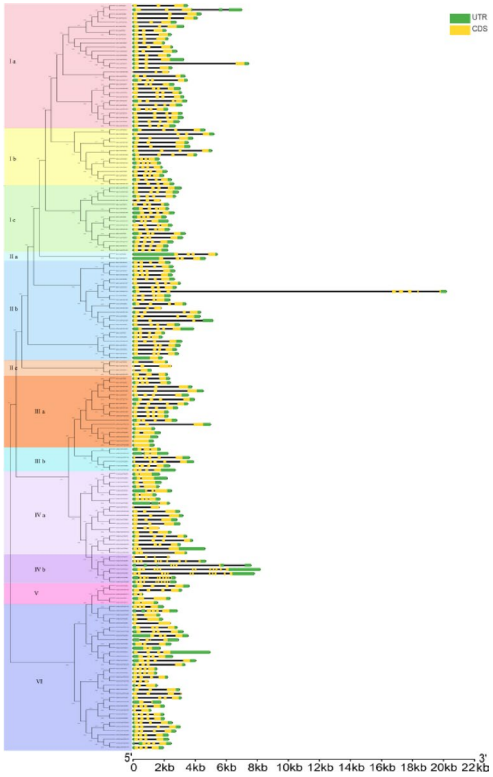

Fig. 6. Phylogenetic affiliation and gene architecture diagram for the soybean GmPOD gene family. With MEGA 11.0, the maximum likelihood approach was used to create the phylogenetic tree, which was backed with 1000 bootstrap repetitions for every branch. TBtools version 1.120 was used to visualize the genetic composition of GmPOD. Black lines represent introns, while a yellow box represents the coding sequence (CDS).

Analysis of cis-acting elements of GmPOD family

Using the PlantCARE database, we looked for cis-acting regulatory elements inside the 2000 bp promoter regions of the GmPOD genes in order to explore their possible functions and regulatory processes. Meja response elements (MeJARE), anaerobic response elements (ARE), ABA response elements (ABRE), drought response elements (DRE), low-temperature response elements (LTRE), gibberellin response elements (GARE), defensin and stress response elements (DSRE), SA response elements (SARE), and growth hormone response elements (AuxRE) are just a few of the cis-acting elements linked to plant hormone signaling and the stress responses that our analysis identified (Fig. 8). In all, 180 GmPOD genes were found to include the following elements: 152 ARE elements, 151 ABA elements, 123 MeJA elements, 95 GARE elements, 79 DRE elements, 76 DSRE elements, 71 SA and AuxRE components, and 50 LTRE elements (Fig. 8). These findings imply that these genes may be involved in several plant hormone signaling pathways and stress responses, suggesting that they are involved in how plants respond to their surroundings.

Analysis of GmPOD family expression in different tissues

We used the SoyMD database (Version v2.0) (available at https://yanglab.hzau.edu.cn/SoyMD/#/) to further explore the expression dynamics of GmPOD in soybean. The transcriptomic data in this database were derived from normally growing tissues under non-stress conditions. GmPOD gene expression profiles of several soybean tissues, for example the roots, sams, young leaves, mature leaves, flowers, seeds, embryos, and emdosperm, were examined. The constitutive expression of GmPOD gene family members varied noticeably among various organs, according to our findings (Fig. 9). With 175 genes found in at least one of the tissues under investigation, the heat map displayed the 180 GmPOD genes' expression patterns throughout these tissues (Fig. 9). Fourteen genes (*POD128*, *POD24*, *POD11*, *POD55*, *POD137*, *POD149*, *POD59*, *POD1*, *POD176*, *POD85*, *POD63*, *POD118*, *POD157*, and *POD141*) showed increased expression (TPM ≥ 3.0) in soybean floral tissues; three genes (*POD38*, *POD13*, and *POD53*) were comparatively high expressed (TPM ≥ 3.0) in soybean endosperm tissues; in soybean leaf tissues, three genes (*POD46*, *POD48*, and *POD105*) had comparatively high expression levels; four genes (*POD147*, *POD26*, *POD57*, and *POD75*) displayed comparatively high expression levels in soybean seed tissues (TPM ≥ 3.0); and, in soybean embryos, two genes (*POD45* and *POD112*) showed increased expression. Moreover, most genes were expressed at comparatively high levels in the tissues of soybean roots and stems (Fig. 9). According to these data, every GmPOD gene demonstrates a unique tissue-specific expression pattern, which may be related to their functions in coordinating plant development, growth, and stress responses.

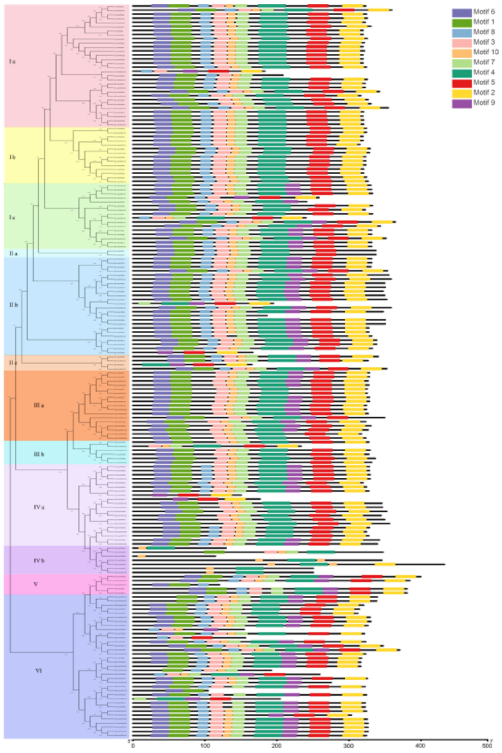

Fig. 7. Examination of the conserved motifs and evolutionary relationships within the soybean GmPOD gene family. Using MEGA 11.0, the maximum likelihood method was used to create the phylogenetic tree, and 1000 bootstrap replicates were used for each branch. Multiple Em Motif Induction (MEME) was used to find conserved motifs. Using TBtools version 1.120 and MEME.XML version 5.5.5, a model representing 10 motifs found in the whole GmPOD amino acid sequence was created. Every conserved motif is shown in the structural diagram as a uniquely colored box, oriented in a way that corresponds to the length of its sequence.

