

Identification and characterization of pathogen recognition genes families in *Gastrodia elata*: Insights into their roles in growth and disease resistance

Linshuang Tang

Guizhou University

Shanshan Luo

Guizhou University

Dachang Wang

Guizhou University

Xinyun Liu

Guizhou University

Xiaofeng Wang

Guizhou University

Mingyan Tan

Guizhou University

Bing He

Guizhou University

Zhibo Zhao

Guizhou University

Mingjin Huang

mjhuang@gzu.edu.cn

Guizhou University

Research Article

Keywords: Pathogen recognition genes, *Gastrodia elata*, Cis-regulatory elements, Transcriptomes

Posted Date: October 23rd, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7424538/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on February 23rd, 2026.

See the published version at <https://doi.org/10.1186/s12870-026-08343-x>.

Abstract

Background *Gastrodia elata* Bl., a perennial herb of the Orchidaceae family whose dried tubers are used as medicine, requires symbiosis with at least two fungal species for growth and development. This study aims to identify disease-resistance gene families in the dark red *G. elata* genome and analyze their structural and expression characteristics, providing a theoretical basis for disease-resistance breeding and cultivation.

Result We identified 68 pathogen recognition genes in *G. elata* (*GePRGs*). This included phylogenetic analysis, classification, chromosomal distribution, protein motifs, gene structure organization, and gene duplication events, revealing substantial evolutionary diversity. The *G. elata* resistance gene family has undergone tandem and segmental duplications, with three identified tandem duplication events. Additionally, 55 *GePRG* genes show syntenic relationships with homologous genes in other plant species, suggesting that these genes have maintained functional stability throughout evolution. A total of 101 *cis*-regulatory elements (CREs) were identified in the 2,000 bp upstream region of the translation start codon (ATG) of *GePRGs*. These CREs include many environmental stress-related elements, such as MYB binding sites, dehydration-responsive element (DRE) core, and low-temperature responsive elements, indicating that these genes play critical roles in responding to abiotic stresses (such as drought, cold, and oxidative stress) and biotic stresses (such as pathogen defense). The expression analysis of *GePRG* genes at different developmental stages of *G. elata* showed that certain genes, such as *GeLRR-RLK12*, *GeLec-RLK15*, and *GeLRR-RLK28*, are highly expressed in early developmental stages (eg. immature tuber, S1). From the S1 to capsules (S5) stages of dark red *G. elata*, the expression levels of genes *GeLRR-RLK24*, *GeLec-RLK3*, *GeLec-RLK2*, and *GeLec-RLK8* are generally upregulated, indicating their potential roles in tissue growth, maturation, and immune regulation. The *G. elata* f. *viridis* with light green stems (TMQG) genotype exhibited the highest expression levels among various *G. elata* genotypes. Specifically, *GeLRR-RLK17*, *GeLRR-RLK36*, *GeLYK1*, and *GeLec-RLK13* were highly expressed in TMQG, making them promising molecular markers for resistance breeding.

Conclusion This study analyzes gene expression profiles and CREs in *G. elata*'s disease resistance genes, highlighting the need for further experiments to confirm their roles in growth and disease resistance, and offering valuable genetic resources for breeding. These genes of *GeLRR-RLK17*, *GeLRR-RLK36*, *GeLYK1*, and *GeLec-RLK13* are likely to play essential roles in enhancing the disease resistance and promoting the growth and development of *G. elata*.

Introduction

Plants in their natural habitats face daily threats from pathogens, and unlike mammals, they lack mobile defensive cells and an adaptive somatic immune system [1]. To survive, plants have developed two primary systems to protect themselves from pathogen invasion. The first is pattern-triggered immunity (PTI), which involves plant pattern recognition receptors (PRRs) located on the cell membrane surface recognizing relatively conserved microbe-associated molecular patterns (MAMPs) or pathogen-

associated molecular patterns (PAMPs), thereby activating broad-spectrum resistance to a variety of pathogens [2, 3]. For instance, this can be induced by lipopolysaccharides and flagellin stimulation [1, 4]. The second is effector-triggered immunity (ETI), where plants utilize the polymorphic intracellular nucleotide-binding/leucine-rich repeat (NLR) receptor family to recognize pathogen effectors, thus activating specific defensive responses in plants. In summary, plant defense is orchestrated by a diverse array of receptors that survey both the extracellular space for conserved molecular patterns and the intracellular environment for specific pathogen effectors. The genetic foundation for this dual-layered surveillance system is provided by a large super-family of genes, known as Pathogen Recognition Genes (PRGs), which encode the specific receptors required to trigger either PTI or ETI [2, 3].

Despite the diversity of structures found within PRGs, it is crucial to emphasize the ubiquity of the leucine-rich repeat (LRR) domain [5]. PRRs containing LRR domain can be divided into two major categories. The first category is the leucine-rich repeat receptor-like protein (*LRR-RLP*) resistance class within the receptor-like protein (*RLP*) gene family, which exclusively contains leucine-rich repeats (LRRs) and transmembrane domains (TM). The second category is the leucine-rich repeat receptor-like kinases (*LRR-RLK*), which are *LRR-RLPs* that include kinase (*KIN*) domains. *LRR-RLK* represents the largest class of *RLKs*, and the basic structure of *RLKs* consists of a variable extracellular domain (ECD), a TM and a conserved KIN domain. Notably, the ECDs of *RLK* proteins are highly diverse and typically comprise different protein domains, such as LRR domains, lectin-like motifs (LECM) domains, and lysin motif (LysM) domains [6, 7]. Additionally, many crucial PRRs in fungal pathogens contain LysM or LECM instead of LRRs, but they exhibit structures similar to those of *RLK* and *RLP* receptors. Examples include LysM domain-containing receptor-like kinase (*LYK*), LysM domain-containing receptor-like protein (*LYP*), and lectin receptor-like kinase (*Lec-RLK*) proteins [8]. The LRR domain is also present in NLR proteins that contain nucleotide-binding site (NBS) domains (*NBS-LRR*) [9]. *NBS-LRR* resistance proteins primarily differ in their N-terminus, which includes either toll/interleukin-1 receptor-like (TIR) domain, a coiled-coil (CC) domain, or a powdery mildew resistance 8 (PMR8) domain. These domains are referred to as the TIR-NBS-LRRs (*TNLs*), CC-NBS-LRRs (*CNLs*) and RPW8-NBS-LRR (*RNL*) classes, respectively [10, 11]. In the Plant Resistance Genes database (PRGdb 4.0, <http://prgdb.org/prgdb4/>), which comprises 199 reference PRGs and 182 proteomes, PRGs are categorized into seven gene families, namely *RLK* (*LRR-RLK*), *RLP* (*LRR-RLP*), *LYK*, *LYP*, *LECRK* (*Lec-RLK*), *CNL*, and *TNL* classes receptors [12].

G. elata is a non-autotrophic orchid species characterized by the absence of roots and leaves, possessing only scales, and incapable of photosynthesis for autotrophic growth. Its seeds are extremely small, dust-like, composed of tens of embryonic cells and a seed coat, without endosperm or other nutrient reserves, and its embryonic tissues lack organ differentiation [13]. *G. elata* depends on symbiotic fungi, such as *Armillaria mellea* or *Armillaria gallica*, to supply essential nutrients, and has concurrently evolved an effective defense system to prevent fungal superinfection [14]. The main diseases of *G. elata* include tuber rot, mixed fungal infections, white sclerotium disease, and black stem disease [15]. For example, in Guizhou and Yunnan, the incidence of *Aspergillus flavus* disease has reached 6%–17%, leading to a yield reduction of 10%–30% [16]. In addition, 13 *Trichoderma* species were isolated

from diseased tubers of *G. elata* collected in Yunnan, confirming that *Trichoderma* is an important fungal pathogen causing black rot disease [17].

With the rapid development of biotechnology, omics-based molecular breeding has been increasingly applied in medicinal plants. Compared to the genomes of *Phalaenopsis equestris*, *Dendrobium officinale*, *Ananas comosus*, and *Arabidopsis thaliana*, the genome of *G. elata* shows a significant reduction in genes related to plant resistance, as well as decreased occurrences of pathogenesis-related gene families and antioxidant protein genes [13]. Notably, the *NLR* gene family is reduced, and the key gene ICS, responsible for regulating salicylic acid-mediated defense responses, is absent in *G. elata*, indicating the loss of an essential component for systemic acquired resistance [13, 18].

Breeding disease-resistant varieties is a sustainable strategy for controlling diseases in *G. elata*, and identifying disease-resistance genes is fundamental for its molecular breeding [19]. In this study, bioinformatics techniques were employed to compare the protein sequences of the *G. elata* genome against the latest plant resistance gene database (prgdb.org/prgdb4/) [20], leading to the initial identification of disease-resistance genes in *G. elata*. Comprehensive analyses were conducted on their physicochemical properties, phylogenetic relationships, conserved motifs, gene structures, chromosomal locations, gene duplications, and conserved cis-acting elements in promoter regions. Furthermore, we analyzed their expression patterns across different developmental stages and in *G. elata* samples with varying stem colors, based on transcriptome data.

A total of five disease-resistance gene families comprising 68 members were identified, and their structural and evolutionary features were systematically elucidated. This study establishes a method for screening disease-resistant materials in *G. elata* based on expression profiles of disease-resistance genes, providing theoretical support for the environmentally friendly cultivation of *G. elata* and offering new insights for molecular breeding of disease resistance in this species.

Materials and methods

Plant material and public data

All experimental materials were collected from Qixingguan District, Bijie City, Guizhou Province, China. Samples were collected from five developmental stages of dark red *G. elata* (identification number: Qian recognized medicine 20230004), including immature tuber (S1), developing tuber (S2), mature tuber (S3), flowers and stems (S4), and capsules (S5). In addition, six flower and stem samples were identified by professor Cai Zhao from the Plant Identification Center of Guizhou University as three forms of *G. elata*, with each variety displaying unique stem colors: (1) *G. elata* f. *elata* (the original variety with red stems), including *G. elata* f. *elata* with dark red stems (TMSR) and *G. elata* f. *elata* with light red stems (TMQR); (2) *G. elata* f. *viridis* (the green-stemmed forms), including *G. elata* f. *viridis* with dark green stems (TMSG) and *G. elata* f. *viridis* with light green stems (TMQG); (3) *G. elata* f. *glauca* (the brown-stemmed

forms), including *G. elata* f. *glauca* with dark brown stems (TMSB) and *G. elata* f. *glauca* with light brown stems (TMQB). For each sample, three biological replicates were used for transcriptome sequencing.

The genome, GFF annotation, and protein sequences of *G. elata* were generated by our research group from a chromosome-level assembly of the dark red *G. elata*, which remains unpublished. The 198 reference resistance amino acid sequences were downloaded from the Plant Resistance Genes database 4.0 (PRGdb 4.0; <http://prgdb.org/prgdb4/>) [12]. The protein sequences and GFF annotation files for *Angraecum sesquipedale*, *Dendrobium chrysotoxum*, *Dendrobium nobile*, *Phalaenopsis guangdongensis*, *Phalaenopsis zijinensis*, and *Vanilla planifolia* were downloaded from the National Genomics Data Center (<https://ngdc.cncb.ac.cn/gwh/>, NGDC). The protein sequences and GFF annotation files for *Oryza sativa* and *A. thaliana* were obtained from Phytozome v13 (<https://phytozome-next.jgi.doe.gov/>) (Table S1).

Sequence Identification and Annotation of the resistance genes of *G. elata*

First, all 198 reference resistance amino acid sequences were used as query sequences and compared to the *G. elata* genome using the BLASTp algorithm (bitscore ≥ 100 , e-value $\leq 1e-10$) [21]. Secondly, pattern recognition gene domains were predicted using Disease Resistance Analysis and Gene Orthology 3 (DRAGO3, <http://prgdb.org/prgdb4/drago3>) with default parameters, an automated tool for annotating and predicting plant resistance genes [12]. Third, all candidate genes were examined using the NCBI-CDD (National Center for Biotechnology Information, Conserved Domain Database, <https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) with a cutoff value of 0.01 to verify the domains. Finally, the pattern recognition resistance genes (GePRGs) of *G. elata* were identified.

The physicochemical property of *GePRGs* members, such as number of amino acids, molecular weight, isoelectric point, aliphatic index, grand average of hydropathicity (GRAVY), were predicted using the ExPasy [22] website (<https://www.expasy.org/>). As well as, the subcellular localization predictions of *GePRGs* members were made using the website WoLF PSORT (<https://wolfpsort.hgc.jp/>) [23].

