The Cucumber Kinome: genome-wide, identification, characterisation and expression analysis of the *Cucumis sativus* protein kinases in response to Powdery Mildew, Alternaria Leaf Spot and Root-Knot Nematode

Francisco Cleilson Lopes Costa[1], Alexandre Hild Aono[2], Edson Mario de Andrade Silva[3] and Welison Andrade Pereira[1]*

[1]Department of Biology, Federal University of Lavras (UFLA), Lavras, MG, Brazil
[2]Molecular Biology and Genetic Engineering Center (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil
[3]Department of Biochemistry and Molecular Biology, Federal University of Viçosa (UFV), Viçosa, MG, Brazil
*welison.pereira@ufla.br

Abstract

Protein kinases make up a family of enzymes that play a key role in regulating metabolism. In the cucumber species (*Cucumis sativus*), this family has not yet been fully characterized. Considering its importance in studies of the botanical family Cucurbitaceae, we identified and characterized the complete set of cucumber kinases, including phylogenetic and classification analyses, their genome positions, chemical characterization, and expression through RNA-seq data in response to powdery mildew and Alternaria leaf spot. The cucumber kinome is made up of 831 proteins, widely distributed across the species's seven chromosomes, with the majority located on chromosome three. The kinases were classified into 21 (9 +11 grupos named “Others”) groups and 125 subfamilies, with the RLK-Pelle group being the most abundant. There was no evidence of gene duplication and it has a smaller genome compared to other species in the botanical family Cucurbitaceae. Our study provides more information to support the understanding about the function PKs and shows how similar PK subfamilies are in composition, their properties and how they are expressed in disease related scenarios.
Keywords: Gene expression, Kinase gene family, Protein evolution, Response to disease.

Introduction

Cucumis sativus (Csa) is an annual crop that belongs to the Cucurbitaceae family and has medicinal effects, which leads to an increase in its demand and cultivation. This demand has emerged due to the cost-effectiveness and plant-based nature of these products, as opposed to commercially available synthetic drugs (Khan et al., 2022). Csa is rich in carbohydrates, proteins, minerals, and secondary metabolites like alkaloids, tannins, flavonoids, saponins, and phenolic compounds, which may be responsible for various therapeutic applications (Khan et al., 2022). Nevertheless, Csa susceptibility to diseases and pests is responsible for reducing their green parts and consequentely, reducing yield. An important pathogen that affects the Csa crop is *Podosphaera xanthii*, causing the powdery mildew. Another pathogen importatly affecting the foliar tissues of the Csa crop is *Alternaria cucumerina*, then reducing phtossynthetic active area and consequently the productivity. Resistant varieties is the most important alternative to control the disease. Understanding the molecular mechanisms of defense is important to develop resistant varieties. Protein kinases (PK) are the main regulators of the plant response to disease processes.

Kinome is the full repertoire of the expressed proteins with kinase activity. The cucumber kinome was previously studied and recognized as a small kinome in relation to other plant kinomes, probably because the Csa experienced no rounds of recent whole genome duplications (Huang et al., 2009a), but no genome-wide study about Csa kinome is available to our knowledge. The Csa kinome is a set of 831 protein kinases, that were retrieved from Csa proteome (CuGenDBv2: Gy14). The first plant kinome was defined in *Arabidopsis* in 2004 (Champion et al., 2004; Zulawski et al., 2014a), after

that many others plant kinomes have been investigated in several plant species, such as, in rice (Dardick et al., 2007), maize (Wei et al., 2014a), tomato (Singh et al., 2014a), soybean (J Liu et al., 2015), wheat (Yan et al., 2017a), *Brassica napus* (Gill et al., 2017), *Gossypium* (Yan et al., 2018a), grapevine (Zhu et al., 2018c), pineapple (Zhu et al., 2018a), strawberry (Liu et al., 2020a), pecan (Zhu et al., 2021a), sugarcane and sorgum (Aono et al., 2021a), cowpea (Ferreira-Neto et al., 2021a), common bean (Aono et al., 2023), rubber tree (Santos et al., 2023).

Some genome-wide analyses have been done in some sub-families of protein kinases (Wei et al., 2019) and dirigent gene family (Yadav et al., 2021) in the cucurbits species, there is no evidence of of comparative studies comprising all the protein kinase loci of Csa. Many genome-wide studies regarding protein kinase families in Csa have been done, such as that Mitogen-Activated Protein Kinase (MAPK) family (Shoresh et al., 2006), Casein Kinase (CK) family (Vidal et al., 2010), Calcium-Dependent Protein Kinase (CDPK) family (Xu et al., 2015); and in others plant species, such as, the CDPK family in maize (Kong et al., 2013), CDPK family in grape (Zhang et al., 2015), CDPK family and calcium-dependent protein kinase-related (CRK) family in tomato (Wang et al., 2016), Valine-glutamine (VQ) motif in melon (Zhang et al., 2017), CRK family in *Gossypium barbadense* (Li et al., 2018), CDPK family in watermelon (Wei et al., 2019), CDPK family in pineapple (Zhang et al., 2020), CDPK family in banana (Li et al., 2019), autophagy genes in Csa (Rui et al., 2020), CDPK and CRK families in *Medicago truncatula* (Zhao et al., 2021), CRK family in *Aeluropus littoralis* L. (Arab et al., 2021), CDPK family in apple (Dai et al., 2021), CDPK and closely related gene families in *Solanum pennellii* (Shi and Du, 2021), PYL gene family in Csa (Zhang et al., 2022). Beside PK families, others genome-wide studies have been done in Csa and other plant species, evidencing the need of this kind of study to Csa breeding: Cyclin-Dependent Kinase (CDK) inhibitor in avocado (Sabag et al., 2013), CDPK and related genes in *Capsicum annuum* (Cai et al., 2015), CDK in pumpkin (Herrera-Pola et al., 2016).

Genome-wide gene families evolutionary analyses give us insight into the conservation and evolution of a gene family important to lineage adaptation, resistance to environmental stress and disease. The protein kinases (PKs) are a family of enzymes that mediate the response of cells to external stimuli, highlighting their importance to the enhancement of Csa species. PKs are responsible for key processes in cell signaling networks, such as phosphorylation, and then they are well poised to contribute to the various cellular processes.There are many PKs which act specifically on diverse substrates and are themselves acted on by diverse regulators. The kinase cross-talk and feedback inhibition mechanisms need to be studied in Csa, as they are checkpoints in virtually every kinase signaling network. In this way, any changes or interruptions in these feedback circuits or in the kinase's crosstalk patterns have effects on the kinome as a whole, which gives support in the understanding of many mechanisms in plant adaptation to environmental stress, such as, response to biotic and abiotic factors. Despite being an important stress responsive gene family, little is known about the evolutionary history and expression patterns of PK genes in Csa.

Methods

Identification and computation prediction of conserved domains of CsaPKs

The protein Ser/Thr/Tyr kinases (PK) in Csa (*Cucumis sativus*) were identified following the methods described bellow. Pkinase (PF00069) and Pkinase_Tyr (PF07714) Hidden Markov Models (HMMs) of eukaryotic protein kinases were retrieved from Pfam database (http://pfam.xfam.org/) (El-Gebali et al., 2019) to perform a HMM search via terminal Ubuntu against the Csa proteome (Gy14_pep_v2.1.fa), with an e-value default of 0.01. Subsequentely, only the protein kinase sequence candidates possessing kinase domains were maintained after a check of kinase domain integrity with

SMART tool (http://smart.embl-heidelberg.de/) (Letunic et al., 2021; Letunic and Bork, 2018). This approach resulted in 831 candidates identified as Csa protein kinase (CsaPK) sequences.

Gene structure analysis and Chromosomal position

The conserved motifs in CsaPK sequences were identified using the MEME Suite 5.5.1 (https://meme-suite.org/meme/tools/meme), selecting 15 motifs to be found. Through NCBI-CDD tool, we identified the number of domains in the PKs candidate sequences. To establish the number of domains to be searched, CDD tool was used (https://www.ncbi.nlm.nih.gov/cdd/), with the searches against the Pfam v33.1 – 18271 PSSMs database (e-value 10e-6), using the CsaPK domains. To visualize and determine the structures of CsaPK genes, such as exon-intron content, Gene Structure Display Server (http://gsds.gao-lab.org/) was used. Number of exons and introns were calculated using the GFF file obtained from the CuGenDBv2. The chromosomal position of the CsaPK genes was predicted using the comand line tblastn software using the CsaPK sequences to align them against the Csa Gy14 genome v2.1 file, obtained from the Cucurbit Genome Database. A map was constructed using the online TBtools (Chen et al., 2020).

Characterization of the CsaPK sequences

Thereafter, 831 ‘‘candidate’’ protein kinases were predicted with transmembrane domains using the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM) and SignalP server 5.0 (https://services.healthtech.dtu.dk/service.php?SignalP-5.0), and the subcellular localization was predicted by Wolf PSORT (https://wolfpsort.hgc.jp/) and CELLO v.2.5 tool (http://cello.life.nctu.edu.tw/). The following attributes isoeletric point (pI), molecular weight (MW), the Grand Average of Hydropathy (GRAVY) of the PK sequences were analysed by the ExPASy ProtParam tool (https://web.expasy.org/protparam/), by using the ProtParam module from the subpackage SeqUtils from the Bio package from the Biopython language (https://biopython.org/docs/1.76/api/Bio.SeqUtils.ProtParam.html). The number of amino acids (aa) and functional annotation of the CsaPK genes were performed by using the Blast2GO tool (Conesa and Götz, 2008) and the GO terms retrieved were analysed using REViGO tool (http://revigo.irb.hr/) (Supek et al., 2011) and treemap was plot using the treemap R package (Tennekes, 2021).

Multiple sequence alignments and Phylogenetic analysis

A protein kinase sequence outgroup (Cre07.g349540.t1.1) from the *Chlamydomonas reinhardtii* v5.6 proteome was included as root in phylogenetic analysis, since this species had no multicelular ancestral (Ratcliff et al., 2013), indecating it evolved previously to the Csa evolution and fullfilled the prerequisite to be a root in the phylogenetic tree. Moreover, the protein sequences were aligned using the MAFFT v7.453 (Katoh and Standley, 2013), with 1000 bootstrap replications. The phylogenetic analysis of the CsaPKs were realized by FastTree 2.1.11 with default setting. Then, a newick file was generated for subsequent analysis. The CsaPK sequences were classifid by iTAK (http://itak.feilab.net/cgi-bin/itak/index.cgi) to confirm the phylogenetic classification results.

Kinase genes expression patterns

To analyse the functional characterization of the protein kinases, we selected and downloaded expression data from the cucumber Gy14 genotype available at the CuGenDBv2 relative to biotic stress conditions, analysed in leaves for powdery mildew and alternaria diseases, and in roots for Meloidogyne infestation. The expression levels were calculated in terms of FPKM (fragments per kilobase of transcript sequence per millions base pairs).

