Genome-wide identification and characterization of lectin receptor-like kinase gene family in cucumber and expression profiling analysis under different treatments

Duo Lv
Shanghai Jiao Tong University
Gang Wang
Shanghai Jiao Tong University
Yue Chen
Shanghai Jiao Tong University
Liang-Rong Xiong
Shanghai Jiao Tong University
Jing-Xian Sun
Shanghai Jiao Tong University
Chun-Li Guo
Shanghai Jiao Tong University
Yao Yu
Shanghai Jiao Tong University
Huan-Le He
Shanghai Jiao Tong University
Run Cai
Shanghai Jiao Tong University
Junsong Pan (jspan71@sjtu.edu.cn)
Shanghai Jiao Tong University https://orcid.org/0000-0003-3983-1285

Research article

Keywords: Cucumber, Expression analysis, Genome-wide analysis, Lectin receptor-like kinase, Phylogenetic analysis

Posted Date: January 3rd, 2020

DOI: https://doi.org/10.21203/rs.2.20023/v1

Version of Record: A version of this preprint was published at Genes on September 2nd, 2020. See the published version at https://doi.org/10.3390/genes11091032.

Abstract

Background

Lectin receptor-like kinases (LecRLKs) are a class of membrane proteins found in plants that are involved in diverse functions, including plant development and stress responses. Although LecRLK families have been identified in a variety of plants, a comprehensive analysis has not yet been undertaken in cucumber (Cucumis sativus L.).

Results

In this study, 46 putative LecRLK genes were identified in cucumber genome, including 23 G-type, 22 L-type and 1 C-type LecRLK genes. They unequally distributed on all 7 chromosomes with a clustering trendency. Most of the genes in the cucumber LecRLK (Cs LecRLK) gene family lacked introns. In addition, there were many regulatory elements associated with phytohormone and stress on these genes' promoters. Transcriptome data demonstrated that distinct expression patterns of CsLecRLK genes in various tissues. Furthermore, we found that each member of the CsLecRLK family had its own unique expression pattern under hormone and stress treatment by the quantitative real time PCR (qRT-PCR) analysis.

Conclusion

This study provides a better understanding of the evolution and function of LecRLK gene family in cucumber, and opens the possibility to explore the roles that LecRLK s might play in the life cycle of cucumber.

Background

Plants are immobile creatures compared with animals, so they have more sensitive signal sensors to cope with the changing environment around them. In the long-term evolutionary process, plants have evolved a complete set of signal receptor proteins. After receiving external stimulus, they would transmit signals to downstream pathways the first time to allow plant respond to the stimulus. Cell surface receptors, a kind of signal receptor proteins, play important roles in receiving and transmitting environmental signals. The receptor-like kinase (RLK) family, one important family of cell surface receptors, contains three kinds of domains, such as extracellular domain, transmembrane domain (TM) and intracellular kinase domain. RLK proteins could be classified into different families based on the structure of the extracellular domains and intracellular kinase domains.

Lectin receptor-like kinases (LecRLKs), a class of RLKs that contain a lectin domain within the extracellular domain, are a gene family that is specialized for sensing external environmental stimuli and transmitting signals. They are localized on the cell membrane, relying on N terminus diverse extracellular ligand recognition domains (also called lectin domain) to recognize various environmental stimulus, and then phosphorylate downstream proteins through their C terminus intracellular kinase domain to pass received signals (Bouwmeester and Govers, 2009).

Based on the identity of lectin domains, the LecRLKs have been divided into three subfamilies (Fig. 1): L-type, G-type, and C-type LecRLKs (Vaid et al., 2012). These subclasses are very distinct from each other, with sugar-binding ability of lectin domain. The G-type LecRLKs possess α-D mannose specific plant lectins, which are also accompanied in most of the proteins (not necessarily all) by both EGF and PAN domain motif or one of them.

The EGF motif is cysteine rich (Shiu and Bleecker, 2001) and probably takes part in the formation of disulfide bonds (Vaid et al., 2012). The PAN motif is believed to be involved in protein-protein and protein carbohydrate interactions (Naithani et al., 2007). As the name suggests the lectin domain of L-type LecRLKs resembles soluble lectin protein found in leguminous plants (Hervé et al., 1999). The third class of LecRLK is C-type lectin kinase. The C-type lectin kinases in plants are thought to be homologues of calcium-dependent lectin motifs which are a large group of mammalian proteins known to be involved in innate immune responses and pathogen recognition (Cambi et al., 2005). Though C-type LecRLKs are present in large number in mammalian system, only one gene encoding for C type LecRLK exists in rice and Arabidopsis respectively (Bouwmeester and Govers, 2009).

The roles of LecRLKs in plants are believed to be primarily involved in plant development, innate immunity and abiotic responses. Previous reports have confirmed that LecRLK family is involved in plant root development (Cheng et al., 2013), pollen development (Wan et al., 2008), cotton fiber development (Zuo et al., 2004), and hormone signal recognition (Deng et al., 2009). At the same time, LecRLKs also play an irreplaceable role in plant resistance to diseases, insect pests and stresses, such as salt stress (Li et al., 2014; He et al., 2004), wounding (Riou et al, 2002), and fungal pathogen (Ohtake, 2000; Desclos et al., 2012). Compared to L-type and G-type LecRLKs, the C-type LecRLK is the most mysterious member of the family. Although there is only one subfamily member in Arabidopsis and rice, the C-type LecRLK had not been reported involving any specific biological traits for a long time until (Guo et al., 2017) identified that a C-type LecRLK may affect morphogenesis of trichome in cucumber (Cucumis sativus). This study suggested that the LecRLKs family may be involved in more biological pathways.

Genome-wide analysis of the LecRLK gene family has been done in Arabidopsis thaliana (Vaid et al., 2012), rice (Vaid et al., 2012), bread wheat (Shumayla et al., 2016), soybean (Liu et al., 2018), and Populus trichocarpa (Yang et al., 2016), but no previous studies have been reported in Cucurbitaceae plant. Cucumber is an important economic crop of Cucurbitaceae, Genome-wide analysis of the LecRLK gene family has been done in Arabidopsis thaliana (Vaid et al., 2012), rice (Vaid et al., 2012), bread wheat (Shumayla et al., 2016), soybean (Liu et al., 2018), and Populus trichocarpa (Yang et al., 2016). Cucumber is an important economic crop of Cucurbitaceae, but there are only a few early studies reporting on the presence of LecRLKs (Wu et al., 2014). A comprehensive understanding of LecRLKs in cucumber is still lacking. Here, we did a complete identification and analysis of the entire LecRLK gene family in cucumber, 46 LecRLK genes (CsLecRLKs) identified. Furthermore, we analyzed their phylogenetic relationship, gene structure, conserved domain, gene duplications, chromosome distribution and cis-acting elements on the promoters. Finally, we profiled the expression of the predicted genes in different tissues and response to gibberellin (GA), abscisic acid (ABA),1-Naphthaleneacetic acid (NAA), indole-3-acetic acid (IAA) and cold treatments in cucumber. Our study provides valuable information for further functional research on the LecRLK gene family in cucumber.

Materials And Methods

Identification of the *LecRLK* genes in cucumber

The whole genome and protein sequence data of cucumber were downloaded from a public database (http://cucurbitgenomics.org/) (Cucumber (Chinese Long) genome v2). The Hidden Markov Model (HMM) was used to identify cucumis sativus *LecRLK* candidates, and the HMM profiles of LecRLKs were downloaded from the Pfam protein database (http://pfam.xfam.org/), the models for these files are L-type (Lectin_legB PF00139),

G-type (B_lectin PF01453) and C-type (Lectin_C PF00059). We used HMMER 3.0 (Johnson et al., 2010) to search three types of genes from cucumber protein sequence data with e-value cutoffs of ≤0.001. Then we further examined these selected genes by PFAM (Finn et al., 2016) and SMART (Letunic) program to ensure that each protein contains conserved lectin domain, transmembrane domain and a kinase domain. The LecRLK family data of rice and *Arabidopsis* were downloaded from a previous research literature (Vaid et al., 2012).

Phylogenetic analysis

The full-length protein sequences of CsLecRLKs were aligned by the MUSCLE program with the default parameters (Edgar, 2004). The phylogenetic tree was constructed through Neighbor-Joining (NJ) method by MEGA 7.0.21, with the following parameters: Poisson model, pairwise deletion, and 1000 bootstrap replications.