Analysis of GmPOD gene expression profile under heat stress

We used qRT-PCR to measure the expression of GmPODs after exposure to stress from high temperatures and humidity in order to explore their possible involvement in response to these circumstances. The research found that the nine genes *GmPOD2*, *GmPOD50*, *GmPOD67*, *GmPOD95*, *GmPOD135*, *GmPOD144*, *GmPOD152*, *GmPOD153* and *GmPOD168* cover the main subfamilies (Ib, Ic, IVb, etc.) of the GmPOD phylogenetic tree (Fig. 5). Ensured the characterization of the entire gene family. It has a complete CDS region and conserved functional motifs (Figs. 6 and 7), indicating potential enzymatic activity. They all exhibit sensitive characteristics to high humidity and high temperature stress. Interestingly, *GmPOD135* showed a quick rise in expression within a brief timeframe (Fig. 10a, b). Among these genes, leaves under high-temperature and -humidity stress showed greater expression levels of *GmPOD50*, *GmPOD67*, *GmPOD153*, and *GmPOD168*, from the Ib, Ic, and IVb families, indicating substantial involvement for these genes in stress response.

Using relative expression values, we then generated relative expression heat maps for these nine GmPOD genes to measure their sensitivity to high-temperature and -humidity stress (Fig. 10c), and all GmPODs showed higher expression after a stress treatment of 40 °C, compared to the 0-hour treatment. The relative expression levels of *GmPOD2* and *GmPOD135* in leaves reached their highest at 24 h under the 40 °C condition, whereas *GmPOD50* and *GmPOD67* reached their peaks at 48 h, as shown in Fig. 10c. In comparison to 0 h, the relative expression of *GmPOD67* rose by roughly 10 times at 24 and 48 h, while the expression of *GmPOD135* increased by nearly eight times at 24 h.

On the other hand, most genes' relative expression levels were mostly steady at 24 and 48 h at 20 °C. The genes *GmPOD50*, *GmPOD67*, *GmPOD153*, and *GmPOD168* had consistently elevated expression levels during the high-temperature and -humidity stress, implying that these genes may have an effect on soybean response to environmental stressors.

MiRNA targeting sites prediction

Using the plant small RNA target server psRNATarget (https://www.zhaolab.org/psRNATarget/analysis?function=2) with default parameters (expectation values set to ≤ 3.0), we identified a total of 49 miRNAs targeting ten out of total GmPOD (Fig. 11). *GmPOD170* was predicted to be targeted by the maximum number of miRNAs, while GmPOD168 was targeted by only one miRNA. Most miRNA is specific to its corresponding GmPOD gene, with the cleavage identified as the primary regulatory effect (Fig. 11). The possibility for miRNA targeting of those GmPOD genes, evaluated by expectation values from psRNATarget, was represented by a positive correlation with circle size in Fig. 11.

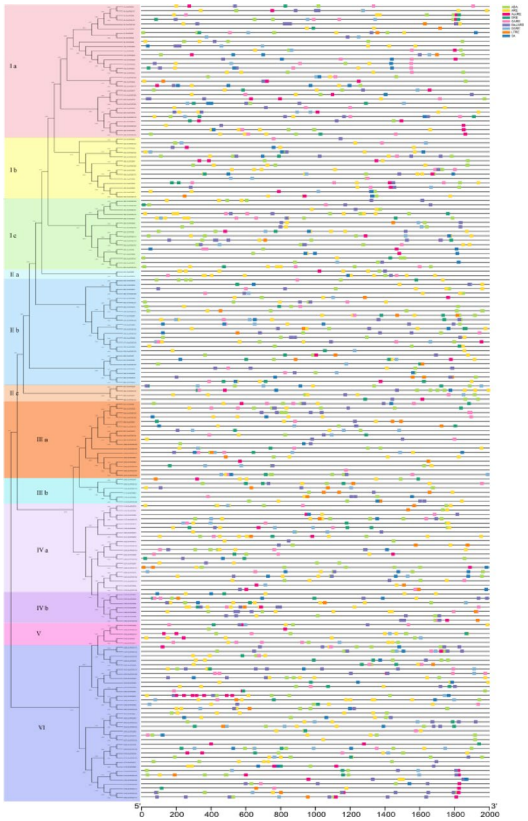

Fig. 8. The relationship between the 2000 bp promoter sequence and the soybean GmPOD gene family was investigated. Using MEGA 11.0, the maximum likelihood method, with 1000 bootstrap repeats per branch, was used to construct the phylogenetic tree. TBtools version 1.098661 was used to display the GmPOD 2000 bp promoter sequence, and the PlantCARE database was used for analysis. Several cis-acting regulatory components are represented with boxes of varying colors.

GmPOD protein structure analysis

We then utilized the AlphaFold2 online platform (https://alphafold.ebi.ac.uk/) to predict the 3D structures of these ten genes proteins (Fig. 12). Tertiary structural similarity among the GmPOD family members was assessed by using the TM-score tool (version v20190822) (http://zhanggroup.org/TM-score/). The results revealed that *GmPOD95*, *GmPOD168*, *GmPOD50*, and *GmPOD152* shared higher similarity (all pairwise TM-scores above 0.5), while *GmPOD135* displayed distinct structural features compared with the other members (Fig. 12). These findings are consistent with their subfamily classification based on protein sequence similarity.

Discussion

Class III POD enzymes play crucial roles in the control of many physiological processes in plants, especially when they are responding to various biotic and abiotic stressors[32,48]. Numerous plant species have been subject to extensive genome-wide investigations into the POD gene family, including *Arabidopsis thaliana*[32], *Populus trichocarpa*[49], *Zea mays*[5], *Pyrus bretschneideri*[48], *Manihot esculenta*[50], *Capsicum annuum*[51], *Litchi chinensis Sonn*[52], and *Brassica napus*[53]. In this study, 180 GmPOD genes were fully discovered in soybean, and their amino acid sequences and evolutionary connections allowed us to group the genes into six subfamilies (Fig. 5); this figure is higher than the numbers previously reported for Arabidopsis (73), capsicum (75), cassava (91), pilocarpa (93), alfalfa (102), and maize (111), suggesting that soybean has a significantly larger GmPOD gene family than other plant species[54]. Our subfamily classification highlights the conservation of these genes throughout many plant species and is consistent with previous findings for POD genes in other plants.