Phylogenetic evolution, conserved motif and gene structure analysis of the GePRGs

The conserved motif of *GePRGs* members was analyzed using the website MEME [24, 25] (<https://meme-suite.org/meme/tools/meme>), including the 42 *GeLRR-RLK* paralogous *GePRGs* proteins with a maximum conserved motif search value of 10, the 20 *GeLec-RLK* paralogous *GePRGs* proteins with 15. In order to explore the phylogenetic evolutionary relationships of *GePRG*, Molecular Evolutionary Genetics Analysis (MEGA v11) [26] was used to conduct phylogenetic analysis with a bootstrap value of 1000 and other default parameter. Firstly, multi-sequence alignment of amino acid sequences of 68 *GePRGs* was performed using the built-in ClustalW. Secondly, the results were trimmed compared using TBtools (v2.037) [27]. Finally, the phylogenetic tree was established using the neighbor-joining (NJ) [28].

Chromosomal localization and syntenic analysis of the GePRGs

Based on the genome and gene annotation data of *G. elata*, we identified and mapped the chromosomal locations of 68 *GePRGs*. Gene duplication events were then analyzed using the MCScanX toolkit (default parameters) [29], and the resulting syntenic relationships were visualized with TBtools (v2.037).

Cis-acting element analysis of the *GePRGs* in *G. elata*

The 2,000 bp promoter sequences upstream of the *GePRGs* members were extracted from the *G. elata* genome using TBtools (v2.037), and subsequently submitted to the PlantCARE [30] website (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) for bioinformatics analysis. Using the default search parameters, we identified the likely presence of various known cis-acting elements. In this analysis, we specifically focused on cis-elements that are likely to be typical and functional, while filtering out common or unnamed ubiquitous elements found in most genes, such as the A-box, CAAT-box, and TATA-box.

RNA-seq data analysis

RNA-seq analysis included cDNA library preparation, paired-end sequencing on the Illumina NovaSeq 6000, quality control using FastQC [31], alignment with the reference genome (dark red *G. elata*) using HISAT2 (v2.0.5) [32], read quantification with StringTie (v1.3.3) [33], and FPKM calculation using Subread software [34].

Result

Identification of pathogen recognition genes in *G. elata*

In this study, BLASTp, DRAGO3 and NCBI-CDD were employed to identify all potential pathogen recognition genes (PRGs) in the *G. elata* genome (*GePRGs*). After removing redundant sequences, a total of 68 *GePRG* members were identified, including 42 *LRR-RLKs*, 20 *Lec-RLKs*, 2 *LYKs*, 2 *LYPs*, and 2 *CNL* genes. These genes were named sequentially based on their chromosomal positions as *GeLRR-RLK* (1–42), *GeLec-RLK* (1–20), *GeLYK* (1–2), *GeLYP* (1–2), and *GeCNL* (1–2) (Table S2). Domain prediction using NCBI-CDD further classified the 20 *GeLec-RLKs* into two subtypes based on structural features: 8 L-type and 12 G-type (Table S3) [35]. Phylogenetic analysis further demonstrated that these two types of *GeLec-RLK* genes form two distinct evolutionary clades.

Additionally, various physicochemical properties of the 68 *GePRGs* were analyzed, including protein length, molecular weight (MW), isoelectric point (pI), aliphatic index, grand average of hydropathicity, and subcellular localization (Table S2). Significant differences were observed in the amino acid sequences of *GeLRR-RLK* and *GeLec-RLK* proteins. Most *GeLRR-RLK* proteins contain approximately 900 to 1,100 amino acids, whereas *GeLec-RLK* proteins generally range from 600 to 800 amino acids. These sequence variations likely originated from genome duplication events in *G. elata*. Notably, longer

proteins often exhibit more complex structures and functional diversity, enabling them to play crucial roles in disease resistance regulation.

Furthermore, 75% (51/68) of the identified *GePRG* proteins have an isoelectric point below 7, suggesting that most *GePRGs* encode acidic proteins. Protein stability was predicted based on the instability index, where a value below 40 indicates stability, while values above 40 suggest potential instability [22]. Among the identified *GePRGs*, 69.1% (47/68) exhibited an instability index below 40, implying that the majority of *GePRG* proteins are potentially stable.

In the *GeLec-RLK* family, the molecular weights of *GePRG* proteins range from 71.78 to 93.40 kDa. Specifically, L-type *GeLec-RLKs* exhibit a lower molecular weight (71.77–82.45 kDa) compared to G-type *GeLec-RLKs* (84.03–93.40 kDa). This difference is attributed to the presence of additional EGF and PAN domains in G-type *GeLec-RLKs*, besides the conserved lectin domain [36]. Variations in physicochemical properties may contribute to differences in disease resistance mechanisms in *G. elata*.

The subcellular localization of proteins plays a critical role in determining their function and interactions within specific cellular environments. *GeLec-RLK* proteins were predicted to be localized in the plasma membrane, vacuole, and mitochondria, whereas *GeLRR-RLK* proteins were predominantly localized to the plasma membrane. Overall, 40 out of 68 *GePRG* proteins were predicted to reside on the plasma membrane. This suggests that *GePRGs* may be involved in pattern-triggered immunity (PTI). Given this localization, damage to the plasma membrane could potentially compromise the defense mechanisms of *G. elata*.

Phylogeny, gene structure and motif composition of *GePRGs*

A phylogenetic tree of the 68 *GePRGs* was constructed using the neighbor-joining method. Based on sequence identification and phylogenetic relationships, these *GePRGs* were classified into eight distinct clades. Six of these clades primarily consisted of RLK subfamilies. For instance, all proteins in Clades I, II, and III belonged to the *LRR-RLK* group, while Clades IV and V comprised L-type and G-type *Lec-RLKs*, respectively. Clade VI was composed of LYK proteins. Notably, genes from the same family clustered together. The phylogenetic branch lengths indicated that LYP and *CNL* genes had the longest branches, suggesting that they diverged earlier than the RLK gene family (Fig. 1).

Since variations in gene structure provide insights into evolutionary relationships and functional divergence, we further examined motifs, exon-intron organization, and structural characteristics of the two major gene subfamilies. Both *GeLRR-RLK* and *GeLec-RLK* genes contained introns, but their numbers varied significantly (Fig. 2b). The number of introns ranged from 1 to 19, with *GeLec-RLKs* generally having fewer introns and simpler gene structures, whereas *GeLRR-RLKs* had more introns and a more complex gene architecture. Moreover, intron length varied considerably among *GeLRR-RLK* members, with *GaLRR-RLK34* harboring the longest intron, further highlighting the structural complexity of *GeLRR-RLKs*.

MEME motif analysis was conducted to examine the conserved protein motifs in *GePRG* sequences. The motif composition of *GeLRR-RLKs* was relatively conserved, with motifs 1–7 and motif 9 present across all members. Among these, motifs 4, 6–7, and 9–10 were associated with the LRR domain, whereas the remaining motifs belonged to the kinase domain (Fig. 2). This structural conservation suggests that *GeLRR-RLKs* may play critical roles in signal transduction and pathogen recognition. Notably, *GeLRR-RLK12* exhibited a longer sequence and more complex structure compared to other members. Within Clade III of *GeLRR-RLKs*, genes displayed higher structural variability, consisting of multiple coding sequences (CDS), indicating potential functional divergence from other genes.

Within the *GeLec-RLK* subfamily, Clade I was identified as L-type, while Clade III belonged to the G-type. Motifs 1–3, 5–6, 9, and 15, corresponding to the Pkinase domain, were widely distributed across all genes. Meanwhile, motifs 8 and 11–13 were associated with the lectin domain, with motif 8 (lectin_legB) being unique to Clades I and II, while motifs 12 and 13 (B_lectin) were exclusive to Clade III. These findings further validate the classification of *GeLec-RLKs* into L-type and G-type genes.

Previous studies have reported that the antifungal protein from *G. elata* (GenBank: AAK59994.1) exhibits antifungal activity and belongs to the mannose-binding protein family with a B_Lectin domain [37]. However, sequence alignment revealed that this protein did not match the G-type *GeLec-RLKs* identified in this study (e-value = 1e-3). This suggests that the G-type *GeLec-RLKs* identified in this study may represent a novel group of disease resistance genes distinct from the previously reported antifungal protein. In summary, the analysis of conserved motifs, gene structures, and phylogenetic relationships of *GeLec-RLK* and *GeLRR-RLK* genes further validated the classification of *GePRGs* and highlighted their potential functional differences in disease resistance.

Chromosomal distribution, gene duplication and synteny analysis of *GeLec-RLK* and *GeLRR-RLK*

Based on the *G. elata* genome, a chromosome map was constructed, revealing that 68 *GePRG* genes were randomly distributed across eighteen chromosomes and scaffold_189 (a scaffold region). Their locations ranged from 135.75 Mb on Chr1 to 36.08 Mb on Chr18 (Fig. 3, Table S4). The *GeLec-RLK* subfamily was localized to multiple chromosomes, specifically Chr5 to Chr7, Chr9 to Chr10, Chr12, Chr14 to Chr15, Chr18, and scaffold_189 (Fig. 3). Notably, Chr7 and Chr14 contained four *GeLec-RLK* genes each, whereas Chr5, Chr9, Chr15, Chr18, and scaffold_189 harbored only one gene. In contrast, *GeLRR-RLKs* were distributed across nearly all chromosomes except Chr14, Chr17, and Chr18. Notably, five *GeLRR-RLK* genes were found on Chr9, while Chr3, Chr7, and Chr8 each contained four RLK genes. Conversely, only a single RLK gene was identified on Chr15 and Chr16. Overall, the distribution of *GePRGs* across chromosomes was uneven, but certain genes appeared to cluster together, such as *GeLec-RLK4* - *GeLec-RLK7* and *GeLRR-RLK19* - *GeLRR-RLK20*. This suggests that tandem duplication events may have contributed to their evolution.

To investigate the expansion of the *GePRG* gene family in *G. elata*, we analyzed duplication events and identified six tandem duplication events involving 12 *GePRG* genes (Fig. 4, Table S5), as well as one

segmental duplication event. Notably, in these six tandem duplications, the paired duplicated genes were located within a 200 kb chromosomal region [38].

To further explore the phylogenetic mechanisms underlying the evolution of *GePRG* genes in *G. elata*, we constructed eight collinearity maps comparing *G. elata* with six orchid species and two model plants (Table S1). Our results revealed that 55 *GePRG* genes in *G. elata* had syntenic relationships with homologous genes in these species. Specifically, the number of orthologous gene pairs between *G. elata* and other species was as follows: *A. sesquipedale* (46 pairs), *D. chrysotoxum* (39 pairs), *D. nobile* (50 pairs), *P. guangdongensis* (33 pairs), *P. zijinensis* (36 pairs), *V. planifolia* (29 pairs), *O. sativa* (20 pairs), and *A. thaliana* (7 pairs) (Fig. 5, Table S6). These results indicate that *GePRG* genes in *G. elata* share a higher degree of homology with genes from orchid species, with the highest number of orthologous pairs observed in *D. nobile* (50 pairs), followed by *A. sesquipedale* (46 pairs). However, fewer orthologous gene pairs were identified between *G. elata* and *O. sativa* (29 pairs) or *A. thaliana* (20 pairs). The collinearity analysis suggests that *GePRG* genes are highly conserved among orchids and monocots, and that the expansion of the *GePRG* gene family in *G. elata* is largely the result of whole-genome duplication events.

Cis-acting element analysis of the *GePRGs* in *G. elata*

A gene promoter is a DNA sequence located upstream of the gene coding region that contains multiple cis-acting elements, which serve as specific binding sites for proteins involved in transcription initiation and regulation [39]. Cis-acting regulatory elements play a crucial role in gene transcription by interacting with trans-acting factors, particularly in response to environmental biotic and abiotic factors. To gain deeper insights into gene regulation, we analyzed the promoter regions of each putative *GePRG* genes.

The identified cis-acting elements were categorized into four major functional classes: light-responsive elements, hormone-responsive elements, developmental regulation elements, and stress-responsive elements [40]. Statistical analysis revealed that the number of cis-acting elements in each gene varied from 102 to 289 (Table S7), and their distribution across the promoter regions of *GePRG* members was uneven.

Among these, light-responsive elements were the most abundant and universally present in all 68 *GePRG* gene promoters. Each *GePRGs* contained at least one light-responsive element, with Box 4 and G-box being the two most prevalent types. This suggests that the transcription of these genes can be induced by light.

Following light-responsive elements, hormone-responsive elements were the second most abundant category. The hormone-related cis-elements identified in *GePRGs* included 50 ERE (Ethylene-responsive), 45 CGTCA-motif (MeJA-responsive element), 45 TGACG-motif (Methyl jasmonate regulatory elements), 39 ABRE (abscisic acid responsive element) and TCA-element (salicylic acid-responsive element) [41, 42]. The widespread presence of these hormone-related motifs indicates the potential involvement of *GePRG* genes in hormonal regulation. The development-related elements were also identified, including

O2-site (involved in zein metabolism regulation), CAT-box (associated with meristem expression), and circadian (participating in circadian rhythm regulation) [35, 43]. These elements suggest that *GePRG* genes may play roles in plant growth, development, and environmental adaptation.