Conserved domain, motif identification and gene structure analysis

The conserved motifs of the CsLecRLK were predicted by the MEME program (Bailey, 1994), the parameters were set as any number of repetitions, optimum motif width of 6–210 residues, and searching for 10 motifs, with other parameters at default. The Gene Structure Display Server (Hu et al., 2014) was used to show the exon–intron structures of *CsLecRLK* genes.

Gene location and duplication analysis of *CsLecRLKs*

The location information of *CsLecRLK*s was from the genome annotation files of Cucurbit Genomics Databases (http://cucurbitgenomics.org/) by a series of in-house Perl scripts. The map of gene location was constructed with MapChart software described previously (Voorrips, 2002). We used two methods to find duplication events among the *CsLecRLK*s, one way was that Gene duplication was confirmed with two criteria by an in-house Perl scripts: (a) the shorter aligned sequence covered >70% of the longer sequence in length; (b) the similarity of aligned sequences was >70% (Gu, 2002; Yang, 2008). Two genes located in the same chromosomal fragment of less than 100 kb and separated by five or fewer genes were identified as tandem duplicated genes (Mehan et al., 2004). Another way was that we used the Multiple Collinearity Scan toolkit (MCScanX) to analyze the gene duplication events, with the default parameters (Wang et al., 2012).

Ks (synonymous substitution rate) and *Ka* (nonsynonymous substitution rate) values of tandem duplicated genes were calculated by the method of Nei and Gojobori as implemented in KaKs_calculator (Zhang et al., 2006) based on the coding sequence alignments. The divergence time was calculated based on the formula T = *Ks*/2*r*, with Ks being the synonymous substitutions per site and *r* being the rate of divergence for nuclear genes from plants. The *r* was taken to be 1.5×10^{-8} synonymous substitutions per site per year for dicotyledonous plants (Olds et al., 2000).

Analysis of *cis*-acting elements

The upstream 1500bp of each *CsLecRLK* was obtained from the genome annotation files of Cucurbit Genomics Databases (http://cucurbitgenomics.org/) by a series of in-house Perl scripts, then scanned *cis*-acting elements contained in these sequences using the Plantcare Databases (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), in the process of analysis, we filtered out the *cis*-acting elements that are ubiquitous in most genes, for example CAAT-box, TATA-box, TATC-box and so on, only showing that may be typical and functional *cis*-acting elements presented in the *CsLecRLK* gene family. Then, the

structures and annotations of the promoters were generated by GSDS (http:// gsds.cbi.pku.edu.cn/;) (Hu et al., 2015).

Expression pattern analysis

For expression profiling of the *CsLecRLK* genes in 10 tissues, we utilized the Illumina RNA-seq data that were previously generated by our lab. Those tissues of cucumber include 10mm floral bud (remove sepals), 10mm ovary (remove sepals), pericarp (one day before flowering and remove the trichomes), pistil (one day before flowering), stamen (one day before flowering), root (seeds were place on a moist sprouting paper for 72 hours), stem, tendril (not stretch), fruit spines (one day before flowering) and cotyledon (extend to 2 cm). The Heatmap of the gene FPKM values in 10 tissues of *CsLecRLK*s was drawn by the R program. We also selected 5 typical tissues, included floral bud, ovary, cotyledon, fruit spines and root, to draw Venn diagrams using the R program.

Hormones and cold treatments

The typical cucumber line '9930' was used as the experimental material to investigate the expression pattern in response to various phytohormones and cold treatments. Cucumber seeds were soaked in 55℃ water for 2h and germinated on petri dish in a growth chamber at 28℃ in the dark for 2d. The germinated seeds were grown in pots containing peat: vermiculite mixture (3:1) in the greenhouse of Shanghai Jiao Tong University, and the controlled environment growth chamber was programed for light 16 h/25°C and dark 8 h/20°C. After germination for 4 weeks, the seedlings were placed into hydroponic boxes with 1/2 MS liquid solution (pH 5.8, without sugar) for 1 week to adapt to the environment (root was lucifugal), and then were treated with 100 mM indole-3-acetic acid (IAA), 100 mM 1-Naphthaleneacetic acid (NAA), 100 mM abscisic acid (ABA), or 100 mM gibberellin (GA) for 3h under the same growth conditions as described earlier, respectively. The 1/2 MS liquid solution without any hormones was used as a control. Another group of seedlings was treated at 4°C for 1h, 25°C was used as a control. Each treatment consisted of three replicates.

RNA extraction and gene expression analysis

Roots were collected from plants treated with different hormones, and young leaves were collected from plants subject to 4°C. Total RNA was extracted using the RNeasy Plant Mini Kit (Cwbio, Beijing, China). The first strand cDNA was prepared according to the PrimeScript RT reagent Kit with gDNA Eraser (Cwbio, Beijing, China) protocol. To identify the relative expression level of different *LecRLK* genes under different treatment, qRT-PCR was conducted using FastStart Essential DNA Green Master (Roche, Mannheim, Germany). *CsActin3* (*Csa6G484600.1*) was used as an internal control. qRT-PCR was performed in a total volume of 20 μL, containing 2 μL of cDNA, 10 μL SYBR mix, 2 μL gene-specific primers (10 μM) and 6 μl ddH_2O, using the CFX Connection Real-Time System (Bio-Rad, California, USA) with 40 cycles of 5s at 95°C, 30s at 60°C. Each experiment was repeated three times, and each experiment included three biological repeats. The data from real-time PCR amplification was analyzed using $2^{-\triangle\triangle CT}$ method. Primers used for qRT-PCR were followed as Table S3. Statistical difference was determined by t-test (*P < 0.05, n = 3) using Microsoft Excel 2010. The Heatmap of the change fold values of *CsLecRLK*s under five treatments was drawn by the R program.

Results

Genome-wide identification of the *LecRLK*s in cucumber

We identified a total of 46 *LecRLK* genes which named *CsLecRLK*s (Table1) in the cucumber genome by the Pfam and Smart search. The total number of *LecRLK*s in cucumber is less than that in Arabidopsis (75 *LecRLK* genes) or rice (173 *LecRLK* genes) (Vaid et al., 2012). The 46 CsLecRLKs were classified into 23 G-type, 22 L-type, and one C-type based on their extracellular lectin domain. The molecular weight (MW) of the proteins ranged from 62.5 kDa (Csa1G056960) to 94.5 kDa (Csa3G733860), and the isoelectric point (Ip) ranged from 4.98 (Csa4G289630) to 9.51 (Csa1G056960), the range of CDS length was 1,803–2,502bp. With the predicted protein structures, it could be considered that most of the CsLecRLKs were localized on the plasma membrane, only Csa7G045520 was located on the extracellular. More information of *CsLecRLK*s, including the length of the gene, the length of CDS, the length of the protein sequence, the protein MW and pI were listed in Table1.

By analyzing the molecular weight of all 46 CsLecRLKs, we found that the weight of G-type CsLecRLKs (83.2 kDa) are generally larger than L-type (62.5 kDa) and C-type (74.6 kDa). This may be mainly due to the fact that in addition to the lectin domain, G-type CsLecRLKs often contain the EGF and PAN domains (Fig. 1). Signal peptides and transmembrane domain (TM) domains are critical for protein localization. The software prediction indicated that not each CsLecRLK had signal peptide and unique TM domain. The loss of signal peptide or TM domain would directly affect the localization of proteins in cells (Table 1). The plasma membrane localization of most of the CsLecRLKs indicated that they are signal receptors which can sense extracellular signals and then transmit the signals to the interior of the cells.

Phylogenetic analysis of the CsLecRLKs

We constructed an unrooted phylogenetic tree by the MEGA 7.0.21 (Fig. 2). As expected, the phylogenetic tree showed that the CsLecRLK family could be classified into three subgroups of L-type, G-type, and C-type. This result is consistent with the domain-based classification of CsLecRLK family. The phylogenetic tree indicated that the L-type and C-type had a closer relationship. This result was different from previous reports in *Arabidopsis* and rice, which revealed a closer genetic relationship between G-type and L-type (Vail et al., 2012). As shown in Fig. 3, the phylogram of G-type and L-type CsLecRLKs could be divided into four and three sub-groups respectively. The division of individual clades was supported by high bootstrap values.