Moreover, investigations of conserved domains, motifs, and gene structures support our GmPOD subfamily categorization, showing that all subgroups have comparable domain arrangements, motif patterns, and exon/intron topologies (Figs. 6 and 7). Members of the same class of the GmPOD family were found to generally have comparable motif types and numbers in terms of conserved domains and motifs (Fig. 7); for example, similar domain configurations and motif compositions of GmPOD proteins tend to cluster into the same subfamily (Fig. 5), and the phylogenetic analysis results are consistent with the structural makeup of each GmPOD gene cluster. Numerous other species, including pears[48], maize[5], and potatoes[55], also exhibit this pattern, suggesting that GmPOD proteins with comparable domain compositions and motif patterns may have similar activities, as a protein's domain structure largely determines its activity.

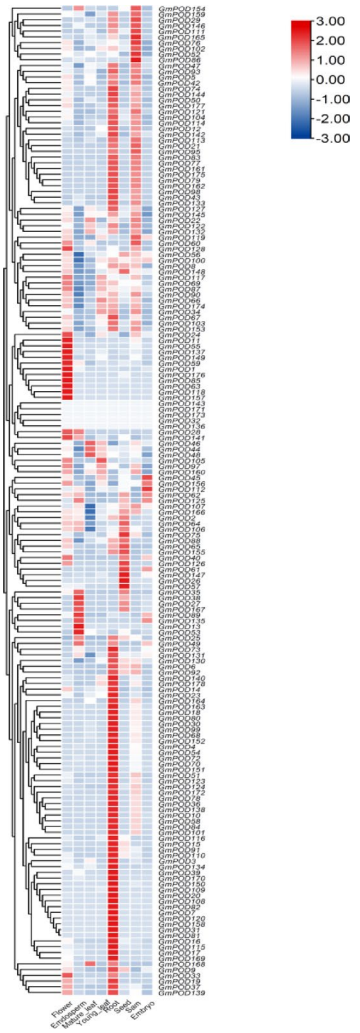

Fig. 9. Expression profiles of GmPOD gene in eight different tissues of soybean: root, stem, leaf, bud, pod, tip, callus and seedling. GmPOD transcriptome analysis of gene expression data from SoyMD database (https://yanglab.hzau.edu.cn/SoyMD/#/) to download. Heat maps were constructed by TBtools (v.1.120) using Log2 (TPM + 1) values for each gene, ranging from low expression levels (blue) to high expression levels (red).

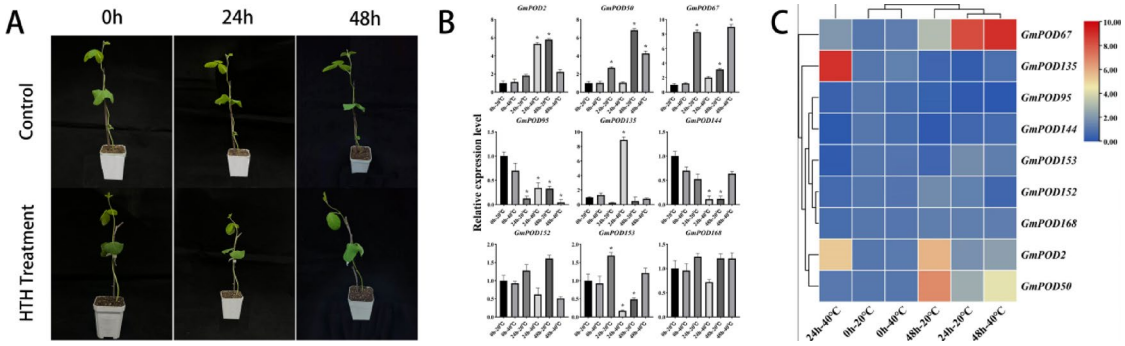

Fig. 10. Dynamics of GmPOD gene expression in soybean. (A) Leaf phenotype of soybean treated with different stress durations. (B) The GmPOD genes' relative expression levels in leaves after 24 and 48 h at 20 and 40 degrees Celsius. The three biological replicates' means and standard deviations are shown in the data. Duncan's multiple range test yielded statistical significance levels, indicated with asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (C) Heat maps showing the expressions of GmPOD genes in leaves at different temperatures (20 °C and 40 °C). The graph's color gradients show the relative expression levels of each gene in several samples.

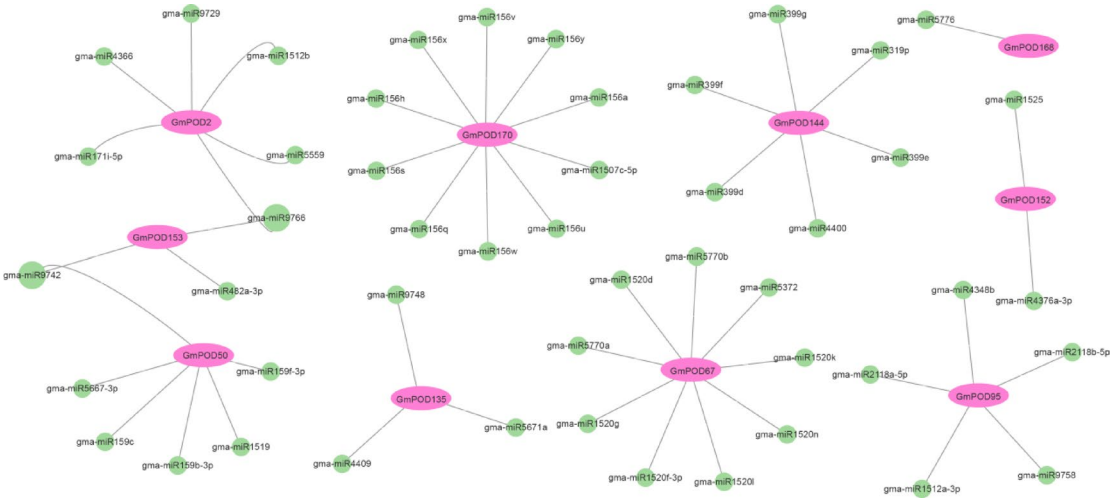

Fig. 11. Prediction of miRNA targeting 10 genes in *Glycine max*. Pink oval shapes indicate those ten family members, and green solid circles denote the targeting miRNAs. Circle size of each miRNA inversely correlates with the expectation values determined by psRNATarget, with smaller circles indicating higher possibility of target accessibility.