In addition to light and hormone-responsive elements, seventeen types of stress-related cis-elements were identified. Among them, the ARE (anaerobic response element) was the most abundant, detected in 46 *GePRGs*, suggesting a role in gene regulation under low-oxygen conditions. Additionally, 26 *GePRG* genes contained MBS elements, which are MYB-binding sites associated with drought-induced responses, implying that these genes may participate in drought stress adaptation.

The W-box, a WRKY protein-binding motif widely reported to be involved in plant defense responses and senescence regulation [44, 45], was found in the promoter regions of 32 *GePRG* genes. Furthermore, within the *GeLRR-RLK* family, three additional stress-related elements were identified, including LTR (low-temperature-responsive element) detected in 23 genes, TC-rich repeats found in 26 genes, and WUN-motif (wound-responsive element) identified in 14 genes and so on (Fig. 6, Table S7).

Analysis of expression patterns of *GePRGs* in the flowers and stems of six *G. elata* forms

Transcriptome sequencing analysis was performed on stem tissues from six color forms of *G. elata* under normal growth conditions, including dark green stem *G. elata* (DGGE), light green *G. elata* (LGG), dark brown *G. elata* (DBGE), light brown *G. elata* (LBGE), dark red *G. elata* (DRGE), and light red *G. elata* (LRGE). Based on the expression patterns of 68 *GePRGs* across these six *G. elata* germplasms, these genes can be classified into five distinct expression patterns (Fig. 7, Table S8). Pattern I comprises seven genes, all of which exhibit a certain level of expression (with an average FPKM of approximately 85) across six *G. elata* germplasms. Among them, five genes have an FPKM greater than 100. These genes are: *GeLec-RLK2* (with FPKM values ranging from 227 to 545 in TMSG, TMQG, and TMQR), *GeLec-RLK3* (with FPKM values ranging from 123 to 140 in TMQG), *GeLec-RLK8* (with FPKM values ranging from 107 to 128 in TMSG and TMQG), *GeLRR-RLK15* (with FPKM values ranging from 139 to 302 in TMQG), and *GeLRR-RLK24* (with FPKM values ranging from 108 to 188 in TMSG and TMQG). Patterns III and V collectively include 20 genes, among which only *GeLRR-RLK17* and *GeLYK1* have an average FPKM exceeding approximately 100 in TMQR. Patterns II and IV encompass 45 genes, with an average FPKM of 5.2.

GeLec-RLK2 and *GeLec-RLK3* are classified as L-type *Lec-RLKs*, while *GeLec-RLK8* is identified as a G-type *Lec-RLK*. These genes were aligned with the *LecRK-VI.2* sequence from *A. thaliana* (with an e-value of 1e-10 and a bitscore ≥ 100). It has been reported that several L-type *Lec-RLKs* were activated upon perception of PAMP signals and trigger PTI as well as other plant responses to biotrophic and necrotrophic pathogens, such as *Xanthomonas* spp. and *Botrytis cinerea*. Examples of such L-type *Lec-RLKs* in *A. thaliana* include *LecRK-IX.2*, *LecRK-I.9*, and *LecRK-VI.2* [46]. In 2023, a G-type lectin receptor-like kinase (*RLK*), designated as NbERK1, was identified in the model plant *Nicotiana benthamiana*. NbERK1 was identified as a novel component of the chitin receptor complex, which enhances plant resistance to fungal pathogens by modulating chitin signaling [47]. Recent studies have shown that G-

type LecRKs mediate the symbiotic interaction between poplar trees and the ectomycorrhizal fungus *Agaricus bisporus* [48]. In plants, *LRR-RLKs* regulate growth, differentiation, and defense responses. For example, FLS2 in *A. thaliana* is responsible for recognizing bacterial flagellin and activating plant immune responses. In plants, *LRR-RLKs* regulate growth, differentiation, and defense responses. For example, FLS2 in *A. thaliana* is responsible for recognizing bacterial flagellin and activating plant immune responses [49]. Therefore, it can be speculated that L-type and G-type *Lec-RLKs*, along with *LRR-RLKs*, play crucial roles in the immunity of *G. elata* germplasms such as TMSG, TMQG, and TMQR. By recognizing pathogen-associated signals, they facilitate defense responses and symbiotic interactions. The diverse functions of disease-resistance genes underscore the complexity and adaptability of the immune system in *G. elata*.

A higher number of PRGs are expressed in TMQG, followed by TMQR and TMSG, suggesting their potential as molecular targets for disease-resistant breeding. In particular, *GeLec-RLK2*, *GeLec-RLK8*, *GeLRR-RLK17*, *GeLRR-RLK24*, and *GeLYK1* exhibit high expression levels (FPKM > 100) in TMQG, TMQR, and TMSG, indicating their potential as molecular markers for selecting resistant germplasms. In summary, the characteristic features of disease-resistance genes in these *G. elata* germplasms can be utilized for molecular breeding aimed at enhancing disease resistance, enabling the screening of *G. elata* accessions with strong disease-resistance characteristics.

Analysis of expression patterns of GePRGs in five developmental stages of dark red *G. elata*

A heatmap of transcriptional expression across five developmental stages of *G. elata* is displayed, including the immature tuber (S1), developing tuber (S2), mature tuber (S3), flower and stem (S4), and capsule (S5). Based on gene expression levels, the genes were clustered into five distinct expression patterns (I to V) (Fig. 8, Table S9). In Pattern I, 22 genes exhibited low or nearly negligible expression levels (with an average FPKM of approximately 1.98). Pattern II comprises four genes, which had an average FPKM of about 54 across the five developmental stages. Among them, *GeLec-RLK2* showed an average expression level of 77 across stages S1 to S5, with an average expression of approximately 98 in stages S2, S3, and S5. *GeLec-RLK8* had an average expression level of about 72 across stages S1 to S5, and its expression reached an average of approximately 177 in stage S5, indicating their critical roles in tissue development.

GeLec-RLK2 was aligned with the L-type lectin receptor kinase LECRK82 from *A. thaliana* in the PRGdb 4.0 database, while *GeLec-RLK8* was compared with the bulb-type lectin S-domain-1 receptor-like kinase LORE (SD129) from *A. thaliana*. These two genes have been verified to be involved in resistance responses against pathogenic oomycetes (*Phytophthora infestans* and *Phytophthora capsici*), and in the susceptibility to *Pseudomonas syringae* infection, respectively [50, 51]. Overall, these four genes are likely to be crucial disease-resistance genes during the transition of *G. elata* from arrow tuber (vegetative tuber) to the flowering and fruiting stages. In other words, these genes are characteristic of the reproductive growth phase in *G. elata*. Monitoring the expression of these genes holds significant implications for disease surveillance and control during the reproductive growth period.

In Pattern III, there are a total of 12 genes with an average FPKM expression of 19 across stages S1 to S5. In stage S1, four genes exhibit an average FPKM expression exceeding 30: *GeLRR-RLK9*, *GeLRR-RLK25*, *GeLRR-RLK26*, and *GeLRR-RLK31*. In stage S2, five genes have an average FPKM expression exceeding 30, including *GeLRR-RLK26* with an FPKM of approximately 55, *GeLRR-RLK30*, *GeLRR-RLK31* (FPKM $\approx$ 55), *GeLRR-RLK34*, and *GeLYP1*. In stage S4, three genes have an average FPKM expression exceeding 30: *GeLRR-RLK30*, *GeLec-RLK9*, and *GeLYP1*, while in stage S5, only *GeLec-RLK9* has an average FPKM expression exceeding 30. Notably, *GeLRR-RLK26* and *GeLRR-RLK31*, which have an average FPKM expression of approximately 55 in stage S2, were aligned with the leucine-rich repeat receptor-like protein kinases ER-Erecta and PEPR1 genes from *A. thaliana* in the database. ER-Erecta plays a crucial role in plant growth and development by regulating the proliferation of epidermal cells in *A. thaliana* to control seed size, and it also confers resistance to the bacterium *Ralstonia solanacearum* [52–54]. The PEPR1 gene acts as a receptor for plant-specific small peptide signaling molecules called Plant Elicitor Peptides (Peps). Unlike typical immune receptors, PEPR1 senses endogenous elicitors that amplify the plant responses induced by Pathogen-Associated Molecular Patterns (PAMPs). It participates in PAMP-Triggered Immunity (PTI) signaling and interacts with BIK1, a central rate-limiting kinase in the PTI signaling transduction pathway, phosphorylating the kinase [55, 56]. Therefore, it is speculated that *GeLRR-RLK26* may have undergone functional degeneration in controlling the size of *G. elata* seed formation, as evidenced by its relatively low average FPKM expression (approximately 17) during the capsule stage (S5) of *G. elata*. However, similar to *GeLRR-RLK31*, it exhibits slightly higher expression levels (FPKM $\approx$ 55) in stage S2 (the white tuber stage), suggesting that it may play a certain role in disease resistance during the reproductive growth of *G. elata*.

In Pattern IV, there is only one gene, *GeLYK1*, which shows certain expression in stage S1 (with an FPKM of approximately 49) and weak expression in stage S5 (with an FPKM of approximately 15). Its expression is primarily notable during the seed germination process (S1, where seeds within the capsule are promoted to germinate by germination fungi, leading to the formation of rice-shaped tubers). *GeLYK1* was aligned with LysM domain receptor-like kinase 3 (LYK3) from *A. thaliana* in the database. LysM motif (Lysin motif) receptor kinases are unique to plants and play crucial roles in plant-microbe interactions. LYK3 has been reported as a potential LysM receptor kinase that recognizes microbial-derived N-acetylglucosamine (NAG) ligands. However, the molecular mechanisms of ligand perception and receptor activation remain unknown [57].

Discussion

In Pattern V, there are a total of 29 genes with an average FPKM expression of approximately 7.8. In stage S3, three genes, namely *GeLRR-RLK2* (FPKM $\approx$ 50), *GeLRR-RLK32*, and *GeLRR-RLK38*, exhibit FPKM expression values exceeding 30. In stage S4, two genes, *GeLRR-RLK15* and *GeLRR-RLK19*, have FPKM values greater than 30. In stage S5, only *GeLRR-RLK15* shows an average FPKM expression level exceeding 30. *GeLRR-RLK2* and *GeLRR-RLK15* were aligned with the senescence-induced receptor-like serine/threonine protein kinase FRK1 (SIRK) and the leucine-rich repeat receptor-like protein kinase PEPR2 from *A. thaliana* in the database, respectively. The SIRK gene has been reported to be involved in

innate immunity [58]. In *A. thaliana*, Peps peptides activate plant immune responses by binding to their specific receptors, such as PEPR1 and PEPR2. These responses include the expression of defense genes, regulation of root hair formation, and growth inhibition, among others [59].

Identifying resistance-related genes is a crucial step in ensuring long-term host genetic resistance. Lectins constitute a vital component of the plant innate immune system [60]. Lectins recognize PAMPs and Damage-Associated Molecular Patterns (DAMPs). On the cell surface, lectin domains (including LecP, LecRLKs, and LecRLPs) form a crucial part of the plant innate immune system [61]. Binding to PAMPs and DAMPs with lectin domains triggers downstream signaling cascades, including Ca^{2+} influx, Reactive Oxygen Species (ROS) production, Mitogen-Activated Protein Kinase pathway activation (MAP kinase pathway), regulation of hormone biosynthesis, modulation of defense-related gene expression, and synthesis of defensive molecules [62]. In plants, *LRR-RLKs* play a significant role in controlling plant growth and development, stress responses and immunity [63]. Many *LRR-RLKs* have been identified as immune receptors that sense pathogens and trigger downstream defense systems. In plant-microbe interactions, some *LRR-RLKs* recognize conserved pathogen/microbe-associated molecular patterns (PAMPs/MAMPs) to mount defense responses against pathogen attacks, which is a key step in host resistance to pathogens through pattern-triggered immunity (PTI) [64, 65]. For instance, *A. thaliana* EF-TU RECEPTOR (EFR) and rice XA21 are typical *LRR-RLKs* that play crucial roles in early immune responses to pathogen-derived elicitors [66].

G. elata, as a plant with both medicinal and edible uses, holds significant nutritional and economic value. However, the disease resistance gene family in *G. elata* has not yet been comprehensively studied. Therefore, a thorough exploration of the disease resistance gene family in *G. elata* was undertaken. This study aims to investigate the importance and diverse roles of the disease resistance gene family in the growth and development of *G. elata*, particularly in varieties with different stem colors, and to provide a molecular basis for the application of disease resistance genes in *G. elata* breeding.