Exon–Intron Structural Analysis of *CsLecRLK*s

The genomic sequence and corresponding cDNA sequence of the *CsLecRLK*s were submitted to GSDS (Gene Structure Display Server) together for analyzing their gene structure (Fig. 3). The genome sequence lengths of *CsLecRLK*s ranged from 1803bp to 6481bp, the lengths of CDS ranged from 1674 bp to 2502 bp. The number of exon of these genes varied from one to nine, 80% *CsLecRLK*s had less than three exons, excepted that *Cs4G296230* contains three exons. All L-type *CsLecRLK*s contained only one or two exons, and the C-type *CsLecRLK* (*Csa1G056960*) contains four exons. The G-type *CsLecRLK*s contain one to nine exons. Among them, the *Csa7G446780* contains nine exons, which has the most exons.

Protein domain and Motif analysis of CsLecRLKs

Through the SMART program prediction, we investigated conserved domains that present in CsLecRLKs. C-type and L-type CsLecRLKs were both only contain three based categories domain, Lectin domain, Transmembrane domain and Kinase domain. But some G-type CsLecRLKs also contained other two categories domains, PAN domain and EGF domain. Among G-type CsLecRLKs, ten proteins contain PAN and EGF domains at the same

time, five proteins only contain PAN domain, eight proteins only contain EGF domain, and only one contains neither PAN domain nor EGF domain. Our result indicated that signal peptide would be not necessary to CsLecRLKs. There are 25 CsLecRLKs without signal peptide and 8 CsLecRLKs with more than two transmembrane domains.

Ten conserved motifs were identified in CsLecRLKs using the MEME program. These motifs were labelled Motif 1 to Motif 10 from the N- to the C-terminus. The details of the conserved motifs were shown in Figure 3. The lengths of these motifs ranged from 15 to 60 residues. Generally, the CsLecRLKs contains 4 to 10 motifs. None of the motifs appeared in all gene family members. Excepted that Motif 8 and Motif 9 were only present in the G-type CsLecRLKs, other motifs were present in three type CsLecRLKs. With the CDD program, we found that the six of these motifs represent different kinase domains (Supplementary table), indicating that there may be multiple phosphorylation catalytic sites in each of CsLecRLKs.

Chromosomal Location and Gene Duplication of *CsLecRLKs*

We extracted the location data of *CsLecRLKs* and the length data of each chromosome from the cucumber genome annotation files by a series of Perl scripts, and constructed gene location map using MapChart software. As shown in Figure 4, all *CsLecRLK*s were unevenly distributed across 7 cucumber chromosomes, and genes from the same subfamily on the same chromosome had a tendency to cluster. The number of *CsLecRLK*s on each chromosome varied from 1 to 12, chromosome 3 contains the largest number of 12 *CsLecRLK*s and chromosome 2 had only one *CsLecRLK*.

During the biological evolution, the generation of gene family could be caused by tandem duplication and segmental duplication (Kent et al., 2003; Mehan et al., 2004). In order to explore whether *CsLecRLK* gene family also have an expansion caused by the two kinds of duplication, we analyzed the duplication events of *CsLecRLK* genes. The result indicated that although many genes were clustered on the chromosomes, only *Csa1G071170* and *Csa1G071160* were a pair of tandem duplicated genes, their divergence time was about 38.606 million of years ago (MYA). The other two pairs of duplicated events, *Csa1G073890* and *Csa7G048050*, and *Csa3G734030* and *Csa4G296230*, may be caused by duplication or ectopia of chromosome fragments during the evolution. These duplicated genes are not in the same chromosome. Their divergence times were 30.96 and 32.35 MYA, respectively. Based on the above results, it could be inferred that tandem duplication contributed to the expansion of *CsLecRLK* gene family.

Cis-acting Elements Analysis on *CsLecRLK*s promoter

Different genes have their own specific or consensus *cis*-acting elements on their promoters. *Trans*-acting factors bind to the *cis*-acting elements to regulate the gene expression. Different *cis*-acting elements may correspond to different biotic or abiotic stress signals which could induce or inhibit the genes expression. Therefore, the *cis*-acting elements analysis on *CsLecRLK*s promoter will help us to further understand these genes' function. We used Plantcare website to analyze the promoters of 1500bp upstream sequence from translation initiation site of *CsLecRLKs*, and found that there were 54 typical and functional *cis*-acting elements (Fig. 5), which could be divided into four types: light response, stress resistance, plant hormone and others. Among them, 24 *cis*-acting elements were related to light response, 11 were related to hormone included salicylic acid (SA), jasmonic acid (JA), ethylene, gibberellin and auxin, and 9 were abiotic stress elements. These results suggested that the *CsLecRLK* gene family may be mainly involved in the biological pathway of stress resistance in cucumber. There

were six developmentally related *cis*-acting elements, five of which were related to seed development, suggesting that this gene family may play a role in seed development. More details were shown in the supplementary table 1.

Expression Pattern Analysis of *CsLecRLK* genes

Little is known about the functions of *LecRLK*s in cucumber. As a first attempt to provide insights into their potential functions, we used RNA-seq data from 10 tissues of cucumber to investigate the expression of each Cs*LecRLK* gene. Most of *CsLecRLK*s were expressed at a low level, some (*Csa6G338050*, *Csa1G071160*, and *Csa3G115090*) were barely expressed in any tissue. The expression pattern of 10 tissues could be divided into two groups based on the expression level of *CsLecRLK*s (Fig. 6A). Group 1 included stamen, most of *CsLecRLK*s were barely expressed in it, just 10 genes had a constitutive expression pattern (FPKM >= 1 in all tissues, Tao et al., 2018). Group 2 included other 9 tissues, there were at least 19 genes had a constitutive expression pattern in each tissue. The expression pattern of all *CsLecRLK*s could be divided into 3 groups based on their expression level in each tissue (Fig. 6A). From group 1 to 3, the range and level of gene expression decreased successively. Group 1 contained 4 genes, which had a high expression level in each tissue with average FPKM of 31.00. There were 14 genes belonged to Group 2, they had an intermediate expression level in each tissue with average FPKM of 7.86. Group 3 included 28 genes expressing at low level in each tissue with average FPKM of 1.81. Excepted that C-type *CsLecRLK* (*Csa1G056960*) belonged to group 1, and that G-type *CsLecRLK*s had higher expression level than L-type.

Thirty-six *CsLecRLK*s were expressed in all tissues (FPKM > 0 in all tissues) (Tao et al., 2018) and 8 genes were constitutively expressed (FPKM >= 1 in all tissues). Then we focused on those genes with relatively high expression (FPKM >2 in all tissues) (Tao et al., 2018) and selected 5 tissues of cucumbers for cluster analysis (Fig. 6B), including 10mm floral bud, 10mm ovary, cotyledon, fruit spines and root. We found a total of 15 genes were expressed in all these tissues. Specially, two genes were expressed only in the roots (*Csa1G605730* and *Csa3G115060*), one gene (*Csa3G048440*) was expressed in ovary, and six genes just expressed in fruit spines (*Csa4G289620*, *Csa4G289630*, *Csa4G289640*, *Csa4G296230*, *Csa7G067410* and *Csa7G446780)*.

Expression analysis of *CsLecRLK* genes in response to different treatments

Gene expression is not only spatiotemporal specific but also can be induced or repressed by hormones and stress. Because most of LecRLKs are receptor proteins on the membrane, they usually can sense those stimuli at the first time and send signals to intracellular receptors. To uncover all the divergence information of *CsLecRLK*s under different environment for a short time, the expression patterns under different hormone treatments, including IAA, GA, ABA, and NAA, and cold stress treatments, were analyzed by qRT-PCR. The result showed that most of *CsLecRLK*s (31/46) responded to at least one treatment (Fold change > 1 than the control group Significance p=0.05). Overall, there were 20 upregulated events and 38 downregulated events totally (Significance p=0.05). In order to show the experimental results more conveniently and intuitively, the change fold under different treatment was displayed in heatmap (Fig. 7) based on the data of qRT-PCR. Firstly, some *CsLecRLK*s (7/46) could be induced or repressed by multiple treatments (treatment number > 3), for instance, *Csa1G071170* could be induced by GA, IAA, NAA and ABA treatments, *Csa4G005510* was repressed by all treatments except ABA. Secondly, different *CsLecRLK*s could be induced or repressed by different treatments. The cold stress induced or repressed the minimal *CsLecRLK*s gene expression, there were 4 genes expression had changed, they were downregulated. On the contrary, NAA induced or repressed the most *CsLecRLK*s gene expression, there were 20 genes that responded to NAA treatment, 6 genes were upregulated and 14 genes were