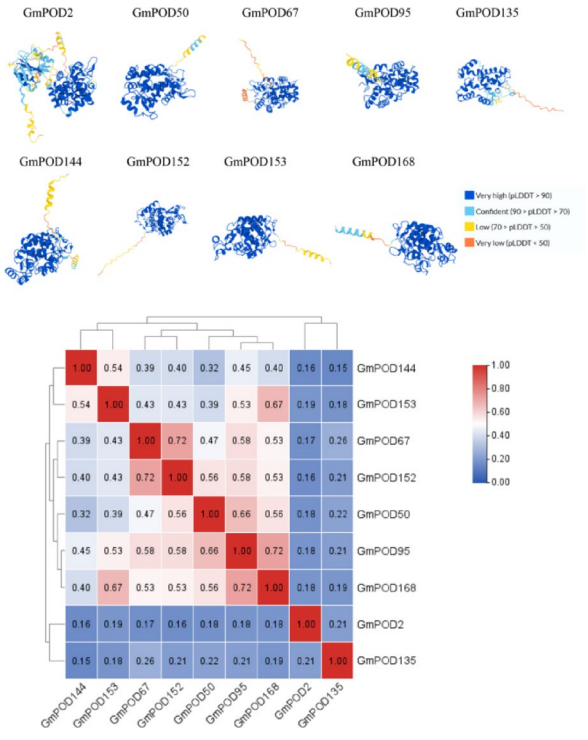

Fig. 12. GmPOD Protein Structure Analysis. Predicted 3D structures of proteins encoded by ten genes and tertiary structural similarity analysis within the GmPOD family.

Several studies have demonstrated that, over the course of evolution, introns are progressively integrated and maintained within the genome[56]. Intron deletion and insertion are frequent occurrences that can lead to a wide range of functional alterations in the evolution of genes. Genes with a comparatively large number of introns may result from old genes being replicated, which might have started with minor exon shuffling[57]. Analysis of the 180 GmPOD genes' structures showed that there was a wide range of variation in the number of introns (zero to six), with most of the genes having two introns. Multiple introns indicate that the GmPOD gene structure has changed over time. As seen in Figs. 6 and 7, proteins belonging to the same subfamily notably differ in intron/exon architecture and motifs, which suggests that exon shuffling may be a common evolutionary feature of GmPODs and a major contributor to the functional variety of the soybean GmPOD family. The development of

multigene families is often influenced by variations in intron/exon structure; of the 180 GmPOD genes, more than half are multigene assemblies. This phenomenon is especially clear in the 4-intron/2-exon model, which is representative of an ancestral intron pattern for the POD gene family and is present in a significant fraction of Arabidopsis[32] and pepper[33] genes, in addition to a large number of GmPOD genes.

Replication is one of the main processes driving genome evolution[57], and segmental and tandem replication are two types of gene replication processes that are essential to the development of gene families[58]. Two gene copies may result from these occurrences, and under lessening evolutionary selection forces, one or both copies may acquire new roles[59], and, because each paralog is specialized for a certain task, gene families are expanding[60]. Tandem replication, which frequently results from uneven crossing over many crossing events, is the duplication of one or more neighboring genes[61]. Large genomic regions can be duplicated as a consequence of chromosomal, fragment, or genome-wide replication, often resulting in multiple losses and rearrangements[60], referred to as fragment replication events. Soybean has had two whole-genome duplications (WGD) and one whole-genome triplication (WGT)[62,63]; as a result, the soybean genome has undergone high levels of replication and has many copies of its genes. Within the GmPOD gene family of soybean, we found 86 duplicate gene pairs in total (Table 1), of which 56 are segmental duplicate gene pairs and 30 are tandem duplicate gene pairs; this number is significantly higher than the numbers of duplicate gene pairs found in pear (26 pairs)[48], maize (28 pairs)[5], and rice (48 pairs)[6]. Given that segmental duplicate gene pairs outnumber tandem duplicate gene pairs, it is likely that segmental duplication had a major role in the GmPOD gene family's proliferation in soybean. The segmental and tandem repeat gene pairs' Ka/Ks ratios were both much lower than one, showing that selection of purification was a major factor in the development of the GmPOD gene family in soybean.

The GmPOD gene family has six subfamilies, according to phylogenetic research. Subgroup VI genes stand out from those of the other five subgroups due to their unique multiple-exon signatures, a discovery consistent with earlier research in watermelon[64], Chinese pears[48], and cassava[50], thus indicating that POD subgroup VI may have a common ancestor with other plants and may have particular purposes in these plants. Gene regulatory networks (GRNs) are complex systems that control growth and development in plants; they comprise transcription factors, regulatory RNAs, and enzymes[65], and the plant's response to various environmental stress conditions is largely mediated by these GRNs. The defense regulatory system of plants is stimulated, and downstream gene transcription is activated, via the interaction of transcription factors with cis-regulatory elements in the promoter region[66,67], a system essential for controlling plant growth, development, and stress tolerance. In order to acquire more understanding of GmPOD regulation, we investigated the cis-acting components located in the 2.0 kb promoter region. MeJARE, ARE, ABRE, DRE, LTRE, GARE, DSRE, SARE, and AuxRE are examples of cis-acting elements whose functional annotations demonstrate their participation in a variety of regulatory processes, including development, hormone signaling, cell cycle regulation, abiotic/biotic stress responses, and transcription; these results provide a basic knowledge of the function of the GmPOD gene in several plant processes and responses to stress. Auxin, gibberellin, brassinosteroid, ABA, and MeJA are important plant hormones that influence how a plant reacts to stressors such as high temperatures and high humidity[68]; for example, MeJA specifically increases the activity of antioxidant enzymes to protect cell membranes and support the cells' capacity for osmotic control when plants are subjected to stressors such as high temperatures and humidity.

