FRANCISCO CLEILSON LOPES COSTA

MULTIOMIC MOLECULAR INTEGRATION IN THE RESPONSE OF *Cucumis sativus* TO BIOTIC STRESS: A COMPREHENSIVE ANALYSIS OF THE EXPRESSION OF PROTEIN KINASES AND CIS-REGULATORY ELEMENTS

LAVRAS – MG

2024

FRANCISCO CLEILSON LOPES COSTA

MULTIOMIC MOLECULAR INTEGRATION IN THE RESPONSE OF *Cucumis sativus* TO BIOTIC STRESS: A COMPREHENSIVE ANALYSIS OF THE EXPRESSION OF PROTEIN KINASES AND CIS-REGULATORY ELEMENTS

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para obtenção do título de Doutor.

Prof. Dr. Welison Andrade Pereira
Orientador

LAVRAS – MG

2024

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Costa, Francisco Cleilson Lopes.
Multiomic molecular integration in the response of Cucumis sativus to biotic stress: a comprehensive analysis of the expression of protein kinases and cis-regulatory elements / Francisco Cleilson Lopes Costa. - 2024.
100 p. : il.

Orientador(a): Welison Andrade Pereira.

Tese (doutorado) - Universidade Federal de Lavras, 2024.
Bibliografia.

1. Genética molecular. 2. Resistência a doenças. 3. Interação planta-patógeno. I. Pereira, Welison Andrade. II. Título.

FRANCISCO CLEILSON LOPES COSTA

MULTIOMIC MOLECULAR INTEGRATION IN THE RESPONSE OF *Cucumis sativus* TO BIOTIC STRESS: A COMPREHENSIVE ANALYSIS OF THE EXPRESSION OF PROTEIN KINASES AND CIS-REGULATORY ELEMENTS

INTEGRAÇÃO MOLECULAR MULTIÔMICA NA RESPOSTA DE *Cucumis sativus* AO ESTRESSE BIÓTICO: UMA ANÁLISE ABRANGENTE DA EXPRESSÃO DE PROTEÍNAS QUINASES E ELEMENTOS CIS-REGULATÓRIOS

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para obtenção do título de Doutor.

APROVADA em 27 de março de 2024.

Dr. Welison Andrade Pereira	UFLA
Dr. Edson Mário de Andrade Silva	UFV
Dr. José Cleydson Ferreira da Silva	UW-Madison
Dr. Renato Coelho de Castro Vasconcellos	N/A
Dr. Victor Hugo Moura de Souza	UC

Prof. Dr. Welison Andrade Pereira
Orientador

LAVRAS – MG

2024

Aos meus pais e irmãos.
Aos meus familiares.

DEDICO

AGRADECIMENTOS

Agradeço a Deus pela vida, bondade e força concedidas, permitindo-me alcançar a finalização do doutorado.

Expresso minha gratidão à Universidade Federal de Lavras e à Pró-Reitoria de Pós-Graduação por possibilitarem a realização desta formação em nível de doutorado.

Um agradecimento especial ao professor Welison Andrade Pereira por seu compromisso em promover e orientar a pesquisa desta tese, pelos conselhos profissionais que seguirão comigo para a vida.

Ao professor Evandro Novaes, sou profundamente grato por me introduzir ao fascinante universo da Genômica e Bioinformática, uma área que tenho almejado desde o ensino médio.

Aos demais professores do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, reconheço o excelente trabalho desenvolvido junto ao Programa para formação de recursos humanos de excelência.

Aos colegas do grupo de estudo de Interação Planta-Patógeno: Elenildo, Laís, Maria Clara e Yasmin, agradeço a amizade, pelo tempo compartilhado e colaboração em pesquisas.

Ao Dr. Carlos Pulcinelli, ao Prof. José Maria Villela Pádua, ao colega Adilson Júnior da UFLA; ao Dr. Ivan Schuster, ao Me. Lander Oliveira e à Me. Alini Gardini da empresa LongPing High-Tech, expresso minha gratidão pela colaboração em outra pesquisa envolvendo mapeamento de caracteres relacionados à heterose e depressão por endogamia no milho, desenvolvida durante meu período de doutorado. Embora não tenha feito parte desta tese, será publicada posteriormente. Ao professor Adriano Bruzi, agradeço pela colaboração e pelo empréstimo de materiais e ferramentas utilizadas no trabalho com milho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) –Código de Financiamento 001.

Muito obrigado a todos!

"O conhecimento é o poder. A informação é libertadora. A educação é a premissa do progresso, em toda sociedade, em toda família."

Kofi Annan

RESUMO

As proteínas quinases (PKs) desempenham um papel fundamental na regulação de diversos processos metabólicos, o que justifica uma caracterização completa desta superfamília gênica. Associar os seus membros a respostas das plantas a estresses bióticos pode contribuir sobremaneira para elucidação das bases moleculares da interação planta patógeno. Este estudo teve por objetivos caracterizar a família PK do pepino; indicar sua distribuição genômica; evidenciar padrões de expressão em resposta a estímulos bióticos desencadeados por patógenos como Oídio (PM), Mancha de Alternaria (ALS) e Nematoide de Galhas (RKN), bem como, analisar a distribuição de cis-elementos na região potencialmente promotora dos genes diferencialmente expressos. A análise de modelos ocultos de Markov (HMMs) possibilitou classificar 835 PKs no Quinoma do pepino, distribuídas em seus sete cromossomos e categorizadas em 20 grupos distintos e 123 famílias, sendo o grupo RLK o mais abundante. Também foram observadas evidências de duplicação em tandem de genes PK, enriquecendo a compreensão da expansão desta superfamília gênica no pepino. Quanto aos perfis de expressão em resposta a PM, ALS e RKN, observou-se um número maior de genes PK expressos diferencialmente (PK DEGs) em genótipos suscetíveis quando submetidos a patógenos biotróficos (PM e RKN), ao contrário do que se observou com o patógeno necrotrófico ALS, onde maior número de PK DEGs foi encontrado no genótipo resistente. Estas descobertas contribuíram para a indicação de possíveis papéis das PKs do pepino na regulação de processos metabólicos, particularmente no contexto das interações planta-patógeno. Entende-se que esta caracterização abrangente abriu caminho para pesquisas subsequentes sobre os intrincados mecanismos subjacentes à expressão gênica, logo, às respostas do pepino a doenças importantes. Sob auxílio da plataforma PlantCARE, bem como da literatura, foram detectados 51,339 CREs (Elementos Cis-Regulatórios), os quais foram classificados em 7 grandes categorias e 125 tipos diferentes. A partir de uma análise posterior, foi possível indicar potenciais CREs que podem ser cruciais para a expressão gênica estimulada pelos estresses bióticos investigados neste estudo. Em conjunto, os resultados desta pesquisa possibilitaram indicar não apenas genes candidatos como também CREs que podem ser determinantes para etapas subsequentes que visam o desenvolvimento de genótipos resistentes a doenças específicas.

PALAVRAS-CHAVE: Genes quinase; regulação do metabolismo; imunidade vegetal; evolução de proteínas; resposta a doenças; redes responsivas a patógenos.

ABSTRACT

Protein kinases (PKs) play a key role in the regulation of several metabolic processes, which justifies a complete characterization of this gene superfamily. Associating its members to plant responses to biotic stresses can greatly contribute to the elucidation of the molecular basis of plant-pathogen interaction. The objectives of this study were to characterize the cucumber PK family; indicate its genomic distribution; to show expression patterns in response to biotic stimuli triggered by pathogens such as powdery mildew (PM), Alternaria leaf spot (ALS) and Root-knot Nematode (RNK), as well as to analyze the distribution of cis-elements in the potentially promoter region of differentially expressed genes. The analysis of hidden Markov models (HMMs) made it possible to classify 835 PKs in the cucumber quinoma, distributed in its seven chromosomes and categorized into 20 distinct groups and 123 families, with the RLK group being the most abundant. Evidence of tandem duplication of PK genes was also observed, enriching the understanding of the expansion of this gene superfamily in cucumber. Regarding the expression profiles in response to PM, ALS and RKN, a higher number of differentially expressed PK genes (PK DEGs) was observed in susceptible genotypes when submitted to biotrophic pathogens (PM and RKN), contrary to what was observed with the necrotrophic pathogen ALS, where a higher number of PK DEGs was found in the resistant genotype. These findings contributed to the indication of possible roles of cucumber PKs in the regulation of metabolic processes, particularly in the context of plant-pathogen interactions. It is understood that this comprehensive characterization paved the way for subsequent research into the intricate mechanisms underlying gene expression, hence cucumber's responses to important diseases. With the help of the PlantCARE platform, as well as the literature, 51,339 CREs (Cis-Regulatory Elements) were detected, which were classified into 7 major categories and 125 different types. From a further analysis, it was possible to indicate potential CREs that may be crucial for gene expression stimulated by the biotic stresses investigated in this study. Taken together, the results of this research made it possible to indicate not only candidate genes, but also CREs that can be determinant for subsequent steps aimed at the development of genotypes resistant to specific diseases.

KEYWORDS: kinase genes; metabolism regulation; plant immunity; protein evolution; disease response; pathogen-responsive networks.

INDICADORES DE IMPACTO

Os indicadores deste estudo são significativos para a biologia vegetal e genética, com implicações práticas e teóricas. A caracterização da superfamília de proteínas quinases em pepino contribui para entender os fundamentos moleculares da resposta biótica ao estresse, elucidando como as plantas detectam e respondem a patógenos, essencial para cultivares de pepino mais resistentes e agricultura sustentável. A análise dos padrões de expressão gênica e a identificação de elementos cis-regulatórios fornecem base para pesquisas futuras, facilitando estudos sobre genes de quinases e cis-elementos. A evidência de duplicação de genes quinase e sua distribuição em cromossomos oferece percepções sobre a evolução da superfamília quinase em pepino e outras plantas. O uso de bioinformática avançada, como Modelos Ocultos de Markov e PlantCARE, demonstra eficácia na caracterização genômica, impulsionando a biotecnologia agrícola. Essa pesquisa pode reduzir a dependência de pesticidas, custos de produção e promover práticas agrícolas sustentáveis, impactando positivamente a produtividade e a economia agrícola.

IMPACT INDICATORS

The indicators of this study are significant for plant biology and genetics, with practical and theoretical implications. The characterization of the protein kinase superfamily in cucumber contributes to understanding the molecular underpinnings of the biotic stress response, elucidating how plants detect and respond to pathogens, essential for more resilient cucumber cultivars and sustainable agriculture. The analysis of gene expression patterns and the identification of cis-regulatory elements provide a basis for future research, facilitating studies on kinase and cis-element genes. Evidence of gene kinase duplication and their distribution on chromosomes offers insights into the evolution of the kinase superfamily in cucumber and other plants. The use of advanced bioinformatics, such as Hidden Markov Models and PlantCARE, demonstrates efficacy in genomic characterization, boosting agricultural biotechnology. This research can reduce pesticide dependency, production costs, and promote sustainable farming practices, positively impacting agricultural productivity and economics.

SUMÁRIO

PART ONE

1 INTRODUCTION

The cucumber (*Cucumis sativus* L.) is a vegetable crop that belongs to the Cucurbitaceae family and has the primary center of origin and domestication in India about 3000 years ago (Candolle, 1886, Whitaker et al., 1962, Pandey et al., 2022). Cucumber has importance in several sectors, with applications in the economy, such as immature fruits for consumption in salads, sandwiches or pickles, input for the cosmetics industry (soaps, facial creams, shampoos), in addition to having components with nutritional and medicinal applications, as a source of vitamin C (2 mg/100 g), protein (0.4%), iron (1.5 mg/100 g), potassium (50-80 mg/100 g) (Pandey et al., 2022).

Powdery Mildew can be caused by a wide range of ascomycete fungi, but in Brazil the disease is caused by *Podosphaera xanthii* (Castagne) U. Braun & Shishkoff 2000, it is the most common disease in cucurbits and for cucumbers, especially when cultivated in a protected environment with high humidity and moderate temperature. Symptoms begin in the leaves from the development of a white, powdery spots that grows covering up to 100% of the leaf surface, that can reach the stem. With the development of the infection, the plant suffers loss of photosynthetically active area, which reflects in loss of productivity and fruit quality (Sharma et al., 2016).

In a review by Chen et al. (2020), studies on *C. sativus* resistance to PM have revealed that various genes with diverse genetic inheritance patterns are involved in this resistance. For example, it was observed that major and minor effect genes exist in the PI 197087 variety (Barnes; Epps, 1956), multiple recessive genes in the Puerto Rico 37 variety (Smith, 1984), and a major recessive gene, a dominant gene (R), and a dominant suppressor gene (I) in the P1212233 and P123514 varieties (Shanmugasundaram et al., 1971, 1972). Additionally, QTLs are related to PM resistance, such as pm5.2 located on chromosome 5 with four identified resistance genes (Zhang et al., 2011). Furthermore, other PM resistance QTLs are directly associated with Downy Mildew (DM) resistance, being colocalized in the PI 197088 variety (Wang et al., 2018), and seven pairs of cosegregated QTLs for PM and DM in the CS-PMR1 variety (Sakata et al., 2006; Fukino et al., 2013; Yoshioka et al., 2014). Thus, PM resistance in *C. sativus* is a complex trait involving the participation of different genes with major and minor effects, as well as a variable number of quantitative loci, conditions that vary according to the variety studied and the approach employed. Currently, the main gene families conferring resistance to PM are Mlo (Jørgensen, 1992), PMR (Schouten et al., 2014), and TCTP (Fan et

al., 2014). The Mlo genes exhibit unique characteristics for several reasons (Jørgensen, 1992). Firstly, they do not conform to the gene-for-gene system but rather confer broader resistance. The recessive Mlo loci are non-complementary even though they are located in the same chromosomal region and within a single gene locus. This means that the Mlo gene that has lost its function will become recessive and cannot complement the dominant allele to restore the phenotype. Additionally, small quantitative differences may be observed, yet the same resistance phenotype will be observed. It is a highly robust and durable form of resistance, as resistance against all pathogen isolates is observed. Finally, large cell wall appositions occur at the sites of contact with the pathogen to block its penetration for this resistance to take place. PMR (Powdery Mildew Resistance) genes have a direct relationship with powdery mildew resistance in cucumber plants (Schouten et al., 2014). Several Mlo-like genes were identified in cucumber, including three genes that cluster in Clade V, the clade that contains all known Mlo-like genes recognized as susceptibility genes for powdery mildew in other dicotyledons. Additionally, homologs of the susceptibility genes PMR4 and PMR5 were identified. These PMR genes and their homologs play a role in plant susceptibility to powdery mildew and exhibit a mechanism similar to that of Mlo genes. When these genes are present and functional, the plant is susceptible to powdery mildew infection. However, when these genes are inactivated or do not function correctly, the plant can become resistant to powdery mildew. The CsTCTP1 and CsTCTP2 genes, which encode translationally controlled tumor proteins, are involved in the defense response of *C. sativus* to *S. fuliginea* (syn. *P. xanthii*) (Meng et al., 2018). These genes have been identified as negative modulators in the plant's defense response. The analysis showed that the CsTCTP1 and CsTCTP2 proteins are located in the cytoplasm of plant cells. The expression of these genes is correlated with the level of cucumber resistance to powdery mildew. When the CsTCTP1 or CsTCTP2 genes are overexpressed in young cucumber leaves, resistance to *S. fuliginea* is impaired. On the other hand, suppressing the expression of these genes increases cucumber resistance to powdery mildew. The research also investigated the relationship between defense genes and genes related to the abscisic acid (ABA) signaling pathway and the target of rapamycin (TOR) pathway in cells, with the overexpression and suppression of CsTCTP1/CsTCTP2 genes in uninfected cucumber plants (Meng et al., 2018). It was found that CsTCTP1 participates in the defense response by regulating the expression of defense-associated genes and/or genes associated with the ABA signaling pathway (Meng et al., 2018). Meanwhile, CsTCTP2 participates in the defense response by regulating the expression of genes associated with the TOR signaling pathway (Meng et al., 2018).

In addition, another very common disease in cucumber is Alternaria Leaf Spot caused by *Alternaria cucumerina* (Ellis & Everh.) J.A. Elliott 1917 and *Alternaria alternata* (Fr.) Keissl. 1912. Symptoms begin as small circular lesions 1-2 mm in diameter on the upper leaf surface of older leaves or crown leaves, which may be surrounded by a yellow halo. These lesions can grow and fuse to form large lesions larger than 10 mm in diameter. As the disease progresses, plants can experience severe leaf loss that results in loss of productivity, reduced fruit quality and size (Sharma et al., 2016).

In a study of expression patterns of differentially expressed genes (DEGs) revealed important insights into cucumber plant responses to stresses caused by *A. cucumerina* infection in two distinct genotypes, D1322 and BJ204 (Sa et al., 2020). These authors emphasized genes related to phenylalanine, which are precursors to phenolic compounds and lignin responsible for cell wall thickening as a resistance mechanism; genes involved in glutathione metabolism; and genes associated with porphyrin and chlorophyll metabolism and photosynthetic antenna proteins. The DEGs associated with phenylalanine metabolism showed a predominant upregulation after infection, especially in D1322, indicating a significant activation of this pathway (Sa et al., 2020). There was higher expression of genes encoding the enzyme phenylalanine ammonia-lyase (PAL) in D1322 compared to BJ204, suggesting a more robust response in this genotype. Additionally, genes encoding 4-coumarate-CoA ligase (4CL), essential for lignin biosynthesis, also showed differential expression between the genotypes (Sa et al., 2020). In terms of glutathione metabolism, both genotypes exhibited a similar number of DEGs but with distinct expression patterns. While most glutathione S-transferase (GST) genes were upregulated in both genotypes, the magnitude of this regulation was more pronounced in BJ204 (Sa et al., 2020). Moreover, genes exclusively upregulated in one of the genotypes were identified, suggesting host-specific responses to infection (Sa et al., 2020). Regarding photosynthetic performance, DEGs associated with various photosynthetic pathways were predominantly downregulated in BJ204, indicating significant suppression of photosynthesis following *A. cucumerina* infection in this genotype (Sa et al., 2020). Finally, the analysis of transcription factors (TFs) revealed differential regulation of several TFs in response to infection, particularly within the MYB, NAC, and WRKY families (Sa et al., 2020). The overlap of regulated TFs between the genotypes indicated both similarities and differences in the gene regulatory networks activated by *A. cucumerina* (Sa et al., 2020).

Another pathogen of economic importance is *Meloidogyne incognita* (Kofold & White) Chitwood, the causal agent of the Root-Knot Nematode (RKN). There are several species of

Root-Knot Nematode, but *M. incognita* has been recognized as the most harmful single pathogen in the world (Trudgill; Blok, 2001, Subedi et al., 2020). RKN causes reduced growth, quality, and yield, along with reduced host resistance against biotic and abiotic stresses (Subedi et al., 2020). The damage begins after the female nematode penetrates the roots and induces the formation of galls, which are nodular-shaped structures that prevent the absorption of water and nutrients by the plant (Castagnone-Sereno; Danchin, 2014). Galls are characterized by the abnormal growth of individual root cells, which have their metabolism directed to feed the nematode. This leads to reduced root growth, yellowing of leaves, wilting, and overall weakening of the plant.

The gene families associated with resistance to root-knot nematodes in cucumbers, as identified by Wang et al. (2018), include: (i) Pathogenesis-Related Proteins (PRs): Specifically, the lipid transfer protein (LTP) Csa6G410090 was highly expressed and may be the main regulator of this response. (ii) Cell Cycle-Related Genes: The F-box Skp2-like domain genes were suppressed in the resistant line IL10–1, which may be related to the abnormal development of giant cells and nematode resistance. (iii) Auxin-Related Genes: Many auxin-related genes were significantly repressed in the resistant line IL10–1, corresponding to a lower level of indole-3-acetic acid (IAA) in the roots compared to the susceptible line CC3. These gene families play important roles in the abnormal development of giant cells, which hinders the development of the nematode *M. incognita* and consequently causes nematode resistance in the IL10–1 line.

Plant diseases caused by pathogens such as Powdery Mildew (*P. xanthii*), Alternaria Leaf Spot (*A. cucumerina*), and Root-Knot Nematode (*M. incognita*) result in significant agricultural production losses. Identifying and exploiting resistance mechanisms are crucial for developing resistant cultivars, ensuring agricultural sustainability. As previously described, many gene families have been studied and detailed in the context of resistance to these diseases. However, other gene families with fundamental roles in resistance still need to be explored.