downregulated. The 16 *CsLecRLK*s changed their expression level under ABA treatment, there were 8 genes upregulated and 8 genes downregulated. IAA and GA caused expression level change in 9 and 8 genes respectively. The IAA treatment caused 1 genes expression upregulated, and 8 genes downregulated. The GA treatment caused 5 gene expression upregulated and 3 downregulated. Thirdly, 14 *CsLecRLK*s had different expression pattern under various treatments, for example, *Csa3G734030* could be induced by NAA, and repressed by ABA, *Csa1G071150* was upregulated under ABA treatment, and downregulated under NAA and IAA treatments. The results indicated that the members of *CsLecRLK*s had their own response characteristics to hormones and stresses and may play an important role in sensing external stimulus signals. For example, although *Csa1G071160* and *Csa3G115090* were not expressed in the root, our experiment showed that they can be induced by NAA and ABA, respectively (Fig. 7B). There were 15 genes did not have significance expression change under different treatment, they were *Csa1G056960*, *Csa7G029930*, *Csa5G550210*, *Csa4G296250*, *Csa7G048050*, *Csa1G073890*, *Csa4G289620*, *Csa3G115060*, *Csa3G099580*, *Csa1G071270*, *Csa6G516770*, *Csa2G439150*, *Csa1G605730*, *Csa6G338050* and *Csa5G648630*.

Discussion

Compared with other plants, cucumber contains fewer members of *LecRLK* gene family, there are only 46 *CsLecRLK*s in cucumber genome, among which 23 G-type, 22 L-type and only one C-type. The number of *LecRLK* family members varies from different plants, this may be caused by the following three reasons. Firstly, the genome sizes of various plants are different, for example, the genome size of soybean is above 1 GB, which contains 52,051 protein-coding genes (Shen et al., 2018), so soybean has more members of *LecRLK* than other plants (soybean contains 189 *LecRLK* genes) (Liu et al., 2018). Secondly, different numbers of *LecRLKs* are also related to the function of *LecRLKs* in different plants' life activities, for instance, G-type LecRLKs are generally considered as a class of proteins that may be involved in self-incompatibility because some G-type LecRLKs contain a S-domain (Bouwmeester and Govers, 2009), which is an essential domain for sporophytic self-incompatibility response-related proteins. The cucumber is a kind of unisexual flower plant that does not have the problem of identifying the source of pollen during reproduction, unlike *Populus* and *Eucalyptus*, which are obligate outcrossing plant, this may be partially explain that why *Populus* and *Eucalyptus* contain more G-types *LecRLK*s, leading to a big increase in the number of members of the entire gene family (Yang et al., 2016). Thirdly, there are fewer duplicated events in *CsLecRLK* gene family. For instance, we only identified three pairs of duplicate genes in cucumber, but there are 36 paralogous gene pairs were generated with duplicate events in soybean (Liu et al., 2018), this lead to a wider range of expansion of the gene family in other plants.

Another interesting finding in our study was that most *CsLecRLK* genes have less introns, it is similar to previous studies in other plants. For example, most members of *LecRLK* gene family in soybean only have one or even no intron (Liu et al., 2018). Previous study in *Arabidopsis* and rice also indicated that there are few genes with introns in this gene family, containing only five and eight in *Arabidopsis* (contains 75 *LecRLK* genes) and rice (contains 173 *LecRLK* genes) respectively (Vaid et al., 2012). The reason for this kind of structure is probably that because the LecRLKs act as the signal receptors in plant, less introns means less selective cutting and splicing, this could save more transcription time, so that the signal could be transmitted at first time (Jeffares et al., 2008).

The *LecRLK* is a special family in plant genome, to date, no homologs of *LecRLK*s have been reported in the genomes of fungus and human (Yang et al., 2016). As the singal station of the immovable creatures, unique characteristic of *LecRLK* family may be closely related to their function on sensing the external environment.

Some *LecRLK*s have been reported that involved in sensing invasion of microorganisms. The *SD1-29* is one kind of G-type *LecRLK* in *Arabidopsis*, it can identify lipopolysaccharide which is a secretion from Gramnegative *Pseudomonas* and *Xanthomona*s (Ranf et al., 2015). Some *LecRLK*s will change their expression pattern with the changes of hormone and nutritional conditions. For example, the *SIT1*, a L-type *LecRLK* in rice, can mediate salt sensitivity. With the increase of NaCl concentration, the *SIT1* was activated rapidly, which reduced the survival of rice (Li et al., 2014). In addition, some *LecRLK*s affect the development and growth of plant, for instance, two L-type *LecRLK*s, *LecRK-IX.1* and *LecRK-IX.2* will induce cell death, thereby increasing plant survival, when infected with phytophthora (Wang et al., 2015). A C-type *LecRLK* mutant changed cell stacking pattern of trichome in cucumber, resulting fruit spines being easy to fall off from pericarp (Guo et al., 2017). It is similar to previous study (Yang et al., 2016), drastic number of *CsLecRLK*s may be expression in root (Fig. 6), this may be because root is one of the most important organ of plant, its role in fixation, temperature sensing, and nutrient absorption is irreplaceable. We also found a number of *CsLecRLK*s whose expression levels were induced or repressed through hormone and stress treatment, implying that the expression level of these *CsLecRLK*s may be affected by biotic or abiotic stimuli not contained in our tested tissues. Of course, there are also some *CsLecRLK*s whose expression levels were not only low in tested tissues, but also not induced or repressed in our hormones and stress treatments, this may be because the stimulation conditions that relate to these genes were not found in our experiment, such as salt stress, insect stress, salicylic acid, ethylene and so on, these stimulation condition represents an exploratory area for further investigation.

Declaration

The experiments comply with the current laws of the country in which we were performed.

Availability of data and materials

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

Competing interests

The authors declare that they have no conflict of interest.

Funding

This study was supported by The National Natural Science Foundation of China (no. 31672148) and the Shanghai Science and Technology Committee program (no. 18Z111220019).

Authors' contributions

DL contributed to bioinformatics analysis and writing of the manuscript. LRX helped with the RNA extraction and qRT-PCR. YC contributed to data of RNA-seq. JXS and CLG helped with growing plants and experiment of hormones. GW and JSP provided critical insights and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Ling Dai (School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China) for helping with bioinformatics analysis, and Wencan Zhang for revising the manuscript (School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University, China).

References

Bailey T L, Nadya W, Chris M, et al. MEME: discovering and analyzing DNA and protein sequence motifs[J]. Nucleic Acids Research (suppl_2):suppl_2.

Bouwmeester K, Govers F. Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles.[J]. Journal of Experimental Botany, 2009, 60(15):4383-4396.

Cambi A, Koopman M, Figdor C G. How C-type lectins detect pathogens[J]. Cellular microbiology, 2005, 7(4):481-488.

Cheng X, Wu Y, Guo J, et al. A rice lectin receptor-like kinase that is involved in innate immune responses also contributes to seed germination[J]. The Plant Journal, 2013, 76(4):687-698.

Christine Hervé, Jérome Serres, Dabos P, et al. Characterization of the Arabidopsis LecRK-a genes: members of a superfamily encoding putative receptors with an extracellular domain homologous to legume lectins[J]. Plant Molecular Biology, 1999, 39(4):671-682.

Deng K, Wang Q, Zeng J, et al. A Lectin Receptor Kinase Positively Regulates ABA Response during Seed Germination and Is Involved in Salt and Osmotic Stress Response[J]. Journal of Plant Biology, 2009, 52(6):493-500.

Desclos Theveniau M, Arnaud D, Huang T Y, et al. The Arabidopsis Lectin Receptor Kinase LecRK-V.5 Represses Stomatal Immunity Induced by Pseudomonas syringae pv. tomato DC3000[J]. PLoS Pathogens, 2012, 8(2):e1002513.

Edgar R C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput[J]. Nucleic Acids Research, 2004, 32(5):1792-1797.

Finn R D, Coggill P, Eberhardt R Y, et al. The Pfam protein families database: towards a more sustainable future[J]. Nucleic Acids Research, 2015:gkv1344.

Gu Z, Cavalcanti A, Chen F C, et al. Extent of Gene Duplication in the Genomes of Drosophila, Nematode, and Yeast[J]. Molecular Biology and Evolution, 2002, 19(3):256-262.

Guo C, Yang X, Wang Y, et al. Identification and mapping of ts (tender spines), a gene involved in soft spine development in Cucumis sativus[J]. Theoretical and Applied Genetics, 2017.