Gene function may often be inferred from patterns of gene expression[69,70]; for a number of plant species, including pears[48], tobacco[71], potatoes[54], and peppers[33], POD gene expression has already been observed and assessed. Thus, we used the SoyMD database (https://yanglab.hzau.edu.cn/SoyMD/#/) to investigate the GmPOD expression levels in different soybean tissues, such as roots, stems, leaves, buds, seed tissues, apical tissues, calli, and seedlings, the results of which were displayed using TBtools software. Out of the 180 GmPOD genes, only 5 were found to be either hardly expressed or not expressed at all in the tissues that were analyzed. The remaining genes appeared to play crucial functions in soybean growth and development, as they were expressed in different tissues and helped to create different plant tissues (Fig. 9), results which support previous studies showing the importance of GmPOD expression in plant growth and development[5,32,33,49,50] [5,33,34,50,51,]. Group IV includes six comparatively high expressed genes in all tissues (*GmPOD8*, *GmPOD90*, *GmPOD107*, *GmPOD112*, *GmPOD153*, and *GmPOD166*), suggesting that these genes may play important maintenance functions in soybean cells grown under typical circumstances; nonetheless, distinct tissues express some GmPOD genes differently. *GmPOD26*, *GmPOD57*, *GmPOD75*, and *GmPOD147*, for example, were comparatively high expressed in seeds but barely expressed at all in leaves, indicating that their primary function is in seed growth and development. It is also noteworthy that the roots included the greatest number of GmPOD genes with comparatively high expression levels (Fig. 9). These findings are in line with research on rice[6], Arabidopsis[32], and maize[5], suggesting that the POD family may be essential to plant root function.

Additionally, it was discovered that some GmPOD genes—particularly, *GmPOD44*, *GmPOD46*, *GmPOD105*, *GmPOD141*, and *GmPOD168*—were strongly expressed in leaves; meanwhile, in blooming tissues, *GmPOD1*, *GmPOD11*, *GmPOD24*, *GmPOD55*, and *GmPOD59* all showed increased expression (Fig. 9). These expression patterns, distinct to different tissues, point to the functional specialization of these genes and the important functions they play in those tissues. Nine GmPOD genes showed notable changes in expression under high-temperature and -humidity stress, according to an assessment of relative gene expression levels. These genes belonged to the subfamilies Ib, Ic, IIa, IIIa, and IVb, and included components that were sensitive to abiotic stressors, including DRE and ABA; additionally, cis-regulatory elements linked to hormones, such as ABA, GARE, and AuxRE, were found in most of these genes.

The results of our expression study showed that, under conditions of high temperature and high humidity, the expression levels of GmPOD50, *GmPOD67*, *GmPOD152*, *GmPOD153*, and *GmPOD168* in leaves increased initially, and subsequently decreased over the course of 24 h, and, over the course of 48 h, this pattern continued for *GmPOD50* and *GmPOD152*. On the other hand, the expression levels of *GmPOD95* and *GmPOD144* first

dropped, and subsequently rose. Under conditions of high-temperature and -humidity stress, the Ib, Ic, and IVb families of genes, including *GmPOD50*, *GmPOD67*, *GmPOD153*, and *GmPOD168*, showed greater expression levels in leaves. After 48 h of stress treatment, the relative expression levels of four genes (*GmPOD50*, *GmPOD67*, *GmPOD153*, and *GmPOD168*) increased in comparison to those of the 0-hour control group, indicating that these genes are important for soybean response to these conditions. Plant hormones, which control important signaling molecules involved in plant growth, development, and stress response, are mostly found in these genes; thus, their reaction to high humidity and high temperature may be related to hormone control, as evidenced by the upregulation of their expression levels.

MicroRNAs (miRNAs), 21–24 nucleotide (nt) short small non-coding RNAs, regulate gene expression at the post-transcriptional level by targeting complementary mRNAs for cleavage or translational repression, playing critical roles in diverse plant developmental processes. According to the prediction results of miRNA targeting sites in Fig. 11, *GmPOD170* was targeted by the largest number of miRNAs, while *GmPOD168* was targeted by only one miRNA. Most mirnas are specific to their corresponding GmPOD gene, and cleavage was identified as the primary regulatory role (Fig. 11). Structural analysis of GmPOD proteins showed that *GmPOD95*, *GmPOD168*, *GmPOD50*, and *GmPOD152* had higher similarity, while GmPOD135 showed different structural features compared to the other members (Fig. 12). These findings are consistent with subfamily classification based on protein sequence similarity.

According to current forecasts, within the next 20 years, the average global temperature of Earth is predicted to rise by 1.5 degrees Celsius[72], while studies have shown that there would be a 3.1% drop in worldwide soybean yield for each degree Celsius the average global temperature rises[21]. For this reason, it is essential to comprehend how soybean plants react to stress from high temperatures and high humidity, as well as to improve crop tolerance to these circumstances, in order to promote human growth and ensure worldwide food security. In order to identify genes for molecular breeding techniques through which to address the problems of soybean growth under high-temperature and -humidity stress in this study, we used the expression pattern of the GmPOD gene family. For example, under the high-temperature and high-humidity stress treatment, *GmPOD50*, *GmPOD67*, *GmPOD153*, and *GmPOD168* consistently showed comparatively high expression levels, suggesting that these genes may be essential for soybean's reactions to such circumstances. A theoretical foundation is provided through the examination of cis-acting components in the promoter region for further research into the regulation processes of different variables on the GmPOD family, which may have complex roles in hormone regulation or signal transduction, as suggested by the existence of many hormone-related cis-acting elements; nevertheless, further investigation is required to verify these theories.