Kinases, a class of enzymes facilitating the transfer of phosphate groups to proteins, are pivotal in cellular communication and managing plant defense mechanisms. These enzymes, functioning as molecular toggles in various biological activities, orchestrate the activation of transcription factors that govern gene expression (Meng; Zhang, 2013; Seo et al., 1995; Zhang; Klessig, 2001). Among the key mechanisms by which kinases contribute to resistance are (i) Pathogen Recognition, (ii) Defense Activation, and (iii) Gene Expression Regulation, along with the activity of Transcription Factors. When a pathogen, such as *P. xanthii*, invades the

plant, Receptor-Like Kinases (RLKs) detect pathogen-associated molecules, triggering defensive signaling cascades (He; Wu, 2016; Morris; Walker, 2003). Additionally, kinases are closely involved in the transduction of defense signals by Mitogen-Activated Protein Kinase (MAPKs). Following pathogen recognition, MAPKs transmit signals downstream RLKs that stimulate various defense responses, such as the generation of reactive oxygen species (ROS), activation of defense gene expression, and strengthening of the cell wall (Meng; Zhang, 2013). Specific kinases, such as Cysteine-Rich Receptor-Like Kinases (CRKs), are activated to modulate the immune response, thereby enhancing the plant's ability to resist infection (Zhang et al., 2023).

Thus, in the context of plant breeding, the use of resistant cultivars is the most economical and efficient form of control, requiring knowledge of resistance genes and their functioning mechanisms. By studying disease resistance, understanding the intricate mechanisms governing plant responses to diseases is of paramount importance. The research in question sought to investigate and evaluate the complete set of cucumber protein kinases (PKs), aiming to understand their behavior in the context of regulating plant metabolism, through structural, evolutionary and expression analyzes in response to three biotic stresses: Powdery Mildew (*P. xanthii*), Alternaria Leaf Spot (*A. cucumerina*) and Root-Knot Nematode (*M. incognita*). With the escalating challenges posed by such pathogens, there is an urgent need to elucidate the underlying genetic and regulatory networks orchestrating plant defense mechanisms.

The research was divided into two articles, the first addressing the cataloging of PKs and the second on the cis-regulatory elements (CREs) of the kinase genes involved in the defense response against these diseases. In the first study, a comprehensive exploration of the cucumber kinome was carried out, focusing on PKs, crucial components in metabolic regulation and signal transduction pathways. Despite their critical roles, the full spectrum of cucumber PKs and their responses to biotic stressors remained largely unknown. Through sophisticated analyzes involving hidden Markov models (HMMs), the research discovered 835 PKs intricately distributed throughout the cucumber genome.

In the second study, building on this initial work, an investigation was carried out into the regulatory landscape governing the expression of disease-responsive kinase genes in cucumber. Recognizing the importance of CREs in modulating gene expression, research has focused on identifying and analyzing these elements within the potential promoter regions (PPRs) of DEGs of the PK superfamily.

Taken together, these two studies represent significant indicators in deciphering the molecular complexities underlying cucumber responses to biotic stress. By elucidating the cucumber kinome and revealing the regulatory elements that govern disease-responsive kinase genes, these efforts provide valuable insights for the development of resilient cucumber cultivars and promise to contribute to global food security in the face of growing agricultural challenges.

2 CONCLUSION

The research presented in these two studies provides a comprehensive understanding of the cucumber kinome and its regulatory elements in response to biotic stress. By cataloging 835 protein kinases (PKs) and investigating their cis-regulatory elements (CREs), the studies offer knowledge into the molecular mechanisms underlying cucumber resistance to Powdery Mildew (*P. xanthii*), Alternaria Leaf Spot (*A. cucumerina*), and Root-Knot Nematode (*M. incognita*). The identification of crucial PKs and CREs advances our knowledge of plant defense responses, facilitating the development of resistant cucumber cultivars. These findings enhance our understanding of cucumber genetics and contribute to broader efforts in agricultural sustainability and global food security amidst escalating pathogen challenges.

REFERENCES

BARNES, W.; EPPS WM. Powdery mildew resistance in South Carolina cucumbers. **Plant Dis Rep**, v. 40, n. 1093, 1956.

CANDOLLE, A. Plants Cultivated for Their Seeds. In: **Origin of Cultivated Plants**. 2. ed. London: Kegan Paul Trench & Co., 1886. XLIX p. 376–384. 1886.

CASTAGNONE-SERENO, P.; DANCHIN, E. G. J. Parasitic success without sex – the nematode experience. **Journal of Evolutionary Biology**, v. 27, n. 7, p. 1323–1333, 15 jul. 2014. doi: 10.1111/jeb.12337.

CHEN, Q.; YU, G.; WANG, X.; MENG, X.; LV, C. Genetics and Resistance Mechanism of the Cucumber (Cucumis sativus L.) Against Powdery Mildew. **Journal of Plant Growth Regulation**, v. 40, n. 1, p. 147–153, 17 fev. 2021. doi: 10.1007/s00344-020-10075-7.

FAN, H.; REN, L.; MENG, X.; SONG, T.; MENG, K.; YU, Y. Proteome-level investigation of Cucumis sativus-derived resistance to Sphaerotheca fuliginea. **Acta Physiologiae Plantarum**, v. 36, n. 7, p. 1781–1791, 24 jul. 2014. doi: 10.1007/s11738-014-1552-6.

FUKINO, N.; YOSHIOKA, Y.; SUGIYAMA, M.; SAKATA, Y.; MATSUMOTO, S. Identification and validation of powdery mildew (Podosphaera xanthii)-resistant loci in recombinant inbred lines of cucumber (Cucumis sativus L.). **Molecular Breeding**, v. 32, n. 2, p. 267–277, 20 ago. 2013. doi: 10.1007/s11032-013-9867-3.

HE, K.; WU, Y. Receptor-Like Kinases and Regulation of Plant Innate Immunity. Em: The Enzymes. p. 105–142. 2016.

JØRGENSEN, I. H. Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. **Euphytica**, v. 63, n. 1–2, p. 141–152, jan. 1992. doi: 10.1007/BF00023919.

MENG, X.; YU, Y.; ZHAO, J.; CUI, N.; SONG, T.; YANG, Y.; FAN, H. The Two Translationally Controlled Tumor Protein Genes, CsTCTP1 and CsTCTP2, Are Negative Modulators in the Cucumis sativus Defense Response to Sphaerotheca fuliginea. **Frontiers in Plant Science**, v. 9, 25 abr. 2018. doi: 10.3389/fpls.2018.00544.

MENG, X.; ZHANG, S. MAPK Cascades in Plant Disease Resistance Signaling. **Annual Review of Phytopathology**, v. 51, n. 1, p. 245–266, 4 ago. 2013. doi: 10.1146/annurev-phyto-082712-102314.

MORRIS, E. R.; WALKER, J. C. Receptor-like protein kinases: the keys to response. **Current Opinion in Plant Biology**, v. 6, n. 4, p. 339–342, ago. 2003. doi: 10.1016/S1369-5266(03)00055-4.

PANDEY, S.; KUJUR, S. N. Importance, Distribution, Botany and Genetics. Em: [s.l: s.n.] p. 1–13. 2022.

SA, R.; LIU, D.; CHEN, L.; LIU, L.; ZHANG, Y.; ZHANG, X.; XU, C. Transcriptome analysis of mechanisms and candidate genes associated with cucumber response to cucumber

alternaria leaf spot infection. **Physiological and Molecular Plant Pathology**, v. 111, p. 101490, ago. 2020. doi: 10.1016/j.pmpp.2020.101490.

SAKATA, Y.; KUBO, N.; MORISHITA, M.; KITADANI, E.; SUGIYAMA, M.; HIRAI, M. QTL analysis of powdery mildew resistance in cucumber (Cucumis sativus L.). **Theoretical and Applied Genetics**, v. 112, n. 2, p. 243–250, 21 jan. 2006. doi: 10.1007/s00122-005-0121-1.

SCHOUTEN, H. J.; KRAUSKOPF, J.; VISSER, R. G. F.; BAI, Y. Identification of candidate genes required for susceptibility to powdery or downy mildew in cucumber. **Euphytica**, v. 200, n. 3, p. 475–486, 9 dez. 2014. doi: 10.1007/s10681-014-1216-z.

SEO, S.; OKAMOTO, M.; SETO, H.; ISHIZUKA, K.; SANO, H.; OHASHI, Y. Tobacco MAP Kinase: A Possible Mediator in Wound Signal Transduction Pathways. **Science**, v. 270, n. 5244, p. 1988–1992, 22 dez. 1995. doi: 10.1126/science.270.5244.1988.

SHANMUGASUNDARAM, S.; WILLIAMS, P.; PETERSON, C. Inheritance of resistance to powdery mildew in cucumber. **Phytopathology** , v. 61, p. 1218–1221, 1971.

SHANMUGASUNDARAM, S.; WILLIAMS, P.; PETERSON, C. A recessive cotyledon marker gene in cucumber with pleiotropic effects. **HortScience** , v. 7, p. 555–556, 1972.

SHARMA, A.; KATOCH, V.; RANA, C. Important Diseases of Cucurbitaceous Crops and Their Management. Em: MOHAMMAD PESSARAKLI (Ed.). **Handbook of Cucurbits: Growth, Cultural Pratices, and Physiology**. Boca Raton London New York: CRC Tress, 2016. p. 301–323. 2016.

SMITH, P. Powdery mildew resistance in cucumber. **Phytopathology**, v. 39, p. 1027–1028, 1948.

SUBEDI, S.; THAPA, B.; SHRESTHA, J. Root-knot nematode (Meloidogyne incognita) and its management: a review. **Journal of Agriculture and Natural Resources**, v. 3, n. 2, p. 21–31, 30 out. 2020. doi: 10.3126/janr.v3i2.32298.

TRUDGILL, D. L.; BLOK, V. C. A. Apomitic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. **Annual Review of Phytopathology**, v. 39, n. 1, p. 53–77, set. 2001. doi: 10.1146/annurev.phyto.39.1.53.

WANG, Y.; VANDENLANGENBERG, K.; WEN, C.; WEHNER, T. C.; WENG, Y. QTL mapping of downy and powdery mildew resistances in PI 197088 cucumber with genotyping-by-sequencing in RIL population. **Theoretical and Applied Genetics**, v. 131, n. 3, p. 597–611, 20 mar. 2018. doi: 10.1007/s00122-017-3022-1.

WHITAKER, T. W.; DAVIS, G. N. **Cucurbits: Botany, Cultivation, and Utilization**. New York: Interscience Publishers Inc., 1962.

YOSHIOKA, Y.; SAKATA, Y.; SUGIYAMA, M.; FUKINO, N. Identification of quantitative trait loci for downy mildew resistance in cucumber (Cucumis sativus L.). **Euphytica**, v. 198, n. 2, p. 265–276, 20 jul. 2014. doi: 10.1007/s10681-014-1102-8.

ZHANG, S.; KLESSIG, D. F. MAPK cascades in plant defense signaling. **Trends in Plant Science**, v. 6, n. 11, p. 520–527, nov. 2001. doi: 10.1016/S1360-1385(01)02103-3.

ZHANG, S.; LIU, M.; MIAO, H.; ZHANG, S.; YANG, Y.; XIE, B.; GU, X. QTL mapping of resistance genes to powdery mildew in cucumber (Cucumis sativus L.). **Scientia Agricultura Sinica**, v. 44, p. 3584–3593, 2011.

PART TWO - ARTICLES

ARTICLE 1 - THE CUCUMBER KINOME: UNCOVERING AND ANALYZING THE EXPRESSION OF C*ucumis sativus* PROTEIN KINASES IN RESPONSE TO BIOTIC STRESS

(References were formatted according to the standards of the Scientia Horticulturae journal)

ABSTRACT

Protein kinases (PKs) play a pivotal role in the regulation of diverse metabolic processes. However, a complete characterization of the PK family in cucumber remains lacking. This study aimed to characterize the cucumber PK family, shedding light on its genomic distribution, classification, and expression patterns in response to biotic stimuli triggered by pathogens like Powdery Mildew (PM), Alternaria Leaf Spot (ALS), and Root-Knot Nematode (RKN). The analysis of hidden Markov models (HMMs) uncovered 835 PKs in the cucumber kinome, distributed across its seven chromosomes, and categorized into 20 distinct groups and 123 families, with the RLK group being the most abundant. Evidence of tandem duplication of PK genes was also observed, enriching our understanding of cucumber PKs. The study delved into the functional characterization of cucumber PKs, scrutinizing their expression profiles in response to PM, ALS, and RKN. Notably, we observed a higher number of differentially expressed PK genes (PK DEGs) in susceptible genotypes when facing biotrophic pathogens (PM and RKN), while the greatest number of PK DEGs was found in the resistant genotype combating the necrotrophic pathogen ALS. We identified PK DEGs per genotype specific to each pathosystem. The number of DEGs ranged from 319 to 2,202, with PK DEGs numbering from 8 to 105. Our findings illuminated the role of cucumber PKs in regulating metabolic processes, particularly in the context of plant-pathogen interactions. This comprehensive characterization paves the way for further research into the intricate mechanisms underlying cucumber's responses to these major diseases.

KEYWORDS: Gene expression; Kinase gene family; Metabolism regulation; Plant immunity; Protein evolution; Response to disease.

1 INTRODUCTION

Cucumber (*Cucumis sativus* L.) (Csa) is a species that belongs to the Cucurbitaceae family, holding significant importance for several industries worldwide (Pandey et al., 2022). This crop holds a prominent position in both large-scale and small-scale agricultural systems, making a substantial contribution to the economy. The demand for cucumbers, whether in their fresh or processed form, drives agricultural production and strengthens the development of international trade alliances, guaranteeing year-round availability. Cucumbers are essential to the food industry due to their nutritional potential and medicinal attributes, particularly their high potassium content, which is known to contribute to the alleviation of blood pressure (Jilani et al., 2009, Pandey et al., 2022). Given their applicability, the cultivation of cucumbers has emerged as an economically viable alternative for different market sectors, including those focused on the production of plant-derived pharmaceuticals (Khan et al., 2022). Furthermore, within the beauty and skincare industry, cucumbers are recognized for their moisturizing, refreshing, and soothing properties (Desam et al., 2021).

Despite its agronomic relevance, cucumber's susceptibility to diseases and pests results in reduced foliage, consequently leading to decreased productivity (Batta, 2003; Mukhtar et al., 2013; Sarhan et al., 2020). For instance, powdery mildew (PM), caused by the fungal pathogens such as *Podosphaera xanthii* (Castagne) U. Braun & Shishkoff (syn. *Sphaerotheca fuliginea* (Schlechtend.:Fr.) Polacci) and *Erysiphe cichoracearum* DC. (syn. *Golovinomyces cichoracearum*) (Abul-Hayja, 1982, Moustafa et al., 1990, Pérez-García et al., 2009), represents a significant threat. Alternaria leaf spot (ALS), caused by the fungus *Alternaria alternata* (Fr.) Keissl. (Vakalounakis, 1990; Vakalounakis et al., 1988) and *Alternaria cucumerina* (Ell. & Ev.) (Batta, 2003; Dey et al., 2022), manifests as dark circular lesions on the leaves, which can expand and result in premature defoliation. The Root-knot nematode (RKN) or gall nematode, specifically *Meloidogyne incognita* (Kofold & White) Chitwood, is a notorious pest that infests cucurbit roots, leading to the development of characteristic galls and impairing nutrient absorption (Cheng et al., 2022; Fassuliotis et al., 1963).

The use of resistant cultivars represents the most efficient approach to disease management, effectively reducing the symptom severity and avoiding the need for chemical control measures (Choudhary et al., 2002). Comprehending the defense mechanisms employed by Csa species is essential for the development of resistant cultivars, providing foundational knowledge to support the development of new plant biotechnology events. In this context,

protein kinases (PKs) (Hanks et al., 1995) receive great attention due to their central role in regulating plant responses to disease defense processes, as well as in modulating plant immunity (Benschop et al., 2007; Peck, 2001; Tena et al., 2011). However, the genetic control of resistance mechanisms and the intricate molecular interactions between plants and pathogens remain critical research targets for elucidating the molecular mechanisms underlying plant immunity (Kamoun et al., 1999; Kliebenstein et al., 2008; Vidhyasekaran, 2014).

No comprehensive or comparative studies have been identified in the existing literature that encompass all families of PKs in Csa, particularly in the context of PM, ALS, and RKN interactions. Previous investigations have characterized the cucumber kinome as relatively small when compared to kinomes of other plant species, which is possibly attributed to the absence of recent whole-genome duplications in Csa (Huang et al., 2009; Lehti-Shiu et al., 2012). It is worth noting that several genomic analyses have already been conducted focusing on specific PK families within this species. To illustrate, the Casein Kinase (CK) family, which includes proteins responsible for phosphorylating casein (Meggio et al., 2003), plays roles related to the circadian cycle (Sugano et al., 1999), plant growth, and the regulation of light-modulated gene expression (Lee et al., 1999). On the other hand, the Mitogen-Activated Protein Kinase (MAPK) family constitutes a group of serine/threonine kinases that participate in diverse signal transduction pathways associated with hormonal responses and substantial developmental changes in organisms (Shoresh et al., 2006).

Conversely, the Lectin Receptor-Like Kinase (LecRLK) family emerges as notably crucial in the innate immunity of plants, especially in defense against biotrophic pathogens such as *P. xanthii*, the causal agent of PM (Haider et al., 2021). Furthermore, the Calcium-Dependent Protein Kinase (CDPK) and CDPK-Related Protein Kinase (CRK) families play roles in calcium-dependent signaling pathways in response to environmental stresses (Wei et al., 2019; Xu et al., 2017). Moreover, these kinase families are involved in pathways related to the inhibition of J2 (second-stage juvenile) invasion of *M. incognita* on cucumber roots, thereby enhancing cultivar resistance (Li et al., 2021). In general, although a comprehensive characterization of the Csa kinome remains limited, specific PK families have received attention in genomic analyses, contributing significantly to our understanding of their functions and potential involvement in various biological processes and stress responses.

PKs play a key role in modulating cellular responses to external stimuli (Peck, 2001; Tena et al., 2011), highlighting their importance in enhancing cucumber genetic improvement.

Primarily, PKs function by facilitating phosphorylation, a critical biological process that enables signaling within gene expression networks in various cellular processes (Johnson et al., 1996). Thus, the comprehensive characterization of a kinome and its analysis holds substantial importance, particularly in the context of plant immunity. Moreover, the identification of PKs offers valuable insights into elucidating their specific functionalities and gives support to the elucidation of the evolutionary processes shaped by environmental pressures within different tissues of the organism.

The objective of this study was to define and evaluate the cucumber kinome, based on structural and evolutionary genomic assessments of the genes and their corresponding encoded proteins. We conducted a comprehensive analysis of cucumber PKs comprising the characterization, classification, and the gene expression evaluation in the context of stress caused by PM, ALS, and RKN. The inferences performed are intended to serve as foundational knowledge for future projects aimed at investigating the functional roles of specific PKs and developing strategies to enhance disease resistance in cucumber and related species of the Cucurbitaceae family.

2 MATERIAL AND METHODS

2.1 Genome-wide identification and classification of cucumber PKs

The identification of PKs was based on the alignment of the Pkinase (PF00069) and Pkinase_Tyr (PF07714) subfamilies against the cucumber Gy14 annotated proteins, available in the Cucurbit Genomics Database v2 (CuGenDbv2) (Yu et al., 2022). Such an alignment was performed with hidden Markov models (HMMs) obtained from the Pfam database (http://pfam.xfam.org/) (El-Gebali et al., 2019) together with the HMMER tool (Finn et al., 2011) considering an E-value cutoff of 0.01. The isoforms of the PKs were included for further analyses.

After identifying the cucumber PKs, we classified them into subfamilies using the HMMER tool together with the subfamily HMMs estimated from PKs of 25 plant species, as described and available as supplementary files by Lehti-Shiu et al. (2012): *Aquilegia coerulea* E. James*, Arabidopsis lyrate* (L.) O'Kane & Al-Shehbaz*, Arabidopsis thaliana* (L.) Heynh.*, Brachypodium distachyon* (L.) P.Beauv.*, Carica papaya* L.*, Citrus clementina* Hort.*, Citrus sinensis* L.*, Chlamydomonas reinhardtii* P.A. Dang.*, Cucumis sativus* L.*, Eucalyptus grandis* W.Hill*, Glycine max* (L.) Merr. *, Manihot esculenta* Crantz*, Medicago truncatula* Gaertn.*,*

Mimulus guttatus (DC.) G.L.Nesom*, Oryza sativa* L.*, Populus trichocarpa* Torr. & A.Gray ex Hook.*, Prunus persica* (L.) Stokes*, Physcomitrella patens* (Hedw.) Mitt.*, Ricinus communis* L.*, Selaginella moellendorffii* Hieron.*, Setaria itálica* (L.) P.Beauv.*, Sorghum bicolor* (L.) Moench.*, Vitis vinifera* L.*, Volvox carteri* F. Stein*,* and *Zea mays* L..