Identification and evolution of the GePRGs

In this study, based on unpublished and a new cultivar of dark red *G. elata* genome data and the plant disease resistance gene database, bioinformatics methods were used to identify 68 disease resistance genes in *G. elata*. These genes were classified into five groups, with the majority belonging to the *LRR-RLK* class (42), followed by *Lec-RLK* (20), *LYK* (2), *LYP* (2), and *CNL* (2). Since *LRR-RLK*, *Lec-RLK*, and *LYK* are subfamilies of *RLKs*, this study primarily identified *RLK* family genes based on the reference database. As one of the largest disease resistance gene families in plants, NBS-LRR genes typically exist in large numbers in most angiosperms [67, 68], with only a few species having fewer than 100 NBS-LRR genes in their genomes [69, 70]. However, in this study, only two *CNL*-type genes were identified, suggesting that *G. elata* may retain a minimal number of NBS-LRR genes among angiosperms, consistent with findings from previous studies on NBS-LRR genes in orchids [70]. In higher plants, the number of *LRR-RLK* genes varies widely, ranging from 200 to 600 genes [71], whereas in *G. elata*, only 64 genes were identified. No C-type lectin kinases were identified in *GeLec-RLKs*. Despite C-type lectin

kinases abundance in mammals, both rice and *A. thaliana* contain only a single gene encoding a C-type LecRLK protein [72]. In summary, the 64 *RLK* genes identified in *G. elata* are substantially fewer than the more than 600 *RLK* genes reported in rice and *Arabidopsis*, likely due to extensive gene loss in *G. elata* [73, 74]. These losses are predominantly concentrated in gene families associated with immunity, photosynthesis, and developmental regulation [75, 76].

This study conducted a comprehensive analysis of resistance genes in *G. elata* and preliminarily investigated their homologous resistance genes. Analysis of gene structure and conserved motifs revealed that *GeLec-RLK* genes exhibit similar arrangements, whereas *GeLRR-RLK* genes exhibit greater diversity. These findings suggest that *GeLRR-RLK* genes may possess more diverse biological functions. Furthermore, physicochemical analysis of *GeLRR-RLK* genes indicated wide variations in gene length, predicted protein molecular weight (MW), and isoelectric point (pI), implying that different *GeLRR-RLK* proteins may play distinct roles in their respective microenvironments. Introns can increase gene length, recombination frequency, and regulatory complexity, making them beneficial for species evolution. Previous studies have suggested that genes responding rapidly to stress often lack introns, as intronless genes enable faster transcriptional regulation and facilitate rapid responses to developmental and defense processes [77, 78]. Each *GePRG* gene contained one to five coding sequences (CDSs) (Fig. 2B).

During the evolution of gene families, tandem duplication, segmental duplication, and transpositional events are considered primary strategies for member expansion and the emergence of new genes in plants [79, 80]. In the *G. elata* genome, three tandem duplication events were identified. Additionally, 55 *GePRG* genes exhibited syntenic relationships with homologous genes in other plant species, suggesting that these resistance genes have highly conserved genomic structures and have maintained functional stability throughout evolution. Previous studies have demonstrated that tandem and segmental/genome duplications played a crucial role in the expansion of LRR-RLK and Lec-RLK genes, as observed in *Brassica*, *Populus trichocarpa*, and *Medicago truncatula* [81, 82]. Gene duplication events often lead to mutational accumulation, which may result in functional divergence [83]. Phylogenetic analysis revealed that *GeLRR-RLKs*, *GeLYKs*, and both G-type and L-type *GeLec-RLK* genes form distinct evolutionary branches, indicating that GeRLK genes have undergone evolutionary differentiation over time.

Cis-acting element of the GePRGs

A total of 101 cis-regulatory elements (CREs) were identified within the 2,000bp upstream region of the translation start codon (ATG). Promoter analysis revealed a diverse and functionally variable set of CREs, with the most abundant categories being light-responsive elements (26.7%), abiotic stress-responsive elements (16.8%), development-related elements (12.9%), and hormone-responsive elements (11.9%).

Cis-acting elements in promoter regions play crucial roles in regulating gene expression under various stress conditions. The identified environmental stress-related elements in *GePRGs* can be classified into different functional categories based on their regulatory roles in plant responses to abiotic and biotic stresses.

Abiotic stress-responsive CREs are key regulatory elements that control drought, cold, oxidative stress, and anaerobic stress responses. For example, MBS (MYB Binding Site) and MYB recognition site elements are MYB transcription factor binding sites involved in drought-induced gene expression. The presence of MBS and MYB-binding sites suggests that *GePRGs* might be regulated by MYB transcription factors under drought stress [84].

DRE core and DRE1 elements are recognized by DREB (Dehydration-Responsive Element-Binding) transcription factors, which regulate responses to drought and cold stress [85]. LTR (Low-Temperature-Responsive Element) is associated with the regulation of cold-induced gene expression [86]. Additionally, anaerobic and oxidative stress-related elements include ARE (Anaerobic Response Element) and STRE (Stress-Responsive Element). Biotic stress-related CREs primarily regulate plant pathogen defense and mechanical damage responses. For example, W-box (WRKY Binding Site) is recognized by WRKY transcription factors and is widely involved in plant disease defense. WRE3 and WUN-motif are associated with plant wound repair and pathogen signaling pathways [87].

The presence of diverse environmental stress-related elements in *GePRG* promoters suggests that this gene family plays crucial roles in stress adaptation, including drought, cold, oxidative stress, pathogen defense, and wound healing. The enrichment of W-box, ARE, MBS, DRE, and LTR elements supports the hypothesis that *GePRGs* are involved in complex regulatory networks responding to both abiotic and biotic stresses [42, 88]. Given that *G. elata* spends over 80% of its life cycle underground and maintains close associations with at least two symbiotic fungi throughout its life, special attention should be given to cis-acting elements involved in hypoxic adaptation and low-temperature responses.

Expression pattern

The transcriptional expression analysis of 68 *GePRG* genes across different developmental stages and in the flowers and stems of six *G. elata* forms provides valuable insights into their roles in plant growth and disease resistance. Early-stage expression (S1): *GeLRR-RLK12*, *GeLec-RLK15*, and *GeLRR-RLK28* exhibit high expression in S1, suggesting their potential roles in early development or initial immune activation. Similar findings have been reported in other plants, where receptor-like kinases (RLKs) play a crucial role in cell differentiation and early immune priming [89, 90]. Gradual increase in expression (S1 to S5): Pattern II genes show steady upregulation, indicating their involvement in tissue growth, maturation, and immune modulation, consistent with previous observations in stress-adaptive receptor-like proteins [91]. *GeLYK1* exhibits high expression in both S1 and S5, potentially contributing to early immune activation and later-stage adaptation [92]. This suggests that certain LYK genes may provide broad-spectrum resistance throughout development. Late-stage expression (S3 to S5): *GeLRR-RLK15* and *GeLec-RLK1* are upregulated in later stages, indicating their potential roles in structural reinforcement, secondary metabolism, or pathogen recognition. Studies have shown that Lec-RLKs are involved in defense signaling and stress adaptation [35, 61]. These findings highlight the dynamic regulation of *GePRG* genes, reinforcing their significance in developmental plasticity and innate immunity.

Expression differences were also observed among *G. elata* forms, with TMQG exhibiting the highest total expression, followed by TMQR and TMSR, suggesting that certain genotypes may possess stronger disease resistance. This is consistent with studies in *A. thaliana* and rice, where natural genetic variation influences immune gene expression and pathogen resistance [93, 94]. *GeLRR-RLK17*, *GeLRR-RLK36*, *GeLYK1*, and *GeLec-RLK13* were highly expressed in TMQG, making them key candidates for molecular marker-assisted selection (MAS) in resistance breeding. Previous studies have shown that RLKs and LYK genes are involved in PAMP-triggered immunity (PTI) and effector-triggered immunity (ETI), making them essential genetic targets for disease resistance breeding.

This study reveals that *GePRG* genes exhibit stage-specific functions. For instance, *GeLRR-RLK12*, *GeLec-RLK15*, and *GeLRR-RLK28* show high expression in S1, suggesting their involvement in early development or initial immune responses. Pattern II genes exhibit a gradual increase from S1 to S5, indicating their role in tissue growth and maturation, which aligns with findings in other plants [89]. Similarly, *GeLYK1* is highly expressed in S1 and S5, suggesting its function in early immune activation and later-stage adaptation [95]. Additionally, *GeLRR-RLK15* and *GeLec-RLK1* show higher expression from S3 to S5, potentially contributing to structural reinforcement in mature tissues or pathogen recognition.

Expression differences were also observed among *G. elata* forms. Overall, TMQG *G. elata* exhibited the highest expression, followed by TMQR and TMSR, suggesting its potential for disease resistance breeding. Notably, *GeLRR-RLK17*, *GeLRR-RLK36*, *GeLYK1*, and *GeLec-RLK13* were highly expressed in TMQG, making them promising molecular markers for resistance selection. Based on these expression patterns, candidate genes can be further validated through gene knockout or overexpression experiments to confirm their roles in growth regulation and disease resistance.

Conclusion

In this study, we first identified 68 *GePRGs* in *G. elata* from the resistance gene database and conducted a comprehensive genomic and transcriptomic overview of these genes. This included phylogenetic analysis, classification, chromosomal distribution, protein motifs, gene structure organization, and gene duplication events, revealing substantial evolutionary diversity. The *G. elata* resistance gene family has undergone tandem and segmental duplications, with three identified tandem duplication events. Additionally, 55 *GePRG* genes show syntenic relationships with homologous genes in other plant species, suggesting that these genes have maintained functional stability throughout evolution. A total of 101 cis-regulatory elements (CREs) were identified in the 2,000 bp upstream region of the translation start codon (ATG) of *GePRGs*. These CREs include many environmental stress-related elements, such as MYB binding sites, DRE core, and low-temperature responsive elements, indicating that these genes play critical roles in responding to abiotic stresses (such as drought, cold, and oxidative stress) and biotic stresses (such as pathogen defense).

The expression analysis of *GePRG* genes at different developmental stages showed that certain genes, such as *GeLRR-RLK12*, *GeLec-RLK15*, and *GeLRR-RLK28*, are highly expressed in early developmental stages (S1), suggesting their involvement in early development or immune responses. The gradual upregulation of Pattern II genes from S1 to S5 indicates their role in tissue growth, maturation, and immune modulation. Among different *G. elata* genotypes, the TMQG type exhibited the highest expression levels, indicating a potentially stronger disease resistance. Specifically, *GeLRR-RLK17*, *GeLRR-RLK36*, *GeLYK1*, and *GeLec-RLK13* were highly expressed in TMQG, making them promising molecular markers for resistance breeding. This study provides a preliminary analysis of gene expression profiles and CREs in *G. elata*'s disease resistance genes, suggesting that further validation through gene knockout or overexpression experiments will confirm the roles of these candidate genes in growth regulation and disease resistance.

In conclusion, the disease resistance gene family in *G. elata* exhibits a wide range of cis-regulatory elements and complex expression patterns. The differential expression of these genes at various developmental stages and across different *G. elata* genotypes provides valuable genetic resources for disease resistance breeding. These genes are likely to play essential roles in enhancing the disease resistance and promoting the growth and development of *G. elata*.

Abbreviations

GePRGs Pathogen recognition genes in *Gastrodia elata*

CREs *Cis*-regulatory elements

ATG Translation start codon

DRE Dehydration-responsive element

TMSR *G. elata* f. *elata* with dark red stems

TMQR *G. elata* f. *elata* with light red stems

TMSG *G. elata* f. *viridis* with dark green stems

TMQG *G. elata* f. *viridis* with light green stems

TMSB *G. elata* f. *glauca* with dark brown stems

TMQB *G. elata* f. *glauca* with light brown stems

NLR Nucleotide-binding/leucine-rich repeat

LRR Leucine-rich repeat

RLP receptor-like protein

TM Transmembrane domains

MAMPs Microbe-associated molecular patterns

PAMPs Pathogen-associated molecular patterns

PRRs Pattern recognition receptors

PTI Pattern-triggered immunity

Declarations

Originality statement

We declare that this manuscript is original, has not been published elsewhere, and is not currently under consideration by any other journal. All authors have approved the final version of the manuscript and agree to its submission.

Availability of data and materials

All data and materials generated or analyzed during this study are included in this article (Supplementary Files) and also available from the corresponding author on reasonable request.

Acknowledgments

We express our profound gratitude to the anonymous reviewers for their invaluable comments on the manuscript. We are deeply appreciative of the State Key Laboratory of Big Data at Guizhou University for providing supercomputing services essential for data analysis.

Funding

This research was supported by funding from the Construction Project of Modern Industry Technology System for Traditional Chinese Medicinal Materials in Guizhou Province, China (No. GZCYTX-02, 2024). The funder had no role in study design, data collection, analysis, interpretation, or manuscript preparation.