He X J, Zhang Z G, Yan D Q, et al. A salt-responsive receptor-like kinase gene regulated by the ethylene signaling pathway encodes a plasma membrane serine/threonine kinase[J]. Theoretical and Applied Genetics, 2004, 109(2):377-383.

Hofberger J A , Nsibo D L , Govers F , et al. A Complex Interplay of Tandem and Whole Genome Duplication Drives Expansion of the L-Type Lectin Receptor Kinase Gene Family in the Brassicaceae[J]. Genome Biology and Evolution, 2015, 7(3):720-734.

Hu B, Jin J, Guo A Y, et al. GSDS 2.0: An upgraded gene feature visualization server[J]. Bioinformatics, 2014, 31(8):1296.

Jeffares D C, Penkett C J, Jürg Bähler. Rapidly regulated genes are intron poor[J]. Trends in Genetics, 2008, 24(8):375-378.

Johnson L S, Eddy S R, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure[J]. Bmc Bioinformatics, 2010, 11(1):431-0.

Kent W J, Baertsch R, Hinrichs A, et al. Evolution\"s cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes[J]. Proceedings of the National Academy of Sciences, 2003, 100(20):11484-11489.

Letunic, I. SMART. Available online: http://smart.embl-heidelberg.de/

Li C H, Wang G, Zhao J L, et al. The Receptor-Like Kinase SIT1 Mediates Salt Sensitivity by Activating MAPK3/6 and Regulating Ethylene Homeostasis in Rice[J]. The Plant Cell, 2014, 26(6):2538-2553.

Liu P L , Huang Y , Shi P H , et al. Duplication and diversification of lectin receptor-like kinases (LecRLK) genes in soybean[J]. Scientific Reports, 2018, 8(1):5861.

Mehan M R, Freimer N B, Ophoff R A. A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture[J]. Human Genomics, 2004, 1(5):335–-344.

Ohtake Y. Salicylic Acid Induces the Expression of a Number of Receptor-Like Kinase Genes in Arabidopsis thaliana[J]. Plant and Cell Physiology, 2000, 41(9):1038-1044.

Olds T, Mitchell M T, Koch M A, et al. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae)[J]. Molecular Biology & Evolution, 2000, 17(10):1483.

Ranf S, Gisch N, SchaFfer M, et al. A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in Arabidopsis thaliana[J]. Nature Immunology, 2015, 16(4):426-433.

Riou C, Christine Hervé, Valérie Pacquit, et al. Expression of an Arabidopsis lectin kinase receptor gene, lecRK-a1, is induced during senescence, wounding and in response to oligogalacturonic acids[J]. Plant Physiology and Biochemistry (Paris), 2002, 40(5):431-438.

Shen Y, Liu J, Geng H, et al. De novo assembly of a Chinese soybean genome[J]. Science China (Life Sciences), 2018, v.61(08):3-16.

Shiu S H, Bleecker A B. Plant receptor-like kinase gene family: diversity, function, and signaling[J]. Sci. STKE. 2001, 2001(113):re22.

Shumayla , Shailesh S , Pandey A K , et al. Molecular Characterization and Global Expression Analysis of Lectin Receptor Kinases in Bread Wheat (Triticum aestivum)[J]. PLOS ONE, 2016, 11(4):e0153925.

Sushma N, Thanat C, Ripoll D R, et al. Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(29):12211-12216.

Timothy Bailey C E. Fitting a Mixture Model By Expectation Maximization to Discover Motifs in Biopolymers[C]// 1994.

Vaid N, Pandey P K, Tuteja N. Genome- wide analysis of lectin receptor-like kinase family from Arabidopsis and rice[J]. Plant Molecular Biology, 2012, 80(4-5): 365-388.

Voorrips R E. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs[J]. Journal of Heredity, 2002, 93(1):77-78.

Wan J, Patel A, Mathieu M, et al. A lectin receptor-like kinase is required for pollen development in Arabidopsis[J]. Plant Molecular Biology, 2008, 67(5):469-482.

Wang Y , Cordewener J H G , America A H P , et al. Arabidopsis Lectin Receptor Kinases LecRK-IX.1 and LecRK-IX.2 Are Functional Analogs in Regulating Phytophthora Resistance and Plant Cell Death[J]. Molecular Plant-Microbe Interactions, 2015, 28(9):1032-1048.

Wang Y, Tang H, Debarry J D, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity[J]. Nucleic Acids Research, 2012, 40(7):e49-e49.

Wu T , Wang R , Xu X , et al. Cucumis sativus L-type lectin receptor kinase (CsLecRK) gene family response to Phytophthora melonis, Phytophthora capsici and water immersion in disease resistant and susceptible cucumber cultivars[J]. Gene, 2014, 549(2):214-222.

Yang S, Zhang X, Yue J X, et al. Recent duplications dominate NBS-encoding gene expansion in two woody species[J]. Molecular Genetics and Genomics, 2008, 280(3):187-198.

Yang Y, Labbé, Jessy, Muchero W, et al. Genome-wide analysis of lectin receptor-like kinases in Populus[J]. BMC Genomics, 2016, 17(1):699.

Yang Y, Labbé, Jessy, Muchero W, et al. Genome-wide analysis of lectin receptor-like kinases in Populus[J]. BMC Genomics, 2016, 17(1):699.

Zhang, Zhang, Xiao Q, et al. KaKs_Calculator: Calculating Ka and Ks through Model Selection and Model Averaging[J]. Genomics Proteomics Bioinformatics, 2006(4):259-263.

Zuo K, Zhao J, Wang J, et al. Molecular Cloning and Characterization of GhlecRK, a Novel Kinase Gene with Lectin-like Domain from Gossypium hirsutum[J]. DNA Sequence-The Journal of Sequencing and Mapping, 2004, 15(1):58-65.

Table

Table 1. Basic character of CsLecRLK family

Group	Gene ID	Chr	Start	End	Strand	Gene (bp)	CDS (bp)	Protein (aa)	MW (KDa)	pI	Localization
C-type	Csa1G056960	1	6245947	6249211	+	3,265	1,674	557	62.5	9.51	PM
L-type	Csa1G004050	1	615237	617749	+	2,513	1,977	658	73.8	5.23	PM
	Csa1G073890	1	7608750	7610584	+	1,835	1,794	597	66.5	6.14	CY
	Csa2G439150	2	22949410	22951362	+	1,953	1,953	650	72.6	6.84	PM
	Csa2G439210	2	22985725	22987680	+	1956	1956	651	72.2	6.73	PM
	Csa3G048440	3	3434259	3436274	-	2,016	2,016	671	73.8	7.39	PM
	Csa3G115060	3	6101289	6103476	-	2,188	2,058	685	76.5	6.35	PM
	Csa3G115090	3	6119460	6122293	-	2,834	1,908	635	71.9	6.66	PM
	Csa3G730910	3	27403916	27405718	+	1,803	1,803	600	66.4	5.25	PM
	Csa3G730920	3	27408398	27410741	+	2,344	2,044	667	74.2	7.04	PM
	Csa3G734030	3	27978179	27980785	+	2,607	2,052	683	75.1	7.41	PM
	Csa3G734040	3	27981843	27984009	+	2,167	1,986	661	73.3	5.62	PM
	Csa3G734050	3	27985087	27987243	+	2,157	2,121	706	78.4	8.13	PM
	Csa3G736960	3	28496446	28499101	-	2,656	2,184	727	78.5	5.62	PM
	Csa4G296230	4	12033379	12039859	-	6,481	2,004	667	75.3	7.41	PM
	Csa4G296250	4	12055822	12057932	+	2,111	1,980	659	73.5	6.71	PM
	Csa5G648630	5	27485882	27487867	+	1,986	1,986	661	73.3	6.68	PM
	Csa6G338050	6	15414546	15416705	+	2,160	2,088	695	76.9	6.45	PM
	Csa7G029930	7	1577321	1580540	+	3,220	2,076	691	78.1	6.55	CY
	Csa7G048050	7	2990460	2992838	+	2,379	2,058	685	76.2	5.99	PM
	Csa7G067400	7	4115966	4118216	-	2,251	2,133	710	79.5	6.82	PM
	Csa7G067410	7	4119365	4121401	-	2,037	2,037	678	75.4	6.04	PM
	Csa7G067430	7	4124769	4127087	+	2,319	2,124	707	79.7	6.3	PM
G-type	Csa1G071150	1	7119234	7121709	+	2,476	1,773	590	66.4	5.18	PM
	Csa1G071160	1	7124699	7127292	-	2,594	1,818	605	68.1	7.46	PM
	Csa1G071170	1	7134918	7137465	-	2,548	1,746	581	65.6	7.26	PM
	Csa1G071270	1	7179289	7181813	+	2,525	1,935	644	73.1	6.98	PM
	Csa1G605730	1	23772913	23775291	-	2,379	2,379	792	89.1	6.24	PM
	Csa1G605740	1	23777925	23780300	-	2,376	2,376	791	89.4	5.46	PM
	Csa1G605750	1	23782466	23785387	-	2,922	2,439	812	91.1	5.28	PM
	Csa3G099580	3	4888747	4892235	-	3,489	2,238	745	84.7	8.97	PM
	Csa3G733860	3	27894955	27899854	-	4,900	2,502	833	94.5	6.32	PM
	Csa3G733880	3	27905835	27908289	-	2,455	2,421	806	90.5	6.22	PM
	Csa4G005510	4	865713	869014	-	3,302	2,460	819	92.1	6.81	PM
	Csa4G288620	4	11220419	11222863	-	2445	2445	736	82.6	5.76	PM
	Csa4G289620	4	11230649	11233131	-	2,483	2,358	785	88.6	7.69	PM
	Csa4G289630	4	11233952	11236384	-	2,433	2,184	727	81.9	4.98	PM
	Csa4G289640	4	11237240	11239680	-	2,441	2,424	807	90.7	5.84	PM
	Csa4G289650	4	11240557	11242883	-	2,327	2,136	711	79.4	5.64	PM
	Csa4G290150	4	11244840	11247660	-	2,821	2,094	697	77.9	7.25	PM
	Csa5G550210	5	19486884	19489396	+	2,513	1,887	628	70.3	8.89	PM
	Csa5G550230	5	19498199	19500774	-	2,576	2,478	825	92.3	6.66	PM
	Csa6G052130	6	3992021	3994920	+	2,900	2,196	731	79.6	5.48	Ec
	Csa6G516770	6	26894261	26897704	-	3,444	2,142	713	79.8	5.91	PM
	Csa7G045520	7	2678627	2681841	-	3,215	2,496	831	93.9	5.37	Ec
	Csa7G446780	7	17724360	17728540	-	4,181	2,421	806	91.9	5.53	PM
Averger						2,674	2,117	703	79	7	