To confirm the subfamily classification obtained, we estimated a phylogenetic tree of PK sequences. PKs were aligned using the MAFFT v7.453 program (Katoh et al., 2013) and the tree was estimated with the FastTree 2.1.11 software (Price et al. 2010) considering 1,000 bootstrap replicates and an approximately-maximum-likelihood approach together with the Whelan Goldman (WAG) model for amino acid evolution (Whelan et al., 2001). Additionally, we included a PK outgroup sequence as a root in the phylogenetic tree. We selected a sequence from the *Chlamydomonas reinhardtii* v5.6 proteome (Cre07.g349540.t1.1) retrieved from Phytozome (https://phytozome-next.jgi.doe.gov) (Goodstein et al., 2012). We selected such a species because it does not have a multicellular ancestor (Ratcliff et al., 2013), which indicates that it must have evolved prior to Csa proteins.

2.2 Characterization of the cucumber PK sequences

The evaluation of cucumber PKs' gene structure was performed considering the number of exons and introns retrieved from the Gy14 GFF file, obtained from the CuGenDBv2. The chromosomal positions of PK genes were also obtained from the Gy14 GFF file, and a physical map of the chromosomes was constructed using the TBtools software (Chen et al., 2020). The domain composition analysis of PKs was performed using the Pfam database and the HMMER tool (Finn et al., 2011). For the physical and chemical characterization of PKs, we estimated using the ProtParam module from the SeqUtils subpackage (Cock et al. 2009): (i) the isoelectric point; (ii) the molecular weight; and (iii) the general average hydropathy (GRAVY). For the number of amino acids (aa), predictions of transmembrane helices and signaling peptides, the TOPCONS server (https://topcons.cbr.su.se/pred/) was employed (Tsirigos et al., 2015). We estimated the potential subcellular locations of PKs using the BUSCA web-server (http://busca.biocomp.unibo.it/) (Savojardo et al., 2018). Finally, for functional annotation, we used the Blast2GO (Conesa et al., 2008) tool and retrieved the associated Gene Ontology (GO) terms to create a treemap using the REViGO tool (http://revigo.irb.hr/) (Supek et al., 2011) together with the treemap R package (Tennekes, 2021).

2.3 Kinase gene expression patterns associated with biotic stress

Initially, we assessed the expression of PK genes using raw RNA sequencing (RNA-Seq) data derived from three independent experiments, all retrieved from the CuGenDBv2 database. Each dataset was chosen based on its response to specific biotic stresses: (i) PM (*P. xanthii*) (Xu et al., 2017); (ii) ALS (*A. cucumerina*) (Sa et al., 2020); and (iii) RKN (*M. incognita*) (Wang et al., 2018). A summary of the data used in this study, focusing on differentially expressed genes (DEGs), is presented in Table 1. Subsequently, after fitting the model with the complete gene set associated with each biotic stress, the dataset of PK-related DEGs underwent filtration for further in-depth analyses.

Table 1. Description of the data considered for the cucumber kinome construction.

Condition	Pathogen lifestyle	Resistant	Susceptible	RNA extraction	Data origin
PM	Biotrophic	SSL508-28	D8	0 × 48 h	Xu et al. 2017
ALS	Necrotrophic	D1322	BJ204	2 × 6 dpi	Sa et al. 2020
RKN	Biotrophic	IL10–1	CC3	0 × 3 dpi	Wang et al. 2018

PM: powdery mildew (*P. xanthii*), ALS: alternaria leaf spot (*A. cucumerina*), RKN: root-knot nematode (*M. incognita*), BJ204: Beijing204, h: hours post-inoculation, dpi: days post-inoculation.

For the evaluation of resistance to PM, assessments were carried out for three consecutive years (spring 2013, spring 2014, and fall 2015) on the SSL508-28 (resistant) and D8 (susceptible) genotypes, conducted under greenhouse conditions (Yangzhou, China). Inoculations were performed by applying conidia collected from naturally infected D8 plants, as described by Xu et al. (2017). A total of 12 standard Illumina paired cDNA libraries were prepared, including three biological replicates for each genotype evaluated at 0- and 48-hours post-inoculation. These libraries were sequenced generating 125 base pairs paired-end reads through the Illumina Hiseq2500 platform. For mapping the reads to the cucumber reference genome, high-quality nucleotides were filtered according to their quality score (Q > 20) and

aligned to the 9930 (Chinese Long) v2 reference genome (Li et al., 2019) using TopHat v2.0.9 (Trapnell et al., 2009). To count the number of reads mapped to reference genes, Cufflinks v2.1.1 (Mortazavi et al., 2008) was employed. The raw RNA-seq data was generated and available by Xu et al. (2017).

For ALS, the genotypes D1322 (resistant) and Beijing204 (BJ204) (susceptible) were evaluated. Spores of *A. cucumerina* were inoculated via spore suspension on the third true leaf of 3-week-old plants, as described by Sa et al. (2020). To obtain samples for RNA extraction and library preparation for sequencing, the first three true leaves were collected at 2- and 6-days post-inoculation (dpi) in three biological replicates, each consisting of one plant. Sequencing was performed using an Illumina HiseqTM-2000 system, which generated 125 bp/150 bp paired-end reads. Count data were obtained through the following steps: (i) removal of low-quality reads and reads containing polynucleotides using the FASTX-toolkit (Gordon et al., 2010); (ii) read mapping to the 9930 (Chinese Long) v2 reference genome (Li et al., 2019) using HISAT 2.0.4 (Kim et al., 2019); and (iii) counting the number of reads mapped to each gene using HTSeq v0.6.1 (Anders et al., 2015). The raw RNA-seq data was generated and available by Sa et al. (2020).

For RKN, 3-4 days old seedlings of two cucumber genotypes were employed, namely IL10-1 (resistant) and CC3 (susceptible). The raw RNA-seq data was generated and available by Wang et al. (2018). The inoculation involved the use of *M. incognita* race 1 eggs were extracted from infected 'Hezuo 903' tomato plants, as described by Wang et al. (2018). Root tips were collected at 0, 1, 2, and 3 dpi from 7 plants for each genotype and then pooled into a single biological sample, with 3 biological replicates for each sample. In total, 24 cDNA libraries were prepared and sequenced on an Illumina Hiseq 2500 platform, generating paired-end reads. To ensure the generation of high-quality clean data, a filtering step was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), including the removal of reads containing adapters, reads with poly-N sequences, and reads of low quality from the raw data. The remaining reads were aligned to the 9930 (Chinese Long) v2 reference genome (Li et al., 2019) using TopHat v2.0.12 (Kim et al., 2013).

To identify the whole set of DEGs, we employed the raw count data as input for DESeq2 package (Love et al. 2014). We considered the following contrasts: PM between genotypes (time as a covariate), D8 between time points (0h vs 48h), SSL508-28 between time points (0h vs 48h); ALS between genotypes (time as a covariate), BJ204 between time points (2 dpi vs 6

dpi), D1322 between time points (2 dpi vs 6 dpi); RKN between genotypes (time as covariate), CC3 between time points (0 vs 3 dpi), and IL10-1 between time points (0 dpi vs 3 dpi). For each comparison, we conducted negative binomial tests to identify DEGs, with a significance threshold set at a maximum adjusted p-value (false discovery rate, FDR) of 0.05 and a minimum absolute log2 fold change (log2fc) of 1.5. After filtering the PK-related DEGs, they were visualized using Volcano Plots, generated using the tidyverse (Wickham et al., 2019), RColorBrewer (Neuwirth, 2014), and ggrepel (Slowikowski et al., 2018) packages in R.

For a better comprehension of the families that respond similarly to each disease, we employed Venn diagrams to illustrate the overlap between the total PK DEGs and those that were either up-regulated and down-regulated. Furthermore, we employed an UpSet plot to analyze PK DEGs that were shared or exclusive across all pathosystems, with a particular emphasis on distinguishing between down-regulated and up-regulated genes within the PK DEGs.

2.4 Kinase duplication and synteny analyses

To conduct the duplication analysis and identify duplication event categories, we employed the Multiple Collinearity Scan toolkit (MCScanX) (Wang et al., 2012). Calculations of synonymous (Ks) and non-synonymous (Ka) substitution rates were performed using the MCScanX module within the TBtools software. The elapsed time for each duplication event was determined using the formula $T = Ks/2\lambda$, where λ represents the mean synonymous substitution rate (6.5×10^{-9}) (Gaut et al., 1996).

A comprehensive synteny analysis was performed to explore the relationships among kinase genes in the Gy14 cucumber and DHL92 melon genomes, both sourced from the CuGenDBv2 database. The identification of syntenic blocks was accomplished through the application of the MCScanX toolkit, while the subsequent visualization of the obtained data was achieved using the Dual Synteny Plot package, seamlessly integrated into the TBtools software. This methodological approach allowed for a detailed examination of conserved genomic regions and facilitated a visually informative representation of the syntenic relationships between kinase genes in the cucumber and melon genomes.

3 RESULTS

3.1 Genome-wide identification and Classification of cucumber PKs

To identify potential cucumber PKs, the alignment of the cucumber proteome (Gy14) against two HMM profiles, Pkinase (PF00069) and Pkinase_Tyr (PF07714) identified 837 putative PKs, and after a domain check two atypical PKs were removed, remaining 835 true PKs. The PK isoforms were kept for further analysis. Our analysis revealed that most kinases (828 PKs) possessed both domains simultaneously, while seven PKs had only one of these domain types (5 Pkinase and 2 Pkinase_Tyr). Specifically, we identified 828 PKs with varying combinations of both types (676 PKs had one of each domain, and 152 PKs had between three to six Pkinase and Pkinase_Tyr domains) (Supplementary Table S1).

The domain composition analysis using the Pfam database and the HMMER tool identified two sequences as atypical PKs due to their absence of canonical kinase domains: (i) CsGy6G033750.2, which belongs to the ABC1 family (PF03109.16); and (ii) CsGy1G024940.2, associated with the Haspin_kinase family (PF12330.8). It is worth noting that atypical kinases, despite lacking canonical kinase domains, may still retain kinase functions, as reported in prior studies (Deshmukh et al., 2010; Manning et al., 2002). However, as our study primarily focuses on typical kinases, we have removed these atypical kinases from further investigations.

We identified PKs that contained multiple kinase domains within their sequences, comprising a total of 558 different domain types, including Pkinase and Pkinase_tyr. The number of distinct domains within each protein ranged from 1 to 97. Among these domains, seven were the most frequently observed across all PKs: Pkinase (PF00069), Leucine Rich Repeat 1 (LRR_1, PF00560), Pkinase_tyr (PF07714), Leucine rich repeat N-terminal domain (LRRNT_2, PF08263), Leucine Rich Repeat 4 (LRR_4, PF12799), Leucine Rich Repeat 6 (LRR_6, PF13516), and Leucine Rich Repeat 8 (LRR_8, PF13855). Other domains were less common (Supplementary Table S2).

Furthermore, we noted the frequent occurrence of other domains, such as Aminoglycoside Phosphotransferase (APH, PF01636) appearing 348 PKs, Kinase-like (PF14531) 297 times, Pkinase_fungal (PF17667) 243 times, and, especially, Leucine-rich Repeat (LRR) domains. The domains LRR_1 (PF00560), LRR_4 (PF12799) and LRR_8 (PF13855) were found 192 times each. Moreover, we observed the Haspin_kinase domain

(PF12330) occurring 188 times. A complete description of these kinase domains is available in Supplementary Table S2.

Based on the HMM approach proposed by Lehti-Shiu et al. (2012), we were able to classify the 835 genes encoding cucumber kinases into 20 distinct groups and 123 families (Figure 1 and Supplementary Figure S1, Supplementary Table S3 and Supplementary Table S4), namely: AGC (cAMP-dependent protein kinases, cGMP-dependent protein kinases, various types of protein kinase C, protein kinase B, 3-phosphoinositide-dependent protein kinase-1, and the ribosomal protein S6 kinases), Aur (aurora kinase), BUB (budding uninhibited by benzimidazoles), CAMK (calcium/calmodulin-dependent protein kinase), CK1 (casein kinase 1), CMGC (cyclin-dependent kinase, mitogen-activated protein kinase, glycogen synthase kinase, and cyclin-dependent-like kinase families), Plant specific (Group-Pl), IRE1 (inositol-requiring enzyme 1), NAK (NF-κB-activating kinase), NEK (never in mitosis gene-A), PEK (pancreatic eukaryotic initiation factor 2 α-subunit kinase), RLK (receptor-like kinase), SCY (*Saccharomyces cerevisiae* [yeast] kinase), STE (serine/threonine kinase), TKL (tyrosine kinase-like kinase); TLK (tousled-like kinase), TTK (threonine/tyrosine kinase), ULK (unc-51-like kinase), WEE (wee1, wee2, and myt1 kinases), and WNK (with no lysine-K). To confirm the subfamily classification, we used a phylogenetic approach that effectively grouped the genes into eight distinct clusters with strong support, as indicated by the 70% bootstrap threshold. These clusters were related to the AGC, CAMK, CK1, CMGC, RLK, SCY, STE, and TKL groups (Figure 1, and Supplementary Figure 1). The remaining proteins exhibited substantial functional diversity and expansion, forming different groupings of families within the same group. Furthermore, 13 genes could not be classified within their respective group (in instances where there was only one gene per family) or within their respective families (in cases where multiple genes belonged to a family). These unclassified genes were considered as belonging to an "unknown" category.

Among the groups defined for the cucumber kinome, the RLK group was the most numerous, with 530 genes (63.5% of the total number of PKs). The genes of this group were subdivided into 57 families. Notably, 33 of these families contained only one gene, while the RLK-Pelle_DLSV family was the most abundant with 68 genes, followed by the RLK-Pelle_LRR-XI-1 (40 genes), RLK-Pelle_LRR-III (38 genes), RLK-Pelle_RLCK-VIIa-2 (32 genes), and RLK-Pelle_L-LEC (28 genes) (Supplementary Table S3 and Supplementary Table S4).

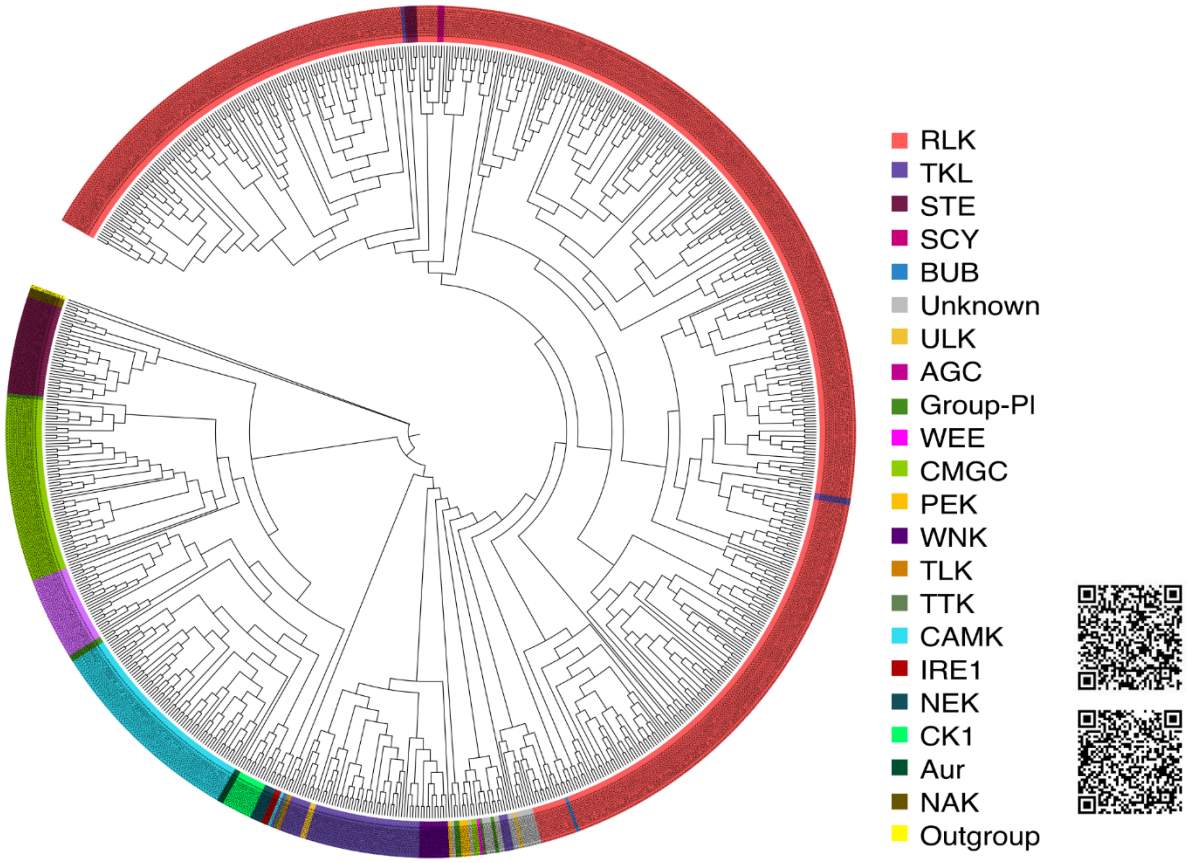

Fig. 1 Phylogenetic tree of *Cucumis sativus* PKs with the approximately-maximum-likelihood Likelihood approach with bootstrap values (> 70%), using the Whelan Goldman (WAG) model with gamma parameter. The clades highlighted are referred to the groups RLK, SCY, TKL, STE, CMGC, AGC1, CK1, and CAMK (from up to down)**.** The members of each family are collapsed to a better visualization. The gene names displayed with square brackets represent branches with single genes. The tree is rooted with a kinase (Cre07.g349540.t1.1) from the *Chlamydomonas reinhardtii* v5.6 proteome. Some genes from expanding groups can be found within other consolidated groups. QR code for a link with a high-quality image from Figure 1 and Supplementary Figure S1.

3.2 Characterization of the cucumber PK sequences

When evaluating the chromosomal distribution of the 835 PK genes along the seven chromosomes, we did not observe any discernible pattern across chromosomes. However, we did observe a noticeable accumulation of PKs in the telomeric regions, aligning with a tendency for resistance genes to be in the distal regions of telomeres (Choulet et al., 2014, The International Barley Genome Sequencing Consortium, 2012). Additionally, genes from the same family tended to cluster together, suggesting potential tandem duplications (Figures 2 and

3A). It is noteworthy that two genes were associated with cucumber scaffolds and were not allocated to any chromosome in the Gy14 genome.

Fig. 2 Chromosome physical localization of kinase genes from *Cucumis sativus*. The kinase genes are presented by their names followed by the family name, and the colors represent different groups. Chromosome zero (0) contains two genes that have not been assigned to specific chromosomes. QR code for a link with a high-quality image.