Conclusions

In the current research, a detailed whole-genome analysis of GmPOD gene families in soybean was carried out, and 180 GmPOD genes were identified. Based on their evolutionary links, these genes were arranged into six subfamilies and found to be distributed throughout 21 chromosomes. We evaluated the structures, conserved motifs, cis-acting regions, protein interaction networks, and homology of these genes, performed miRNA targeting site prediction and GmPOD protein structure analysis to better understand the evolutionary connections among members of the GmPOD gene family. In addition, we used qRT-PCR to explore the GmPOD gene expression patterns in different organs under different stress scenarios, providing information for further studies on the GmPOD genes' functions. The finding of this research supply a theoretical framework for investigating the molecular regulatory systems of soybean under stress from high temperatures and high humidity and lay the groundwork for discovering important GmPOD candidate genes in soybean that are sensitive to these stressors.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 5 June 2025; Accepted: 29 October 2025
Published online: 24 November 2025

References

1. Hiraga, S., Ichinose, C., Onogi, T., Niki, H. & Yamazoe, M. Bidirectional migration of SeqA-bound hemimethylated DNA clusters and pairing of OriC copies in *Escherichia coli*. *Genes Cells* **5**(5), 327–341. https://doi.org/10.1046/j.1365-2443.2000.00334.x (2000).
2. Mathé, C., Barre, A., Jourda, C. & Dunand, C. Evolution and expression of class III peroxidases. *Arch. Biochem. Biophys.* **500** (1), 58–65. https://doi.org/10.1016/j.abb.2010.04.007 (2010).
3. Cosio, C. & Dunand, C. Specific functions of individual class III peroxidase genes. *J. Exp. Bot.* **60**(2), 391–408. https://doi.org/10.1093/jxb/ern318 (2009).
4. Almagro, L. et al. Class III peroxidases in plant defence reactions. *J. Exp. Bot.* **60** (2), 377–390. https://doi.org/10.1093/jxb/ern277 (2008).
5. Wang, Y., Wang, Q., Zhao, Y., Han, G. & Zhu, S. Systematic analysis of maize class III peroxidase gene family reveals a conserved subfamily involved in abiotic stress response. *Gene* **566** (1), 95–108. https://doi.org/10.1016/j.gene.2015.04.041 (2015).
6. Passardi, F., Longet, D., Penel, C. & Dunand, C. The class III peroxidase Multigenic family in rice and its evolution in land plants☆☆☆. *Phytochemistry* **65**(13), 1879–1893. https://doi.org/10.1016/j.phytochem.2004.06.023 (2004).
7. Kidwai, M., Ahmad, I. Z. & D Chakrabarty Class III peroxidase: an indispensable enzyme for biotic/abiotic stress tolerance and a potent candidate for crop improvement. *Plant Cell Rep.* **39** (11), 1381–1393. https://doi.org/10.1007/s00299-020-02588-y (2020).
8. Pandey, V. P., Awasthi, M., Singh, V. P., Tiwari, S. & Dwivedi, U. N. A comprehensive review on function and application of plant peroxidases. *Biochem. Anal. Biochem.* https://doi.org/10.4172/2161-1009.1000308 (2017).
9. González, A. M. et al. Peroxidase profiling reveals genetic linkage between peroxidase gene clusters and basal host and Non-Host resistance to rusts and mildew in barley. *PLOS ONE.* **5** (8), e10495–e10495. https://doi.org/10.1371/journal.pone.0010495 (2010).