To the powdery mildew infection (*Podosphaera xanthii*, an obligate parasite), a study was realized by Xu et al. (2017) utilizing a segment substitution line (SSL508-28) containing a segment introgressed in the chromosome 5 named Pm5.1, confering resistance to the powdery mildew to the genotype Jin5-508 (resistant). A differentially expressed genes analysis was realized using the genotypes Jin5-508 and its recurrent parent D8 (susceptible). To identify the DEGs, the total RNA was extracted at 0 and 48 hours post-innoculation (HPI), and the genes were ranked according to size and sequence coverage was normalized in fragments per kilobase of exon per million. From this work, the total transcriptomes containing including the 831 kinases transcripts from the resistant and susceptible genotypes were retrieved. The public data of this work (PRJNA321023) is available at the CuGenDBv2. The results were tested statistically in terms of log2 fold change (log2fc) of gene FPKM between the contrasting genotypes by the Student's t-test. We found this study was validated and, in doing so, we filtered the kinase genes to assess only their expression. The kinase set was contrasted two times: R48 x R0, S48 x S0. The log2fc > 1.0 was thresolded and p-value was set to -log10 (0.05). The DEG data were visualised through a Volcano plot using the R packages tidyverse (Wickham et al., 2019), RColorBrewer (Neuwirth, 2014) and ggrepel (Slowikowski et al., 2018).

To the alternaria leaf spot infection (*Alternaria cucumerina*, a necrotroph pathogen), two cucumber homozyguous lines were used, D1322 (resistant to Alternaria leaf spot) and Beijin204 (BJ204, susceptible to Alternaria leaf spot), to a differentially expression gene analysis (Sa et al., 2020a). The FPKM was used to quantify the gene expression levels at 2 and 6 days post-inoculation (DPI). The public data of this work (PRJNA563186) is available at the CuGenDBv2. Alike the powdery mildew study, we found the Alternaria leaf spot gene differential expression analysis validated and filtered the kinase genes to assess their expression, that was made by contrasting the kinases two times: R2 x R6, S2 x S6. The log2fc was thresolded to > 1.0 and the p-value was set to -log10 (0.05). Volcano plots were constructed using the R packages tidyverse (Wickham et al. 2019), RColorBrewer (Neuwirth et al. 20114) and ggrepel (Slowikowski et al. 2018).

We also studied the root-knot nematode (*Meloidogyne incognita*) in for our kinase genes differentially expression analysis using two cucumber lines IL10–1 (resistant) and 'beijingjietou'(CC3, susceptible), and *M. incognita* race 1 (Wang et al., 2018). We found this study valid for the purposes of the study of kinase gene expression, and retrieved the public data (PRJNA419665) available at the CuGenDBv2, and filtered the kinase genes to assess their expression contrasting the resistant (R) and susceptible (S) lines after 0, 1, 2 and 3 DPI: R0 x R1, R0 x R2, R0 x R3, S0 x S1, S0 x S2, S0 x S3. The log2fc was thresolded to > 1.0 and the p-value was set to -log10 (0.05). Moreover, the Volcano plots were constructed using the R packages tidyverse (Wickham et al. 2019), RColorBrewer (Neuwirth et al. 20114) and ggrepel (Slowikowski et al. 2018).

Kinase gene expression was evaluated by a heatmap constructed using the pheatmap R package (Kolde and Kolde, 2015). For the construction of the heatmap, the average FPKM values were considered and the hierarchy was performed by the complete linkage method using Euclidean distances.

Cis-elements prediction

Cis-elements were predicted with 1500 bp genomic region upstream the ATG of CsaPK genes with Plant-CARE server (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot, 2002).

Results

Gene Structure and Chromosomal Location

The PK genes of the *C. sativus* are distributed in seven chromosomes and two genes (CsGy0G000160.2 and CsGy0G002040.1) were mapped to the linkage group zero (Chr0) (Figure 1 and Supplementary file S1). The PK genes are distributed across all the chromosomes, and tending to the

telomeric regions in the chromosomes 1 and 3. The chromosome 3 hosts the highest number (169) of PKs, and the least (82) are hosted in the chromosome 7. Strucuturally, the genes organization is very variable, with number of introns is ranging from 0 to 28. Most kinases (127 or 15.2%) has a single intron, while 123 kinases (14.8%) have none, 94 kinases (11.3%) have six introns, 49 kinases (5.8%) have two introns, and 47 kinases (5.6%) with five introns. The number of kinases ranges up to a single kinase (0.1%) with 28 introns (Supplementary file S2). The disposition of exon-intron in the sequences can be visualised in Supplementary file S3 and the functional domain structures in Supplementary file S4.

[Figure 1]

Genome-Wide Identification of Csa PKs

The Csa kinome is a collection of 831 protein kinases obtained from the Csa Gy14 genome. The full characterization of Csa kinase enzymatic properties, interactomes, phosphorylomes, biological functions depend on the construction of the Csa kinome. The Csa proteome was aligned against the eukaryotic kinase domains HMM profile, resulting in 831 candidates identified, after the exclusion sequences containing non-integral kinase domains (Supplementary file S5). The HMM profile of kinase domains was defined based on the Hidden Markov Models of eukaryotic kinases utilised to search for CsaPK domains. To get insight into the evolutionary process of the protein kinase family of Csa, a comprehensive phylogenetic analysis was carried out. This resulted in a phylogeny of the CsaPKs, comprising about 3.7% of 22,626 protein-coding genes based on the Csa genome annotation (v2.1 hosted in CuGenDBv2). This value is very related to the plant kinomes (Manning et al., 2002; Shiu et al., 2004). Indeed, dispite being a small proportion of the Csa genome, it highlights the importance of the kinase superfamily in the regulation of the plant developmental processes.

Csa Protein Kinases Properties

A complete analysis of the CsaPKs is ilustrated in Figure 2. The CsaPKs were characterized regarding the presence of functional domains by using the CDD/NCBI database, and were identified 116 different domains and accounting for 1898 domains distributed across the CsaPKs (Supplementary file S6). The top five most frequent domains belong to the PKc_like superfamily (642 domains), LRR_8 (324), Pkinase (156), LRRNT_2 (119) e LRR_8 superfamily (91) (Supplementary file S12). Most kinases (331) contained a single kinase domain, and increasing number of domains from 1 to 8, 12 ou 13 kinase domains per protein (Figure 2). The most abundant protein in domain number predicted is CsGy6G023150.2, belonging to the subfamily TKL-Pl-1, whose function is carried out in the cytoplasma as an RING-type E3 ubiquitin transferase. The structural motifs were predicted with the MeMe Suite and were found 15 motifs (Figure 3, Supplementary file S8). Most frequently found motifs are the motif 3 (present in 788 kinases), motif 10 (722 kinases), motif 6 (684 kinases), motif 4 (652 kinases) and motif 5 (652 kinases).

[Figure 2]

[Figure 3]

Another important protein property is the amino acid number. We predicted it by using ExPASy ProtParam online tool, that ranged from 57 to 1,694. The molecular weight (MW) from 6,419.35 to 191,012.84 kDa. The isoeletric points of the protein kinases ranged from 9.85 to 4.38 and the GRAVY, from -1.19 to 0.39, indicating protein activity in both hydrophilic and hydrophobic microenvironments (Figure 2, Supplementary file S9).

The presence of transmembrane helices was predicted with TMHMM tool. Transmembrane domains are membrane-spanning domains that perform a great number of funtions, such as, signal transduction, transport of proteins and ions into the cell. In our dataset, the protein kinases were predicted to have 0, 1, 2, 3 or 12 transmembrane helices (TH), and most kinases in this study lack transmembrane domains (427 PKs or 51.3 %) (Figure 2, Supplementary file S9).

The CsaPKs are widely distributed in subcellular localisations. About 41.0% of the CsaPKs were localised in the plasma membrane, 27.4 % were predicted in the nucleus, in the cytoplasm (17.7 %), in the chloroplast (6.5 %), mitochondrial (5.9 %) and extracellular (1.6 %). The plasma membrane protein kinases set was composed mostly of RLKs, but other kinases from other groups were also found in plasma membrane (BUB, CAMK_OST1L, CMGC, CMGC_CDKL-Cr, CMGC_GSKL, SCY1_SCYL1, STE_STE11, TKL-Pl-3, TKL-Pl-4, ULK_Fused, UNKNOWN, and WNK_NRBP). Out of 527 RLKs, 37 in chloroplast, 46 in cytoplasm, 12 extracellular, 20 in mitochondrion, 89 in nucleus and 323 in plasma membrane.

In our study of gene ontology, we identified 4260 GO terms in total, 372 GO different terms, from which 2008 related to protein function, 1459 related to metabolic process and 793 related to cellular component (Figure 4).

[Figure 4]

Phylogenetic Analysis of CsaPKs

The phylogenetic analysis (Figure 5) resulted in a clear division of 527 (RLK), 296 protein kinases of different subfamilies and eight protein kinases that had different classification from the phylogenetic analysis (Table 1) according to the system of kinase classification proposed by Lehti-Shiu and Shiu (2012a), therefore named as “unknown” in the phylogenetic tree (Supplementary file S11), and so, these unknown proteins need to be annotated. Basing on the plant protein kinase superfamily system classification (Lehti-Shiu and Shiu, 2012), the entire set of protein kinases was composed of following groups, with the most abundant was the RLK-Pelle (527), CAMK (66), CMGC (64), TKL (53), STE (37), OTHERS (32), AGC (28), CK1 (10), UNKNOWN (8), GROUP-PI/PLANT-SPECIFIC (6), and SCY (1), accounting for 125 subfamilies, set apart in different clades defined by the bootstrap value > 0.7 (Supplementary file S12). The group OTHERS contains the following subfamilies: AGC-Pl, Aur, BUB, IRE1, NAK, NEK, PEK_GCN2, SCY1_SCYL1, SCY1_SCYL2, TLK, TTK, ULK_Fused, ULK_ULK4, WEE, and WNK_NRBP. By the phylogeny, we classified 10 CsaPKs in new subfamilies, and eight CsaPKs were not grouped in any subfamily clades, described as “Unknown” in Table 1.

[Figure 5]

[Table 1]

The RLK is the largest group, with 55 subfamilies in the Csa kinome, accounting for about 62.7 % of the CsaPKs. This highlights their extensive responses to environmental stimuli and cell regulation processes. The subfamilies with the greatest number of proteins are RLK-Pelle_DLSV (69), RLK-Pelle_LRR-XI-1 (40), and RLK-Pelle_LRR-III (38). The second kinase richest group is CAMK, accounting for six only subfamilies, with the most abounding proteins in the subfamily CAMK_CDPK (29), CAMK_CAMKL-CHK1 (21), and CAMK_OST1L (12); and the third is CMGC, with the more kinase richest subfamilies CMGC_CDK-CRK7-CDK9 (16), CMGC_MAPK (15), and CMGC_GSK (7).

CsaPK Differential Gene Expression

Regarding the Csa DEG to powdery mildew (Xu et al., 2017a), comparing the down-regulated kinases of 48 HPI versus those of 0 HPI (Figure 6, Supplementary file S13), six were common to the genotypes, in addition to 34 and 28 kinases exclusive to the resistant and susceptible genotypes, respectively. Regarding the up-regulated kinases, 47 were exclusively expressed by the resistant genotype and another 48 by the susceptible genotype; with 14 kinases common to both. Down-regulated kinases were classified into 32 sub-families, 21 (65.6%) of the RLK group; other groups also occurred, such as AGC, CAMK, CMGC, Group-PI-3 (exclusive for down-regulated), NAK, STE, TKL and WNK. Among the up-regulated kinases, 49 different subfamilies were observed, with 31 (63.3%) belonging to the RLK group; the other groups observed were AGC, CAMK, CMGC, NAK, STE, TKL, WNK, in addition to the Group-PI-2 and IRE1 groups, exclusive to the up-regulated ones. In total, 177 kinases were disease-responsive, a result that supports the understanding of the plant defense metabolic reactions to the pathogen, allowing further research in the scope of Csa genetic improvement.