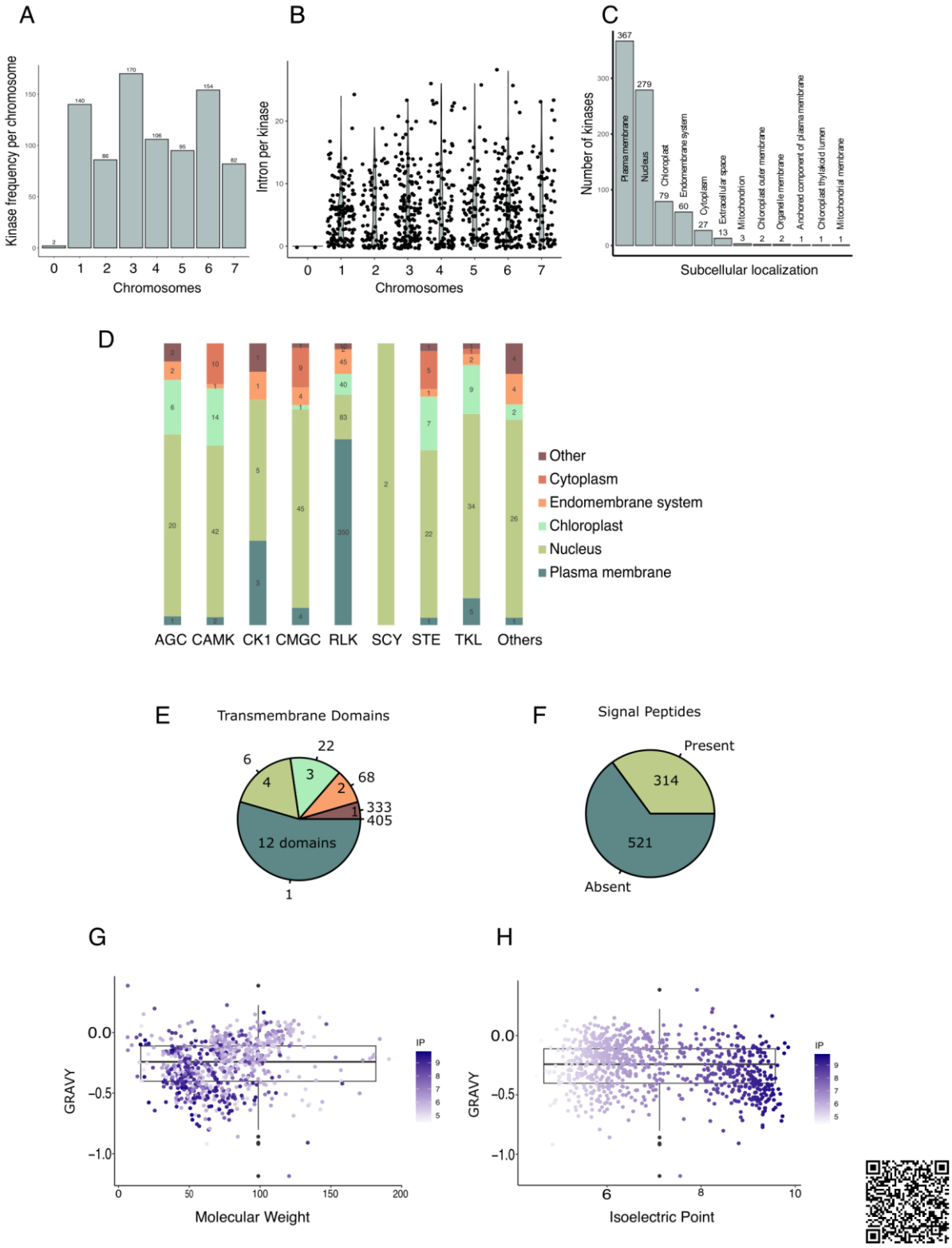

Fig. 3 (A) Number of cucumber kinases per chromosome. **(B)** Number of introns per kinase per chromosome. **(C)** Number of kinases per subcellular location. **(D)** Distribution of kinase per subcellular location per kinase group. **(E)** Number of transmembrane domains and **(F)** signal peptides. **(G)** molecular weight (MW), and **(H)** isoelectric point (IP) in function of grand average of hydropathy (GRAVY). QR code for a link with a high-quality image.

To analyze the gene organization of the 835 cucumber PKs, we assessed the number of introns. This analysis of gene structure revealed a variable number of introns, ranging from 0 to 28, with a median of 5 introns per gene (standard deviation of 5.46). The highest number of introns was observed in the CsGy6G003600 gene, a member of the PEK_GCN2 family, which presented 28 introns. Among the total of 835 PK genes, 123 (14.7%) presented no introns, while 554 (66.3%) contained between 1 and 10 introns, 139 (16.7%) had between 11 and 20 introns, and 19 (2.3%) exhibited a substantial range of 21 to 28 introns (Figure 3-B, Supplementary Table S5).

The subcellular location prediction results revealed that approximately 44% of PKs are potentially located in the plasma membrane, 33% in the nucleus, 9% in the chloroplast, 7% in the endomembrane system, 3% in the cytoplasm, and 2% in the extracellular space (Figure 3-C-D). Additionally, each of the following locations accounts for less than 1%: mitochondria, chloroplast outer membrane, organelle membrane, anchored component of plasma membrane, chloroplast thylakoid lumen, and mitochondrial membrane (Figure 3-C-D, and Supplementary Table S6). These results illustrate the importance of PKs in the plasma membrane and nucleus, since 77% of the proteins analyzed are in these locations.

The number of transmembrane (TM) domains per kinase exhibited significant variation (Figure 3-E). Specifically, we observed 405 PKs lacking TM domains, 333 PKs with one TM domain, 68 PKs with two TM domains, 22 PKs with three TM domains, 6 PKs with four TM domains, and a single PK with 12 TM domains. We also were able to predict the presence of a signal peptide in 314 kinases, while it was absent in 521 kinases (Figure 3-F, and Supplementary Table S6). The number of TM domains not only includes proteins located within the plasma membrane, but also encompasses those found in other internal membranes. Furthermore, the presence of a signal peptide suggests that the PK action is related to the secretory pathway, implying their likely extracellular site of action.

The physicochemical characteristics of the PKs were highly variable (Figure 3-G-H). The number of amino acids ranged from 57 to 1694 (median: 645); molecular weight spanned from 6.42 to 191.01 kDa (median: 71.27 kDa); the isoelectric point ranged from 4.38 to 9.85 (median: 6.57); and the hydropathy (GRAVY) varied from -1.19 to 0.39 (median: -0.24). This diversity in physical and chemical properties highlights the wide range of sizes, charges, and hydrophobicity characteristics of the PKs, providing insights into the intricate mechanisms by

which proteins harness a highly complex enzymatic system to modulate responses to environmental stress.

Functional annotation revealed a total of 386 different GO terms occurring 4,273 times. These GO categories encompass molecular function (F) (95 terms), biological processes (P) (227 terms), and cellular components (C) (64 terms). These annotations provide valuable insights into the functions and locations of PKs, highlighting their involvement in a wide range of metabolic processes and cellular components (Figure 4-A-B, and Supplementary Table S7). Notably, cucumber PKs play pivotal roles in processes such as protein phosphorylation, signal transduction, regulation of stress responses, and various biological processes, with molecular functions primarily associated with binding activities.

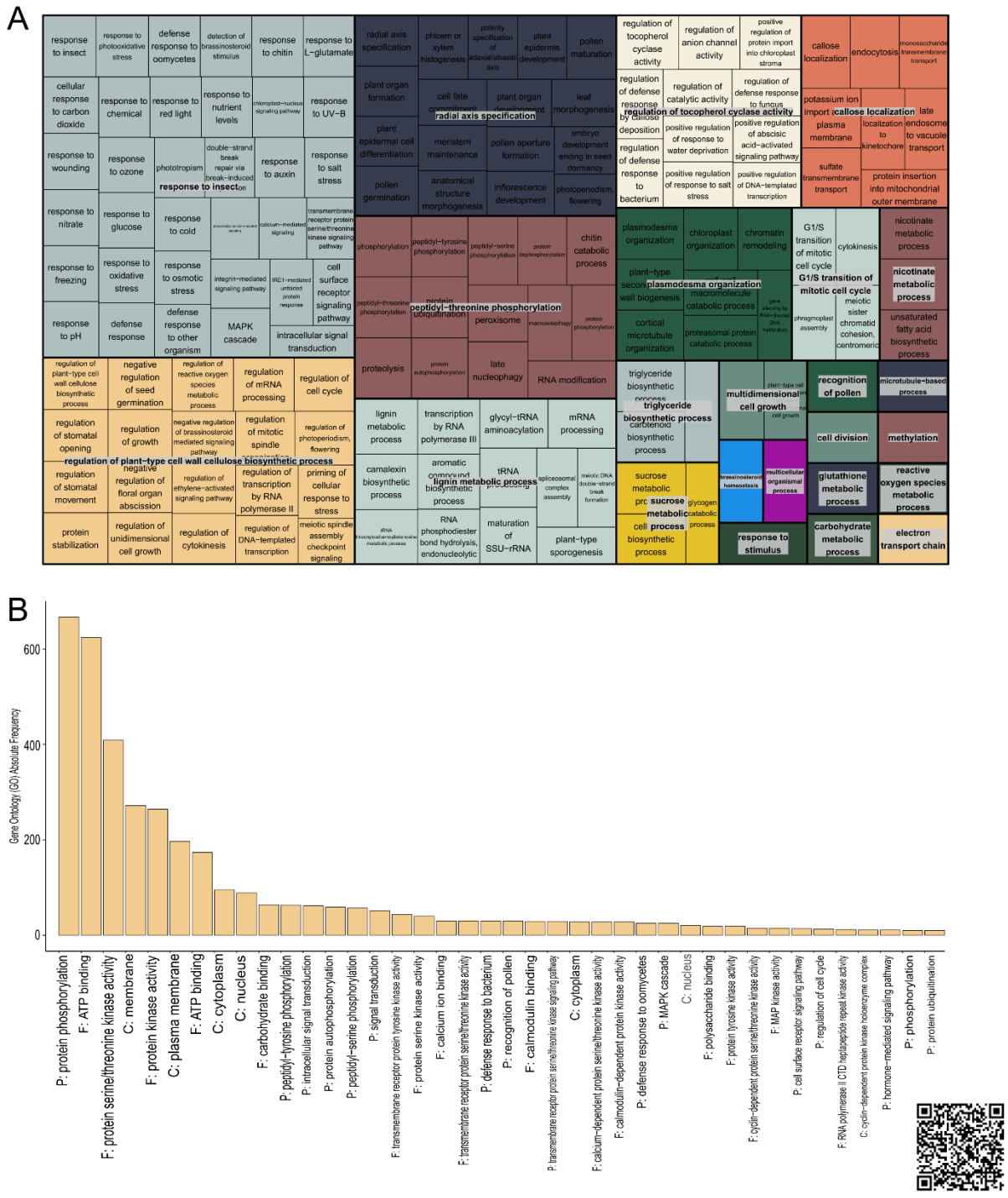

Fig. 4 (A) Biological processes (227) predicted in the Gene Ontology analysis. **(B)** P: biological process, F: molecular function, C: cellular component. Figure shows the 40 top GO terms; a full list is available in Supplementary Table S7. QR code for a link with a high-quality image.

3.3 Kinase gene expression patterns associated with biotic stress

In the analysis of all 22,626 genes, the identification of DEGs involved the execution of three specific types of contrasts, as detailed in Table 2. These included: (i) a comparison between resistant and susceptible genotypes while incorporating time as a covariate; (ii) an assessment of gene expression changes in the resistant genotype at both initial and final time points; and (iii) an evaluation of gene expression alterations in susceptible genotypes at both initial and final time points.

For each comparison, we estimated a distinct model using the DESeq2 package and retained different numbers of genes for differential expression evaluation, ranging from 17,839 to 21,490 (Table 2). Following this, we identified DEGs specific to each pathosystem and subsequently narrowed our focus to the PK DEGs for further in-depth analyses. The number of DEGs varied from 319 to 2,202, while PK DEGs numbered from 8 to 105.

Among susceptible genotypes, we observed a higher number of DEGs in diseases caused by biotrophic pathogens (*P. xanthii* and *M. incognita*), whereas resistant genotypes exhibited more DEGs in response to the necrotrophic pathogen *A. cucumerina*. When considering time as a covariate, i.e., contrasting the gene expression between genotypes, we identified 319 DEGs (8 PKs) for the PM pathosystem, 929 DEGs (26 PKs) for ALS, and 498 DEGs (14 PKs) for RKN. These DEG numbers were lower in comparison to the other contrasts performed.

Table 2. DEG analysis of the compatible and incompatible interactions of cucumber genotypes against the pathogenic fungi *P. xanthii* (PM) and *A. cucumerina* (ALS), and the root-knot nematode *M. incognita* (RKN).

Condition	Contrast	Contrast time	Genes for modeling	DEGs	PK DEGs	Down	Up
PM	SSL508-28^{R} × D8^{S}	covariate	17,839	319	8	2	6
	SSL508-28 × SSL508-28	0h × 48h	17,522	1,428	87	50	37
	D8 × D8	0h × 48h	18,314	1,690	105	86	19
ALS	D1322^{R} × BJ204^{S}	covariate	19,317	929	26	15	11
	D1322 × D1322	2 dpi × 6 dpi	19,024	2,202	98	66	32
	BJ204 × BJ204	2 dpi × 6 dpi	19,001	1,847	64	40	24
RKN	IL10-1^{R} × CC3^{S}	covariate	21,490	498	14	9	5
	IL10-1 × IL10-1	0 dpi × 3 dpi	19,897	1,612	50	21	29
	CC3 × CC3	0 dpi × 3 dpi	21,188	1,863	63	41	22

DEG: differentially expressed genes, PK: protein kinase, PM: powdery mildew (*P. xanthii*), ALS: alternaria leaf spot (*A. cucumerina*), RKN: root-knot nematode (*M. incognita*), (R): resistant, (S): susceptible, dpi: days post-inoculation. Down: down-regulated, Up: up-regulated.

Of the general expression data, only values related to the expression of PK DEGs were considered among the total set of DEGs. To identify PK DEGs responsive to PM, three approaches were performed, in which the differential expression between resistant (SSL508-28) and susceptible (D8) genotypes at evaluation times (0 and 48 h, using the variable time as a covariate) were analyzed (Figure 5, Table 2). The second and third approaches were considered the expression of susceptible and resistant genotypes in relation to the initial and final expression times. Among genotypes, only 8 PK DEGs were significant, of which 2 were down-regulated, and 6 were up-regulated (Figure 5-A). In the resistant genotype, 87 significant

PK DEGs were observed (50 down-regulated and 37 up-regulated). In the susceptible genotype, a total of 105 significant PK DEGs were identified (86 were down-regulated and 19 were up-regulated).

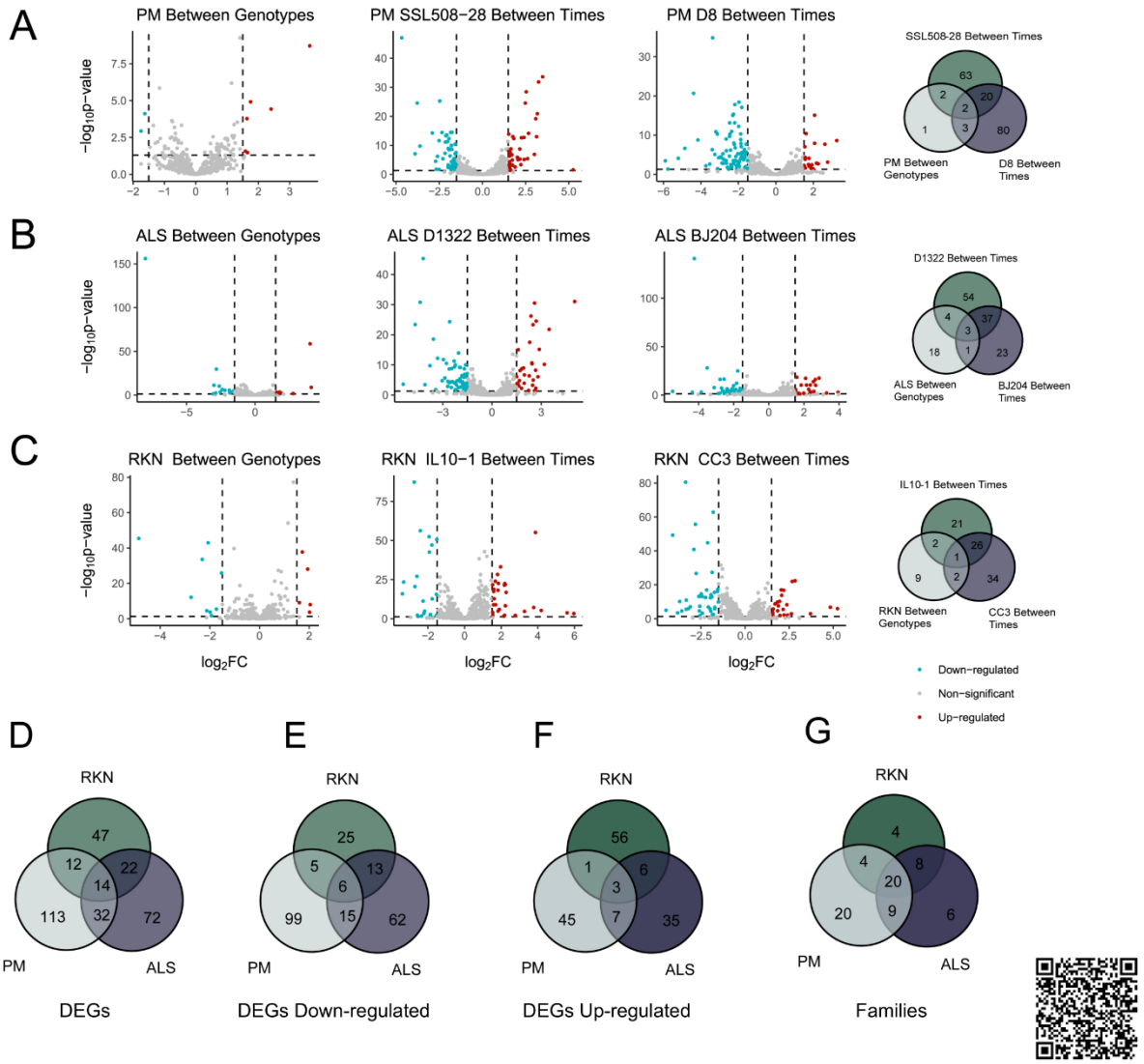

Fig. 5 Differential expression of kinase genes (kinase DEGs) of the **(A)** the genotypes resistant (SSL508-28) and susceptible (D8) to powdery mildew (PM) (*Podosphaera xanthii*), **(B)** the genotypes resistant (D1322) and susceptible (BJ204) to alternaria leaf spot (ALS) (*Alternaria cucumerina*); and **(C)** the genotypes resistant (IL10-1) and susceptible (CC3) to root-knot nematode (RKN) (*Meloidogyne incognita*). **(D)** DEGs in common to all pathosystems; **(E)** DEGs down-regulated in common to all pathosystems; **(F)** DEGs up-regulated in common to all pathosystems; **(G)** Family DEGs in common to all pathosystems. QR code for a link with a high-quality image.

To identify the DEGs responsive to ALS infection, PM-like approaches were used. First, between the resistant (D1322) and susceptible (BJ204) genotypes with time as a covariate, 26 significant DEGs were identified (15 were down-regulated and 11 were up-regulated) (Figure 5-B). Subsequently, 98 significant DEGs were observed in the resistant genotype (66 DEGs down-regulated and 32 DEGs up-regulated). Finally, 64 DEGs were identified in the susceptible genotype at evaluation times 2 and 6 dpi (40 DEGs down-regulated and 24 DEGs up-regulated).

In case of RKN-responsive DEGs, 14 significant DEGs were identified between resistant (IL10–1) and susceptible (CC3) genotypes with the time of evaluation as a covariate (9 DEGs down-regulated and 5 DEGs up-regulated) (Figure 5-C). In the resistant genotype, 50 DEGs were observed in 0-3 dpi (21 down-regulated and 29 up-regulated). In the susceptible genotype, 63 significant DEGs were observed in 0-3 dpi (41 down-regulated DEGs and 22 up-regulated DEGs).

To study the DEGs in common, Venn diagrams were constructed with the PK genes common to the three pathosystems, with 312 DEGs in common among all of them, of which 225 down-regulated and 153 up-regulated (Figure 5-D-F). To better understand the down- and up-regulated families common to each pathosystem, an UpSet plot (Figure 6-A) support the identification of 203 families (71 non-redundant PK families) distributed across all conditions analyzed. Members of the RLK families are most frequently observed, with the RLK-Pelle_DLSV family appearing 17 times (Figure 6-B-C, Supplementary Table S8, and Supplementary Table S9). The lists were categorized based on different combinations of proteins related to specific biological processes, containing proteins associated with responses to PM, RKN, and ALS. The number of families is varied, ranging from 1 to 41 members, and they contain proteins that belong to different RLK families, among others. These lists may represent significant groupings of proteins that play important roles in specific biological pathways or cellular regulatory processes. According to our results, 20 families were exclusive to PM, 6 were exclusive to ALS, and 4 were exclusive to RKN, 9 in common to PM and ALS, 4 common to PM and RKN, 8 in common to ALS and RKN, and 20 families in common to all (Figure 6, Supplementary Tables S10-S13).