Abbreviations: *Cy* cytosol, *Ec* extracellular, *PM* plasma membrane, *CDS* coding sequence
Localization prediction by CELLO v.2.5 (_http://cello.life.nctu.edu.tw/_)

Figures

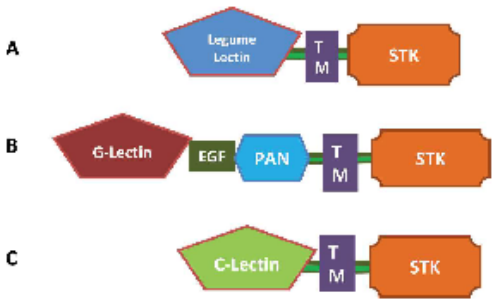

Figure 1

1 Line Model of the Three LecRLK Classes (Vaid et al. 2013). On the basis of extracellular domain, the LecRLKs are classified into three classes: L-type, G-type, and C-type. (A) L-type LecRLK with legume lectin-like extracellular domain, (B) G-type LecRLK with α-mannose binding bulb lectin domain. This class of proteins also possesses a PAN and/or an EGF domain, (C) C-type LecRLKs characterized by the presence of calcium-dependent carbohydrate-binding domains. TM: transmembrane region, STK: a cytoplasmic Serine/Threonine kinase domain.

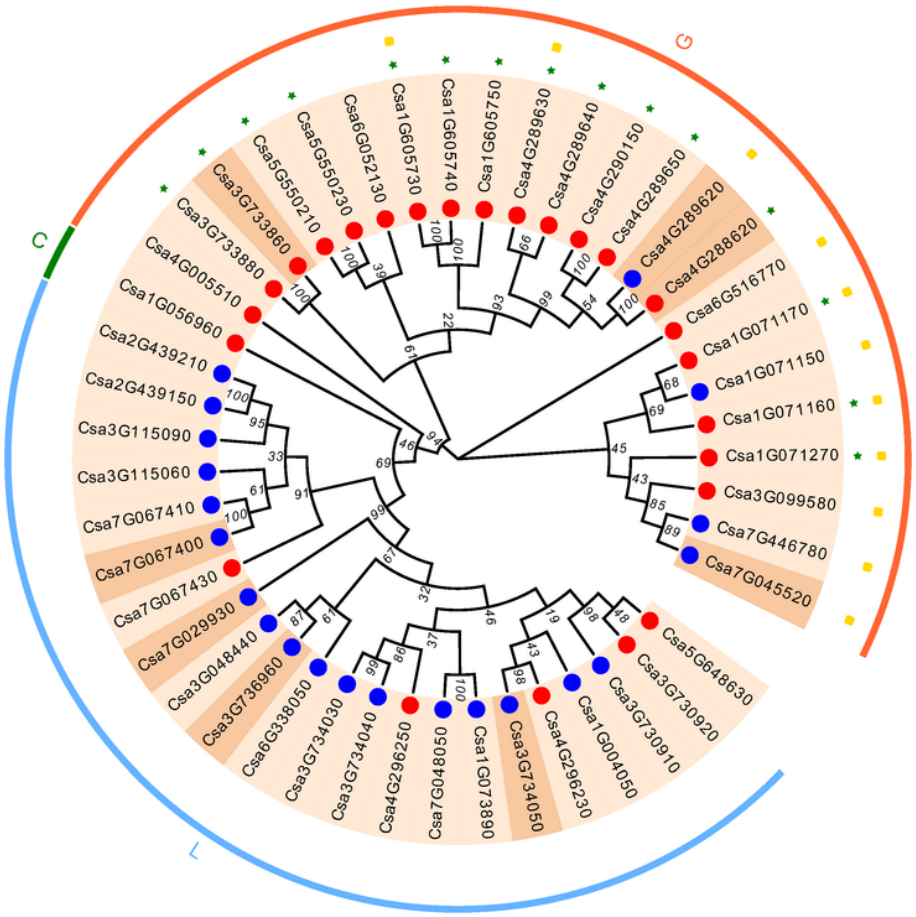

Figure 2

Unrooted Phylogenetic tree construction and structure analysis of CsLecRLKs The L, G and C respectively represented different L-type, G-type and C-type LecRLK subfamilies. The blue circles represented proteins with a signal peptide, the red circles proteins without a signal peptide. The green stars represented proteins contained EGF domains, the yellow squares represented proteins contained PAN domains. The sandybrown leaves represented proteins only contained one transmembrane domain, the peachpuff leaves represented proteins containing more than one transmembrane domain.