When comparing DEG results in relation to Alternaria leaf spot (Figure 6, Supplementary file S13), in relation to down-regulated genes, 19 DEGs (AGC, CAMK, CMGC, OTHERS, RLK, STE, TKL) were commonly expressed between the resistant and susceptible genotypes, 24 DEGs (AGC, CAMK, CMGC, OTHERS, PLANT-SPECIFIC, RLK, STE, TKL) by the resistant genotype and 23 DEGs (CMGC, AGC, CAMK, RLK, TKL, OTHERS) by the susceptible genotype. Regarding the up-regulated genes, seven DEGs (CAMK, CMGC, RLK, TKL) were common among the genotypes, 33 DEGs (AGC, CAMK, OTHERS, RLK, STE) for the resistant genotype and 10 DEGs (AGC, CAMK, RLK, OTHERS) for the susceptible genotype. In all, 116 DEGs were responsive to Alternaria leaf spot.

In the differential gene expression study involving *Meloidogyne incognita* (Figure 7, Supplementary file S13), three comparisons were performed: 0 x 1 DPI, 0 x 2 DPI and 0 x 3 DPI. At 1 DPI, regarding the Down regulated genes, the resistant genotype (Rijecheng) presented 37 DEGs (related to the groups AGC, CAMK, RLK, STE, OTHERS), the susceptible genotype (Superina) presented 15 DEGs (AGC, RLK, STE, TKL) and 15 common DEGs among them (AGC, RLK, STE and TKL). In relation to Up regulated genes, the resistant genotype presented 8 DEGs (CMGC, RLK, TKL), the susceptible genotype presented 37 DEGs (AGC,CAMK, CMGC, PLANT-SPECIFIC, RLK, STE, UNKNOWN, OTHERS) and 12 DEGs common among them (AGC, CAMK, CMGC, PLANT-SPECIFIC, RLK, UNKNOWN). When analyzing the down and up regulated DEGs, we observed that after 2 DPI the susceptible genotype did not show any regulated DEG Up. At 2 DPI, the resistant genotype had 27 down regulated DEGs (AGC, CAMK, RLK, OTHERS), and the susceptible genotype had 13 DEGs (CAMK, RLK, STE), and among them 16 common DEGs (RLK,STE,TKL). Regarding Up regulated genes, the resistant genotype presented 28 DEGs (AGC, CAMK, CMGC, PLANT-SPECIFIC, RLK, UNKNOWN, OTHERS). At 3 DPI, the resistant genotype presented 33 down-regulated DEGs (AGC, CAMK, RLK, STE, TKL), the susceptible genotype presented 6 DEGs (RLK, STE), and 13 common DEGs among them (CAMK, RLK, STE, TKL, OTHERS). As for the expression of regulated Up genes, only the resistant genotype presented expression, 33 DEGs (AGC, CAMK, CMGC, PLANT-SPECIFIC, RLK, UNKNOWN, OTHERS).

Proximal cis-regulatory elements

Cis-regulatory elements (cis-elements) may be classified in three main types (promoters, enhancers and silencers), regulating gene expression by direct interaction or proximity. The putative promoters were successifuly extracted from the 1.5 Kbp (proximal) upstream of the 831 CsaPK genes. The identification of the cis-elements was performed by searching the sequences against the PlantCARE database. A total of 107,978 cis-elements were identified, among which only 127 different elements. The most abundant cis-elements were TATA-box (35,184) and CAAT-box (27,161). Twelve elements appeared only once: 4cl-CMA1b, 4cl-CMA2b, GC-repeat, GGA-motif, GRA, HMG-TATA-region, OCT, Pc-CMA2c, Unnamed__16, chs-Unit, plant_AP-2-like and silencer. The number of cis-elementss

in the genes ranged from 38 to 246, since the highest number of cis-elements is the CsGy1G026280 gene. TATA-box is consensus sequence found in archea and eukaryotes, and it is a site for the TATA-binding protein (TFIID) and other transcription factors (Xu et al., 2016), having 4 to 13 bp. In our dataset, the TATA-box elelment was found since the first to the 1497th position upstream the transcription iniciation site, appearing in variable order, and in all 831 genes. This confirms the elementar function of the TATA-box in the regulation of the initiation of the transcription of CsaPKs. The same occurred to CAAT-box and Unnamed__4, occurring in all 831 genes. Twelve (12) cis-elements were exclusive, appearing only once (plant_AP-2-like, OCT, Unnamed__16, Pc-CMA2c, chs-Unit 1 m1, GC-repeat, GRA, 4cl-CMA1b, GGA-motif, 4cl-CMA2b, HMG-TATA-region, and silencer). Three (3) cis-elements have the greatest sequence size (Unnamed__5, 5UTR_Py-rich_stretch, and silencer) with 18 bp. Motif enrichment analysis determines whether a given promoter has a significantly higher level of affinity for a regulatory protein or microRNA for which the DNA binding motif is known. If significant, this motif is said to be enriched. A full list of the cis-elements retrieved are provided in the Supplementary file S14.

Discussion

The protein kinases are core regulators of the plant metabolism and are considered important modulators of biotic stress. The number of protein kinases among plant species is a quite variable information, in average 4.0% for the species common bean (3.25%), maize (3.8%), Arabidopsis (3.4%), grapevine (3.7%), cowpea (4.3%), soybean (4.7%) and rubber tree (4.9%) (Aono et al., 2023; Ferreira-Neto et al., 2021b; Jinyi Liu et al., 2015; Wei et al., 2014b; Zhu et al., 2018d; Zulawski et al., 2014b). In cucumber, 831 PK were found and classified into 20 different groups and 125 subfamilies. Although the number of protein kinases in the species *C. sativus* is quite similar, its genome lacks recent WGD (Huang et al., 2009b), which may limit the evolution and expansion of these protein families. They have a very diverse protein structure, with a variable number of kinase domains, transmembrane domains and signal peptide. Signal peptide is a sequence of amino acids present in newly synthesized proteins indicating that the protein must be secreted from one point to another in the cell, i.e., that protein belongs to the secretory pathway (Heijne, 1990). The blast-based approach could be used to determine such proteins in genome-wide analysis studies (Abbas et al., 2022). However, the profile-HMM-based approach seems to be the most frequente approach in such studies (Aono et al., 2023, 2021b; Ferreira-Neto et al., 2021b; Liu et al., 2020b; Jinyi Liu et al., 2015; Lv et al., 2020; Singh et al., 2014b; Wei et al., 2014b; Yan et al., 2018b, 2017b; Zhu et al., 2021b, 2018d, 2018b; Zulawski et al., 2014b).

A concise analysis was performed regarding the protein structure, including motifs, structural and transmembrane domains, the exon-intron structure and the construction of a chromosomal map identifying the position of the genes in each chromosome. Protein motifs and domains are their functional modules which are manifest in different levels, comprising since whole-protein elements (domains and complexes) to protein parts (binding motifs and modification sites (Tompa et al., 2014). In our study, wide set of proteins motifs and domains were detected, highlighting that their functioning in phosphorylating targets is much more than a formidable task. The maximum number of introns corroborated with those observed in cowpea (Ferreira-Neto et al., 2021b), soybean (Jinyi Liu et al., 2015) and common bean (Aono et al., 2023). Introns are unique components of eukaryotic organisms, and it is assumed that they participate in complex and specific characteristics, so they are not eliminated from the gene structure (Chorev and Carmel, 2012; Lynch, 2007; Wahl et al., 2009). Introns perform direct (regulation of alternative splicing, enhancement of gene expression, control of transport of mRNA or chromatin assembly and interfere nonsense mediated decay in the 5' and 3' UTR) and indirect functions (they provide a source of new genes, the length of introns affects the efficiency of natural selection, harbor various types of functional non-coding RNA genes, trait-associated SNPs are enriched

in introns, different ordinal position of introns within the gene may have a different functional role between lineages of the same species, intron sequences regulate alternative splicing) (Jo and Choi, 2015). Furthermore, researches suggest that introns make up cis-acting regulatory elements, such as intronic splicing silencers and intronic splicing enhancers. In addition, the presence of introns in certain genes directly interferes with the evolution of other genes in cases where there is a genetic linkage between sites, as this connection reduces the efficiency of natural selection, which acts differently in genes with a differentiated exon-intron structure according to the Hill-Robertson interference model (Hill-Robertson effect), which increases recombination between two sites in two neighboring exons (Comeron et al., 2008).

In our study of DEGs, we used RNA-seq data involving three diseases, Powdery mildew (Xu et al., 2017b), Alternaria leaf spot (Sa et al., 2020b) and Root-knot nematode (Wang et al., 2018). Including the three cases of our study, in total 366 genes were differentially expressed, of which, for the most part, proteins from the RLK group were prevalent (67.5% or 247 PKs), followed by PKs from the CAMK group (9.6% or 35 PKs), TKL (5.2% or 19 PKs), CMGC (4.6% or 17 PKs), STE (4.1% or 15 PKs), OTHERS (3.6% or 13 PKs), AGC (1.4% or 5 PKs), PLANT-SPECIFIC (1.4 % or 5 PKs), and UNKNOWN (1.1% or 4 PKs). Notably, the PKs of the RLK group stood out in the differential expression studies, with 49 subfamilies, in 13.5% of the cases occurring PKs of the RLK-Pelle_DLSV subfamily, 8.9% PKs of the RLK-Pelle_LRR-XI-1 subfamily, and 7.2% PKs of the subfamily RLK-Pelle_LRR-I-1. Like the Phaseolus kinome differential gene expression (Aono et al., 2023), the first two subfamilies also stood out. However, the groups with the highest correlations between Aurora kinase subfamily means and a UNKNOWN subfamily (R = 0.99, Supplementary file S15).

References

Abbas, A., Shah, A.N., Shah, A.A., Nadeem, M.A., Alsaleh, A., Javed, T., Alotaibi, S.S., Abdelsalam, N.R., 2022. Genome-Wide Analysis of Invertase Gene Family, and Expression Profiling under Abiotic Stress Conditions in Potato. Biology (Basel) 11, 539. https://doi.org/10.3390/biology11040539

Aono, A.H., Pimenta, R.J.G., Dambroz, C.M. da S., Costa, F.C.L., Kuroshu, R.M., de Souza, A.P., Pereira, W.A., 2023. Genome-wide characterization of the common bean kinome: Catalog and insights into expression patterns and genetic organization. Gene 855, 147127. https://doi.org/10.1016/j.gene.2022.147127

Aono, A.H., Pimenta, R.J.G., Garcia, A.L.B., Correr, F.H., Hosaka, G.K., Carrasco, M.M., Cardoso-Silva, C.B., Mancini, M.C., Sforça, D.A., dos Santos, L.B., Nagai, J.S., Pinto, L.R., Landell, M.G. de A., Carneiro, M.S., Balsalobre, T.W., Quiles, M.G., Pereira, W.A., Margarido, G.R.A., de Souza, A.P., 2021a. The Wild Sugarcane and Sorghum Kinomes: Insights Into Expansion, Diversification, and Expression Patterns. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.668623

Aono, A.H., Pimenta, R.J.G., Garcia, A.L.B., Correr, F.H., Hosaka, G.K., Carrasco, M.M., Cardoso-Silva, C.B., Mancini, M.C., Sforça, D.A., dos Santos, L.B., Nagai, J.S., Pinto, L.R., Landell, M.G. de A., Carneiro, M.S., Balsalobre, T.W., Quiles, M.G., Pereira, W.A., Margarido, G.R.A., de Souza, A.P., 2021b. The Wild Sugarcane and Sorghum Kinomes: Insights Into Expansion, Diversification, and Expression Patterns. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.668623

Arab, M., Zarrini, H.N., Nematzadeh, G., Hashemipetroudi, S.H., 2021. Bioinformatics analysis of calcium-dependent protein kinase (CPK) in Aeluropus littoralis L. Crop Biotechnology 10, 1–15. https://doi.org/10.30473/CB.2021.58258.1834

Cai, H., Cheng, J., Yan, Y., Xiao, Z., Li, J., Mou, S., Qiu, A., Lai, Y., Guan, D., He, S., 2015. Genome-wide identification and expression analysis of calcium-dependent protein Kinase and its closely related Kinase genes in capsicum Annuum. Front Plant Sci 6. https://doi.org/10.3389/fpls.2015.00737

Champion, A., Kreis, M., Mockaitis, K., Picaud, A., Henry, Y., 2004. Arabidopsis kinome: After the casting. Funct Integr Genomics 4, 163–187. https://doi.org/10.1007/s10142-003-0096-4

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., Xia, R., 2020. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant 13, 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chorev, M., Carmel, L., 2012. The Function of Introns. Front Genet 3. https://doi.org/10.3389/fgene.2012.00055

Comeron, J.M., Williford, A., Kliman, R.M., 2008. The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity (Edinb) 100, 19–31. https://doi.org/10.1038/sj.hdy.6801059

Conesa, A., Götz, S., 2008. Blast2go: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics.