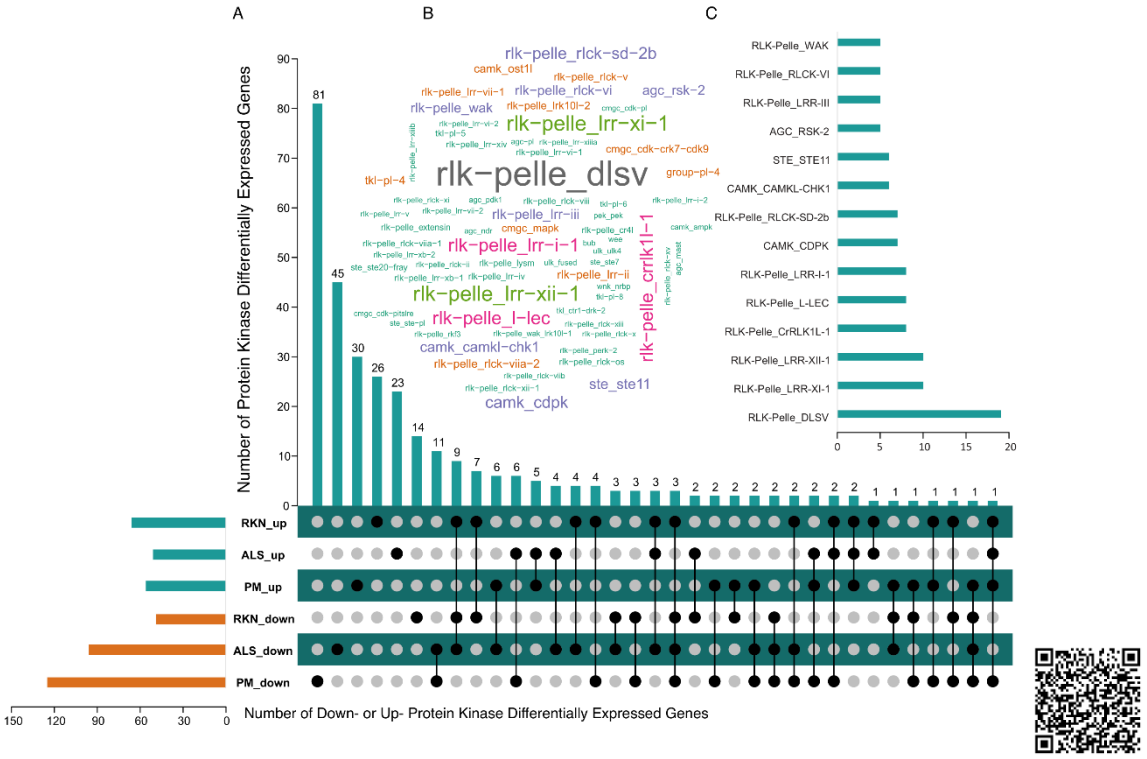

Fig. 6 (A) UpSet plot illustrating the shared 312 Protein Kinases Differentially Expressed Genes (312 PK DEGs) across three pathosystems: RKN (Root-knot nematode, *M. incognita*), ALS (Alternaria leaf spot, *A. cucumerina*), and PM (powdery mildew, *P. xanthii*), for both down-regulated (RKN_down, ALS_down, and PM_down) and up-regulated (RKN_up, ALS_up, and PM_up) genes. In the plot, black dots connected by vertical black lines represent the count of PK DEGs shared among the respective conditions. Vertical bars indicate the number of DEGs in specific conditions, and horizontal bars show the sum of down- or up-regulated DEGs. **(B)** Word cloud representation for the more frequent protein kinase families. **(C)** Chart for the more frequent protein kinase families. QR code for a link with a high-quality image.

3.4 Kinase duplication and synteny analyses

Through an in-depth examination of duplication events, our study yielded estimations for 833 protein kinases (PKs), considering that two kinases are still discernible at the scaffold level in the genome. The predominant origin of these PKs was attributed to whole genome duplication (WGD) events, encompassing approximately 69.4% (578 PKs) (Figure 7-A). Tandem duplications contributed to ~28.5% (237 PKs), proximal duplications accounted for ~1.6% (13 PKs), singleton events represented ~0.36% (3 PKs), and dispersed duplications were observed at a minor frequency of 0.24% (2 PKs). For a subset of 40 genes, we calculated both the Ka and Ks substitution rates (Figure 7-B). The Ka/Ks ratio, indicative of the interplay

between purifying selection, neutral mutations, and advantageous mutations, ranged from 0.146 to 2.198, with a mean ratio of 0.146 (Supplementary Table S14). Ratios below 1 suggest selective pressure, values around 1 denote neutral effects, while ratios exceeding 1 signify advantageous mutations. Furthermore, the estimated ages of these duplication events were determined based on the Ks rate, spanning from ~ 50.42 to ~422.81 million years ago (MYA). This temporal insight provides a comprehensive perspective on the evolutionary dynamics of the PK landscape, shedding light on the selective forces and mutational processes that have shaped this vital component of the genome over extensive periods.

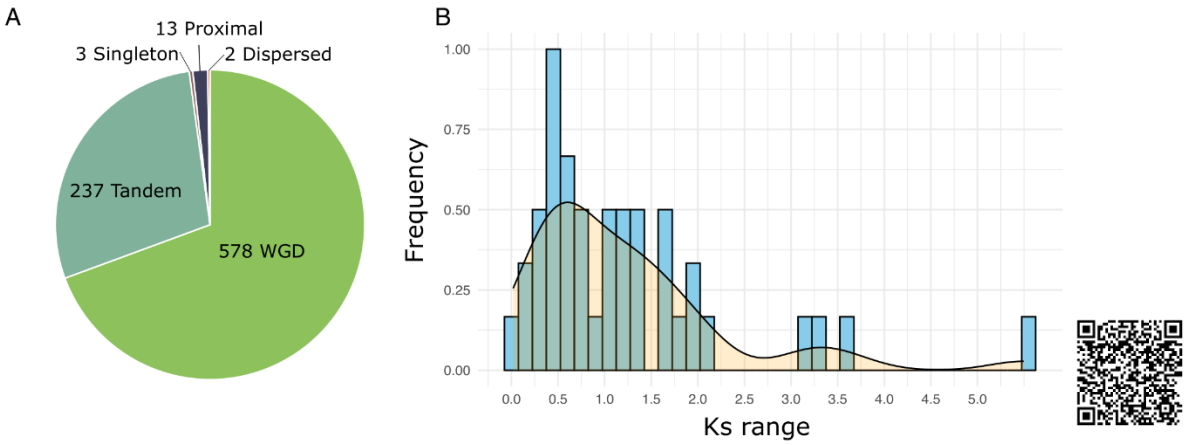

Fig. 7 (A) Classification of duplication types of 833 Protein Kinases of Cucumis sativus (Gy14). **(B)** Range of Ks values for the Protein Kinases of Cucumis sativus (Gy14). QR code for a link with a high-quality image.

A total of 237 PKs exhibiting tandem duplications were identified across 54 families. Among these, one family belonged to the CAMK group, two to CMGC, 43 to RLK, 5 to TKL, and singular representatives from TLK, WEE, and WNK. Tandem duplications manifest when a chromosomal segment is replicated and integrates adjacent to the original segment, a result of the crossing-over process between chromatids. This mechanism plays a crucial role in amplifying the count of genes sharing identical or similar functions, contributing significantly to genomic diversity and functional specialization.

To investigate the evolutionary relationships between cucumber and melon species, we conducted a detailed synteny analysis, focused on cucumber kinase genes. Of the 835 kinase genes identified in cucumber, we identified remarkable syntenic relationships for 670 of these genes on all chromosomes of the two species (Figure 8, Supplementary Table S15). This finding encompasses 821 syntenic relationships with melon genes, of which 11 genes showed three

relationships, 129 genes showed two relationships, while the remaining genes showed a single relationship. Analyzing the melon genome in relation to the cucumber PK array, we identified that 672 genes exhibited syntenic relationships, providing a comprehensive view of the genetic interconnections between these two species. These results provide insights into evolution and genetic relationships, highlighting complex patterns of conservation and divergence. The richness of these syntenic relationships reveals the intricate web of genetic connections between cucumber and melon, indicating not only shared ancestry but also specific duplication and rearrangement events throughout their evolutionary histories.

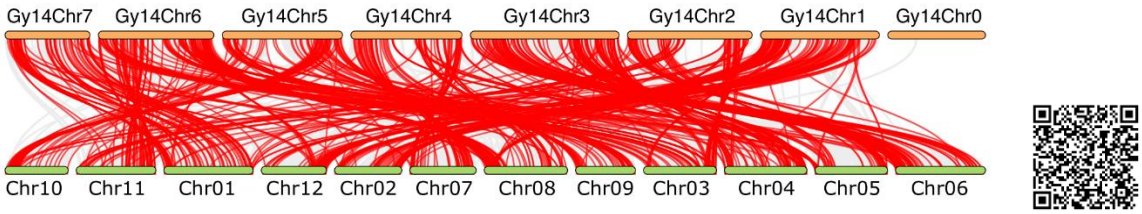

Fig. 8 Synteny analysis of Protein Kinases between Cucumis sativus (Gy14) and Cucumis melo (DHL92). The chromosomes of C. sativus are represented in orange color and the prefix Gy14, and the chromosomes of C. melo represented in green and the prefix chr. QR code for a link with a high-quality image.

4 DISCUSSION

4.1 Genome-wide identification and Classification of cucumber PKs

Cucumber was one of the first economically important crop to have its genome published. Over five years the genomes of the lineages '9930' (Huang et al., 2009; Li et al., 2011), 'PI 183967' (Qi et al., 2013; Wóycicki et al., 2011; Yang et al., 2012), 'Gy14' and 'B10' were made available, which provided a significant scientific advance in the field of cucumber genetic research. In addition, the utilization of recently developed high-throughput long-read sequencing technology has improved the quality of genome assembly for the 'Gy14' v2.1, '9930' v3.0, and 'B10' v3.0 (Weng, 2022). This improvement enabled us to conduct a genome-wide analysis, resulting in the construction of the cucumber 'Gy14' kinome.

Some projects have already been carried out considering some families of PKs in Csa species (Lehti-Shiu et al., 2012; Xu et al., 2015). In this study, we performed analyses of the cucumber proteome to identify and characterize the PK family (Hanks et al., 1995). Using the profiles of Hidden Markov Models (HMM) PF00069 (Pkinase) and PF07714 (Pkinase_Tyr) to

scan the cucumber annotated proteome, a total of 835 proteins containing kinase domains were identified, which provided the opportunity for a holistic view of the biological processes played by PKs. Among these, 517 sequences were classified as Serine/Threonine kinase (Pkinase) and 318 sequences as Tyrosine kinase (Pkinase_Tyr) This number represents 3.69% of the total proteins of the Csa species, slightly below the average (3.85%) of other species already studied, such as Arabidopsis, common beans, corn, cowpea, soybean and vine (Aono et al., 2023; Champion et al., 2004; Ferreira-Neto et al., 2021; Liu et al., 2015; Wei et al., 2014; Zhu et al., 2018; Zulawski et al., 2014).

For a better understanding of the evolutionary relationships between cucumber PKs, a phylogenetic analysis was obtained, which successfully grouped the members of eight well-established groups: AGC, CAMK, CK1, CMGC, RLK, SCY, STE and TKL (Lehti-Shiu et al., 2012). These groups represent distinct families of PKs with known functions in several biological processes. However, there was significant functional diversity and expansion within these groups, as evidenced by the presence of different families within the groups. The expansion of these families may be related to the evolutionary process, since PKs are central regulators of stress in plants (Benschop et al., 2007; Peck, 2001; Tena et al., 2011), acting like a “central processor unit” (CPU), processing external signals and converting them into specific response in the cell metabolism (Hardie, 1999) and, therefore, need to stand fine-tuned to recognize a plethora of pathogenic effectors that trigger the immune response processes. As observed by Lehti-Shiu et al., (2012), some PKs of the remaining groups (AUR, BUB, PLANT-SPECIFIC, IRE1, NAK, NEK, PEK, TLK, TTK, ULK, UNKNOWN, WEE, and WNK) observed in our analysis are composed of PKs that probably arose after events that led to the divergence of plants, fungi and animals, with these families having few members, as they are absent in the algae *Chlamydomonas reinhardtii*, such as BUB, IRE1, PEK. This is an aspect that needs to be elucidated, regarding which families arose before and after the divergence of eukaryotes into fungi, plants and animals. Some specific families still have not their functions clarified in some plant species, such as the PKs of the Group-Pl (Lehti-Shiu et al., 2012), NEK in both plants and animals (Baker et al., 2023, Dahan et al., 2009, Zhang et al., 2011).

According to the rules proposed by Lehti-Shiu et al. (2012), the 835 PKs of the cucumber were classified into 20 groups. Out of these, the RLK group was the most numerous, comprising 527 genes, representing approximately 63.5% of the total kinases identified. This result is consistent with previous studies highlighting the abundance of RLKs in common bean, corn, Arabidopsis, vine, cowpea, and soybean plants, averaging 66.9%, where RLKs play

crucial roles in signal transduction and various developmental and stress response processes (Aono et al., 2023; Champion et al., 2004; Ferreira-Neto et al., 2021; Liu et al., 2015; Wei et al., 2014; Zhu et al., 2018; Zulawski et al., 2014). Notably, Lehti-Shiu et al. (2012) did not detect the AGC_PKA-PKG, BUB, PEK_PEK, and TKL-PL-8 families in the Csa species; however, our study identified their presence. Discrepancies in analytical methodologies and the absence of genome collinearity among the studied cucumber lineages may account for these differences.

According to Lehti-Shiu et al. (2012), the PK groups could be subdivided into 123 families, with 34 of them composed of only one PK gene (Supplementary table S6). Notably, in our studies, the RLK-Pelle_DLSV family was the most numerous, consisting of 68 proteins, a result compatible with that obtained for the Csa species by Lehti-Shiu et al. (2012), this being one of the most numerous families of the species reported. Similar results have also been found in other species, such as Gossypium (Yan et al., 2018), rice (Dardick et al., 2007), Arabidopsis (Champion et al., 2004; Zulawski et al., 2014) and corn (Wei et al., 2014). The expansion and establishment of this family probably occurred before the divergence of Viridiplantae species (Lehti-Shiu et al., 2012), as no evidence that the Csa species has undergone whole genome duplication (Huang et al., 2009). However, after a new assembly of the cucumber genome, Li et al. (2019) reported 239 tandemly duplicated genes, and Yu et al. (2022), reporting on 15 pairs of tandem duplication genes and 31 pairs of segmental duplication in the cucumber genome.

Not only does the RLK-Pelle_DLSV family stand out, but within the PKs of the RLK group, the RLK-LRR forms the most abundant subgroup, comprising 202 genes and further subdivided into 23 families (refer to Supplementary Table S4). These findings align with those presented by Yu et al. (2022), who identified 189 genes distributed across 22 families RLK-LRR in the Chinese Long cucumber genome v3. The substantial presence of genes within this family underscores their significance in stress response mechanisms, as highlighted in studies by Dambroz et al. (2023), Fischer et al. (2016), Magalhães et al. (2016), and Sakamoto et al. (2012).

4.2 Characterization of the cucumber PK sequences

This investigation delves into various facets of PK genes, providing insights into the diversity and functional characteristics of cucumber's PK gene repertoire. The analysis commenced by scrutinizing the structural attributes of the 835 PK genes, with a particular focus on the determination of intron numbers. The presence of introns holds significance as it may be

intricately linked to the regulation of gene expression, facilitating processes such as alternative splicing. Notably, intronic genes, which encompass specific intron fragments undergoing expression, as well as factors such as intron length and intron phase, assume crucial roles in the regulation of gene expression (Baulin et al., 2020). These findings underscore the structural diversity inherent in cucumber's PK genes and shed light on the intricacies of their regulatory landscapes (Rose, 2008). Moreover, the presence of introns in a gene can exert an influence on its susceptibility to silencing through microRNAs (miRNAs). Furthermore, miRNAs have the ability to bind to mRNA sequences that include introns, and the quantity of introns can impact the efficiency of miRNA-mediated silencing (Huberdeau et al., 2018; Lin et al., 2006). Hence, the number of introns plays a role in the regulatory process. Then, the occurrence of introns in miRNA target genes can impact mRNA maturation and translation efficiency, thereby exerting a downstream effect on overall gene expression dynamics (Lin et al., 2006).

The uneven distribution of PK genes on the seven chromosomes of the cucumber genome represents a non-random pattern, which may be a result of the evolutionary process undergone by the species (Yang, et al., 2012). However, the distribution of RLK genes in clusters was observed, supporting the hypothesis that this family originated by tandem duplication and later subfunctionalization (Holub, 2001), although no duplication was evidenced in cucumber species (Huang et al., 2009). On the other hand, the results of Li et al. (2019), and Yu et al. (2022) contradict previous studies, giving evidence of the cucumber genome duplication.

Considerable variation was observed in all the characteristics examined, indicating specialized functions in their respective microenvironments, similar to patterns also observed by Dambroz et al. (2023). Most PK functions are associated with the plasma membrane and nucleus, comprising about 77% of the PK genes in cucumber. The kinase domains of nuclear and membrane proteins are intricately linked to pathogen identification and the plant defense response. The predominant functions of PKs are associated with the plasma membrane and the nucleus, playing a pivotal role in the regulation of gene expression and the response to environmental stimuli, such as stress and disease resistance in plants (Delormel et a., 2019; Luan et al., 2002). This dual functionality, spanning both nuclear and membrane realms, underscores the versatility of PKs in performing essential cellular processes (Romeis, 2001). At the plasma membrane, PKs are instrumental in perceiving external signals, transducing them into intracellular responses, and modulating various physiological pathways crucial for plant growth and adaptation (Ni et al., 2017). Simultaneously, within the nucleus, PKs contribute

significantly to the intricate machinery of gene regulation, influencing transcriptional processes that ultimately shape the plant's responses to environmental challenges (Carbonnel et al., 2010). A pattern of specialized functions localized predominantly in the plasma membrane and nucleus emerges, aligning with observations in cucumber, common beans (Aono et al., 2023), *Hevea brasiliensis* and *Manihot esculenta* (Santos et al., 2023). This consistency across diverse plant species suggests a fundamental role for PKs in essential cellular processes and stress responses. However, the observed diversity in subcellular locations among PKs also hints at their involvement in specific signaling pathways, highlighting the versatile and intricate nature of their functions. Similar results were also observed in common beans (Aono et al., 2023), *Hevea brasiliensis*, and *Manihot esculenta* (Santos et al., 2023). However, other genes can be found in various subcellular locations, suggesting their involvement in specific cellular processes, and signaling pathways in various cellular biological processes together with the stress responses.

In principle, PKs are responsible for the processes of phosphorylation and binding to specific targets in the cell related to activation processes and can be listed as interesting targets for genetic breeding, in the hypothesis of improvement of central metabolic processes, such as increased efficiency of enzymatic conversion or modifications at the binding site that recognize pathogen effectors, preventing its infection. As an example, regulation of gene expression through transcription factors led to overexpression of mutant rice genotypes, increasing nitrogen use efficiency and productivity (Wei et al., 2022). However, studies detailing these processes involving phosphorylation remain scarce.

4.3 Kinase gene expression patterns

The study of differential gene expression is one of the main methodologies for understanding the nuances of pathogen-plant interaction (PPI). This interaction involves several processes that include the proteins phosphorylation and the expression of various defense-related genes, among others. In this study, we investigated the expression profiles of the 22,626 genes encoded in Gy14 genome in response to biotic stress conditions of PM, ALS, and RKN, using transcriptome data from different cucumber genotypes interacting with different pathogens. After modeling with DESeq2, some genes did not fit and were removed to further analyses. Then, PK DEGs were identified from the DESeq2 fitted pool. It is possible to observe that susceptible genotypes present a greater number of down-regulated DEGs when the pathogens have a biotrophic lifestyle (PM and RKN); on the other hand, in the necrotrophic

pathogen (ALS), a greater number of down-regulated genes is observed in the resistant genotype (Table 2, and Figure 5).