Author information

Author details

College of Agriculture, Guizhou University, Guiyang, Guiyang, Guizhou, 550025, China
Linshuang Tang, Shanshan Luo, Dachang Wang, Xinyun Liu, Xiaofeng Wang, Mingyan Tan, Bing He, Zhibo Zhao, Mingjin Huang

Author Contributions

Mingjin Huang and Linshuang Tang designed the study. Minjing Huang, Linshuang Tang, Shanshan Luo and Dachang Wang prepared the plant samples. Linshuang Tang conducted experiments, analyzed the data, drafted the manuscript. Minjing Huang, Zhibo Zhao, Linshuang Tang, Shanshan Luo, Dachang Wang, Xinyun Liu, Xiaofeng W, Mingyang Tan, Bing He revised the manuscript. All authors approved the final version of the manuscript.

Corresponding author

Correspondence to Zhibo Zhao, Mingjin Huang.

Ethics declarations

Ethics approval and consent to participate

The plant materials utilized in this study were sourced from publicly accessible or commercially available collections, and their procurement did not require specific licensing or permits under national or institutional regulations. All experimental methodologies strictly adhered to relevant international, national, and institutional guidelines for plant research, including biosafety protocols and environmental protection standards. It is explicitly stated that this research did not involve human participants, animal experimentation, or any form of tissue sampling from protected species.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Clinical Trial

Not applicable.

References

1. Zipfel C, Felix G. Plants and animals: a different taste for microbes? *Curr Opin Plant Biol.* 2005;8:353–60.
2. Jones JDG, Dangl JL. The plant immune system. *Nature.* 2006;444:323–9.
3. Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell.* 2006;124:803–14.
4. Nürnberger T, Brunner F, Kemmerling B, Piater L. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol Rev.* 2004;198:249–66.

5. Abdul Malik NA, Kumar IS, Nadarajah K. Elicitor and receptor molecules: orchestrators of plant defense and immunity. *Int J Mol Sci.* 2020;21:963.
6. Liu J, Li W, Wu G, Ali K. An update on evolutionary, structural, and functional studies of receptor-like kinases in plants. *Front Plant Sci.* 2024;15.
7. Shiu S-H, Bleecker AB. Receptor-like kinases from arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci.* 2001;98:10763–8.
8. Torii KU. Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. *International Review of Cytology. Academic;* 2004. pp. 1–46.
9. Belkhadir Y, Subramaniam R, Dangl JL. Plant disease resistance protein signaling: NBS–LRR proteins and their partners. *Curr Opin Plant Biol.* 2004;7:391–9.
10. Pan Q, Wendel J, Fluhr R. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J Mol Evol.* 2000;50:203–13.
11. Fluhr R. Sentinels of disease. Plant resistance genes. *Plant Physiol.* 2001;127:1367–74.
12. Calle García J, Guadagno A, Paytuyi-Gallart A, Saera-Vila A, Amoroso CG, D'Esposito D, et al. PRGdb 4.0: an updated database dedicated to genes involved in plant disease resistance process. *Nucleic Acids Res.* 2022;50:D1483–90.
13. Yuan Y, Jin X, Liu J, Zhao X, Zhou J, Wang X, et al. The *gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat Commun.* 2018;9:1615.
14. Tsai C-C, Wu K-M, Chiang T-Y, Huang C-Y, Chou C-H, Li S-J, et al. Comparative transcriptome analysis of *gastrodia elata* (orchidaceae) in response to fungus symbiosis to identify gastrodin biosynthesis-related genes. *BMC Genomics.* 2016;17:212.
15. Yu ChangJun YC, Wang ShaoBai WS. Identification of pathogen of *gastrodia elata* white silk disease. 2010.
16. Zhang J-Q, Tang X, Xiao C-H, Xu J, Yuan Q-S, Wang X, et al. Investigation, analysis and identification of disease of *gastrodia elata* f. *glauca*. *Zhongguo Zhong Yao Za Zhi = Zhongguo Zhongyao Zazhi = China. J Chin Mater Med.* 2020;45:478–84.
17. Ye C, You Y, Li W, Jing T, Mo M, Qiao M et al. Diversity of *Trichoderma* species associated with the black rot disease of *Gastrodia elata*, including four new species. *Front Microbiol.* 2024;15.
18. Zhang G-Q, Xu Q, Bian C, Tsai W-C, Yeh C-M, Liu K-W, et al. The *dendrobium catenatum* lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci Rep.* 2016;6:19029.
19. Hasan N, Choudhary S, Naaz N, Sharma N, Laskar RA. Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J Genet Eng Biotechnol.* 2021;19:128.
20. Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, et al. PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucleic Acids Res.* 2009;38 suppl1:D814–21.

21. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
22. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana; 2005. pp. 571–607.
23. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007;35 suppl2:W585–7.
24. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res.* 2015;43:W39–49.
25. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME suite: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37 suppl2:W202–8.
26. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021;38:3022–7.
27. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13:1194–202.
28. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4:406–25.
29. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40:e49–49.
30. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30:325–7.
31. Wingett SW, Andrews S. FastQ screen: a tool for multi-genome mapping and quality control. *F1000Research.* 2018;7:1338.
32. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12:357–60.
33. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33:290–5.
34. Liao Y, Smyth GK, Shi W. The R package rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 2019;47:e47–47.
35. Vaid N, Pandey PK, Tuteja N. Genome-wide analysis of lectin receptor-like kinase family from arabidopsis and rice. *Plant Mol Biol.* 2012;80:365–88.
36. Lv D, Wang G, Xiong L-R, Sun J-X, Chen Y, Guo C-L, et al. Genome-wide identification and characterization of lectin receptor-like kinase gene family in cucumber and expression profiling analysis under different treatments. *Genes.* 2020;11:1032.
37. Wang X, Bauw G, Van Damme EJ, Peumans WJ, Chen Z-L, Van Montagu M, et al. Gastrodianin-like mannose-binding proteins: a novel class of plant proteins with antifungal properties. *Plant J.*

- 2001;25:651–61.
38. Holub EB. The arms race is ancient history in arabidopsis, the wildflower. *Nat Rev Genet.* 2001;2:516–27.
 39. Hernandez-Garcia CM, Finer JJ. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* 2014;217:109–19.
 40. Su Y, Peng X, Shen S. Identification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in paper mulberry and their potential roles in response to cold stress. *Comput Biol Chem.* 2022;97:107622.
 41. Ge H, Lu J, Han M, Lu L, Tian J, Zheng H, et al. Identification and functional analysis of 1-deoxy-D-xylulose-5-phosphate synthase gene in tomatoes (*solanum lycopersicum*). *Horticulturae.* 2024;10:304.
 42. ain-Ali Q, Mushtaq N, Amir R, Gul A, Tahir M, Munir F. Genome-wide promoter analysis, homology modeling and protein interaction network of dehydration responsive element binding (DREB) gene family in *solanum tuberosum*. *PLoS ONE.* 2021;16:e0261215.
 43. Li C, Qiao Z, Qi W, Wang Q, Yuan Y, Yang X, et al. Genome-wide characterization of cis-acting DNA targets reveals the transcriptional regulatory framework of opaque2 in maize. *Plant Cell.* 2015;27:532–45.
 44. Liu L, Xu W, Hu X, Liu H, Lin Y. W-box and G-box elements play important roles in early senescence of rice flag leaf. *Sci Rep.* 2016;6:20881.
 45. Wang Z, Yang P, Fan B, Chen Z. An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with the plant defence response. *Plant J.* 1998;16:515–22.
 46. Sun Y, Qiao Z, Muchero W, Chen J-G. Lectin receptor-like kinases: the sensor and mediator at the plant cell surface. *Front Plant Sci.* 2020;11.
 47. Pi L, Zhang Y, Wang J, Wang N, Yin Z, Dou D. A G-type lectin receptor-like kinase in *nicotiana benthamiana* enhances resistance to the fungal pathogen *sclerotinia sclerotiorum* by complexing with CERK1/LYK4. *Phytopathol Res.* 2023;5:27.
 48. Liu L, Liu J, Xu N. Ligand recognition and signal transduction by lectin receptor-like kinases in plant immunity. *Front Plant Sci.* 2023;14.
 49. Gómez-Gómez L, Boller T. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Mol Cell.* 2000;5:1003–11.
 50. Ranf S, Gisch N, Schäffer M, Illig T, Westphal L, Knirel YA, et al. A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in *arabidopsis thaliana*. *Nat Immunol.* 2015;16:426–33.
 51. Wang Y, Bouwmeester K, Beseh P, Shan W, Govers F. Phenotypic analyses of *arabidopsis* T-DNA insertion lines and expression profiling reveal that multiple L-type lectin receptor kinases are involved in plant immunity. *Mol Plant-Microbe Interact.* 2014;27:1390–402.
 52. Qu X, Zhao Z, Tian Z. ERECTA regulates cell elongation by activating auxin biosynthesis in *arabidopsis thaliana*. *Front Plant Sci.* 2017;8:1688.

53. Godiard L, Sauviac L, Torii KU, Grenon O, Mangin B, Grimsley NH, et al. ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant J.* 2003;36:353–65.
54. Wu X, Cai X, Zhang B, Wu S, Wang R, Li N, et al. ERECTA regulates seed size independently of its intracellular domain via MAPK-DA1-UBP15 signaling. *Plant Cell.* 2022;34:3773–89.
55. Yamaguchi Y, Pearce G, Ryan CA. The cell surface leucine-rich repeat receptor for At Pep1, an endogenous peptide elicitor in arabidopsis, is functional in transgenic tobacco cells. *Proc Natl Acad Sci.* 2006;103:10104–9.
56. Liu Z, Wu Y, Yang F, Zhang Y, Chen S, Xie Q, et al. BIK1 interacts with PEPRs to mediate ethylene-induced immunity. *Proc Natl Acad Sci.* 2013;110:6205–10.
57. Willmann R, Nürnberger T. How plant lysin motif receptors get activated: lessons learned from structural biology. *Sci Signal.* 2012;5:pe28–28.
58. Asai T, Tena G, Plotnikova J, Willmann MR, Chiu W-L, Gomez-Gomez L, et al. MAP kinase signalling cascade in arabidopsis innate immunity. *Nature.* 2002;415:977–83.
59. Okada K, Kubota Y, Hirase T, Otani K, Goh T, Hiruma K, et al. Uncoupling root hair formation and defence activation from growth inhibition in response to damage-associated pep peptides in arabidopsis thaliana. *New Phytol.* 2021;229:2844–58.
60. Katoch R, Tripathi A. Research advances and prospects of legume lectins. *J Biosci.* 2021;46:104.
61. Bellande K, Bono J-J, Savelli B, Jamet E, Canut H. Plant lectins and lectin receptor-like kinases: how do they sense the outside? *Int J Mol Sci.* 2017;18:1164.
62. Lannoo N, Van Damme EJ. Lectin domains at the frontiers of plant defense. *Front Plant Sci.* 2014;5:397.
63. Wang Q, Zhao X, Sun Q, Mou Y, Wang J, Yan C, et al. Genome-wide identification of the LRR-RLK gene family in peanut and functional characterization of AhLRR-RLK265 in salt and drought stresses. *Int J Biol Macromol.* 2024;254:127829.
64. Man J, Gallagher JP, Bartlett M. Structural evolution drives diversification of the large LRR-RLK gene family. *New Phytol.* 2020;226:1492–505.
65. Boller T, Felix G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu Rev Plant Biol.* 2009;60:379–406.
66. Pruitt RN, Schwessinger B, Joe A, Thomas N, Liu F, Albert M, et al. The rice immune receptor XA21 recognizes a tyrosine-sulfated protein from a Gram-negative bacterium. *Sci Adv.* 2015;1:e1500245.
67. McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 2006;7:212.
68. Bezerra-Neto JP, Araújo FC, Ferreira-Neto JR, Silva RL, Borges AN, Matos MK, et al. NBS-LRR genes—plant health sentinels: structure, roles, evolution and biotechnological applications. *Applied Plant Biotechnology for Improving Resistance to Biotic Stress.* Elsevier; 2020. pp. 63–120.