Dai, Y., Zhang, X., Tian, J., Zhu, L., Li, D., 2021. Genome-Wide Characterization of CDPK Gene Family in Apple (Malus Domestica) and Its Transcriptional Expression During Apple Fruit Development. Res Sq. https://doi.org/10.21203/rs.3.rs-445130/v1

Dardick, C., Chen, J., Richter, T., Ouyang, S., Ronald, P., 2007. The rice kinase database. A phylogenomic database for the rice kinome. Plant Physiol 143, 579–586. https://doi.org/10.1104/pp.106.087270

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47, D427–D432. https://doi.org/10.1093/nar/gky995

Ferreira-Neto, J.R.C., Borges, A.N. da C., da Silva, M.D., Morais, D.A. de L., Bezerra-Neto, J.P., Bourque, G., Kido, E.A., Benko-Iseppon, A.M., 2021a. The Cowpea Kinome: Genomic and Transcriptomic Analysis Under Biotic and Abiotic Stresses. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.667013

Ferreira-Neto, J.R.C., Borges, A.N. da C., da Silva, M.D., Morais, D.A. de L., Bezerra-Neto, J.P., Bourque, G., Kido, E.A., Benko-Iseppon, A.M., 2021b. The Cowpea Kinome: Genomic and Transcriptomic Analysis Under Biotic and Abiotic Stresses. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.667013

Gill, R.A., Ali, B., Yang, S., Tong, C., Islam, F., Gill, M.B., Mwamba, T.M., Ali, S., Mao, B., Liu, S., Zhou, W., 2017. Reduced glutathione mediates pheno-ultrastructure, kinome and transportome in chromium-induced Brassica napus L. Front Plant Sci 8, 1–24. https://doi.org/10.3389/fpls.2017.02037

Heijne, G. von, 1990. The Signal Peptide. J Membr Biol 115, 195–201.

Herrera-Pola, P.S., Xoconostle-Cázares, B., Toscano-Morales, R., Ruiz-Medrano, R., 2016. Characterization of a pumpkin mRNA encoding a Cyclin-Dependent Protein Kinase (CDK) potentially involved in phloem development. Plant Omics 9, 213–219. https://doi.org/10.21475/poj.16.09.03.pne114

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., Ren, Yuanyuan, Zhu, H., Li, Jun, Lin, K., Jin, W., Fei, Z., Li, Guangcun, Staub, J., Kilian, A., van der Vossen, E.A.G., Wu, Y., Guo, J., He, J., Jia, Z., Ren, Yi, Tian, G., Lu, Y., Ruan, J., Qian, W., Wang, M., Huang, Q., Li, B., Xuan, Z., Cao, J., Asan, Wu, Z., Zhang, J., Cai, Q., Bai, Y., Zhao, B., Han, Y., Li, Ying, Li, X., Wang, S., Shi, Q., Liu, Shiqiang, Cho, W.K., Kim, J.-Y., Xu, Y., Heller-Uszynska, K., Miao, H., Cheng, Z., Zhang, S., Wu, J., Yang, Y., Kang, H., Li, M., Liang, H., Ren, X., Shi, Z., Wen, M., Jian, M., Yang, Hailong, Zhang, G., Yang, Z., Chen, R., Liu, Shifang, Li, Jianwen, Ma, L., Liu, H., Zhou, Y., Zhao, J., Fang, X., Li, Guoqing, Fang, L., Li, Yingrui, Liu, D., Zheng, Hongkun, Zhang, Y., Qin, N., Li, Z., Yang, G., Yang, S., Bolund, L., Kristiansen, K., Zheng, Hancheng, Li, Shaochuan, Zhang, X., Yang, Huanming, Wang, Jian, Sun, R., Zhang, B., Jiang, S., Wang, Jun, Du, Y., Li, Songgang, 2009a. The genome of the cucumber, Cucumis sativus L. Nat Genet 41, 1275–1281. https://doi.org/10.1038/ng.475

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., Ren, Yuanyuan, Zhu, H., Li, Jun, Lin, K., Jin, W., Fei, Z., Li, Guangcun, Staub, J., Kilian, A., van der Vossen, E.A.G., Wu, Y., Guo, J., He, J., Jia, Z., Ren, Yi, Tian, G., Lu, Y., Ruan, J., Qian, W., Wang, M., Huang, Q., Li, B., Xuan, Z., Cao, J., Asan, Wu, Z., Zhang, J., Cai, Q., Bai, Y., Zhao, B., Han, Y., Li, Ying, Li, X., Wang, S., Shi, Q., Liu, Shiqiang, Cho, W.K., Kim, J.Y., Xu, Y., Heller-Uszynska, K., Miao, H., Cheng, Z., Zhang, S., Wu, J., Yang, Y., Kang, H., Li, M., Liang, H., Ren, X., Shi, Z., Wen, M., Jian, M., Yang, Hailong, Zhang, G., Yang, Z., Chen, R., Liu, Shifang, Li, Jianwen, Ma, L., Liu, H., Zhou, Y., Zhao, J., Fang, X., Li, Guoqing, Fang, L., Li, Yingrui, Liu, D., Zheng, Hongkun, Zhang, Y., Qin, N., Li, Z., Yang, G., Yang, S., Bolund, L., Kristiansen, K.,

Zheng, Hancheng, Li, Shaochuan, Zhang, X., Yang, Huanming, Wang, Jian, Sun, R., Zhang, B., Jiang, S., Wang, Jun, Du, Y., Li, Songgang, 2009b. The genome of the cucumber, Cucumis sativus L. Nat Genet 41, 1275–1281. https://doi.org/10.1038/ng.475

Jo, B.-S., Choi, S.S., 2015. Introns: The Functional Benefits of Introns in Genomes. Genomics Inform 13, 112. https://doi.org/10.5808/GI.2015.13.4.112

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772–780. https://doi.org/10.1093/molbev/mst010

Khan, A., Mishra, A., Hasan, S.M., Usmani, A., Ubaid, M., Khan, N., Saidurrahman, M., 2022. Biological and medicinal application of Cucumis sativus Linn. – review of current status with future possibilities. J Complement Integr Med 19, 843–854. https://doi.org/10.1515/jcim-2020-0240

Kolde, R., Kolde, M.R., 2015. Package 'pheatmap'. R package 1.

Kong, X., Lv, W., Jiang, S., Zhang, D., Cai, G., Pan, J., Li, D., 2013. Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. BMC Genomics 14. https://doi.org/10.1186/1471-2164-14-433

Lehti-Shiu, M.D., Shiu, S.H., 2012. Diversity, classification and function of the plant protein kinase superfamily. Philosophical Transactions of the Royal Society B: Biological Sciences 367, 2619–2639. https://doi.org/10.1098/rstb.2012.0003

Lescot, M., 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30, 325–327. https://doi.org/10.1093/nar/30.1.325

Letunic, I., Bork, P., 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46, D493–D496. https://doi.org/10.1093/nar/gkx922

Letunic, I., Khedkar, S., Bork, P., 2021. SMART: Recent updates, new developments and status in 2020. Nucleic Acids Res 49, D458–D460. https://doi.org/10.1093/nar/gkaa937

Li, M., Hu, W., Ren, L., Jia, C., Liu, J., Miao, H., Guo, A., Xu, B., Jin, Z., 2019. Identification, Expression, and Interaction Network Analyses of the CDPK Gene Family Reveal Their Involvement in the Development, Ripening, and Abiotic Stress Response in Banana. Biochem Genet 58, 40–62. https://doi.org/10.1007/s10528-019-09916-2

Li, T.G., Zhang, D.D., Zhou, L., Kong, Z.Q., Hussaini, A.S., Wang, D., Li, J.J., Short, D.P.G., Dhar, N., Klosterman, S.J., Wang, B.L., Yin, C.M., Subbarao, K. V., Chen, J.Y., Dai, X.F., 2018. Genome-wide identification and functional analyses of the crk gene family in cotton reveals gbcrk18 confers verticillium wilt resistance in gossypium barbadense. Front Plant Sci 9, 1–16. https://doi.org/10.3389/fpls.2018.01266

Liu, H., Qu, W., Zhu, K., Cheng, Z.M., 2020a. The wild strawberry kinome: Identification, classification and transcript profiling of protein kinases during development and in response to gray mold infection. BMC Genomics 21, 1–14. https://doi.org/10.1186/s12864-020-07053-4

Liu, H., Qu, W., Zhu, K., Cheng, Z.M., 2020b. The wild strawberry kinome: Identification, classification and transcript profiling of protein kinases during development and in response to gray mold infection. BMC Genomics 21. https://doi.org/10.1186/s12864-020-07053-4

Liu, J, Chen, N., Grant, J.N., Cheng, Z.-M. (Max), Stewart Jr, C.N., Hewezi, T., 2015. Soybean kinome: functional classification and gene expression patterns. J Exp Bot 66, 1919–1934. https://doi.org/10.1093/jxb/eru537

Liu, Jinyi, Chen, N., Grant, J.N., Cheng, Z.M., Stewart, C.N., Hewezi, T., 2015. Soybean kinome: Functional classification and gene expression patterns. J Exp Bot 66, 1919–1934. https://doi.org/10.1093/jxb/eru537

Lv, D., Wang, G., Xiong, L.-R., Sun, J.-X., Chen, Y., Guo, C.-L., Yu, Y., He, H.-L., Cai, R., Pan, J.-S., 2020. Genome-Wide Identification and Characterization of Lectin Receptor-Like Kinase Gene Family in Cucumber and Expression Profiling Analysis under Different Treatments. Genes (Basel) 11, 1032. https://doi.org/10.3390/genes11091032

Lynch, M., 2007. The Origins of Genome Architecture, 1st ed. Sinauer , Sunderland.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The Protein Kinase Complement of the Human Genome. Science (1979) 298, 1912–1934. https://doi.org/10.1126/science.1075762

Neuwirth, E., 2014. RColorBrewer: ColorBrewer Palettes. R package version 1.1-3.