10. Jin, J., Hewezi, T. & T J. Baum Arabidopsis peroxidase AtPRX53 influences cell elongation and susceptibility ToHeterodera schachtii. *Plant Signal. Behav.* **6** (11), 1778–1786. https://doi.org/10.4161/psb.6.11.17684 (2011).
11. Jäggi, M., Kumar, S. & Sinha, A. Overexpression of an apoplastic peroxidase gene CrPrx in Transgenic hairy root lines of catharanthus roseus. *Appl. Microbiol. Biotechnol.* **90**(3), 1005–1016. https://doi.org/10.1007/s00253-011-3131-8 (2011).
12. Wu, Y. et al. Overexpression of a peroxidase gene (AtPrx64) of Arabidopsis Thaliana in tobacco improves plant's tolerance to aluminum stress. *Plant Mol. Biol.* **95** (1–2), 157–168. https://doi.org/10.1007/s11103-017-0644-2 (2017).
13. Choi, H. W., Kim, Y. J., Lee, S. C., Hong, J. K. & Hwang, B. K. Hydrogen peroxide generation by the pepper extracellular peroxidase CaPO2 activates local and systemic cell death and defense response to bacterial pathogens. *Plant Physiol.* **145** (3), 890–904. https://doi.org/10.1104/pp.107.103325 (2007).
14. Mei, W., Qin, Y-M., Song, W-Q., Li, J. & Zhu, Y. Cotton GhPOX1 encoding plant class III peroxidase May be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. *J. Genet. Genomics.* **36** (3), 141–150. https://doi.org/10.1016/s1673-8527(08)60101-0 (2009).
15. Dong, L. et al. Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. *Nat. Commun.* https://doi.org/10.1038/s41467-021-25800-3 (2021).
16. Liu, Y. et al. Heat stress in legume seed setting: Effects, Causes, and future prospects. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2019.00938 (2019).
17. Nadeem, M. A. et al. Unraveling field crops sensitivity to heat stress: Mechanisms, Approaches, and future prospects. *Agronomy* **8** (7), 128–128. https://doi.org/10.3390/agronomy8070128 (2018).
18. Mourtzinis, S. et al. Climate-induced reduction in US-wide soybean yields underpinned by region- and in-season-specific responses. *Nat. Plants* https://doi.org/10.1038/nplants.2014.26 (2015).
19. Saha, R. R. & Sultana, W. Influence of seed ageing on growth and yield of soybean. *Bangladesh J. Bot.* **37** (1), 21–26 (2008).
20. Kebede, H., Smith, J. R. & Ray, J. D. A new gene that controls seed coat wrinkling in soybean. *Euphytica* **189** (2), 309–320. https://doi.org/10.1007/s10681-012-0818-6 (2012).
21. Zhao, C. et al. Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci. U.S.A.* **114** (35), 9326–9331. https://doi.org/10.1073/pnas.1701762114 (2017).
22. Li, J. et al. Physiological and molecular mechanisms of leaf response to high-temperature stress in high-temperature-resistant soybean varieties. *Res. Square (Research Square)* https://doi.org/10.21203/rs.3.rs-3844375/v1 (2024).
23. Wistrand, M. & Sonnhammer, E. L. L. Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER. *BMC Bioinform.* https://doi.org/10.1186/1471-2105-6-99 (2005).
24. Wilkins, M. R. et al. Protein identification and analysis tools in the expasy server. *Methods Mol. Biology (Clifton N J)*. **112**, 531–552. https://doi.org/10.1385/1-59259-584-7:531 (1999).
25. Gasteiger, E. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31** (13), 3784–3788. https://doi.org/10.1093/nar/gkg563 (2003).
26. Chen, C. et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant.* **13**, 1194–1202 (2020).
27. Krzywinski, M. et al. An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
28. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40** (7), e49–e49. https://doi.org/10.1093/nar/gkr1293 (2012).
29. Wang, D., Liu, F., Wang, L. & Huang, S. Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biol. Direct* https://doi.org/10.1186/1745-6150-6-13 (2011).
30. Aleem, M. et al. Genome-wide characterization and functional analysis of class III peroxidase gene family in soybean reveal regulatory roles of GsPOD40 in drought tolerance. *Genomics* **114** (1), 45–60. https://doi.org/10.1016/j.ygeno.2021.11.016 (2022).
31. Bailey, T. L., Johnson, J. E., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**(W1), W39–W49. https://doi.org/10.1093/nar/gkv416 (2015).
32. Tognolli, M., Penel, C., Greppin, H. & Simon, P. Analysis and expression of the class III peroxidase large gene family in Arabidopsis Thaliana. *Gene* **288** (1–2), 129–138. https://doi.org/10.1016/s0378-1119(02)00465-1 (2002).
33. González-Gordo, S., Muñoz-Vargas, M. A., Palma, J. M. & Corpas, F. J. Class III peroxidases (POD) in pepper (Capsicum annuum L.): Genome-Wide identification and regulation during nitric oxide (NO)-Influenced fruit ripening. *Antioxidants* **12** (5), 1013–1013. https://doi.org/10.3390/antiox12051013 (2023).
34. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**(6), 1547–1549. https://doi.org/10.1093/molbev/msy096 (2018).
35. Shiu, S-H. & Bleecker, A. B. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U.S.A.* **98** (19), 10763–10768. https://doi.org/10.1073/pnas.181141598 (2001).
36. Letunić, I. & Bork, P. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**(suppl), W475–W478. https://doi.org/10.1093/nar/gkr201 (2011).
37. Lescot, M. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in Silico analysis of promoter sequences. *Nucleic Acids Res.* **30** (1), 325–327. https://doi.org/10.1093/nar/30.1.325 (2002).
38. Yu, J. et al. Genome-Wide identification and expression profiling of tomato Hsp20 gene family in response to biotic and abiotic stresses. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2016.01215 (2016).
39. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. MiRBase: from MicroRNA sequences to function. *Nucleic Acids Res.* **47** (D1), D155–D162. https://doi.org/10.1093/nar/gky1141 (2019).
40. Pandey, A. et al. Genome-wide identification of the fibrillin gene family in Chickpea (*Cicer arietinum* L.) and its response to drought stress. *Int. J. Biol. Macromol.* **234**, 123757–123757. https://doi.org/10.1016/j.ijbiomac.2023.123757 (2023).
41. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using Real-Time quantitative PCR and the 2− ΔΔCT method. *Methods* **25** (4), 402–408. https://doi.org/10.1006/meth.2001.1262 (2001).
42. Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* **18** (9), 486–487. https://doi.org/10.1016/s0168-9525(02)02722-1 (2002).
43. Hu, R. et al. Genome-Wide Identification, evolutionary Expansion, and expression profile of Homeodomain-Leucine zipper gene family in Poplar (Populus trichocarpa). *PLOS ONE.* **7** (2), e31149–e31149. https://doi.org/10.1371/journal.pone.0031149 (2012).
44. Rensing, S. A. Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant. Biol.* **17**, 43–48. https://doi.org/10.1016/j.pbi.2013.11.002 (2014).
45. Panchy, N., Lehti-Shiu, M. D. & Shiu, S-H. Evolution of gene duplication in plants. *Plant Physiol.* **171** (4), 2294–2316. https://doi.org/10.1104/pp.16.00523 (2016).
46. Zhou, J., Xiong, W. & Yang, W. Protein function prediction based on PPI networks: network reconstruction vs edge enrichment. *Front. Genet.* https://doi.org/10.3389/fgene.2021.758131 (2021).
47. Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13** (11), 2498–2504. https://doi.org/10.1101/gr.1239303 (2003).
48. Cao, Y. et al. Structural, Evolutionary, and functional analysis of the class III peroxidase gene family in Chinese Pear (*Pyrus bretschneideri*). *Front. Plant Sci.* https://doi.org/10.3389/fpls.2016.01874 (2016).
49. Lin, R. et al. Subcellular relocalization and positive selection play key roles in the retention of duplicate genes ofpopulusclass III peroxidase family. *Plant. Cell.* **26** (6), 2404–2419. https://doi.org/10.1105/tpc.114.124750 (2014).