69. Shao Z-Q, Xue J-Y, Wu P, Zhang Y-M, Wu Y, Hang Y-Y, et al. Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol.* 2016;170:2095–109.
70. Xue J-Y, Zhao T, Liu Y, Zhang Y-X, Zhang G-Q, Chen H et al. Genome- Wide Analysis of the Nucleotide Binding Site Leucine-Rich Repeat Genes of Four Orchids Revealed Extremely Low Numbers of Disease Resistance Genes. *Front Genet.* 2020;10.
71. Soltabayeva A, Dauletova N, Serik S, Sandybek M, Omondi JO, Kurmanbayeva A, et al. Receptor-like kinases (LRR-RLKs) in response of plants to biotic and abiotic stresses. *Plants.* 2022;11:2660.
72. Cambi A, Koopman M, Figdor CG. How C-type lectins detect pathogens. *Cell Microbiol.* 2005;7:481–8.
73. Shiu S-H, Karlowski WM, Pan R, Tzeng Y-H, Mayer KF, Li W-H. Comparative analysis of the receptor-like kinase family in arabidopsis and rice. *Plant Cell.* 2004;16:1220–34.
74. Gao L-L, Xue H-W. Global analysis of expression profiles of rice receptor-like kinase genes. *Mol Plant.* 2012;5:143–53.
75. Jiang Y, Hu X, Yuan Y, Guo X, Chase MW, Ge S, et al. The gastrodia menghaiensis (orchidaceae) genome provides new insights of orchid mycorrhizal interactions. *BMC Plant Biol.* 2022;22:179.
76. Xu Y, Lei Y, Su Z, Zhao M, Zhang J, Shen G, et al. A chromosome-scale *Gastrodia elata* genome and large-scale comparative genomic analysis indicate convergent evolution by gene loss in mycoheterotrophic and parasitic plants. *Plant J.* 2021;108:1609–23.
77. Rogozin IB, Carmel L, Csuros M, Koonin EV. Origin and evolution of spliceosomal introns. *Biol Direct.* 2012;7:11.
78. Jeffares DC, Penkett CJ, Bähler J. Rapidly regulated genes are intron poor. *Trends Genet.* 2008;24:375–8.
79. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. *Proc Natl Acad Sci U S A.* 2003;100:15682–7.
80. Mehan MR, Freimer NB, Ophoff RA. A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture. *Hum Genomics.* 2004;1:1–10.
81. Hofberger JA, Nsibo DL, Govers F, Bouwmeester K, Schranz ME. A Complex Interplay of Tandem- and Whole-Genome Duplication Drives Expansion of the L-Type Lectin Receptor Kinase Gene Family in the Brassicaceae. *Genome Biol Evol.* 2015;7:720–34.
82. Zhu K, Wu Q, Huang Y, Ye J, Xu Q, Deng X. Genome-wide characterization of cis-acting elements in the promoters of key carotenoid pathway genes from the main species of genus citrus. *Hortic Plant J.* 2020;6:385–95.
83. Liu P-L, Wan J-N, Guo Y-P, Ge S, Rao G-Y. Adaptive evolution of the chrysanthemyl diphosphate synthase gene involved in irregular monoterpene metabolism. *BMC Evol Biol.* 2012;12:214.
84. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell.*

2003;15:63–78.

85. Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M, et al. Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol.* 2006;47:141–53.
86. Dolferus R, Jacobs M, Peacock WJ, Dennis ES. Differential interactions of promoter elements in stress responses of the arabidopsis adh gene. *Plant Physiol.* 1994;105:1075–87.
87. Wu W, Yang J, Yu N, Li R, Yuan Z, Shi J et al. Evolution of the WRKY family in angiosperms and functional diversity under environmental stress. *Int J Mol Sci.* 2024;25.
88. Kaur D, Schedl A, Lafleur C, Martinez Henao J, van Dam NM, Rivoal J et al. Arabidopsis transcriptomics reveals the role of lipoxygenase2 (AtLOX2) in wound-induced responses. *Int J Mol Sci.* 2024;25.
89. Gao L-L, Xue H-W. Global analysis of expression profiles of rice receptor-like kinase genes. *Mol Plant.* 2012;5:143–53.
90. Sun X, Wang G-L. Genome-Wide Identification, Characterization and Phylogenetic Analysis of the Rice LRR-Kinases. *PLoS ONE.* 2011;6:e16079.
91. Tang D, Wang G, Zhou J-M. Receptor kinases in plant-pathogen interactions: more than pattern recognition. *Plant Cell.* 2017;29:618–37.
92. Song J, Liu Q, Hu B, Wu W. Comparative transcriptome profiling of Arabidopsis Col-0 in responses to heat stress under different light conditions. *Plant Growth Regul.* 2016;79:209–18.
93. Corwin JA, Copeland D, Feusier J, Subedy A, Eshbaugh R, Palmer C, et al. The quantitative basis of the arabidopsis innate immune system to endemic pathogens depends on pathogen genetics. *PLOS Genet.* 2016;12:e1005789.
94. Yuan M, Jiang Z, Bi G, Nomura K, Liu M, Wang Y, et al. Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature.* 2021;592:105–9.
95. Liu Y, Cai Y, Li Y, Zhang X, Shi N, Zhao J et al. Dynamic changes in the transcriptome landscape of arabidopsis thaliana in response to cold stress. *Front Plant Sci.* 2022;13.

Figures

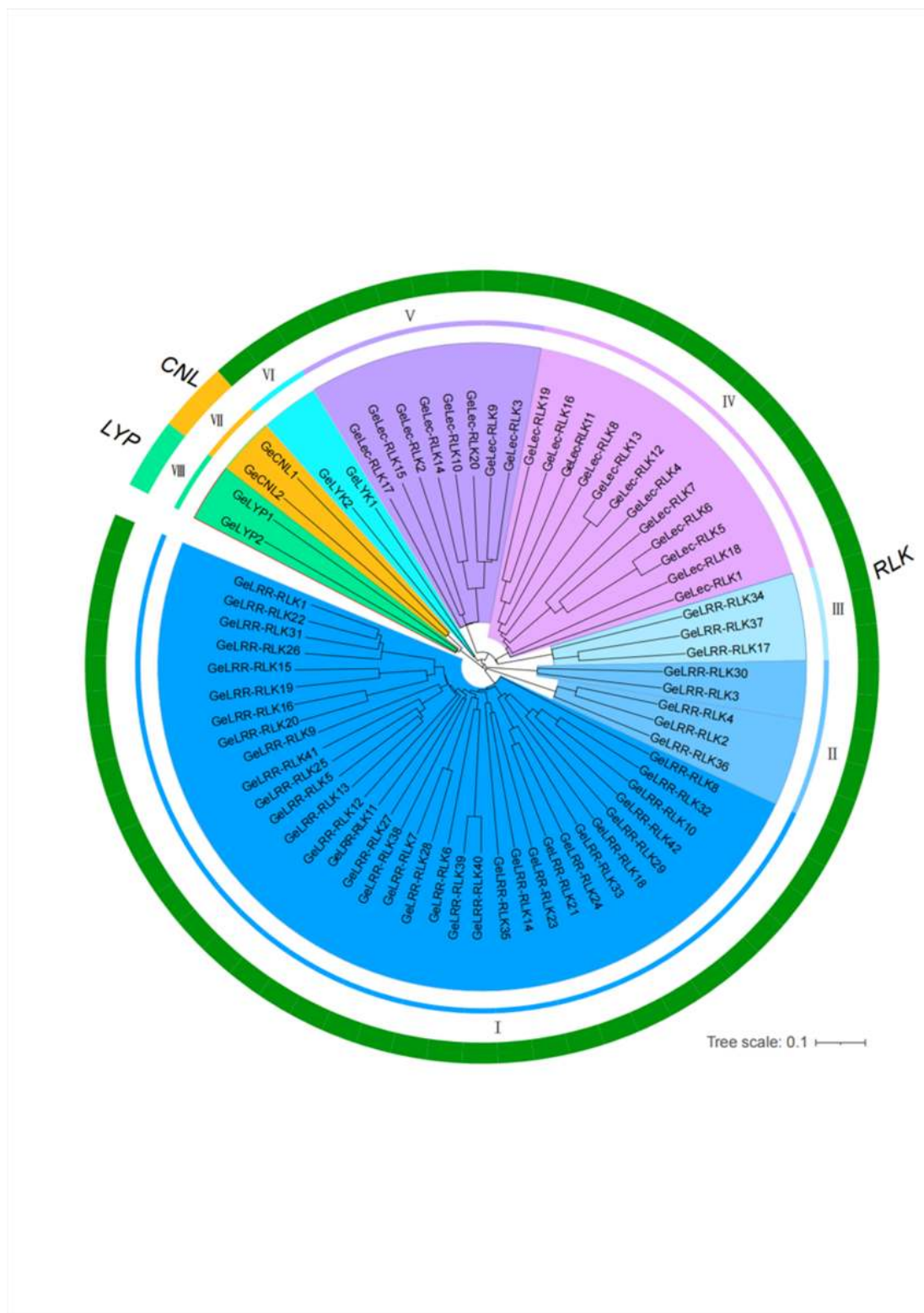


Figure 1

phylogenetic tree of 68 *GePRGs*. The 68 genes were clustered into eight branches, among which branches 1 to 3 belong to the *LRR-RLK* gene family, branches 4 and 5 to the *Lec-RLK* gene family, and branches 6, 7, and 8 to the *LYK*, *CNL*, and *LYP* gene families, respectively. The outer ring indicates that *LRR-RLK*, *Lec-RLK*, and *LYK* all belong to subfamilies under the *RLK* gene family.

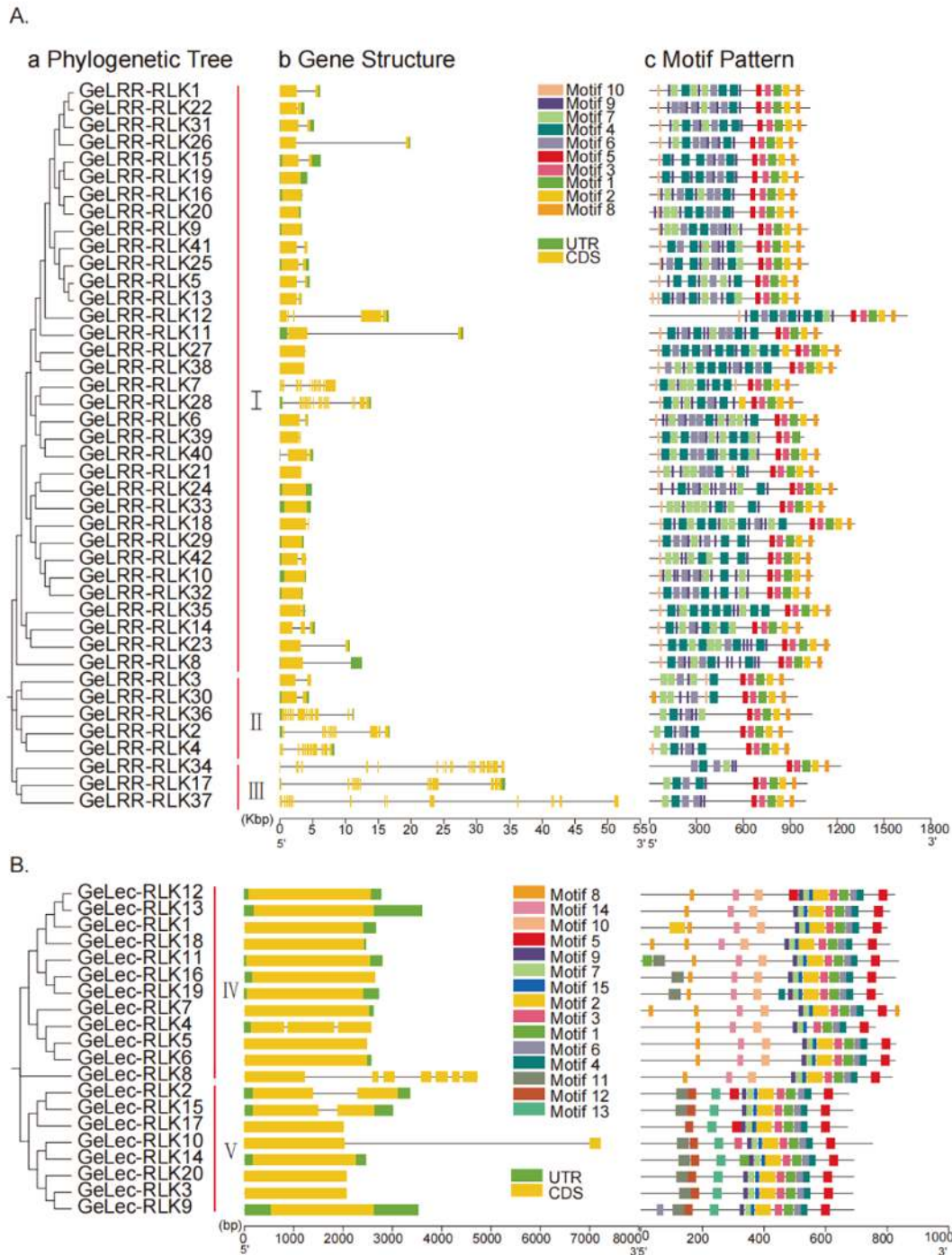


Figure 2

Phylogenetic relationships, gene structure, motif distribution of 62 *GePRG* proteins in *G. elata*. **A** is *GeLRR-RLK* genes. **B** is *GeLec-RLK* genes. **a.** Phylogenetic tree with 1000 replicates per node. **b.** Gene structure map. Green rectangles, yellow rectangles and black lines respectively indicate UTR (untranslated), CDS (coding sequence) and introns. **c.** Motif composition. Amino acid conserved motif respectively is Motif 1-10 and 1-15 in *GeLRR-RLKs* and *GeLec-RLKs*, with different color blocks

corresponding to different conserved motifs, and black lines indicating the relative length of corresponding proteins.