Our analysis identified PK DEGs in response to *P. xanthii* infection (PM), showing similar expression patterns but at different response times. However, the process is still elusive and needs to be investigated, since some genes with differential expression are present in both resistant and susceptible genotypes. Some DEGs act at the cell nucleus level (~25% of DEGs) in processes related to protein phosphorylation and mitotic cell cycle regulation; while most of the DEGs (~55%) function in the nucleus, with general functions related to phosphorylation and participation in some more specific processes such as Peptidyl-Tyrosine Phosphorylation, MAPK Cascade, Hormone-Mediated Signaling Pathway, Phragmoplast Assembly (Supplementary Table S10). Some proteins in smaller numbers were also located in chloroplasts, endomembrane system, extracellular space, mitochondrion, and mitochondrion membrane with functions related to Polysaccharide Binding, Nucleotide Binding, Transmembrane Receptor Protein Tyrosine Kinase Activity, Polysaccharide Binding; and metabolic processes, such as Defense Response to Bacterium, and Peptidyl-Tyrosine Phosphorylation (Figures 5 and 6, and Supplementary Table S6). These findings suggest that PKs play important roles in the defense response against PM, with similar expression patterns observed between resistant and susceptible genotypes and major expression differences related to response time to infection, because in general, susceptible genotypes show an immediate response, while resistant genotypes show a later response after infection. The shallow response time may be related to the time of cell maturation, time of development of infection-induced apoptosis, rates of infection spread, and recovery of infected cells. In addition, these factors may express themselves with different intensity among different genotypes, suggesting that uninfected cells do not have sufficient acquired immunity to recover the host from infection (Goyal et al., 2022). In this way, resistant genotypes have a more effective immune response responding more quickly against infection, due to several factors including localized immune efficacy (Escobar et al., 2003, Groenenboom et al., 2005). The difference in stress response time also involves the perception of changes in the environment by plant tissues and directly interferes with the cell cycle. This perception involves the activation of CDKs, which are the central regulators of this cycle. Cyclin-dependent kinases (CDKs) play important roles as it relates to cell division and response to many intracellular and extracellular signals (Malumbres, 2014). This process results in the activation of a signaling cascade that prolongs the S phase

with delayed entry into mitosis (Kitsios et al., 2011). This reveals the importance of the PKs of the CMGC families to cucumber resistance to PM and RKN.

We then analyzed the expression profiles of the PK genes in response to ALS caused by *A. cucumerina*. Among our findings, different patterns of DEGs were observed in different families of PK. The resistant genotype showed a small number of both down- and up-regulated DEGs compared to the susceptible genotype (Figure 5-C). This pathogen exhibits a necrotrophic lifestyle, alike from the others. This means that the plant's response pattern is mainly based on the synthesis of toxins, such as, phenolic compounds, phytoalexins, and hydrogen peroxide, whose purpose is to stop the pathogen's development more quickly (Akter et al., 2015, Debona et al., 2012, Kaur et a., 2022). In fact, the response of the resistant genotype is related to the genes of quantitative effect, which involves processes related to controlling physical traits and developmental processes, enhancing the plant's fundamental defense mechanisms, encoding enzymes for neutralizing phytotoxins produced by pathogens, aiding in the transmission of defense signals, including genes associated with the circadian clock, and incorporating weaker variants of R genes (Zhang et al., 2013). The predicted function of the genes is related to chloroplast, chloroplast thylakoid lumen, cytoplasm, endomembrane system, extracellular space, nucleus and plasma membrane in biological processes, such as, activation of protein kinase activity, anatomical structure morphogenesis, aromatic compound biosynthetic process, blue light signaling pathway, brassinosteroid mediated signaling pathway, defense response to bacterium/fungus, positive regulation of abscisic acid-activated signaling pathway, proteolysis, reactive oxygen species metabolic process, and MAPK cascade (Supplementary Table S6).

Finally, we also analyzed the expression of PK genes in response to RKN (*M. incognita*). The response to this disease showed a pattern like that caused by PM, in which the same families participated in the responses, with differences observed only concerning response time (Figures 5 and 6). Genes of RLK family are more important in the response to RKN infection, with similar response patterns between resistant and susceptible genotypes, but with greater differences related to response time. The kinases of the RLK-Pelle_DLSV family play an important role in the regulation of gene expression, as it presented a greater number of genes, facilitating signal transduction when pathogens threaten. However, RLK genes were the only ones in common between the two pathosystems whose pathogens are biotrophic, PM and RKN. In addition to the RLK family, other DEGs only observed in response to RKN belonged to the AGC_PDK1 and TKL-Pl-6 families (Supplementary Table S13). AGC_PDK1 acts positively

in regulating basal disease resistance in rice upstream of the OsOxi1-OsPti1a phosphorylation cascade (Matsui et al., 2010). CMGC genes were predicted only in the nucleus, unlike PM-responsive CMGC genes, which were also observed in the cytoplasm. They activate immune responses such as hypersensitivity reactions and systemic acquired resistance, phosphorylating immunity-related proteins to bolster defenses. In addition, these kinases regulate gene expression, favoring the synthesis of defense proteins. By phosphorylating defense-related proteins, including pathogen receptors, CMGC kinases increase a plant's ability to recognize and neutralize pathogens (Benschop et al., 2007; Tena et al., 2011), but it is important to note that other components of the cell determine the response time to the disease triggered by PK signaling, which still needs to be investigated. Further studies in this area are needed to elucidate these processes involving the response to diseases.

4.4 Kinase duplication and synteny analyses

Our detailed analysis of the duplication events revealed significant insights into the cucumber kinome. The estimates performed for 833 kinases identified the predominance of doubling events, highlighting notable contributions of total genomic duplications (WGD) and tandem duplications. The predominance of WGD events, which account for approximately 69.4% of kinases in cucumber, suggests a key role of these duplications in the evolution of the kinome. These events can be associated with specific spikes in synonymous substitution rates (Ks), indicating significant duplications around 50.41 and 422.81 million years ago (MYA). Notably, this last interval coincides with polyploidization events in the eudicot lineage, pointing to a possible influence of these events on the diversification of these kinases.

Analysis of the non-synonymous substitution rate (Ka) revealed that most kinases in cucumber are under purifying selection, indicating an evolutionary pressure to preserve the structure and stabilize the function of these proteins. This pattern is consistent with other studies on different gene families in common bean (Aono et al., 2023), suggesting a general trend in functional conservation across the genome. The identification of 237 kinases with tandem duplications is especially interesting. Among these, notable families, such as RLK-Pelle_DLSV, exhibited a significant number of protein duplications. These tandem duplications are known to be associated with stress responses, indicating a possible adaptive expansion of these subfamilies to address environmental challenges (Freeling, 2009; Hu et al., 2022; Leister et al., 2004; Panchy et al., 2016).

When comparing our results with studies in other plant species, we observed similar patterns regarding the presence of duplications under purifying selection. However, we highlight the uniqueness of our findings, especially in terms of high rates of Ks and distant dates of doubling for the kinases. These traits may be specific to the evolution of kinases in this species and may influence functional diversity and long-term adaptation.

5 CONCLUSION

We underscore the pivotal role of Protein Kinases (PKs) in cucumber immunity and disease resistance. Our comprehensive analysis delineates the characteristics of cucumber's 835 PKs, categorized into 123 families and 20 groups. Notably, only 312 PKs demonstrated significant relevance in response to biotic stress, further classified into 10 groups: AGC, BUB, CAMK, CMGC, RLK, STE, TKL, ULK, WEE, and WNK. This investigation establishes a solid groundwork for future exploration into the specific functional roles of these PKs, paving the way for the development of targeted strategies to enhance disease resistance in cucumbers and other Cucurbitaceae species. The outcomes of this research contribute significantly to the progression of agricultural practices, offering valuable insights for bolstering crop resilience against diverse disease challenges.

REFERENCES

(References were formatted according to the standards of the Scientia Horticulturae journal)

Abul-Haya, Z.M., 1982. Growth and sporulation of Sphaerotheca fuliginea on resistant and susceptible cucumber genotypes. Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz / Journal of Plant Diseases and Protection 89, 671–677.

Akter, S., Huang, J., Waszczak, C., Jacques, S., Gevaert, K., Van Breusegem, F., Messens, J., 2015. Cysteines under ROS attack in plants: a proteomics view. J Exp Bot 66, 2935–2944. https://doi.org/10.1093/jxb/erv044

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

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

Baker, B.M., Uchenna, H., White, A.L., Edwards, J.J., Jamison, K.M., Abdellaoui, Y., Henderson, J.S., Goodrich, S., Newman, R.H., Drewery, D.H., 2023. Development of a Genetically Encodable Fluorescent Biosensor to Examine the Spatiotemporal Regulation of NEK Kinases in Living Cells, in: ASPET 2023 Annual Meeting Abstract - Cellular and Molecular

Pharmacology. American Society for Pharmacology and Experimental Therapeutics, p. 548. https://doi.org/10.1124/jpet.122.266850

Batta, Y., 2003. Alternaria Leaf Spot Disease on Cucumber: Susceptibility and Control Using Leaf Disk Assay. An-Najah University Journal for Research - A (Natural Sciences) 17, 269–279. https://doi.org/10.35552/anujr.a.17.2.635

Baulin, E.F., Kulakovskiy, I. V., Roytberg, M.A., Astakhova, T. V., 2020. Brain-related genes are specifically enriched with long phase 1 introns. PLoS One 15, e0233978. https://doi.org/10.1371/journal.pone.0233978

Benschop, J.J., Mohammed, S., O'Flaherty, M., Heck, A.J.R., Slijper, M., Menke, F.L.H., 2007. Quantitative Phosphoproteomics of Early Elicitor Signaling in Arabidopsis. Molecular & Cellular Proteomics 6, 1198–1214. https://doi.org/10.1074/mcp.M600429-MCP200

Carbonnel, M., Davis, P., Roelfsema, M.R.G., Inoue, S., Schepens, I., Lariguet, P., Geisler, M., Shimazaki, K., Hangarter, R., Fankhauser, C., 2010. The Arabidopsis PHYTOCHROME KINASE SUBSTRATE2 Protein Is a Phototropin Signaling Element That Regulates Leaf Flattening and Leaf Positioning. Plant Physiol 152, 1391–1405. https://doi.org/10.1104/pp.109.150441

Champion, A., Kreis, M., Mockaitis, K., Picaud, A., Henry, Y., 2004. Arabidopsis kinome: after the casting. Funct Integr Genomics 4. 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

Cheng, C., Chen, J., 2022. Cyto-Molecular Genetics of the Interspecific Hybridization in Cucumber. pp. 121–144. https://doi.org/10.1007/978-3-030-88647-9_10

Choudhary, B., Fageria, M., 2002. Breeding for multiple disease resistance in cucurbits (water melon, musk melon cucumber and squash) – a review. Agricultural Reviews 23, 300–304.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J.L., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Conesa, A., Götz, S., 2008. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics 2008, 1–12. https://doi.org/10.1155/2008/619832

Dahan, J., Wendehenne, D., Ranjeva, R., Pugin, A., Bourque, S., 2010. Nuclear protein kinases: still enigmatic components in plant cell signalling. New Phytologist 185, 355–368. https://doi.org/10.1111/j.1469-8137.2009.03085.x

Dambroz, C.M. da S., Aono, A.H., Silva, E.M. de A., Pereira, W.A., 2023. Genome-wide analysis and characterization of the LRR-RLK gene family provides insights into anthracnose resistance in common bean. Sci Rep 13, 13455. https://doi.org/10.1038/s41598-023-40054-3

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

Debona, D., Rodrigues, F.Á., Rios, J.A., Nascimento, K.J.T., 2012. Biochemical Changes in the Leaves of Wheat Plants Infected by *Pyricularia oryzae*. Phytopathology 102, 1121–1129. https://doi.org/10.1094/PHYTO-06-12-0125-R

Delormel, T.Y., Boudsocq, M., 2019. Properties and functions of calcium-dependent protein kinases and their relatives in *Arabidopsis thaliana*. New Phytologist 224, 585–604. https://doi.org/10.1111/nph.16088

Desam, N.R., Al-Rajab, A.J., 2021. The Importance of Natural Products in Cosmetics. pp. 643–685. https://doi.org/10.1007/978-3-030-54027-2_19

Deshmukh, K., Anamika, K., Srinivasan, N., 2010. Evolution of domain combinations in protein kinases and its implications for functional diversity. Prog Biophys Mol Biol 102, 1–15. https://doi.org/10.1016/j.pbiomolbio.2009.12.009

Dey, S.S., Singh, S., Munshi, A.D., Behera, T.K., 2022. Classical Genetics and Traditional Breeding. pp. 159–183. https://doi.org/10.1007/978-3-030-88647-9_12

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

Escobar, M., Dandekar, A., 2003. Post-transcriptional gene silencing in plants, in: Noncoding RNAs: Molecular Biology and Molecular Medicine. Kluwer Academic, Dordrecht.

Fassuliotis, G., Rau, G., 1963. Evaluation of Cucumis spp. for resistance to the cotton rootknot nematode, Meloidogyne incognita acrita. Plant Disease Reporter 47.

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., 2021. The Cowpea Kinome: Genomic and Transcriptomic Analysis Under Biotic and Abiotic Stresses. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.667013

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

Fischer, I., Diévart, A., Droc, G., Dufayard, J.-F., Chantret, N., 2016. Evolutionary Dynamics of the Leucine-Rich Repeat Receptor-Like Kinase (LRR-RLK) Subfamily in Angiosperms. Plant Physiol 170, 1595–1610. https://doi.org/10.1104/pp.15.01470

Freeling, M., 2009. Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. Annu Rev Plant Biol 60, 433–453. https://doi.org/10.1146/annurev.arplant.043008.092122

Gaut, B.S., Morton, B.R., McCaig, B.C., Clegg, M.T., 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proceedings of the National Academy of Sciences 93, 10274–10279. https://doi.org/10.1073/pnas.93.19.10274

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40, D1178–D1186. https://doi.org/10.1093/nar/gkr944

Goyal, N., Bhatia, G., Garewal, N., Upadhyay, A., Singh, K., 2021. Identification of defense related gene families and their response against powdery and downy mildew infections in Vitis vinifera. BMC Genomics 22, 776. https://doi.org/10.1186/s12864-021-08081-4

Groenenboom, M.A.C., Marée, A.F.M., Hogeweg, P., 2005. The RNA Silencing Pathway: The Bits and Pieces That Matter. PLoS Comput Biol 1, e21. https://doi.org/10.1371/journal.pcbi.0010021

Haider, M.S., De Britto, S., Nagaraj, G., Gurulingaiah, B., Shekhar, R., Ito, S., Jogaiah, S., 2021. Genome-Wide Identification, Diversification, and Expression Analysis of Lectin Receptor-Like Kinase (LecRLK) Gene Family in Cucumber under Biotic Stress. Int J Mol Sci 22, 6585. https://doi.org/10.3390/ijms22126585

Hanks, S.K., Hunter, T., 1995. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification [1]. The FASEB Journal 9, 576–596. https://doi.org/10.1096/fasebj.9.8.7768349

Hardie, D.G., 1999. PLANT PROTEIN SERINE/THREONINE KINASES: Classification and Functions. Annu Rev Plant Physiol Plant Mol Biol 50, 97–131. https://doi.org/10.1146/annurev.arplant.50.1.97

Holub, E.B., 2001. The arms race is ancient history in Arabidopsis, the wildflower. Nat Rev Genet 2, 516–527. https://doi.org/10.1038/35080508

Hu, X., Hao, J., Pan, L., Xu, T., Ren, L., Chen, Y., Tang, M., Liao, L., Wang, Z., 2022. Genome-wide analysis of tandem duplicated genes and their expression under salt stress in seashore paspalum. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.971999

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, 2009. The genome of the cucumber, Cucumis sativus L. Nat Genet 41, 1275–1281. https://doi.org/10.1038/ng.475

Huberdeau, M.Q., Simard, M.J., 2019. A guide to micro <scp>RNA</scp> -mediated gene silencing. FEBS J 286, 642–652. https://doi.org/10.1111/febs.14666

Jilani, M., Bakar, A., Waseem, K., Kiran, M., 2009. Effect of different nitrogen levels on growth and yield of cucumber (Cucumis sativus L.). Journal of Agriculture & Social Sciences 5, 99–101.

Johnson, L.N., Noble, M.E.M., Owen, D.J., 1996. Active and Inactive Protein Kinases: Structural Basis for Regulation. Cell 85, 149–158. https://doi.org/10.1016/S0092-8674(00)81092-2

Kamoun, S., Huitema, E., Vleeshouwers, V.G.A.A., 1999. Resistance to oomycetes: a general role for the hypersensitive response? Trends Plant Sci 4, 196–200. https://doi.org/10.1016/S1360-1385(99)01404-1

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

Kaur, S., Samota, M.K., Choudhary, Manoj, Choudhary, Mukesh, Pandey, A.K., Sharma, A., Thakur, J., 2022. How do plants defend themselves against pathogens-Biochemical mechanisms and genetic interventions. Physiology and Molecular Biology of Plants 28, 485–504. https://doi.org/10.1007/s12298-022-01146-y

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

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kitsios, G., Doonan, J.H., 2011. Cyclin dependent protein kinases and stress responses in plants. Plant Signal Behav 6, 204–209. https://doi.org/10.4161/psb.6.2.14835

Kliebenstein, D.J., Rowe, H.C., 2008. Ecological costs of biotrophic versus necrotrophic pathogen resistance, the hypersensitive response and signal transduction. Plant Science 174, 551–556. https://doi.org/10.1016/j.plantsci.2008.03.005

Lee, Y., Lloyd, A.M., Roux, S.J., 1999. Antisense Expression of the CK2 α-Subunit Gene in Arabidopsis. Effects on Light-Regulated Gene Expression and Plant Growth1. Plant Physiol 119, 989–1000. https://doi.org/10.1104/pp.119.3.989

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

Leister, D., 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends in Genetics 20, 116–122. https://doi.org/10.1016/j.tig.2004.01.007

Li, Q., Li, H., Huang, W., Xu, Y., Zhou, Q., Wang, S., Ruan, J., Huang, S., Zhang, Z., 2019. A chromosome-scale genome assembly of cucumber (Cucumis sativus L.). Gigascience 8. https://doi.org/10.1093/gigascience/giz072

Li, X., Sun, Y., Yang, Y., Yang, X., Xue, W., Wu, M., Chen, P., Weng, Y., Chen, S., 2021. Transcriptomic and Histological Analysis of the Response of Susceptible and Resistant Cucumber to Meloidogyne incognita Infection Revealing Complex Resistance via Multiple Signaling Pathways. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.675429

Li, Z., Zhang, Z., Yan, P., Huang, S., Fei, Z., Lin, K., 2011. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. BMC Genomics 12, 540. https://doi.org/10.1186/1471-2164-12-540

Lin, S.-L., Miller, J.D., Ying, S.-Y., 2006. Intronic MicroRNA (miRNA). J Biomed Biotechnol 2006, 1–13. https://doi.org/10.1155/JBB/2006/26818

Liu, J., Chen, N., Grant, J.N., Cheng, Z.-M. (Max), 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

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Luan, S., Kudla, J., Rodriguez-Concepcion, M., Yalovsky, S., Gruissem, W., 2002. Calmodulins and Calcineurin B–like Proteins. Plant Cell 14, S389–S400. https://doi.org/10.1105/tpc.001115

Magalhães, D.M., Scholte, L.L.S., Silva, N. V., Oliveira, G.C., Zipfel, C., Takita, M.A., De Souza, A.A., 2016. LRR-RLK family from two Citrus species: genome-wide identification and evolutionary aspects. BMC Genomics 17, 623. https://doi.org/10.1186/s12864-016-2930-9

Malumbres, M., 2014. Cyclin-dependent kinases. Genome Biol 15, 122. https://doi.org/10.1186/gb4184

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

Matsui, H., Miyao, A., Takahashi, A., Hirochika, H., 2010. Pdk1 Kinase Regulates Basal Disease Resistance Through the OsOxi1–OsPti1a Phosphorylation Cascade in Rice. Plant Cell Physiol 51, 2082–2091. https://doi.org/10.1093/pcp/pcq167

Meggio, F., Pinna, L.A., 2003. One-thousand-and-one substrates of protein kinase CK2? The FASEB Journal 17, 349–368. https://doi.org/10.1096/fj.02-0473rev

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5, 621–628. https://doi.org/10.1038/nmeth.1226

Moustafa, M., Abd-el-Shahid, Y., Ez-Eldin, I., Anwar, H., 1990. Factors affecting acquired resistance in Erysiphe cichoracearum the causal organism of cucumber powdery mildew. Agricultural Research Review 68, 521–528.