Ratcliff, W.C., Herron, M.D., Howell, K., Pentz, J.T., Rosenzweig, F., Travisano, M., 2013. Experimental evolution of an alternating uni- and multicellular life cycle in Chlamydomonas reinhardtii. Nat Commun 4. https://doi.org/10.1038/ncomms3742

Rui, T., Han, N., Hu, Y., Wang, Lixue, Ren, Z., Wang, Lina, 2020. Genome-wide Identification and Expression Analysis of GRAS Genes in Cucumber. International Journal of Horticulture. https://doi.org/10.5376/ijh.2020.10.0001

Sabag, M., Ari, G. Ben, Zviran, T., Biton, I., Goren, M., Dahan, Y., Sadka, A., Irihimovitch, V., 2013. PaKRP, a cyclin-dependent kinase inhibitor from avocado, may facilitate exit from the cell cycle during fruit growth. Plant Science 213, 18–29. https://doi.org/10.1016/j.plantsci.2013.08.007

Santos, L.B. dos, Aono, A.H., Francisco, F.R., da Silva, C.C., Souza, L.M., Souza, A.P. de, 2023. The rubber tree kinome: Genome-wide characterization and insights into coexpression patterns associated with abiotic stress responses. Front Plant Sci 14. https://doi.org/10.3389/fpls.2023.1068202

Sa, R., Liu, D., Chen, L., Liu, L., Zhang, Y., Zhang, X., Xu, C., 2020a. Transcriptome analysis of mechanisms and candidate genes associated with cucumber response to cucumber alternaria leaf spot infection. Physiol Mol Plant Pathol 111, 101490. https://doi.org/10.1016/j.pmpp.2020.101490

Sa, R., Liu, D., Chen, L., Liu, L., Zhang, Y., Zhang, X., Xu, C., 2020b. Transcriptome analysis of mechanisms and candidate genes associated with cucumber response to cucumber alternaria leaf spot infection. Physiol Mol Plant Pathol 111. https://doi.org/10.1016/j.pmpp.2020.101490

Shi, J., Du, X., 2021. Basic Analysis of Calcium-dependent Protein Kinase Gene and Its Closely Related Gene Families in Solanum Pennellii Genome. Res Sq.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F.X., Li, W.H., 2004. Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell 16, 1220–1234. https://doi.org/10.1105/tpc.020834

Shoresh, M., Gal-On, A., Leibman, D., Chet, I., 2006. Characterization of a mitogen-activated protein kinase gene from cucumber required for trichoderma-conferred plant resistance. Plant Physiol 142, 1169–1179. https://doi.org/10.1104/pp.106.082107

Singh, D.K., Calvino, M., Brauer, E.K., Fernandez-Pozo, N., Strickler, S., Yalamanchili, R., Suzuki, H., Aoki, K., Shibata, D., Stratmann, J.W., Popescu, G. v., Mueller, L.A., Popescu, S.C., 2014a. The tomato kinome and the tomato kinase library orfeome: Novel resources for the study of kinases and signal transduction in tomato and solanaceae species. Molecular Plant-Microbe Interactions 27, 7–17. https://doi.org/10.1094/MPMI-08-13-0218-TA

Singh, D.K., Calvino, M., Brauer, E.K., Fernandez-Pozo, N., Strickler, S., Yalamanchili, R., Suzuki, H., Aoki, K., Shibata, D., Stratmann, J.W., Popescu, G. v., Mueller, L.A., Popescu, S.C., 2014b. The tomato kinome and the tomato kinase library orfeome: Novel resources for the study of kinases and signal transduction in tomato and solanaceae species. Molecular Plant-Microbe Interactions 27, 7–17. https://doi.org/10.1094/MPMI-08-13-0218-TA

Slowikowski, K., Schep, A., Hughes, S., Dang, T.K., Lukauskas, S., Irisson, J.-O., Kamvar, Z.N., Ryan, T., Christophe, D., Hiroaki, Y., Gramme, P., Abdol, A.M., Barrett, M., Cannoodt, R., Krassowski, M., Chirico, M., Aphalo, P., 2018. ggrepel: Automatically Position Non-Overlapping Text Labels with “ggplot2.”

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS One 6, e21800. https://doi.org/10.1371/journal.pone.0021800

Tennekes, M., 2021. treemap: Treemap Visualization.

Tompa, P., Davey, N.E., Gibson, T.J., Babu, M.M., 2014. A Million Peptide Motifs for the Molecular Biologist. Mol Cell 55, 161–169. https://doi.org/10.1016/j.molcel.2014.05.032

Vidal, D., Alvarez-Flórez, F., Simón, E., 2010. Casein kinase activity in etiolated Cucumis sativus cotyledons. Plant Biol 12, 134–144. https://doi.org/10.1111/j.1438-8677.2009.00212.x

Wahl, M.C., Will, C.L., Lührmann, R., 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell 136, 701–718. https://doi.org/10.1016/j.cell.2009.02.009

Wang, J.P., Xu, Y.P., Munyampundu, J.P., Liu, T.Y., Cai, X.Z., 2016. Calcium-dependent protein kinase (CDPK) and cdpk-related kinase (CRK) gene families in tomato: Genome-wide identification and functional analyses in disease resistance. Molecular Genetics and Genomics 291, 661–676. https://doi.org/10.1007/s00438-015-1137-0

Wang, X., Cheng, C., Zhang, K., Tian, Z., Xu, J., Yang, S., Lou, Q., Li, J., Chen, J.-F., 2018. Comparative transcriptomics reveals suppressed expression of genes related to auxin and the cell cycle contributes to the resistance of cucumber against Meloidogyne incognita. BMC Genomics 19, 583. https://doi.org/10.1186/s12864-018-4979-0

Wei, C., Zhang, R., Yang, X., Zhu, C., Li, H., Zhang, Y., Ma, J., Yang, J., Zhang, X., 2019. Comparative analysis of calcium-dependent protein kinase in cucurbitaceae and expression studies in Watermelon. Int J Mol Sci 20, 1–18. https://doi.org/10.3390/ijms20102527

Wei, K., Wang, Y., Xie, D., 2014a. Identification and expression profile analysis of the protein kinase gene superfamily in maize development. Molecular Breeding 33, 155–172. https://doi.org/10.1007/s11032-013-9941-x

Wei, K., Wang, Y., Xie, D., 2014b. Identification and expression profile analysis of the protein kinase gene superfamily in maize development. Molecular Breeding 33, 155–172. https://doi.org/10.1007/s11032-013-9941-x

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J Open Source Softw 4, 1686. https://doi.org/10.21105/joss.01686

Xu, M., Gonzalez-Hurtado, E., Martinez, E., 2016. Core promoter-specific gene regulation: TATA box selectivity and Initiator-dependent bi-directionality of serum response factor-activated transcription. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1859, 553–563. https://doi.org/10.1016/j.bbagrm.2016.01.005

Xu, Q., Xu, X., Shi, Y., Qi, X., Chen, X., 2017a. Elucidation of the molecular responses of a cucumber segment substitution line carrying Pm5.1 and its recurrent parent triggered by powdery mildew by comparative transcriptome profiling. BMC Genomics 18, 21. https://doi.org/10.1186/s12864-016-3438-z

Xu, Q., Xu, X., Shi, Y., Qi, X., Chen, X., 2017b. Elucidation of the molecular responses of a cucumber segment substitution line carrying Pm5.1 and its recurrent parent triggered by powdery mildew by comparative transcriptome profiling. BMC Genomics 18, 21. https://doi.org/10.1186/s12864-016-3438-z

Xu, X., Liu, M., Lu, L., He, M., Qu, W., Xu, Q., Qi, X., Chen, X., 2015. Genome-wide analysis and expression of the calcium-dependent protein kinase gene family in cucumber. Molecular Genetics and Genomics 290, 1403–1414. https://doi.org/10.1007/s00438-015-1002-1

Yadav, V., Wang, Z., Yang, X., Wei, C., Changqing, X., Zhang, X., 2021. Comparative analysis, characterization and evolutionary study of dirigent gene family in cucurbitaceae and expression of novel dirigent peptide against powdery mildew stress. Genes (Basel) 12, 1–26. https://doi.org/10.3390/genes12030326

Yan, J., Li, G., Guo, X., Li, Y., Cao, X., 2018a. Genome-wide classification, evolutionary analysis and gene expression patterns of the kinome in Gossypium. PLoS One 13, 1–28. https://doi.org/10.1371/journal.pone.0197392

Yan, J., Li, G., Guo, X., Li, Y., Cao, X., 2018b. Genome-wide classification, evolutionary analysis and gene expression patterns of the kinome in Gossypium. PLoS One 13. https://doi.org/10.1371/journal.pone.0197392

Yan, J., Su, P., Wei, Z., Nevo, E., Kong, L., 2017a. Genome-wide identification, classification, evolutionary analysis and gene expression patterns of the protein kinase gene family in wheat and Aegilops tauschii. Plant Mol Biol 95, 227–242. https://doi.org/10.1007/s11103-017-0637-1

Yan, J., Su, P., Wei, Z., Nevo, E., Kong, L., 2017b. Genome-wide identification, classification, evolutionary analysis and gene expression patterns of the protein kinase gene family in wheat and Aegilops tauschii. Plant Mol Biol 95, 227–242. https://doi.org/10.1007/s11103-017-0637-1

Zhang, H., Wei, C., Yang, X., Chen, H., Yang, Y., Mo, Y., Li, H., Zang, Y., Ma, J., Yang, J., Zhang, X., 2017. Genome-wide identification and expression analysis of calcium-dependent protein kinase and its related kinase gene families in melon (Cucumis melo L.). PLoS One 12, 1–23. https://doi.org/10.1371/journal.pone.0176352

Zhang, K., Han, Y.T., Zhao, F.L., Hu, Y., Gao, Y.R., Ma, Y.F., Zheng, Y., Wang, Y.J., Wen, Y.Q., 2015. Genome-wide Identification and Expression Analysis of the CDPK Gene Family in Grape, Vitis spp. BMC Plant Biol 15. https://doi.org/10.1186/s12870-015-0552-z

Zhang, M., Liu, Yanhui, He, Q., Chai, M., Huang, Y., Chen, F., Wang, X., Liu, Yeqiang, Cai, H., Qin, Y., 2020. Genome-wide investigation of calcium-dependent protein kinase gene family in pineapple: Evolution and expression profiles during development and stress. BMC Genomics 21, 1–16. https://doi.org/10.1186/s12864-020-6501-8

Zhang, Z., Luo, S., Liu, Z., Wan, Z., Gao, X., Qiao, Y., Yu, J., Zhang, G., 2022. Genome-wide identification and expression analysis of the cucumber PYL gene family. PeerJ 10. https://doi.org/10.7717/peerj.12786

Zhao, P., Liu, Y., Kong, W., Ji, J., Cai, T., Guo, Z., 2021. Genome-wide identification and characterization of calcium-dependent protein kinase (CDPK) and CDPK-related kinase

(CRK) gene families in medicago truncatula. Int J Mol Sci 22, 1–16. https://doi.org/10.3390/ijms22031044

Zhu, K., Fan, P., Liu, H., Zhao, J., Tan, P., Mo, Z., Peng, F., 2021a. Pecan kinome: classification and expression analysis of all protein kinases in *Carya illinoinensis*. Forestry Research 1, 1–12. https://doi.org/10.48130/fr-2021-0014

Zhu, K., Fan, P., Liu, H., Zhao, J., Tan, P., Mo, Z., Peng, F., 2021b. Pecan kinome: classification and expression analysis of all protein kinases in *Carya illinoinensis*. Forestry Research 1, 1–12. https://doi.org/10.48130/FR-2021-0014