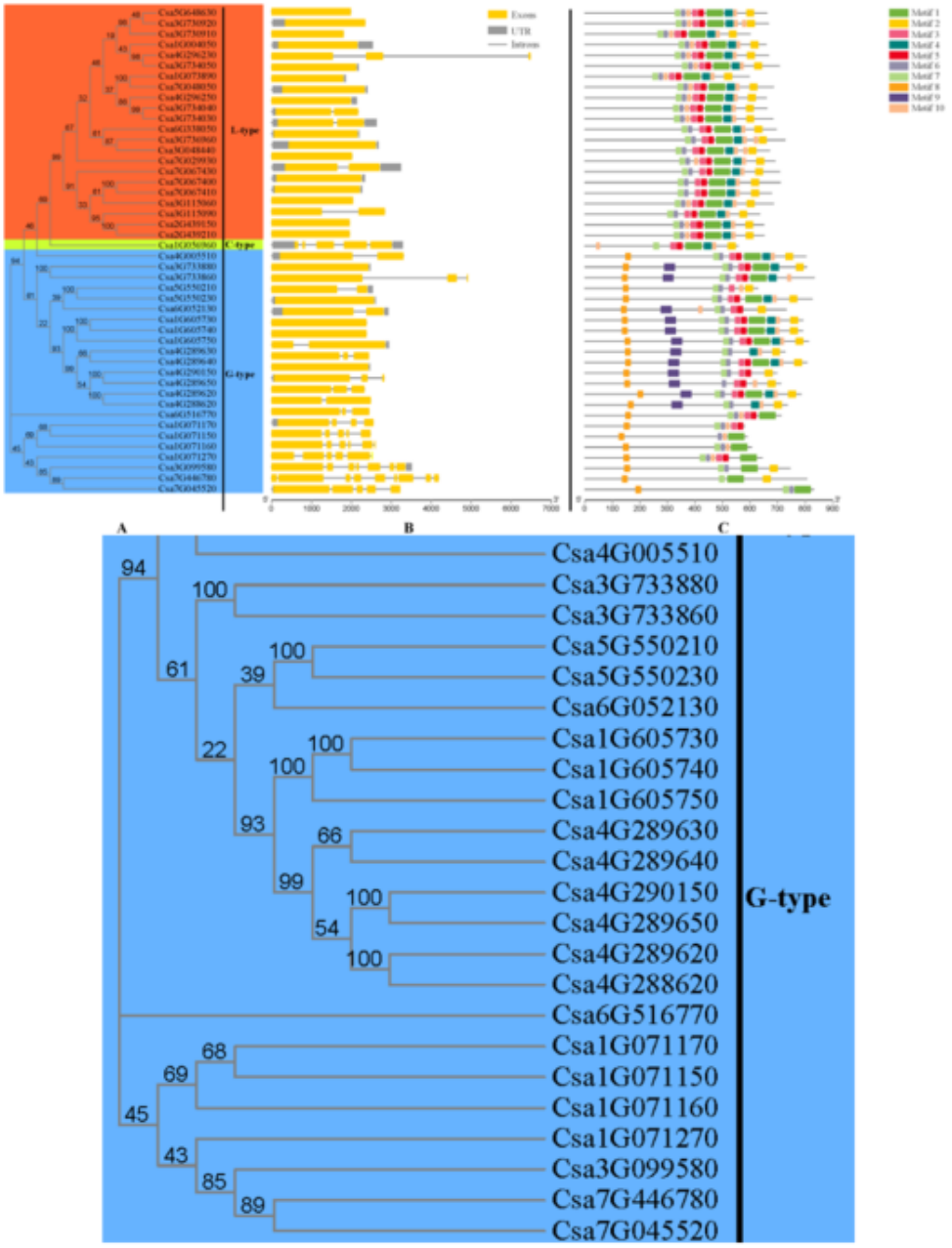

Figure 3

An analytical view of the CsLecRLK genes The following parts are shown from left to right. (A) Protein maximum likelihood tree: The tree was constructed using a maximum-likelihood method, and bootstrap values were calculated with 1,000 replications using MEGA 7.0.21. (B) Gene structure: The lines represented the intron. The grap square represented the 3’ UTR and the 5’ UTR. The yellow square represented the exon. (C) Protein structure: The search for 10 common motifs shared among the CsLecRLK proteins, different color represented different motifs.

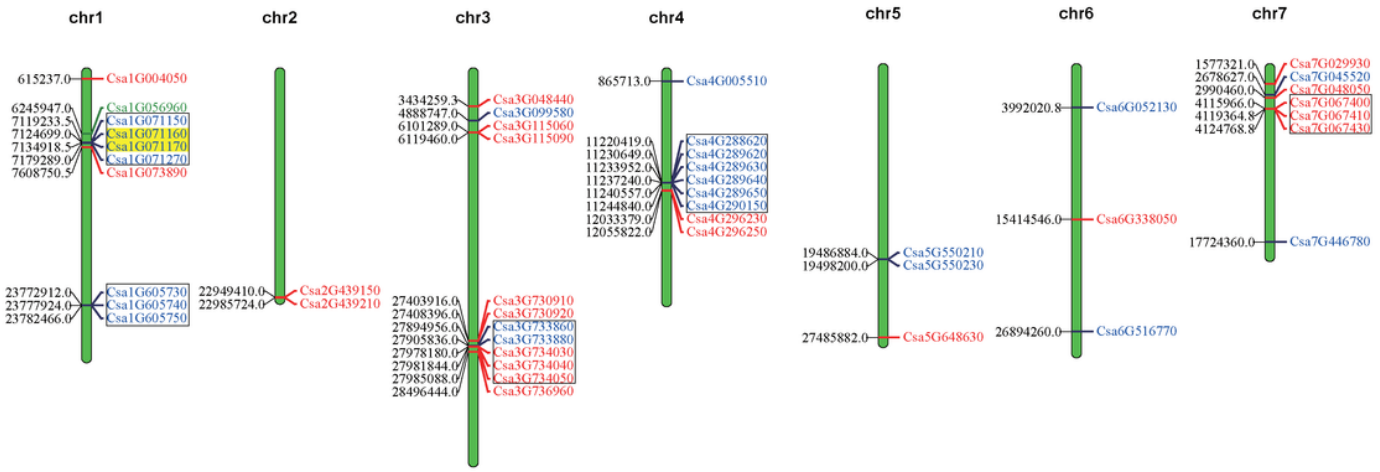

Figure 4

Chromosomal distribution and gene duplications of CsLecRLK genes The number at the chromosome (left) represented the position of CsLecRLK genes, and the right information on the chromosome (right) represented the gene ID of CsLecRLKs. The tandem duplicated genes are represented by yellow background, and the gene clusters are boxed together by black lines. The red represented L-type CsLecRLKs, the yellow represented C-type CsLecRLKs, the blue represented G-type CsLecRLKs.

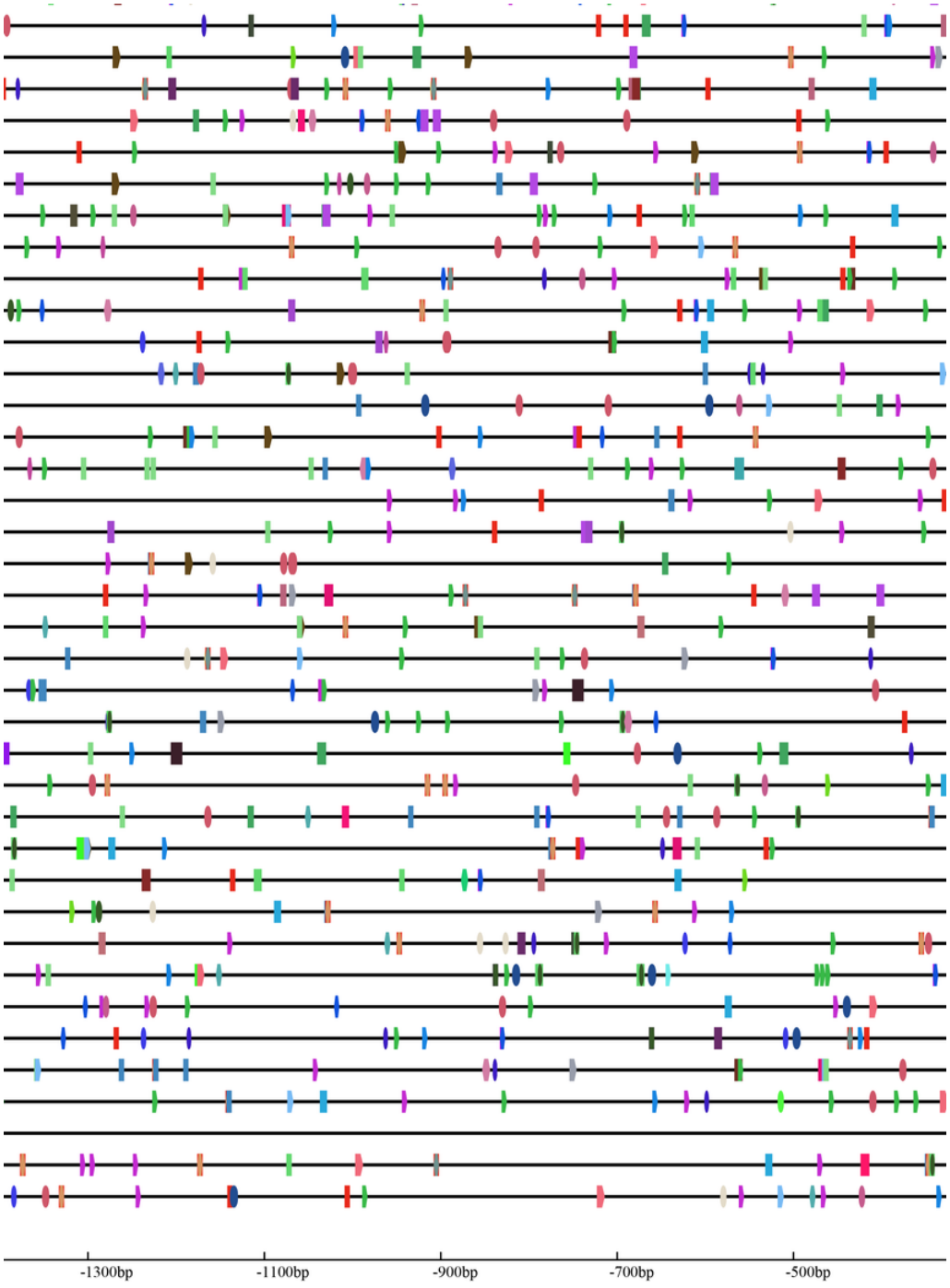

Figure 5

Promoter Analysis of CsLecRLKs Different color and shape represent different cis-acting elements.