Mukhtar, T., Kayani, M.Z., Hussain, M.A., 2013. Response of selected cucumber cultivars to Meloidogyne incognita. Crop Protection 44, 13–17. https://doi.org/10.1016/j.cropro.2012.10.015

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

Ni, W., Xu, S.-L., González-Grandío, E., Chalkley, R.J., Huhmer, A.F.R., Burlingame, A.L., Wang, Z.-Y., Quail, P.H., 2017. PPKs mediate direct signal transfer from phytochrome photoreceptors to transcription factor PIF3. Nat Commun 8, 15236. https://doi.org/10.1038/ncomms15236

Panchy, N., Lehti-Shiu, M., Shiu, S.-H., 2016. Evolution of Gene Duplication in Plants. Plant Physiol 171, 2294–2316. https://doi.org/10.1104/pp.16.00523

Pandey, S., Kujur, S.N., 2022. Importance, Distribution, Botany and Genetics. pp. 1–13. https://doi.org/10.1007/978-3-030-88647-9_1

Peck, S.C., 2001. Directed Proteomics Identifies a Plant-Specific Protein Rapidly Phosphorylated in Response to Bacterial and Fungal Elicitors. THE PLANT CELL ONLINE 13, 1467–1475. https://doi.org/10.1105/tpc.13.6.1467

PÉREZ-GARCÍA, A., ROMERO, D., FERNÁNDEZ-ORTUÑO, D., LÓPEZ-RUIZ, F., DE VICENTE, A., TORÉS, J.A., 2009. The powdery mildew fungus *Podosphaera fusca* (synonym *Podosphaera xanthii*), a constant threat to cucurbits. Mol Plant Pathol 10, 153–160. https://doi.org/10.1111/j.1364-3703.2008.00527.x

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One 5, e9490. https://doi.org/10.1371/journal.pone.0009490

Qi, J., Liu, X., Shen, D., Miao, H., Xie, B., Li, X., Zeng, P., Wang, S., Shang, Y., Gu, X., Du, Y., Li, Y., Lin, T., Yuan, J., Yang, X., Chen, J., Chen, H., Xiong, X., Huang, K., Fei, Z., Mao, L., Tian, L., Städler, T., Renner, S.S., Kamoun, S., Lucas, W.J., Zhang, Z., Huang, S., 2013. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet 45, 1510–1515. https://doi.org/10.1038/ng.2801

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, 2742. https://doi.org/10.1038/ncomms3742

Romeis, T., 2001. Protein kinases in the plant defence response. Curr Opin Plant Biol 4, 407–414. https://doi.org/10.1016/S1369-5266(00)00193-X

Rose, A.B., 2008. Intron-Mediated Regulation of Gene Expression. pp. 277–290. https://doi.org/10.1007/978-3-540-76776-3_15

Sa, R., Liu, D., Chen, L., Liu, L., Zhang, Y., Zhang, X., Xu, C., 2020. 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

Sakamoto, T., Deguchi, M., Brustolini, O.J., Santos, A.A., Silva, F.F., Fontes, E.P., 2012. The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRII-RLK subfamily in antiviral defense. BMC Plant Biol 12, 229. https://doi.org/10.1186/1471-2229-12-229

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

Sarhan, E.A.D., Abd-Elsyed, M.H.F., Ebrahiem, A.M.Y., 2020. Biological control of cucumber powdery mildew (Podosphaera xanthii) (Castagne) under greenhouse conditions. Egypt J Biol Pest Control 30, 65. https://doi.org/10.1186/s41938-020-00267-4

Savojardo, C., Martelli, P.L., Fariselli, P., Profiti, G., Casadio, R., 2018. BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acids Res 46, W459–W466. https://doi.org/10.1093/nar/gky320

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

Slowikowski, K., Schep, A., Hughes, S., Dang, T.K., Lukauskas, S., Irisson, J.-O., 2018. ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2."

Sugano, S., Andronis, C., Ong, M.S., Green, R.M., Tobin, E.M., 1999. The protein kinase CK2 is involved in regulation of circadian rhythms in *Arabidopsis*. Proceedings of the National Academy of Sciences 96, 12362–12366. https://doi.org/10.1073/pnas.96.22.12362

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

Tena, G., Boudsocq, M., Sheen, J., 2011. Protein kinase signaling networks in plant innate immunity. Curr Opin Plant Biol 14, 519–529. https://doi.org/10.1016/j.pbi.2011.05.006

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

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Tsirigos, K.D., Peters, C., Shu, N., Käll, L., Elofsson, A., 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res 43, W401–W407. https://doi.org/10.1093/nar/gkv485

Vakalounakis, D.J., 1990. Host Range of Alternaria alternata f. sp. cucurbitae Causing Leaf Spot of Cucumber. Plant Dis 74, 227. https://doi.org/10.1094/PD-74-0227

Vakalounakis, D.J., Malathrakis, N.E., 1988. A Cucumber Disease Caused by Alternaria alternata and its Control. Journal of Phytopathology 121, 325–336. https://doi.org/10.1111/j.1439-0434.1988.tb00976.x

Vidhyasekaran, P., 2014. Protein Phosphorylation and Dephosphorylation in Plant Immune Signaling Systems. pp. 385–408. https://doi.org/10.1007/978-94-007-7426-1_9

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, 2527. https://doi.org/10.3390/ijms20102527

Wei, K., Wang, Y., Xie, D., 2014. 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, S., Li, X., Lu, Z., Zhang, H., Ye, X., Zhou, Yujie, Li, J., Yan, Y., Pei, H., Duan, F., Wang, D., Chen, S., Wang, P., Zhang, C., Shang, L., Zhou, Yue, Yan, P., Zhao, M., Huang, J., Bock, R., Qian, Q., Zhou, W., 2022. A transcriptional regulator that boosts grain yields and shortens the growth duration of rice. Science (1979) 377. https://doi.org/10.1126/science.abi8455

Weng, Y., 2022. The Cucumber Genome—An Update. pp. 25–35. https://doi.org/10.1007/978-3-030-88647-9_3

Whelan, S., Goldman, N., 2001. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol Biol Evol 18, 691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851

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

Wóycicki, R., Witkowicz, J., Gawroński, P., Dąbrowska, J., Lomsadze, A., Pawełkowicz, M., Siedlecka, E., Yagi, K., Pląder, W., Seroczyńska, A., Śmiech, M., Gutman, W., Niemirowicz-Szczytt, K., Bartoszewski, G., Tagashira, N., Hoshi, Y., Borodovsky, M., Karpiński, S., Malepszy, S., Przybecki, Z., 2011. The Genome Sequence of the North-European Cucumber (Cucumis sativus L.) Unravels Evolutionary Adaptation Mechanisms in Plants. PLoS One 6, e22728. https://doi.org/10.1371/journal.pone.0022728

Xu, Q., Xu, X., Shi, Y., Qi, X., Chen, X., 2017. 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

Yan, J., Li, G., Guo, X., Li, Y., Cao, X., 2018. Genome-wide classification, evolutionary analysis and gene expression patterns of the kinome in Gossypium. PLoS One 13, e0197392. https://doi.org/10.1371/journal.pone.0197392

Yang, L., Koo, D.-H., Li, Y., Zhang, X., Luan, F., Havey, M.J., Jiang, J., Weng, Y., 2012. Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. The Plant Journal 71, 895–906. https://doi.org/10.1111/j.1365-313X.2012.05017.x

Yu, J., Wu, S., Sun, H., Wang, X., Tang, X., Guo, S., Zhang, Z., Huang, S., Xu, Y., Weng, Y., Mazourek, M., McGregor, C., Renner, S.S., Branham, S., Kousik, C., Wechter, W.P., Levi, A., Grumet, R., Zheng, Y., Fei, Z., 2023. CuGenDBv2: an updated database for cucurbit genomics. Nucleic Acids Res 51, D1457–D1464. https://doi.org/10.1093/nar/gkac921

Yu, J., Zhang, B., Liu, S., Guo, W., Gao, Y., Sun, H., 2022. Genome-wide characterisation, evolution and expression analysis of the leucine-rich repeat receptor-like kinase (LRR-RLK) gene family in cucumbers. Plant Protection Science 58, 125–138. https://doi.org/10.17221/131/2021-PPS

Zhang, B., Chen, H.-W., Mu, R.-L., Zhang, W.-K., Zhao, M.-Y., Wei, W., Wang, F., Yu, H., Lei, G., Zou, H.-F., Ma, B., Chen, S.-Y., Zhang, J.-S., 2011. NIMA-related kinase NEK6 affects plant growth and stress response in Arabidopsis. The Plant Journal 68, 830–843. https://doi.org/10.1111/j.1365-313X.2011.04733.x

Zhu, K., Wang, X., Liu, J., Tang, J., Cheng, Q., Chen, J.-G., Cheng, Z.-M., 2018. The grapevine kinome: annotation, classification and expression patterns in developmental processes and stress responses. Hortic Res 5, 19. https://doi.org/10.1038/s41438-018-0027-0

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

ABBREVIATIONS LIST

AGC (cAMP-dependent protein kinases cGMP-dependent protein kinases various types of protein kinase C protein kinase B 3-phosphoinositide-dependent protein kinase-1 and the ribosomal protein S6 kinases)

ALS (Alternaria Leaf Spot)

Aur (aurora kinase)

BUB (budding uninhibited by benzimidazoles)

C (cellular component)

CAMK (calcium/calmodulin-dependent protein kinase)

CDK (Cyclin-dependent kinase)

CDPK (Calcium-Dependent Protein Kinase)

CK1 (casein kinase 1)

CMGC (cyclin-dependent kinase mitogen-activated protein kinase glycogen synthase kinase and cyclin-dependent-like kinase families)

CPU (central processor unit)

CRK (CDPK-Related Protein Kinase)

Csa (Cucumis sativus)

CuGenDbv2 (Cucurbit Genomics Database v2)

DEG (Differentially Expressed Gene)

F (molecular function)

FDR (false discovery rate)

GRAVY (Grand Average of Hydropathy)

Gene Ontology (GO)

HMM (Hidden Markov Model)

IRE1 (inositol-requiring enzyme 1)

Ka (Non-synonymous substitution rates)

Ks (Synonymous substitution rates)

LRR (Leucine Rich Repeat)

LecRLK (Lectin Receptor-Like Kinase)

MAPK (Mitogen-Activated Protein Kinase)

MYA (million years ago)

NAK (NF-κB-activating kinase)

NEK (never in mitosis gene-A)

P (biological process)

PEK (pancreatic eukaryotic initiation factor 2 α-subunit kinase)

PK (Protein Kinase)

PM (Powdery Mildew)

PPI (pathogen-plant interaction)

Pkinase (Protein Serine/Threonine Kinase)

Pkinase_Tyr (Protein Tyrosine Kinase)

Plant specific (Group-Pl)

RKN (Root-Knot Nematode)

RLK (receptor-like kinase)

RNA-Seq (RNA sequencing)

SCY (Saccharomyces cerevisiae [yeast] kinase)

STE (serine/threonine kinase)

TKL (tyrosine kinase-like kinase)

TLK (tousled-like kinase)

TM (transmembrane domain)

TTK (threonine/tyrosine kinase)

ULK (unc-51-like kinase)

WAG (Whelan Goldman)

WEE (wee1 wee2 and myt1 kinases)

WGD (Whole genome duplication)

WNK (with no lysine-K)

aa (amino acid)

iP (Isoelectric Point)

log2fc (log2 fold change)

ARTICLE 2 - UNVEILING THE COMPLEXITY OF THE CIS-REGULATORY ELEMENTS ACTING ON THE MODULATION OF THE RESPONSE OF CUCUMBER DISEASE-RESPONSIVE KINASE GENES

(References were formatted according to the standards of the Scientia Horticulturae journal)

ABSTRACT

Understanding how plants regulate responses to diseases is crucial for developing resistant cultivars and ensuring food security. This study aimed to identify and analyze cis-regulatory elements (CREs) in promoter regions of protein kinase superfamily genes in Cucumis sativus challenged by Powdery Mildew, Alternaria Leaf Spot, and Root-Knot Nematode. Differentially expressed genes (DEGs) were identified, and their 2 Kbp upstream sequences were analyzed for CREs. These sequences were analyzed for the availability of CREs, which were classified, quantified, and analyzed. Overall, 51,339 CREs of 125 different types were obtained. These, in turn, were distributed into seven categories, with CREs from the Core category being the most abundant. The Core includes the following CREs: TATA-box, CAAT-box, AT-TATA-box, Unnamed__1, TATA, AT-rich element, A-box, and AT-rich sequence. Among the most represented CREs, TATA-box, CAAT-box and Unnamed__4 stand out. On the other hand, among the less frequent are 4cl-CMA2b, DRE, GGA-motif, H-box, JERE, NON, Pc-CMA2c, chs-CMA2c, chs-Unit_1_m1, plant_AP-2-like, silencer and telo-box. The similarity between the PPRs was evaluated, revealing 70 promoters (35 pairs) with similarity greater than 0.5, as well as 12 promoters who present similarity with other promoters, evidenced in the range of 62 to 89 connections. It was possible to identify and highlight a set of CREs potentially essential for modulating the expression of key genes that may be involved in plant defense pathways. In addition, the coexpression of these genes was investigated, identifying communities (hubs) under conditions of control and stress, evidencing the existence of putative metabolic pathways or immunological contexts in which these genes can act in a coordinated manner. Based on the results of this study, a series of possibilities of gene silencing based on CREs is considered as a methodological possibility for research, to potentiate or suppress the expression of specific genes, without them being directly affected, and, thus, evaluate the dynamics and subsequent results for specific genotypes.

KEYWORDS: Genomic Regulation; Plant Immunity; Pathogen-Responsive Networks; Plant Immunology; Cucumber Defense Mechanisms; Disease-Responsive Kinase Genes.

1 INTRODUCTION

As sessile organisms, plants' main challenge is to overcome the most varied types of environmental stresses and maintain a productivity that provides them with evolutionary success. Under adverse conditions, plants show marked phenotypic plasticity, both at the physiological and morphological levels, resulting from transcriptional, translational, and post-translational adaptations, which allow them to survive, even in unfavorable environments (Jacinto Junior & Lucena, 2022; Windram et al., 2014). Transcriptional responses to external stimuli often encompass a large group of gene promoters, which can be analyzed and categorized in a variety of ways. This categorization can consider induction kinetics, cell type, signal specificity, and response duration (Pope & Medzhitov, 2018). This complexity shows the diversity and adaptability of molecular processes underlying the interactions between cells and their environment.

At the molecular level, adaptability translates into the form of reprogramming of gene transcription in terms of space, time, and quantity. It can be stated that at the base of this modulation are cis-regulatory elements (CREs), transcription factors (TFs), and all specific interactions, which result in the transcription of genes responsible for the molecular response to stress. All these adaptations underscore the collective and multifactorial nature of the response and adaptation of plants to stress, involving a precise reprogramming of genetic transcription from the participation of various CREs and TFs.

TFs are elements that bind to CREs to modulate the expression of the genes that contain them (Peremarti et al., 2010; Porto et al., 2014). In other words, TFs are a type of proteins responsible for the regulation of gene expression by binding to specific CREs in the promoter region of genes (Latchman, 2013). They function by activating or repressing transcription, and can be regulated by mechanisms, such as protein-protein interactions and phosphorylation (Latchman, 2013). TFs can be categorized into families based on their DNA-binding regions (Bulky & Walhout, 2013). Several types of DNA-binding domains are available, some of which are specific to lineages, while others are more widely distributed. Intriguingly, only three types of DNA-binding domains are common in all kingdoms of life: cold shock, Helix-turn-helix (HTH) type 3, and HTH psq (Charoensawan et al., 2010).

Each gene has a TF that acts alone or in combination with other TFs by activating its transcription in a specific environmental condition through binding to specific DNA sequence (DNA-binding sites) that exert a cis-regulatory effect on one or more genes (FitzPatrick, 2017).

However, environmental conditions imprint several types of stresses on plants simultaneously, leading to reprogramming that may involve complex relationships between CREs, such as presence or absence, combinatorial relationships, location in relation to the transcription start site (TSS), and number of copies of the elements (Pilpel et al., 2001; Zou et al., 2011). Identifying all genes in a genome that encode TFs is not a trivial task because it can be challenging to recognize protein domains based on sequence alone, and because of ambiguity in the functionality of the proposed protein (Bulky & Walhout, 2013). On the other hand, it is possible to identify CREs and, by this means, predict potential effects of their presence or absence.

Significant advances have been made in understanding the mechanisms underlying gene expression reprogramming, highlighting the interaction between TFs and a specific set of CREs (Peremarti et al., 2010; Porto et al., 2014). For the Gy14 lineage, of the species *Cucumis sativus* L., about 1,476 genes encoding TFs are predicted (Zheng et al., 2016), many of which are associated with the regulation of gene transcription in response to various stresses. Among these TFs, the bHLH, ERF, and MYB families deserve to be highlighted for playing an essential role in *Cucumis melo* resistance to Powdery Mildew (Zhao et al., 2022). However, the CREs associated with most of these TFs remain unknown in *C. sativus*, raising questions about the ability of known CREs to explain variations in gene expression as a function of stress. The complexity of CREs responsive to biotic stresses is evidenced by their classification into two distinct superfamilies (α and β) in *Arabidopsis thaliana* (Zou et al., 2011). In the superfamily of CREs responsive to biotic stresses, 12% have a sequence (TTGAC) like that found in the W-box, targets of WRKY TFs (Zou et al., 2011; Yu et al., 2001).

It is essential to consider the simultaneous evolution of CREs along with the evolution of stress-responsive genes (Purugganan, 1998). This becomes especially relevant due to the origin of resistance genes, which have their origin in Whole Genome Duplication (WGD) events, thus providing an essential raw material in the form of CREs. An extremely relevant factor to be considered lies in the coevolution of CREs in line with the evolutionary development of genes, notably those related to resistance. Understanding this process is essential, given the origin of these resistance genes through WGD in conjunction with the activity of transposable elements (TE), which, in turn, generate an essential raw material for the formation of CREs (Sundaram et al., 2014; Chuong et al., 2017; Cosby et al., 2019; Diehl et al., 2020; Sundaram & Wysocka, 2020).

Numerous transcriptome studies have been conducted in *Cucumis sativus* species (Berg et al., 2020; Dong et al., 2020; Sá et al., 2020; Xu et al., 2017; Wang et al., 2018), highlighting the significance of these investigations for understanding diseases affecting the species. This interest has prompted the publication of genomes from diverse lineages, thereby facilitating the characterization of gene families. In view of the above, the objective of this study was to identify and analyze the cis-regulatory elements (CREs) found in the putative promoter regions (PPRs) of genes of the protein kinase superfamily, which have been differentially expressed by *Cucumis sativus* when challenged by Powdery Mildew (PM), Alternaria Leaf Spot (ALS), and Root-Knot Nematode (RKN). To this end, we sought to identify the set of kinase genes differentially expressed in three important cucumber diseases, as well as the CREs related to the biotic stress response, investigating their association with upregulation or downregulation. In addition, we sought to characterize the regulatory elements underlying kinase genes in response to biotic stress, thus offering a deeper understanding of the gene regulation processes involved in this context. Finally, we built gene coexpression networks to reveal interconnections (communities or hubs) between kinase genes in response to different biotic stresses, aiming to identify CREs in each hub potentially linked to metabolic pathways in which these genes participate.

2 MATERIAL AND METHODS

2.1 Kinase DEGs and Cucumber 5'-UTR sequences

The identification of kinase genes was conducted by identifying protein kinases (PKs) through an alignment with Hidden Markov Models (HMM) in relation to the complete set of annotated proteins of the cucumber Gy14 genome. To this end, the consensus sequences Pkinase (PF00069) and Pkinase_Tyr (PF07714) obtained from PFAM (http://pfam.xfam.org/) (El-Gebali et al., 2019) were used as queries, and the HMMER tool (Finn et al., 2011) was used, with an E-value cutoff of 0.01. The annotated proteins were obtained from Cucurbit Genomics Database v2 (CuGenDbv2) (Yu et al., 2023). Subsequently, the identified PKs were investigated to define which of them were differentially expressed genes (DEGs) in the pathosystems Powdery Mildew (PM) (Xu et al., 2017), Alternaria Leaf Spot (ALS) (Sa et al., 2020), and Root-Knot Nematode (RKN) (Wang et al., 2018) (Table 1). This filtering was based on log2 fold change (log2fc) criteria > 1.5 and p-value < 0.05 with False Discovery Rate (FDR) correction. DEG analysis was carried out using the DESeq2 package (Love et al., 2014). These pathosystems were chosen to represent biotic stress caused by biotrophic and non-biotrophic

pathogens, and the data were obtained from previous studies by Costa et al. (2023, unpublished data).

Tab. 1 Description and origin of transcriptomic data used in this study.

Condition	Resistant Genotype	Susceptible genotype	Times of evaluation	Data origin
Powdery Mildew	SSL508-28	D8	0, 48 h	Xu et al. 2017
Alternaria Leaf Spot	D1322	BJ204	2, 6 dpi	Sa et al. 2020
Root-Knot Nematode	IL10-1	CC3	0, 3 dpi	Wang et al. 2018

h: hours post-inoculation, dpi: days post-inoculation.