Zhu, K., Liu, H., Chen, X., Cheng, Q., Cheng, Z.M.M., 2018a. The kinome of pineapple: Catalog and insights into functions in crassulacean acid metabolism plants. BMC Plant Biol 18, 1–16. https://doi.org/10.1186/s12870-018-1389-z

Zhu, K., Liu, H., Chen, X., Cheng, Q., Cheng, Z.M.M., 2018b. The kinome of pineapple: Catalog and insights into functions in crassulacean acid metabolism plants. BMC Plant Biol 18. https://doi.org/10.1186/s12870-018-1389-z

Zhu, K., Wang, X., Liu, J., Tang, J., Cheng, Q., Chen, J.G., Cheng, Z.M., 2018c. The grapevine kinome: Annotation, classification and expression patterns in developmental processes and stress responses. Hortic Res 5. https://doi.org/10.1038/s41438-018-0027-0

Zhu, K., Wang, X., Liu, J., Tang, J., Cheng, Q., Chen, J.G., Cheng, Z.M., 2018d. The grapevine kinome: Annotation, classification and expression patterns in developmental processes and stress responses. Hortic Res 5. https://doi.org/10.1038/s41438-018-0027-0

Zulawski, M., Schulze, G., Braginets, R., Hartmann, S., Schulze, W.X., 2014a. The Arabidopsis Kinome: Phylogeny and evolutionary insights into functional diversification. BMC Genomics 15. https://doi.org/10.1186/1471-2164-15-548

Zulawski, M., Schulze, G., Braginets, R., Hartmann, S., Schulze, W.X., 2014b. The Arabidopsis Kinome: phylogeny and evolutionary insights into functional diversification. BMC Genomics 15, 548. https://doi.org/10.5061/dryad.pq7d7

Figures and Tables

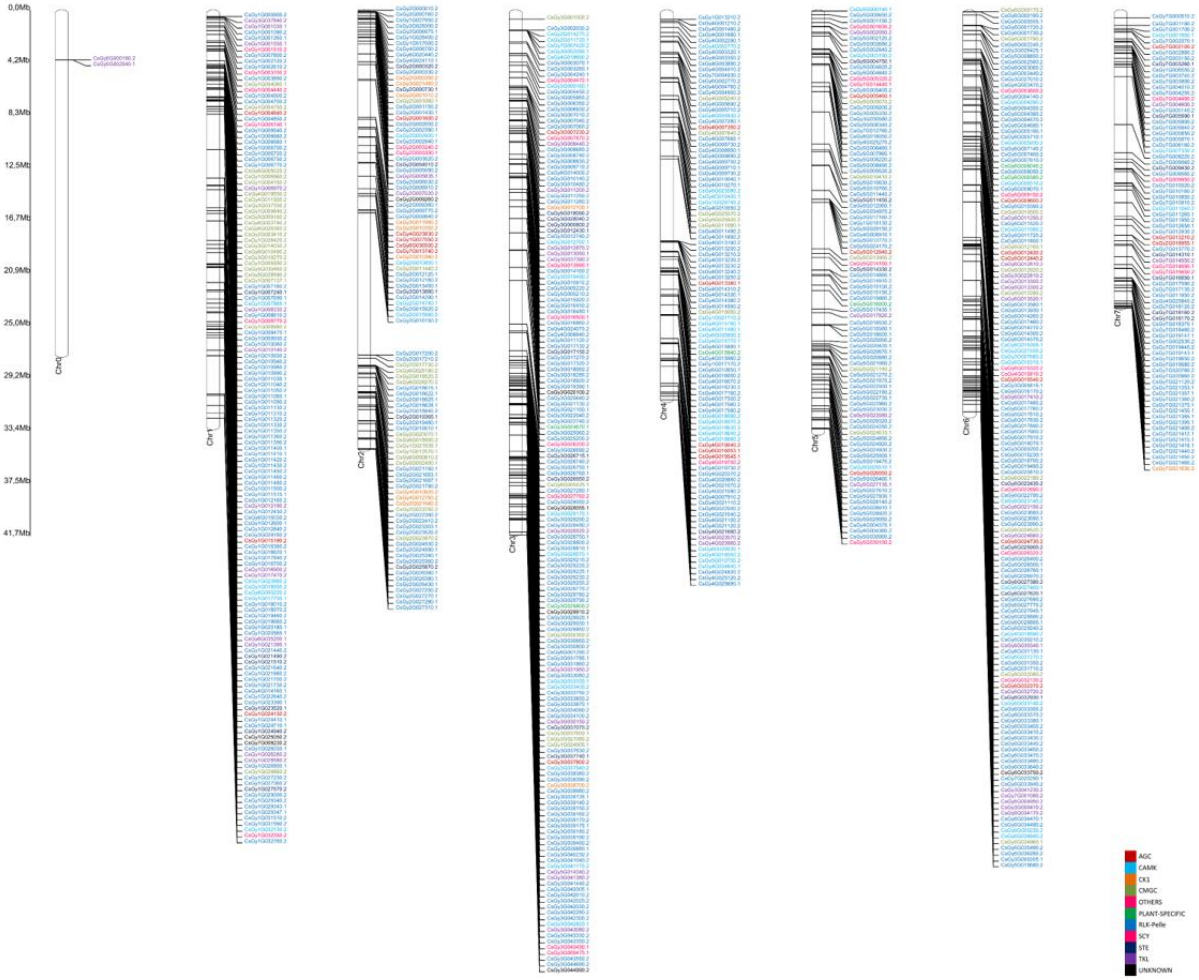

Figure 1. Chromosomal map of *Cucumis sativus* protein kinases. Different colors represent protein kinases by group.

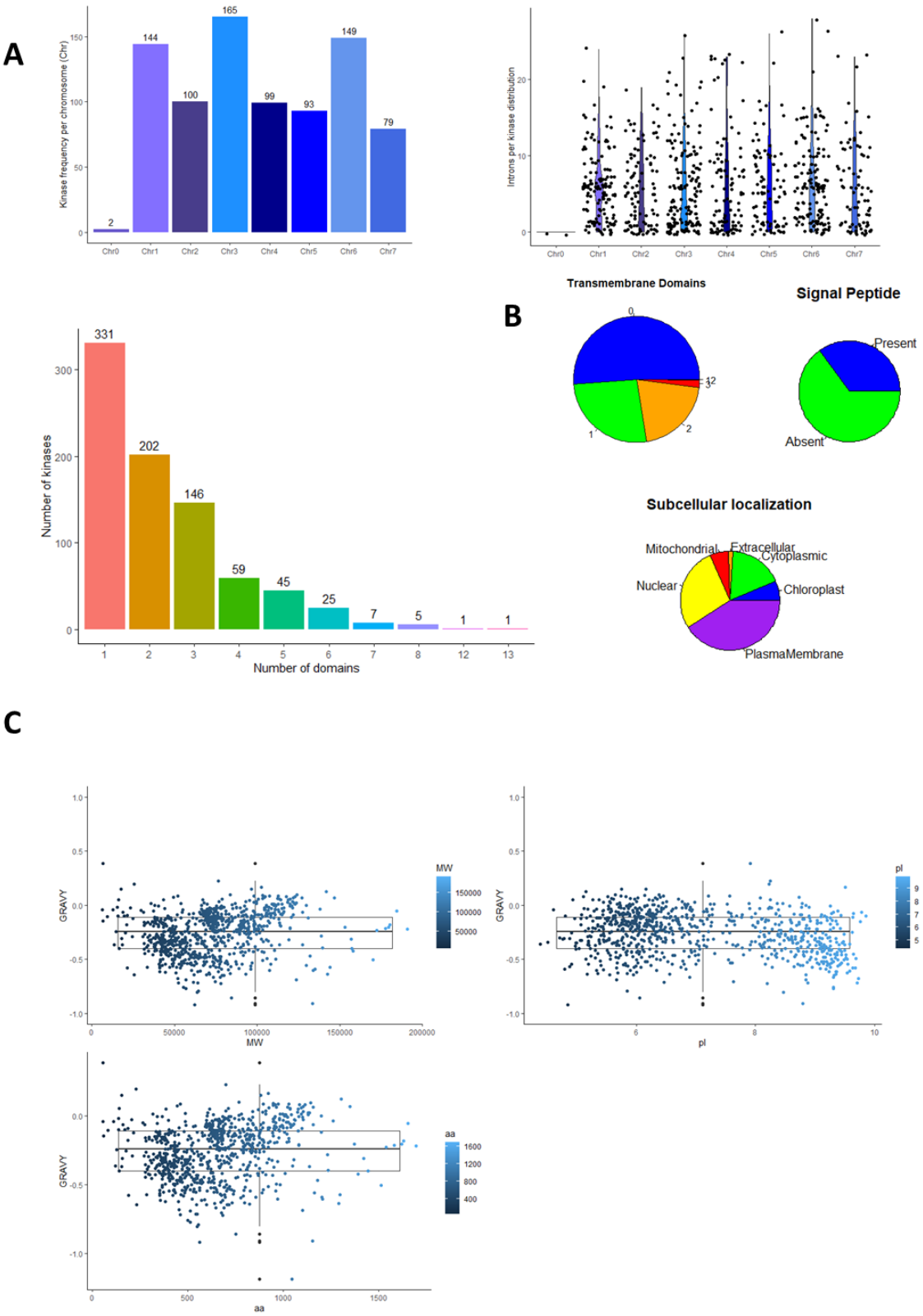

Figure 2. (A) Distribution of kinases per chromosomes, distribution of introns per kinase per chromosome and Number of domains per kinase. (B) Subcellular localization, presence of transmembrane domains and signal peptides. (C) Molecular properties of the protein kinases: Molecular weight (MW), isoelectric point (pI), Grand Average of Hydropathicity (GRAVY) and amino acid number per kinase (aa).