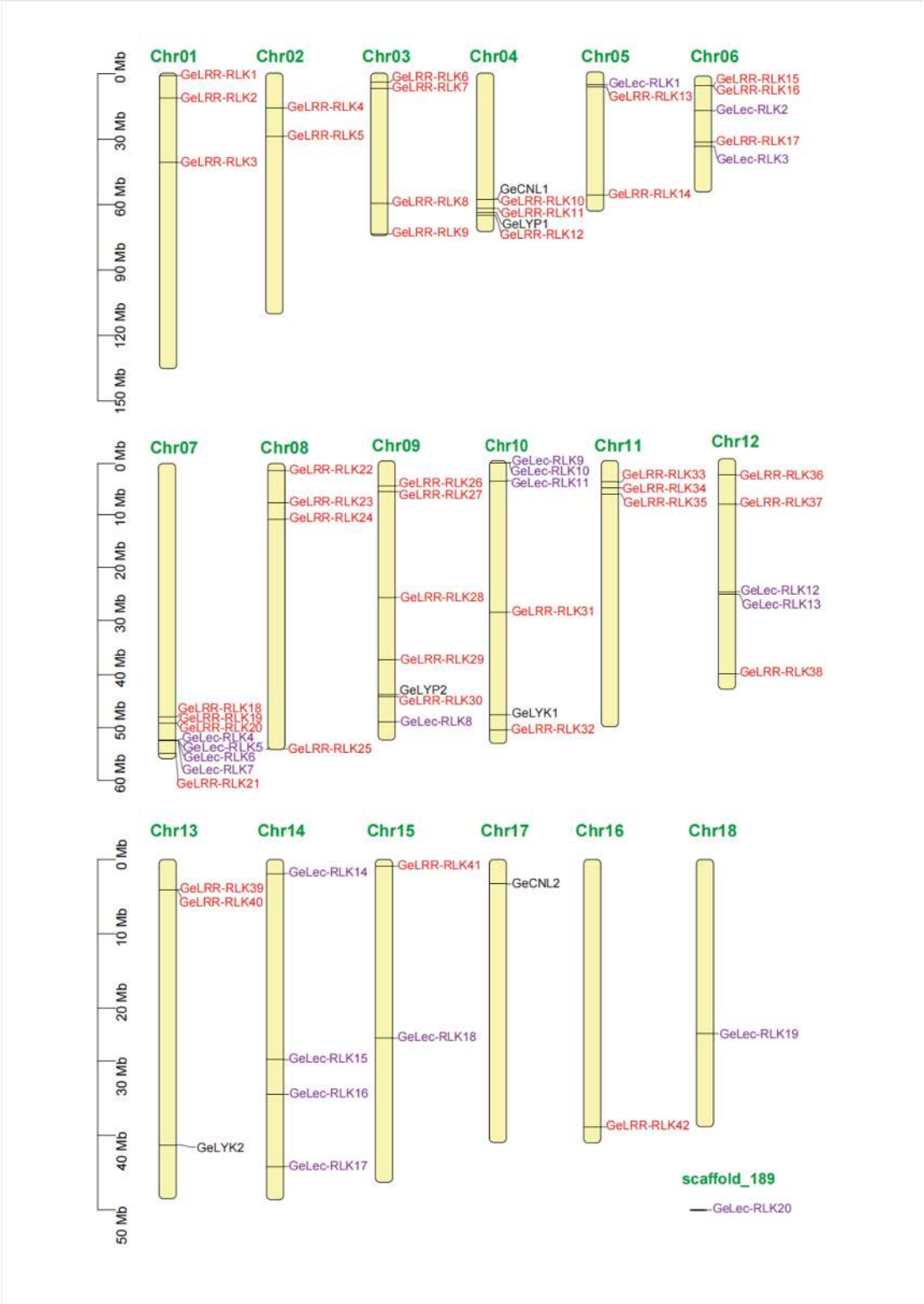


Figure 3

Chromosomal mapping of 68 GePRGs in the genome of dark red *Gastrodia elata*. The yellow bar chart shows the chromosomes of *G. elata*, with the chromosome number marked at the top of each

chromosome. The scale on the left indicates chromosome length. Red and purple correspond to the *GeLRR-RLKs* and *GeLec-RLKs* genes respectively, while black is the *GeLYKs*, *GeLYPs* and *CNLs* genes.

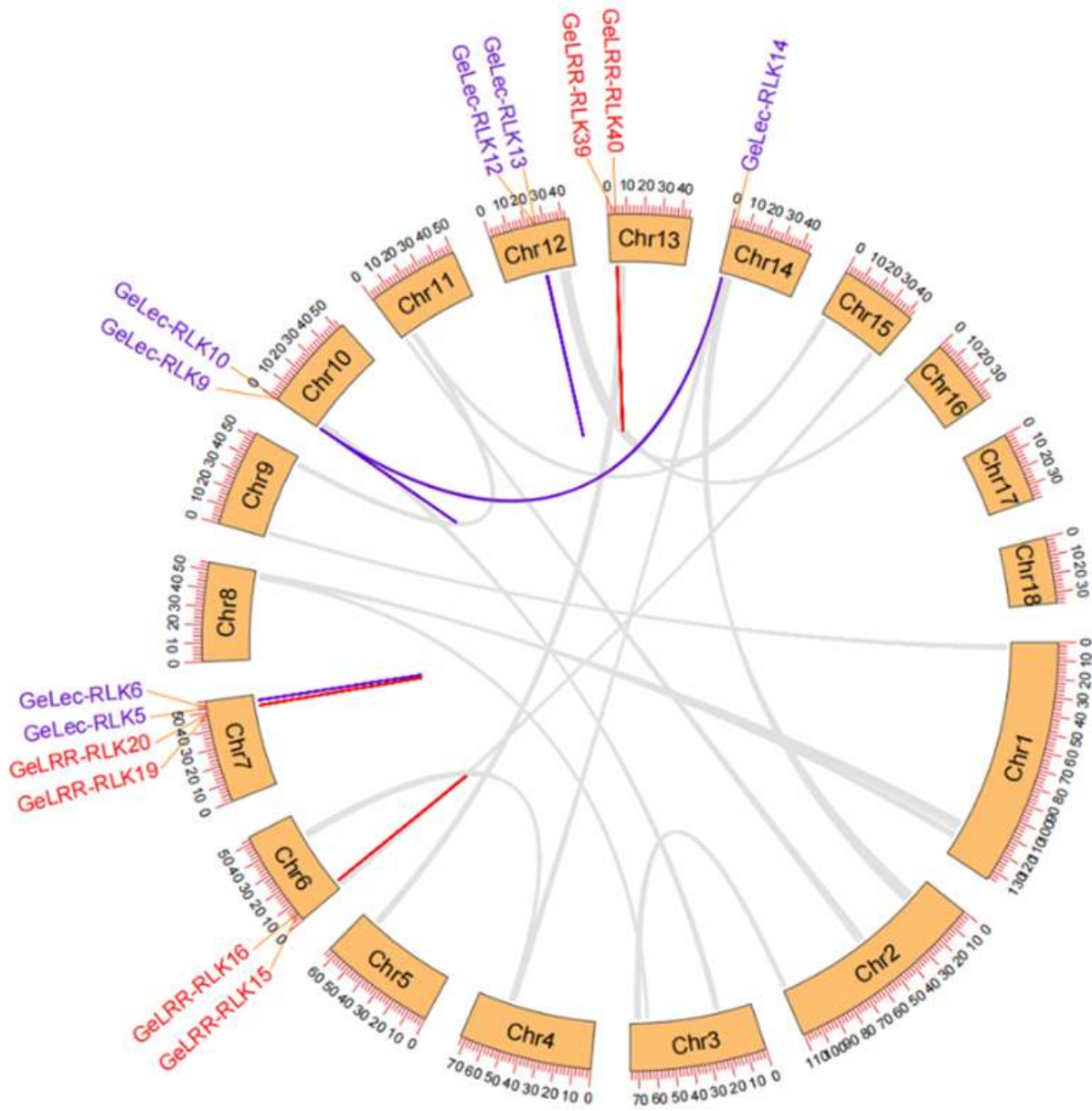


Figure 4

Collinear region of *GePRG* genes in *G. elata* genome. The gray lines represent the collinear blocks in the genome of *G. elata*, the red lines and purple lines respectively represent *GeLRR-RLKs* and *GeLec-RLKs* gene pairs or tandem duplication. Chromosome numbers are shown on the inside of each chromosome.

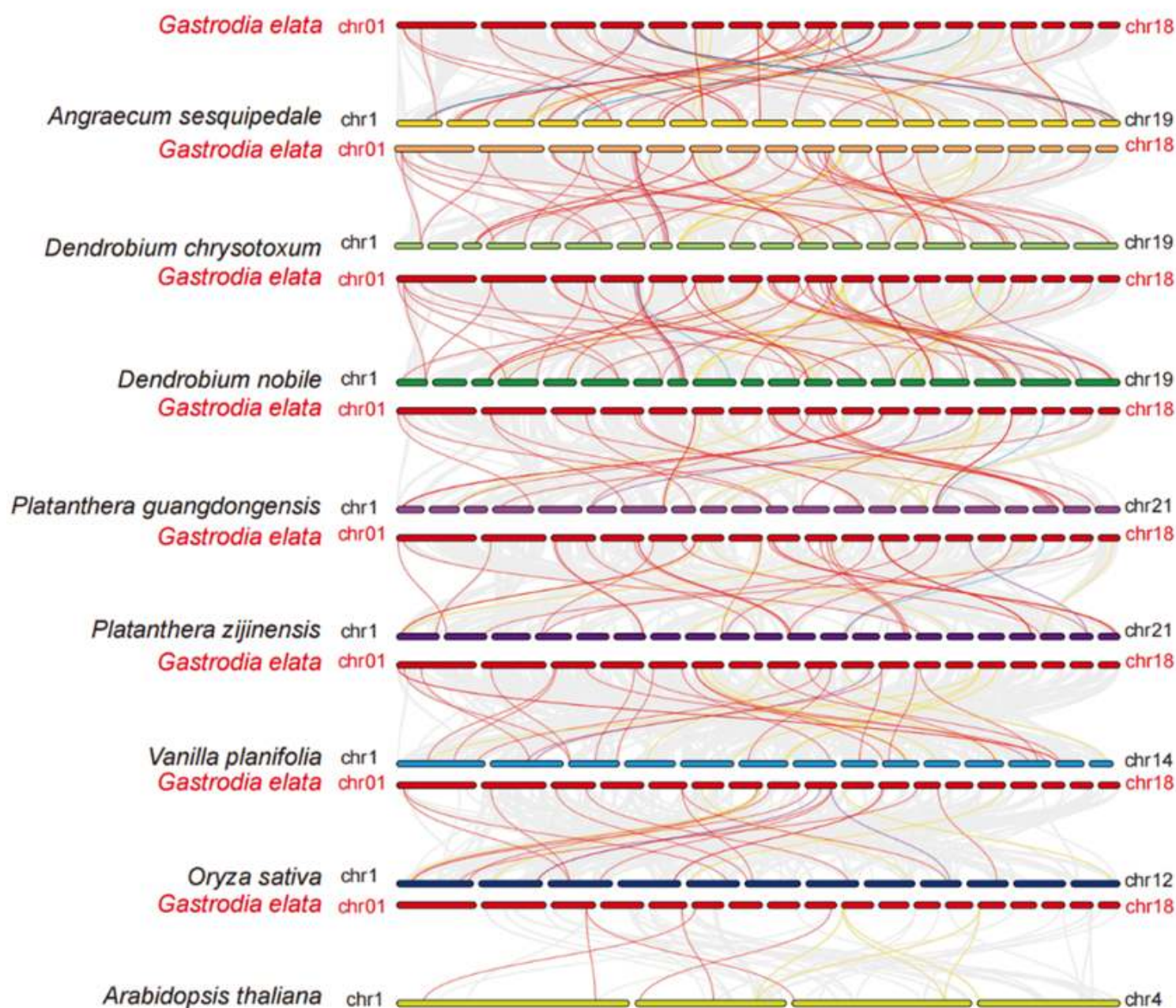


Figure 5

Synteny analysis of the *GePRG* genes between *G. elata* and eight representative plants (*A. sesquipedale*, *D. chrysotoxum*, *D. nobile*, *P. guangdongensis*, *P. zijinensis*, *V. planifolia*, *O. sativa* and *A. thaliana*). The red, yellow, purple, and blue lines respectively represent the *GePRG* genes of the *GeLRR-RLKs*, *GeLec-RLK*, *LYK/LYP* and *CNL* gene families in the genomes of *G. elata* and other plants, while the gray lines depict collinear blocks.