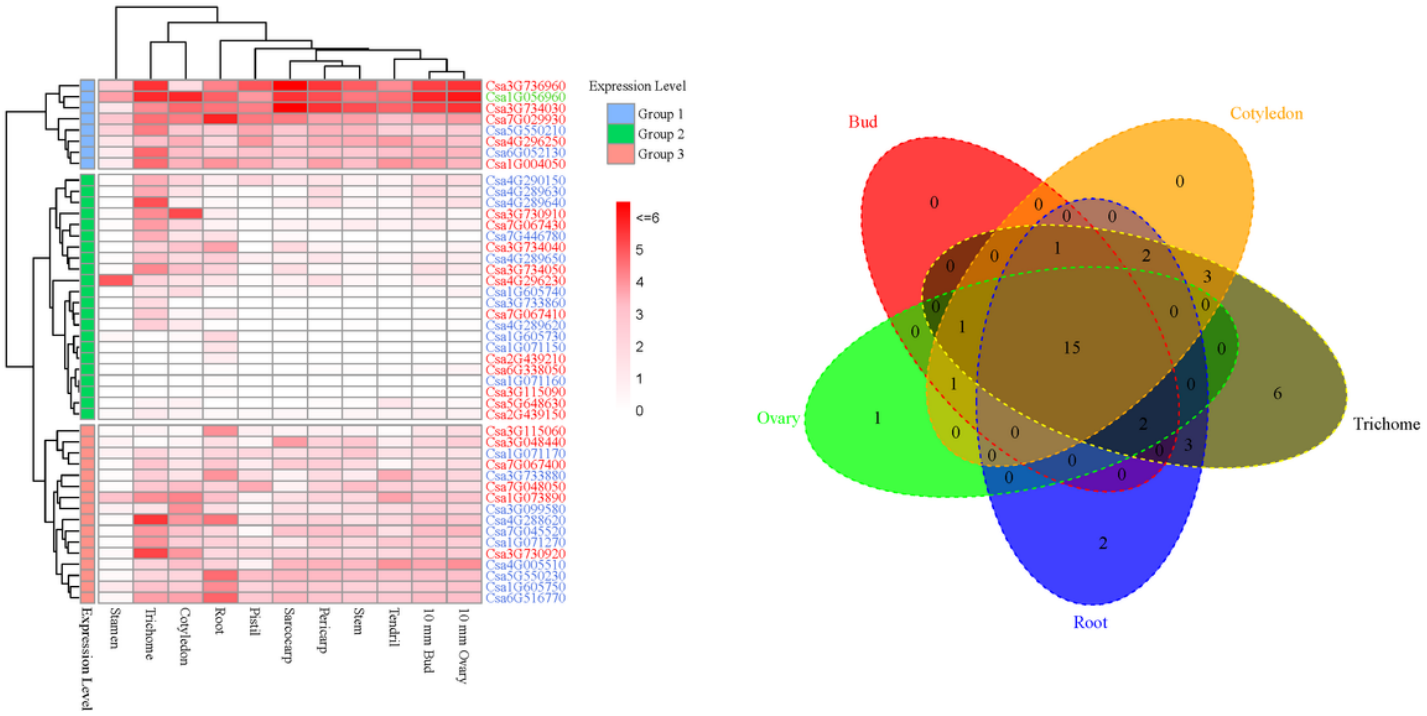

Figure 6

Expression profile of CsLecRLKs gene family in different developmental conditions (A) Heat map depicting the expression profile of CsLecRLK family in different developmental conditions. The red represents L-type CsLecRLKs, the black represents C-type CsLecRLKs, the blue represents G-type CsLecRLKs; (B) Venn diagram depicting the distribution of shared expression of CsLecRLKs under 5 typical tissues, 10mm floral bud, 10mm ovary, cotyledon, fruit spines and root

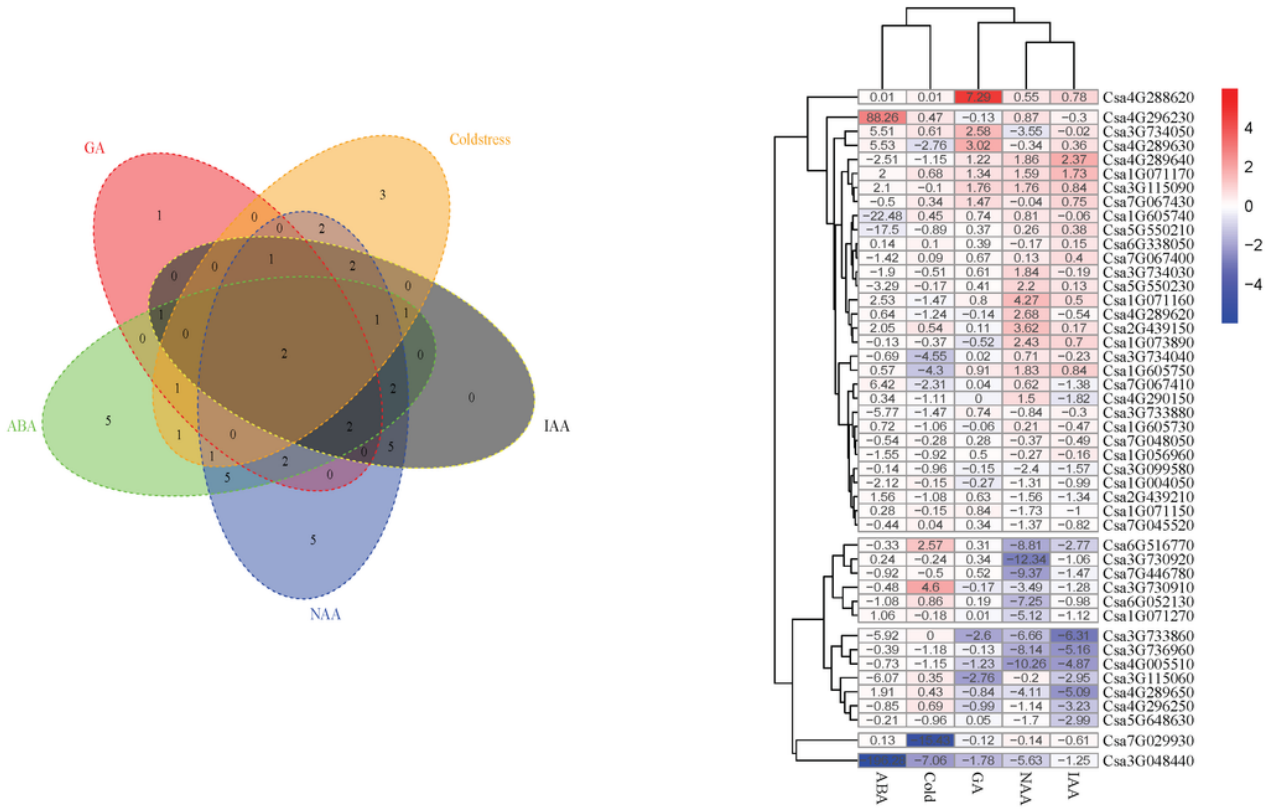

Figure 7

Expression analysis of CsLecRLK genes in response to different treatments The heatmap represented the RNA expression quantity of CsLecRLKs under 5 treatments, GA, ABA, IAA, NAA and cold stress.

Supplementary Files

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

- Gtypeprotein.pdf
- supplementarytable3PrimersforqRTPCR.pdf
- LtypeCDS.pdf
- supplementarytable1.xlsx
- CtypeCDS.pdf
- Ltypeprotein.pdf
- MotifStracture.pdf
- Ctypeprotein.pdf
- supplementarytable2.xlsx
- GtypeCDS.pdf