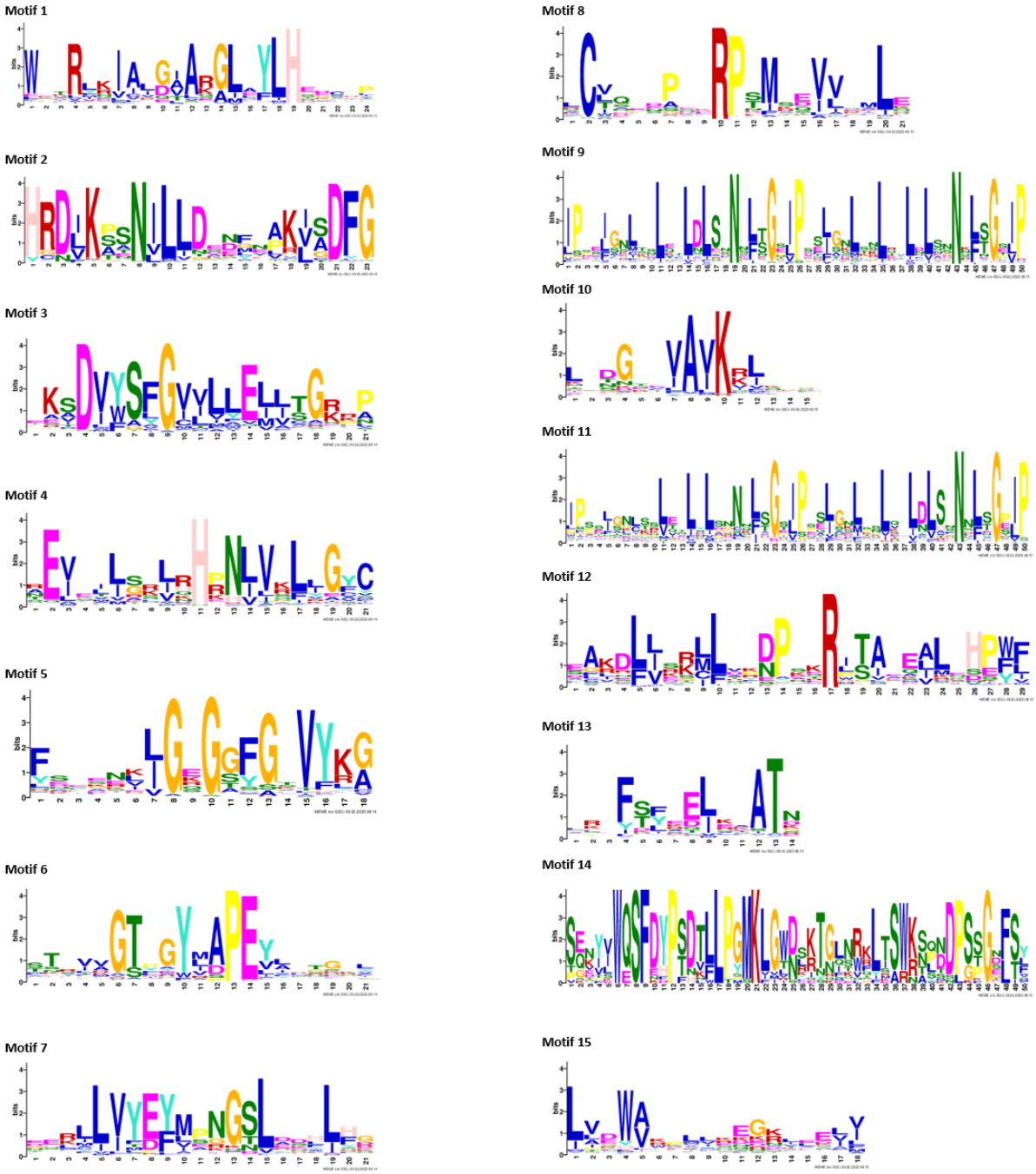

Figure 3. Conserved motifs of Cucumis sativus protein kinases.

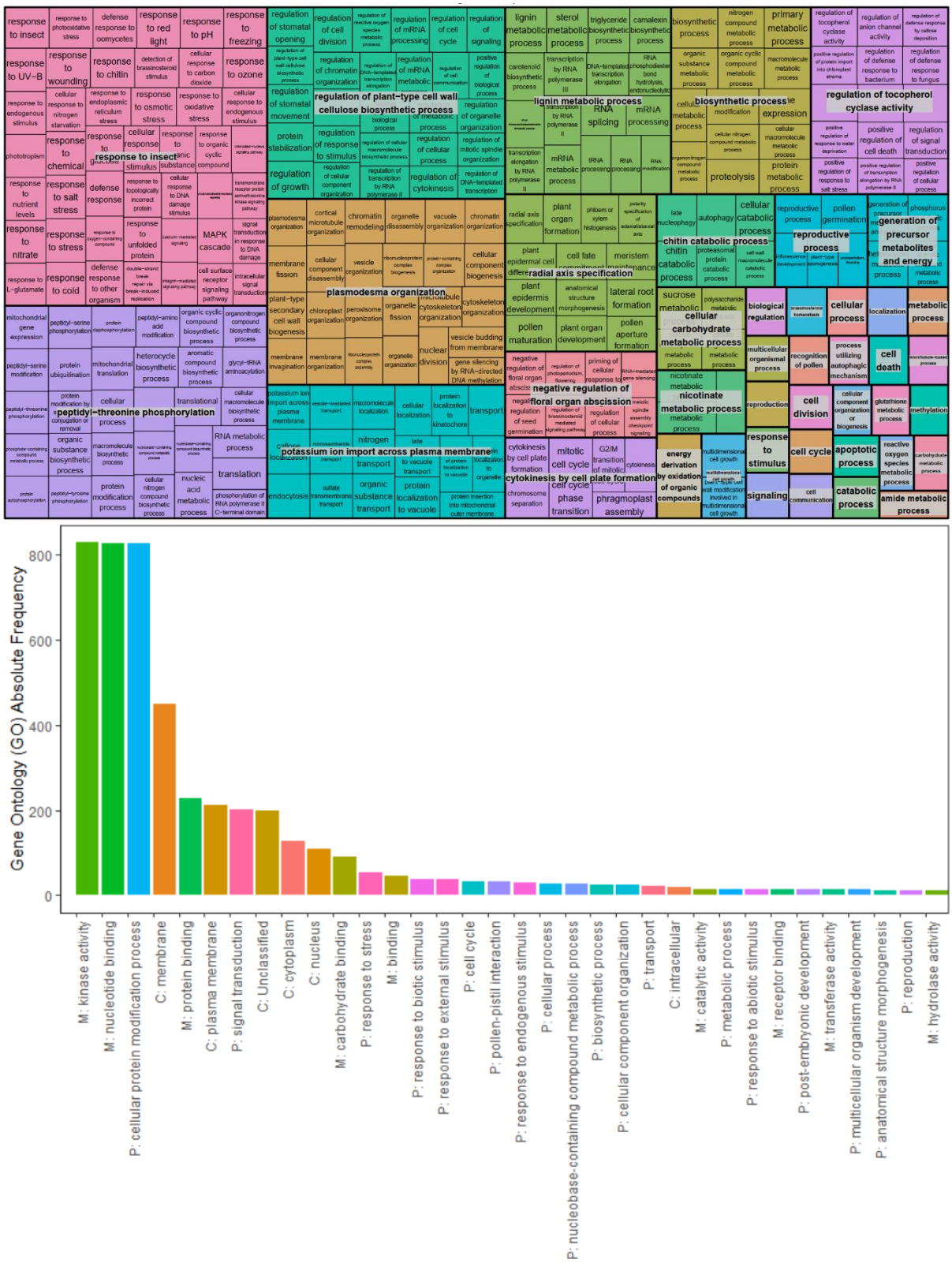

Figure 4. Gene Ontology of CsaPKs. A) Biological process. B) M: molecular function, P: biological process, C: cellular component.

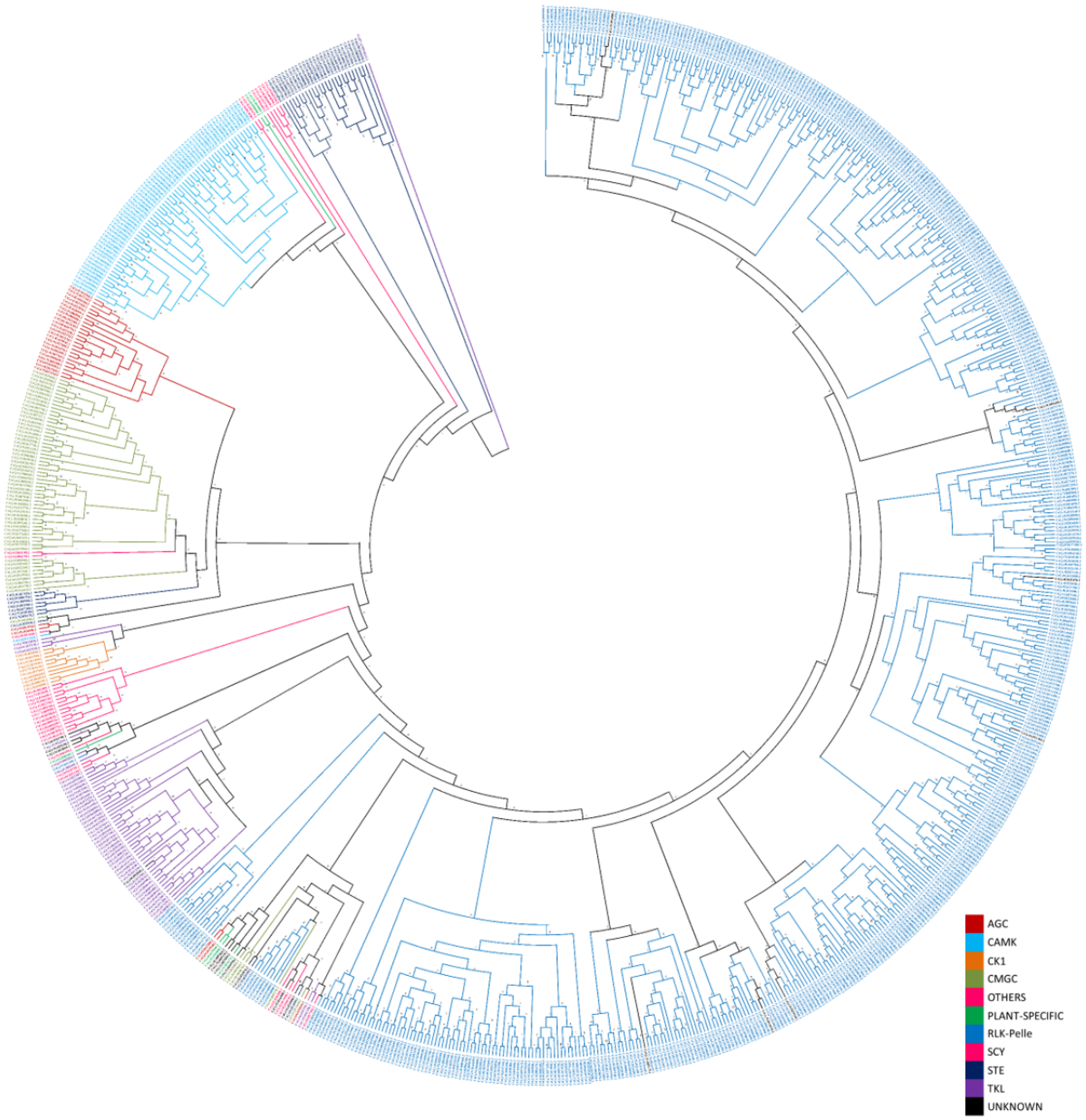

Figure 5. Phylogenetic analysis of *Cucumis sativus* (Csa) protein kinases (PKs). RLK-group (521 PKs in blue taxa) and non-RLK-group (310 PKs in other colours). Clade of the diferent PK families are represented in different colors.