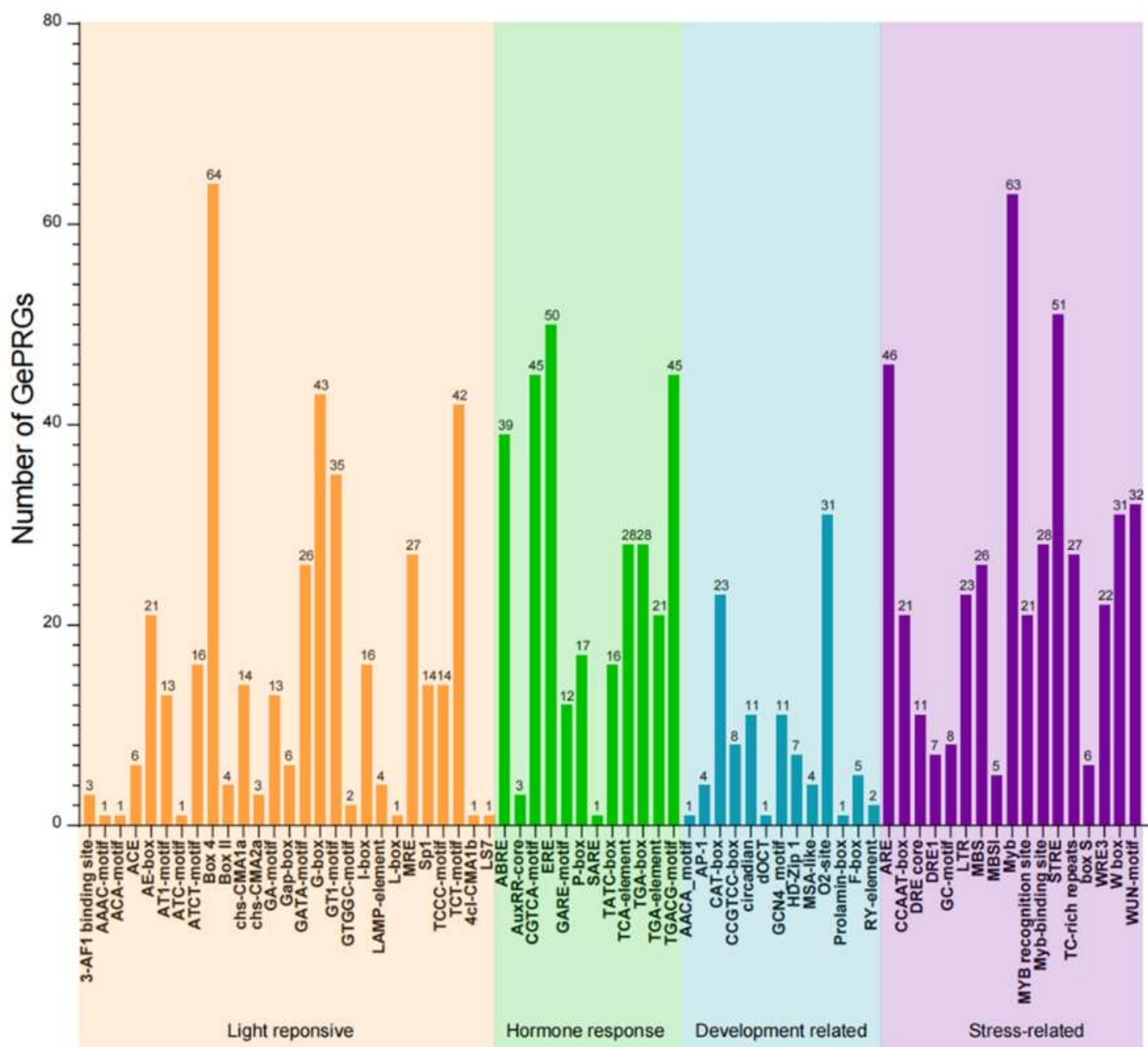


Figure 6

Four categories of *cis*-acting elements in *GePRGs*. The x-axis represents *cis*-acting elements, which are classified into four categories: yellow for light-responsive elements, green for hormone-responsive elements, blue for development-related elements, and purple for stress-related elements. The y-axis indicates the number of each *cis*-acting element present in the *GePRGs*.

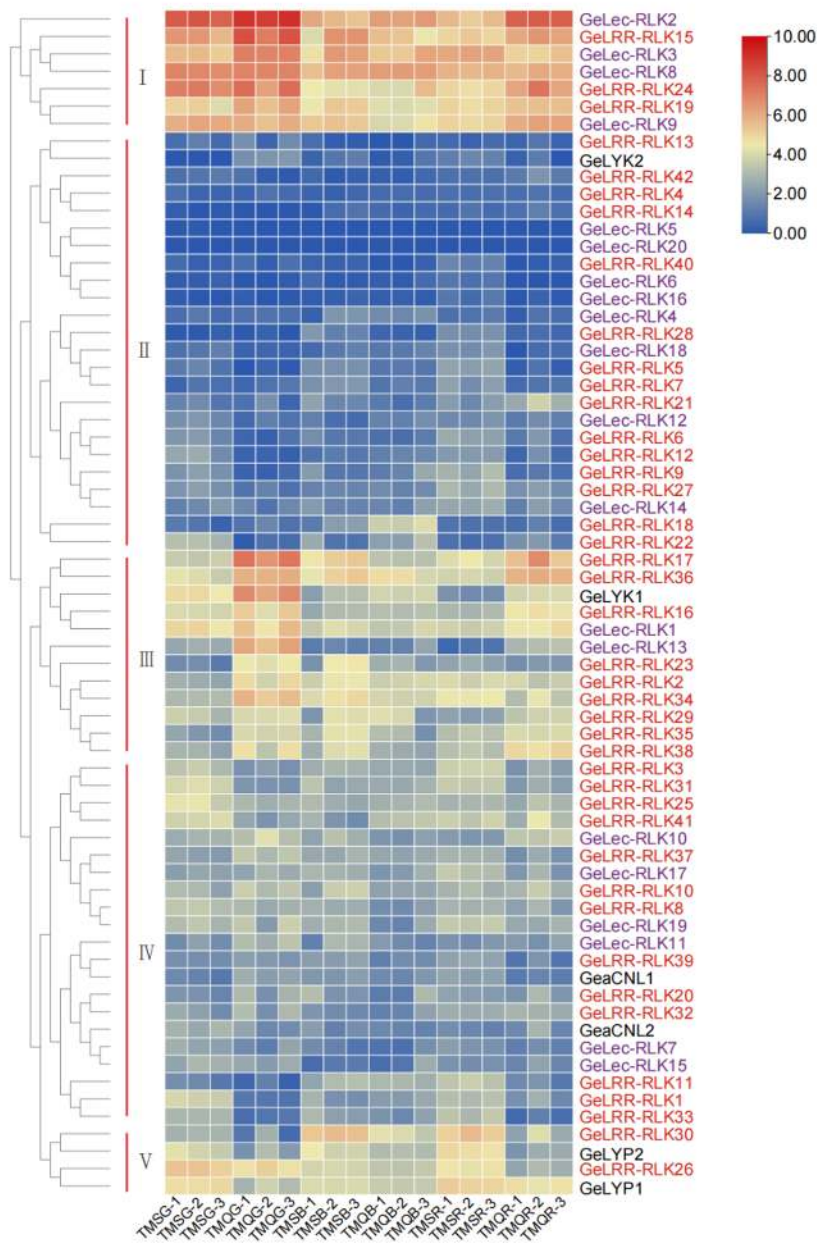


Figure 7

Heatmap of transcriptional expression in the flowers and stems of six *G. elata* forms. DGGE: dark green *G. elata*, LGGE: light green *G. elata*, DBGE: dark brown *G. elata*, LBGE: light brown *G. elata*, DRGE: dark red *G. elata*, LRGE: light red *G. elata*. Based on the expression profiles of 68 *GePRGs* across six samples of the flower and stem, the genes were clustered into five distinct expression patterns. Different colors

represent genes from different families: red and purple correspond to *GeLRR-RLK* and *GeLec-RLK* genes, respectively, while black denotes *GeLYK*, *GeLYP*, and *GeCNL* genes.

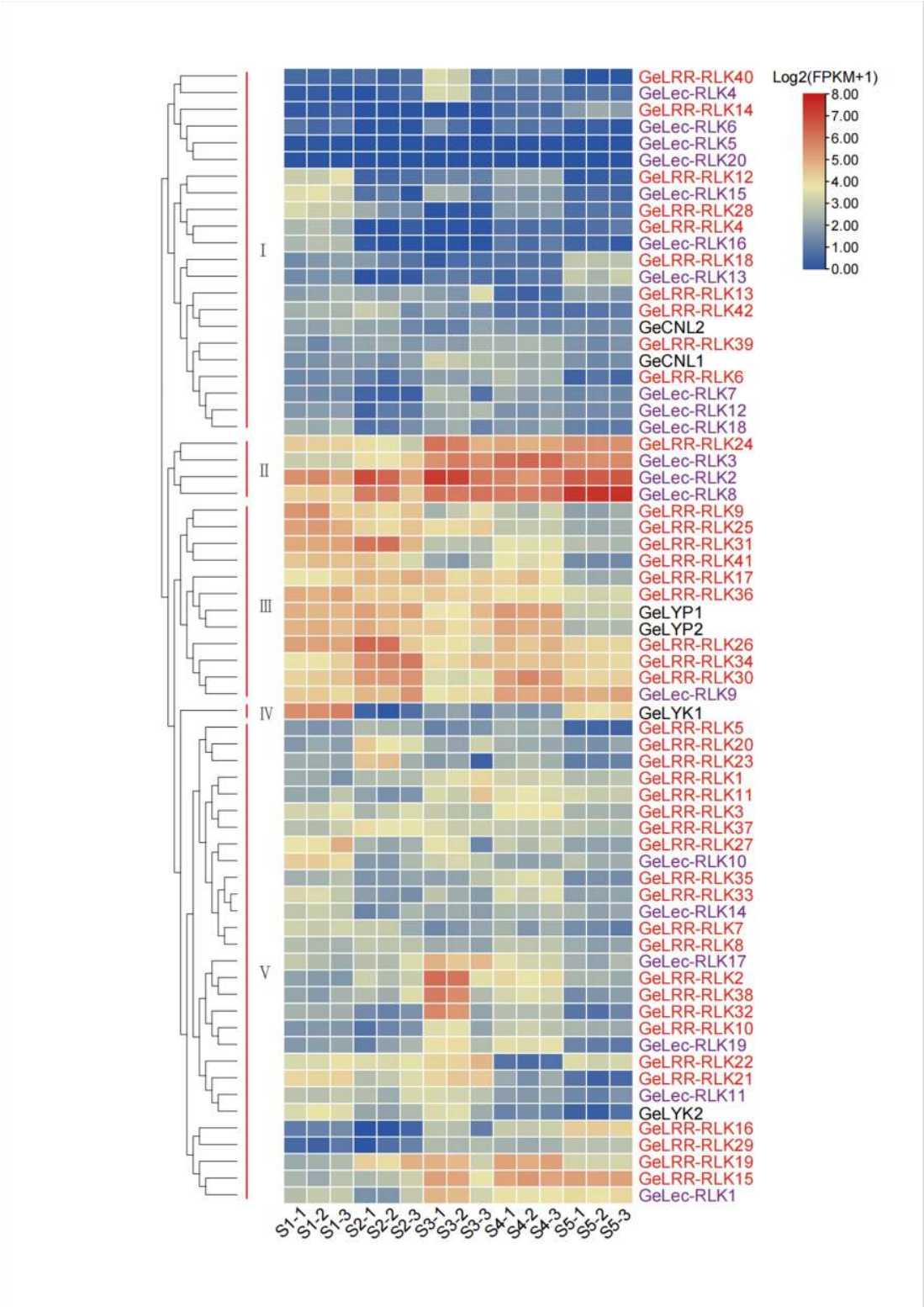


Figure 8

Heatmap of expression data for 68 *GePRG* genes in five developmental stages of dark red *G. elata*, including immature tuber (S1), developing tuber (S2), mature tuber (S3), flowers and stems (S4) and capsule (S5). Based on gene expression profiles across five developmental stages, the genes were

clustered into five expression patterns. Different colors represent genes from different families: red and purple indicate *GeLRR-RLK* and *GeLec-RLK* genes, respectively, while black represents *GeLYK*, *GeLYP*, and *GeCNL* genes.

Supplementary Files

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

- [Supplementarytables.xlsx](#)