50. Wu, C. et al. The class III peroxidase (POD) gene family in cassava: Identification, Phylogeny, Duplication, and expression. *Int. J. Mol. Sci.* **20** (11), 2730–2730. https://doi.org/10.3390/ijms20112730 (2019).
51. Lüthje, S., Meisrimler, C-N., Hopff, D. & Möller, B. Phylogeny, topology, structure and functions of membrane-bound class III peroxidases in vascular plants. *Phytochemistry* **72** (10), 1124–1135. https://doi.org/10.1016/j.phytochem.2010.11.023 (2011).
52. Yang, C. & Fan,. Genome-Wide Identification and Expression Analysis of the Class III Peroxidase Gene Family under Abiotic Stresses in Litchi (*Litchi chinensis* Sonn). *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms25115804 (2024).
53. Shah, O. U. et al. Genome-Wide investigation of class III peroxidase genes in brassica Napus reveals their responsiveness to abiotic stresses. *Plants* **13** (7), 942. https://doi.org/10.3390/plants13070942 (2024).
54. Yang, X. et al. Genome-Wide identification and expression analysis of the class III peroxidase gene family in potato (*Solanum tuberosum* L). *Front. Genet.* https://doi.org/10.3389/fgene.2020.593577 (2020).
55. Rogozin, I. B., Wolf, Y. I., Sorokin, A. V., Mirkin, B. & Koonin, E. V. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* **13**(17), 1512–1517. https://doi.org/10.1016/s0960-9822(03)00558-x (2003).
56. Gilbert, W., de Souza, S. J. & Long, M. Origin of genes. *Proc. Natl. Acad. Sci. U.S.A.* **94** (15), 7698–7703. https://doi.org/10.1073/pnas.94.15.7698 (1997).
57. Moore, R. C. & Purugganan, M. D. The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. U.S.A.* **100** (26), 15682–15687. https://doi.org/10.1073/pnas.2535513100 (2003).
58. Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D. & May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis Thaliana. *BMC Plant Biol.* **4**, 10. https://doi.org/10.1186/1471-2229-4-10 (2004).
59. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**(10), 725–732. https://doi.org/10.1038/nrg2600 (2009).
60. Zhang, J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18** (6), 292–298. https://doi.org/10.1016/s0169-5347(03)00033-8 (2003).
61. Achaz, G., Coissac, É., Viari, A. & Netter, P. Analysis of intrachromosomal duplications in yeast Saccharomyces cerevisiae: A possible model for their origin. *Mol. Biol. Evol.* **17** (8), 1268–1275. https://doi.org/10.1093/oxfordjournals.molbev.a026410 (2000).
62. Severin, A. J., Cannon, S. B., Graham, M. M., Grant, D. & Shoemaker, R. C. Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy. *Plant. Cell.* **23** (9), 3129–3136. https://doi.org/10.1105/tpc.111.089573 (2011).
63. Schmutz, J. et al. Genome sequence of the palaeopolyploid soybean. *Nature* **463** (7278), 178–183. https://doi.org/10.1038/nature08670 (2010).
64. Yang, T. et al. Genome-Wide analysis of the peroxidase gene family and verification of lignin Synthesis-Related genes in watermelon. *Int. J. Mol. Sci.* **23** (2), 642–642. https://doi.org/10.3390/ijms23020642 (2022).
65. Wils, C. R. & Kaufmann, K. Gene-regulatory networks controlling inflorescence and flower development in *Arabidopsis thaliana*. *Biochim. Biophys. Acta (BBA) - Gene Regul. Mech.* **1860**(1), 95–105. https://doi.org/10.1016/j.bbagrm.2016.07.014 (2017).
66. Ibraheem, O., Botha, C. E. J. & Bradley, G. In Silico analysis of cis-acting regulatory elements in 5′ regulatory regions of sucrose transporter gene families in rice (*Oryza sativa* Japonica) and *Arabidopsis thaliana*. *Comput. Biol. Chem.* **34**(5–6), 268–283. https://doi.org/10.1016/j.compbiolchem.2010.09.003 (2010).
67. Su, L., Wan, S., Zhou, J., Shao, Q. & Xing, B. Transcriptional regulation of plant seed development. *Physiol. Plant.* **173** (4), 2013–2025. https://doi.org/10.1111/ppl.13548 (2021).
68. Waadt, R. et al. Plant hormone regulation of abiotic stress responses. *Nat. Rev. Mol. Cell Biol.* **23** (10), 680–694. https://doi.org/10.1038/s41580-022-00479-6 (2022).
69. Guo, M. et al. Genome-wide analysis of the CaHsp20 gene family in pepper: comprehensive sequence and expression profile analysis under heat stress. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2015.00806 (2015).
70. Zhang, H.-X. et al. Transcription factor CaSBP12 negatively regulates salt stress tolerance in pepper (*Capsicum annuum* L). *Int. J. Mol. Sci.* **21**(2), 444–444. https://doi.org/10.3390/ijms21020444 (2020).
71. Cheng, L., Ma, L., Meng, L., Shāng, H. & Cao, P. Genome-Wide identification and analysis of the class III peroxidase gene family in tobacco (*Nicotiana tabacum*). *Front. Genet.* https://doi.org/10.3389/fgene.2022.916867 (2022).
72. Intergovernmental Panel on Climate Change. Climate change 2021: The physical science basis. https://doi.org/10.1017/9781009157896 (2023).

Acknowledgements

The authors thank Bread of Life Farm for the experimental materials provided.

Author contributions

Sushuang Liu: Writing – review & editing, Writing – original draft, Conceptualization. Yizhou Wu: Writing – review & editing, Writing – original draft, Conceptualization, Software. Yingsheng Qiu: Software, Methodology. Dandan He: Software, Methodology. Kangqi Zhao: Methodology, Investigation. Liuyang Yang: Supervision, Investigation. Yang Li: Supervision, Conceptualization. Zaibao Zhang: Formal analysis, Conceptualization. Huasong Zou: Methodology, Supervision. Yanmin Liu: Data curation, Project administration, Funding acquisition.

Funding

This research was financially supported by the Huzhou Natural Science Foundation (Grant No. 2022YZ32), the 2024 Higher Education Research Program by the Zhejiang Province Association of Higher Education (Grant No. KT2024035), the Scientific Research Fund of the Zhejiang Provincial Education Department (Grant No. Y202248468), the National Training Programs of Innovation and Entrepreneurship for Undergraduates (Grant No. 202313287004; 202313287010), and Zhejiang Students' Technology and Innovation Program (Xin Miao Talents Program) (Grant No. 2024R438A003).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-26477-0.

Correspondence and requests for materials should be addressed to Y.L.