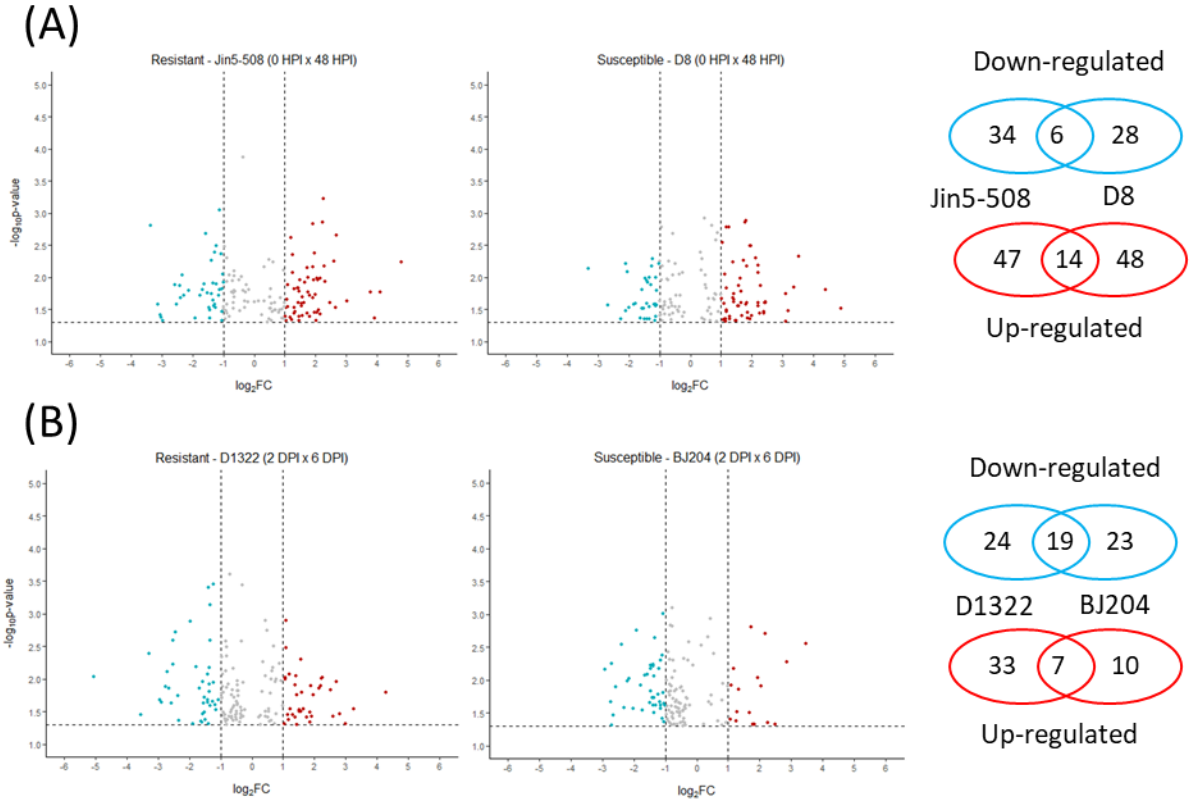

Figure 6. (A) Diffetential expression of kinase genes of the genotypes resistant (Jin5-508) and susceptible (D8) to Powdery mildew. (B) Diffetential expression of kinase genes of the genotypes resistant (D1322) and susceptible (BJ204) to Alternaria leaf spot.

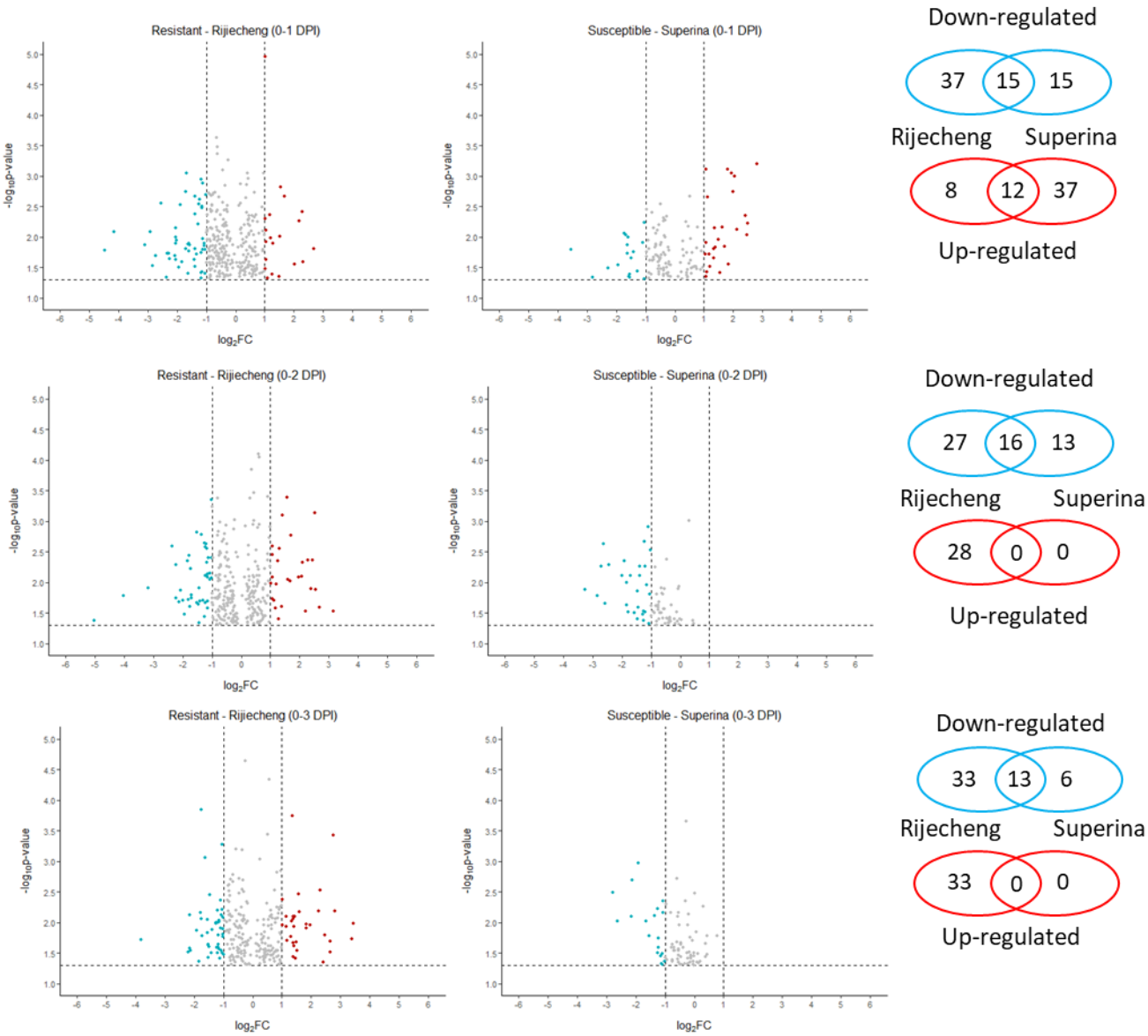

Figure 7. Diffetential expression of kinase genes of the genotypes resistant (Rijecheng) and susceptible (Superina) to *Meloidogyne incognita* at 1, 2 and 3 days post-innoculation (DPI).

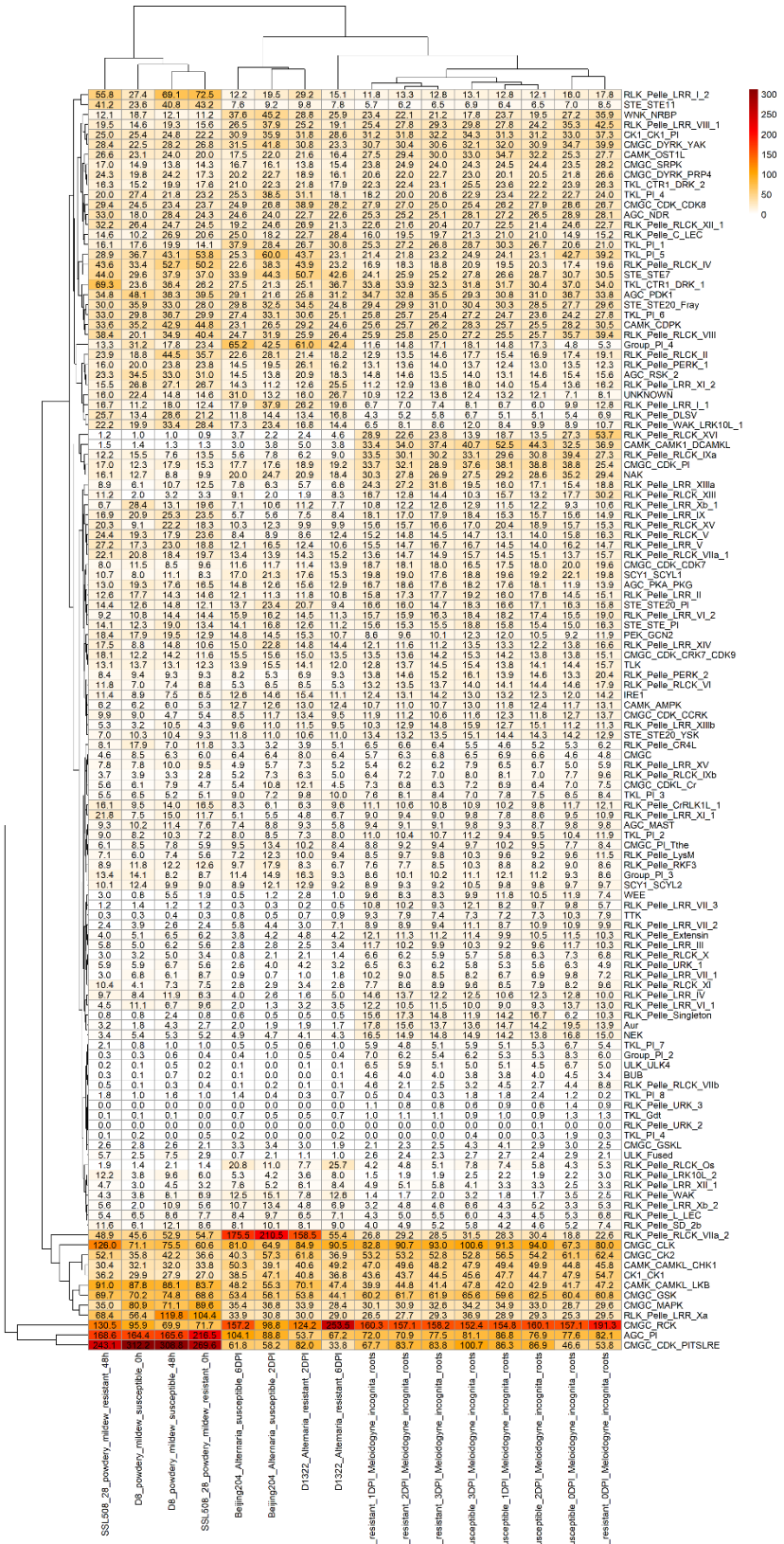

Figure 8. Expression Profile of *Cucumis sativus* kinase genes. Columns indicate different biotic stress conditions shown considering leaf and root tissues, and rows indicate the average of subfamilies.

Table 1. Novel indication of Csa PKs to subfamily not found in iTAK.

Kinase	Classification by phylogeny	Classification by iTAK
CsGy0G002040.1	TKL-PI-4	Absent
CsGy1G016380.2	RLK-Pelle_LRR-XII-1	Absent
CsGy1G024940.2	STE_STE20-YSK	Absent
CsGy2G023870.2	CMGC	Absent
CsGy3G005450.2	RLK-Pelle_LRR-VI-2	Absent
CsGy3G016500.1	SCY1_SCYL2	Absent
CsGy3G039175.1	RLK-Pelle_LRR-I-1	Absent
CsGy4G021070.2	RLK-Pelle_DLSV	Absent
CsGy5G014910.2	RLK-Pelle_RLCK-V	Absent
CsGy6G031350.2	RLK-Pelle_DLSV	Absent
CsGy1G025050.2	Unknown	Absent
CsGy2G009260.2	Unknown	Absent
CsGy3G020100.2	Unknown	Absent
CsGy3G028055.1	Unknown	Absent
CsGy3G029910.2	Unknown	Absent
CsGy3G044990.2	Unknown	Absent
CsGy6G033750.2	Unknown	Absent
CsGy7G009230.2	Unknown	Absent

Unknown: non grouped by the phylogeny by the 0.7 bootstrap threshold.