After the identification of the DEGs, the genomic sequences of the potentially promoter regions (PPRs) of these genes were obtained. To this end, the extension of 2 kbp upstream of the TSS [-2000, -1] was established for this analysis, and these sequences were downloaded from the Cucurbit Genomic Database v2 (CuGenDBv2) (Yu et al., 2023). To predict CREs, we used the Plant Cis-Acting Regulatory Elements (PlantCARE) database (Lescot et al., 2002). Subsequently, the CREs obtained were analyzed according to categories previously known in the literature, with which these cis elements are related. This process resulted in a "consensual classification" of the CREs categories, integrating contributions from previous studies (Ablazov et al., 2016; ain-Ali et al., 2021; Chen et al., 2021; Chowdhury et al., 2023; Deshmukh et al., 2017; Howlader et al., 2017; Mongkolsiriwatana et al., 2009; Pépin et al., 2021; Shariatipour et al., 2020; Sun et al., 2018; Tripathi et al., 2018; Wang et al., 2023). Information on isoelectric point (pI) and molecular weight (MW) of DEG kinases were retrieved from Costa et al. (2023, unpublished data) and from them a digital gel was built that simulates the dispersion of PKs based on iP and MW, considering the upregulated DEGs.

2.2 Analysis of Cis-Regulatory Elements

Subsequently, CREs were studied in the context of upregulation and downregulation in relation to three diseases affecting cucumber, either jointly or separately. The CREs associated with genes that are upregulated and downregulated have been thoroughly examined, considering each disease separately. Subsequently, to deepen our understanding of the impact of these DEGs on plant resistance, we performed a comparative analysis focused on the suspicion of the R (resistant) and S (susceptible) genes. In this approach, we categorized genes as R genes if they were upregulated in the resistant genotype or downregulated in the susceptible genotype. Conversely, genes were classified as S genes if they were upregulated in the susceptible genotype or downregulated in the resistant genotype. Genes that exhibited common upregulation or downregulation in both resistant and susceptible genotypes were excluded from the analysis, as they did not conform to the expected expression profiles of R and S genes.

To assess similarity in the PPRs of the DEGs, an approach based on graph theory was adopted (Andolfo et al., 2019; Qi et al., 2011). Initially, an alignment of the 2 kbp was performed using the MAFFT software (https://www.ebi.ac.uk/jdispatcher/msa/mafft) (Katoh et al., 2013). From this alignment, a similarity matrix was generated using the PHYLIP v3.697 software (PHYLIP, 2024). To provide a deeper understanding of similarity, a non-directed weighted network (graph) was constructed using the igraph package (Csardi et al., 2006; Csárdi et al., 2024). Network visualization was performed using Cytoscape v3.10.1 (Shannon et al., 2003). For the construction of the network, a limit of similarity greater than 0.4 was considered.

2.3 Integrating Coexpression Network and Cis-Regulatory Elements in PK Genes

Considering the premise that coexpressed genes are simultaneously co-regulated and share the same CREs (Ge, 2011; Priest et al., 2009; Wang et al., 2009; Zou et al., 2011), our objective in analyzing coexpression networks was to identify CREs associated with biotic stress-responsive kinase genes. This identification was based on coexpression patterns observed in *C. sativus* under different experimental conditions, which encompassed six distinct genotypes and addressed three significant diseases for cucumber crop (Table 1).

We performed a Coexpression Network Analysis (CNA) of the gene set using the igraph package (Csardi et al., 2006; Csárdi et al., 2024). The total set of genes analyzed in the transcriptomes (Xu et al., 2017; Sá et al., 2020; Wang et al., 2018) had their counts normalized in terms of transcripts per million (TPM). From this normalized set, 312 Differentially Expressed Genes (DEGs) were selected for the construction of the coexpression network. In the

graph, genes are represented as nodes in the structure, while edges represent Pearson's correlation coefficients (r-values), with a minimum value set at 0.6. A binary adjacency matrix was generated from the correlation matrix, in which values of $r > 0.6$ were converted to 1, and values of $r < 0.6$ were converted to 0. This provided a clear representation of the coexpression relationships between the genes, highlighting those with more robust correlations in the network.

The parameters hub scores, betweeness centrality and edge betweeness were calculated and the structure of two distinct networks were analyzed, which represent control and stress conditions. To evaluate and compare these networks, we used a clustering algorithm with the Louvain method (Blondel et al., 2008). The hub scores were calculated using Kleinberg's hub centrality (Kleinberg, 1999), while the betweeness centrality was determined as proposed by Freeman (1979) and Brandes (2001). The edge betweeness values were quantified by the number of geodesics (shortest paths) that cross a connection (specific edge), according to the approach of Brandes (2001). Following the delineation of the communities within the two networks, coexpressed genes were pinpointed, and their corresponding CREs were scrutinized to ascertain their involvement in orchestrating responses amidst network disturbances.

3 RESULTS

3.1 Kinase DEGs and Cucumber 5'-UTR sequences

In this study, the identification of protein kinases (PKs) from the complete set of annotated proteins of *Cucumis sativus* was initially performed using HMM profiles. From this step, 835 PK-encoding genes were recruited, and from this set were extracted those responsive to biotic stress in cucumber, using the differential gene expression criteria. This filtering process, which included data from three gene expression studies, PM (Xu et al., 2017), ALS (Sa et al., 2020), and RKN (Wang et al., 2018), resulted in 515 differentially expressed genes (DEGs). In summary, 200 genes (~23.95%) showed differential expression in response to PM, 188 genes (22.51%) were affected by ALS, and, finally, 127 genes (15.21%) exhibited responsiveness to RKN. When comparing this data set, duplicates of the names of the genes in common between the pathosystems were removed, leaving 312 kinase genes (71 families) responsive to biotic stress (Figure 1, Supplementary Table S1). It is pertinent to acknowledge that removing duplicates was imperative for our analysis, given our focus on examining CREs associated with gene expression rather than assessing the gene set itself. Failure to remove duplicates would have led to redundant analysis of CREs, providing no additional insights.

Therefore, this step was essential to ensure the integrity and specificity of our investigation into CREs' involvement in biotic stress responses.

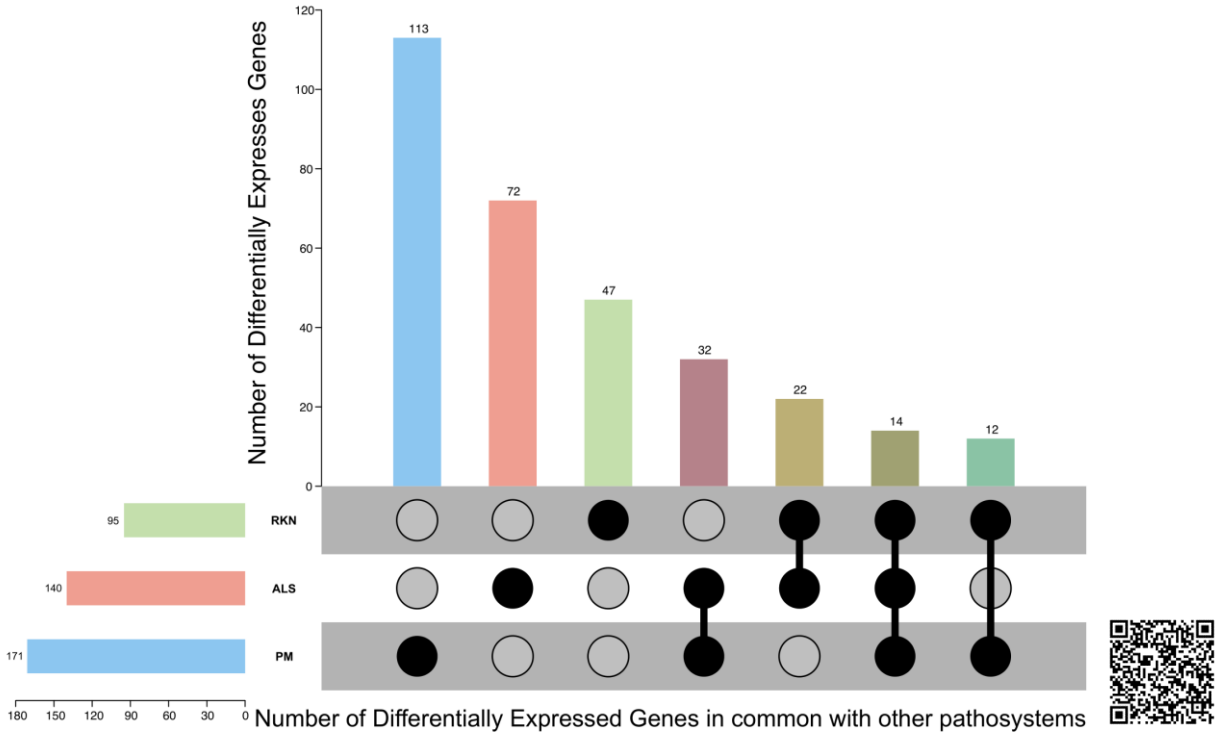

Fig. 1 Overlap of Biotic Stress-Responsive Kinase Genes Across Multiple Pathosystems in Cucumber. PM: Powdery Mildew (*Podosphaera xanthii*), ALS: Alternaria Leaf Spot (*Alternaria cucumerina*), RKN: Root-Knot Nematode (*Meloidogyne incognita*). The values represent the number of kinase genes with differential expression value delimited by the log2 fold change > 1.5 and p-value < 0.05 with FRD correction. The vertical columns represent the number of differentially expressed kinase genes. The horizontal columns represent the number of kinase genes in each stress condition separately. Dark circles represent the number of kinase genes in each subset. The circles connected with a black line represent the number of DEGs in common with two or three conditions. QR code for a link with a high-quality image.

In pursuit of a better understanding of differentially expressed protein kinase, a digital gel was simulated to visualize the distribution of upregulated kinases (Figure 2, Figure S1). The figure illustrates how these proteins are dispersed based on their molecular weight (MW - Y-axis) and isoelectric point (iP - X-axis) across three pathosystems, each comprising both resistant and susceptible cultivars. The visualization helps to understand the distribution pattern of upregulated protein kinases in response to these pathosystems.

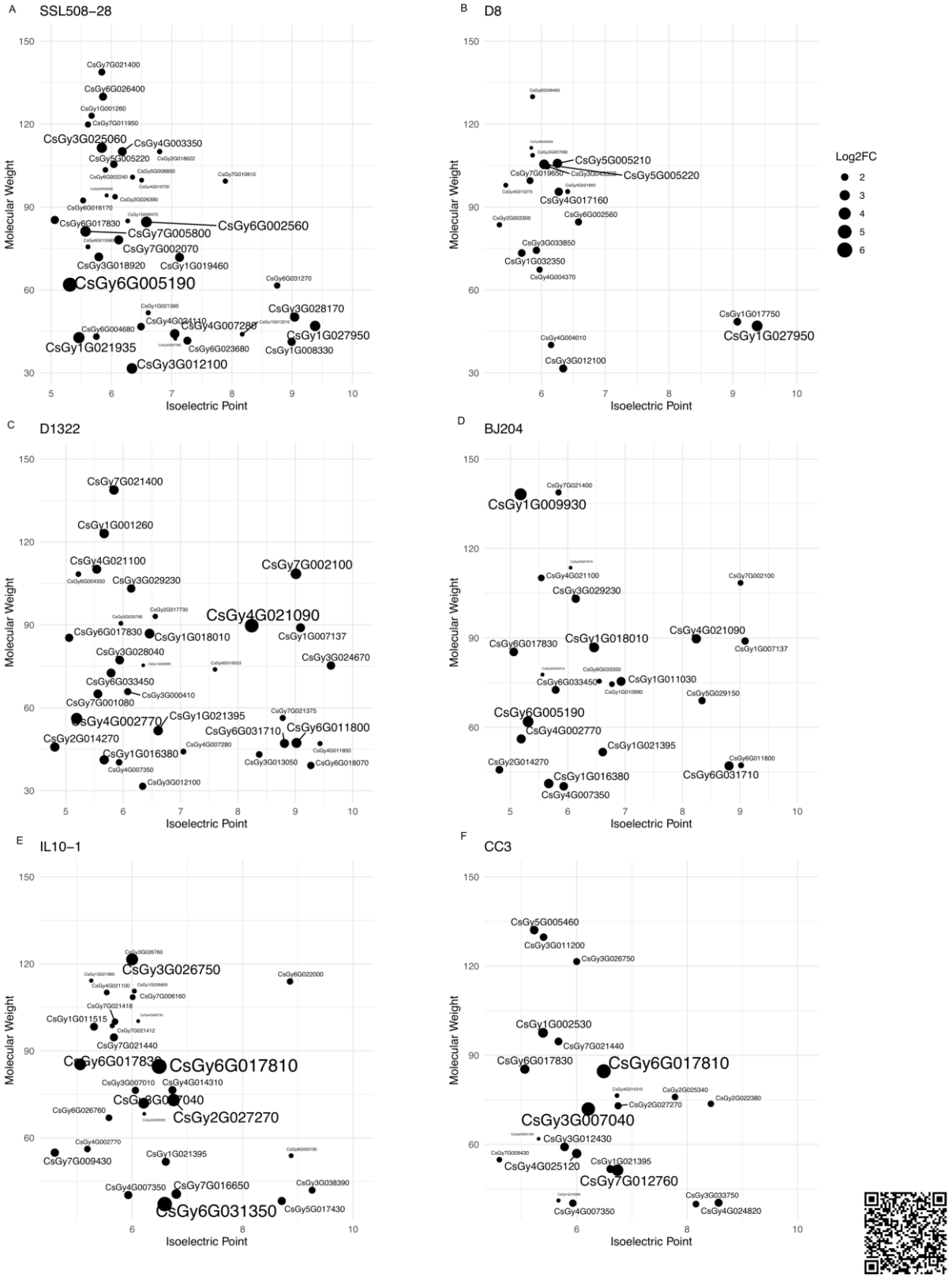

Fig. 2 Digital gel simulating the dispersion of upregulated differentially expressed kinase genes (DEGs) according to the molecular weight (MW – Y-axis) and isoelectric point (iP – X-axis) for three pathosystems. **(A)** SSL508-28 (resistant to Powdery Mildew), **(B)** D8 (susceptible to Powdery Mildew), **(C)** D1322 (resistant to Alternaria Leaf Spot), **(D)** BJ204 (susceptible to

Alternaria Leaf Spot), **(E)** IL10-1 (resistant to Root-Knot Nematode), and **(F)** CC3 (susceptible to Root-Knot Nematode). Powdery Mildew (PM, *Podosphaera xanthii*), Alternaria Leaf Spot (ALS, *Alternaria cucumerina*), and Root-Knot Nematode (RKN, *Meloidogyne incognita*). QR code for a link with a high-quality image.

The prediction of CREs was carried out from the region of 2 kbp of extension, upstream of the TSS [-2000, -1], using PlantCARE (Lescot et al., 2002). The analysis revealed that most of the predicted CREs were associated with transcriptional CORE, as well as, with different aspects of plant life, such as biotic stress response (wounds and diseases), light reactions, vegetative development, hormonal responses, and abiotic stress (Supplementary Table S2). In addition to this categorization, an analysis of the comparison of the results obtained by PlantCARE with information published in scientific articles was carried out, which resulted in a consensus classification for the categories of CREs (Ablazov et al., 2016; ain-Ali et al., 2021; Chen et al., 2021; Chowdhury et al., 2023; Deshmukh et al., 2017; Howlader et al., 2017; Mongkolsiriwatana et al., 2009; Pépin et al., 2021; Shariatipour et al., 2020; Sun et al., 2018; Tripathi et al., 2018; Wang et al., 2023).

3.2 Prediction of Cis-Regulatory Elements

In all, 51,339 CREs were found in the PPR of the 312 kinase genes studied. These elements were categorized according to the consensus classification into seven groups: Core (8 non-redundant CREs), Abiotic stress (18), Biotic stress (4), Development (30), Hormone (20), Light (38), and Unknown (7). The total numbers of CREs are illustrated in Figure 3 and Supplementary Table S2, and the "Core" category, responsible for initiating the transcription, is presented separately (Figure 3A) from the others (Figure 3B).

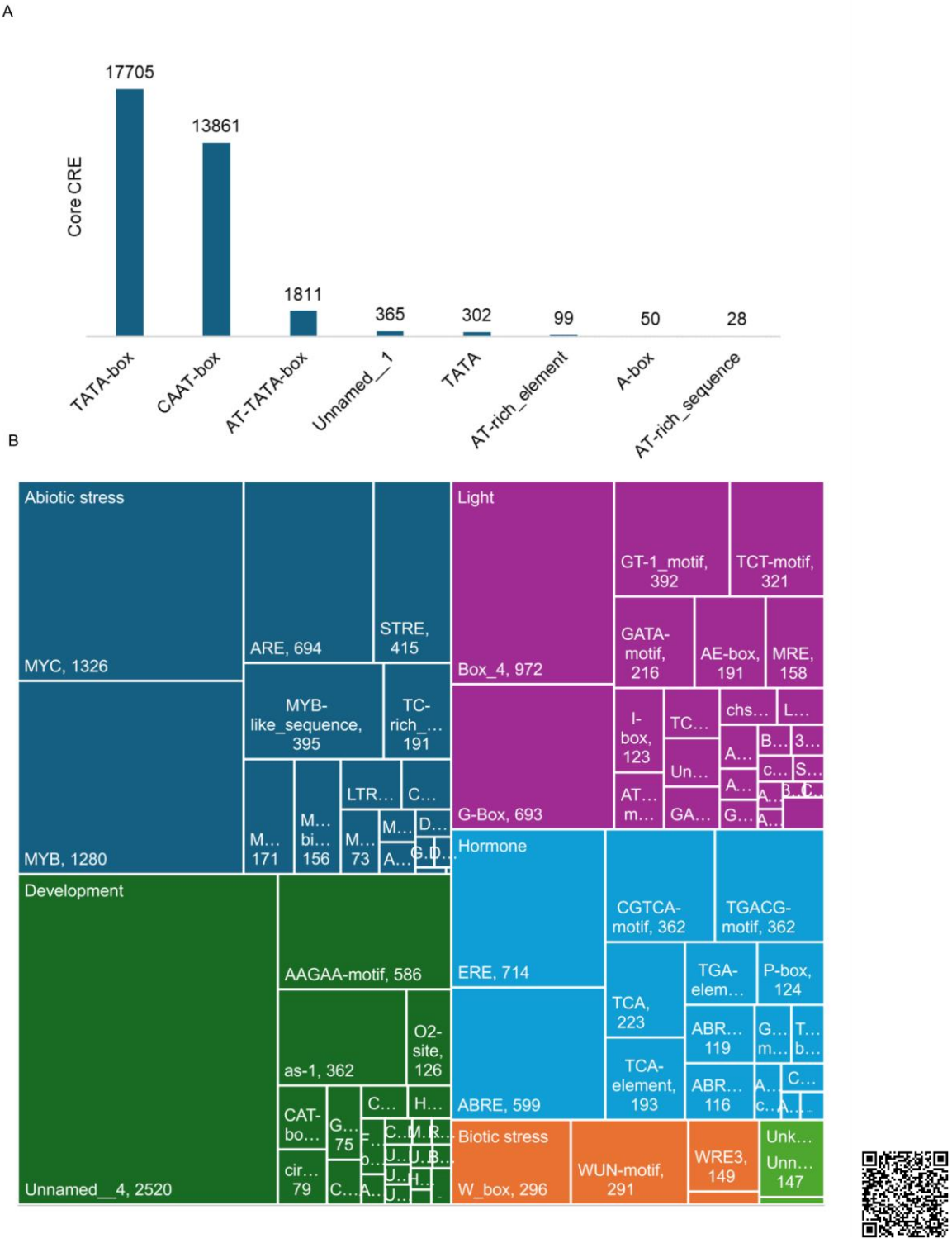

Fig. 3 Cis-Regulatory Elements (CREs) identified in the promoter region of kinase genes differentially expressed in response to Powdery Mildew (*Podosphaera xanthii*), Alternaria Leaf Spor (*Alternaria cucumerina*), and Root-Knot Nematode (*Meloidogyne incognita*). **(A)** Core promoter elements. **(B)** CRE categories classified based on the literature review that resulted in a consensus classification of the more recent papers. QR code for a link with a high-quality image.

After this discovery, we filtered to identify unique CREs of genes that were downregulated and upregulated in the contexts of the three diseases studied (Figure 4, Supplementary Table S3). Regarding the differentially expressed genes by PM, only the Pc-CMA2c element was exclusively observed in PPR of upregulated DEGs. In the case of downregulated DEGs, 18 unique CREs were observed. In the case of ALD-associated DEGs, 16 CREs were exclusively detected in the PPRs of downregulated genes, while 4 were in upregulated genes. As for RKN, the largest number of unique CREs (10) were found in the PPRs of upregulated genes. Four unique CREs were detected in PPRs of downregulated genes.

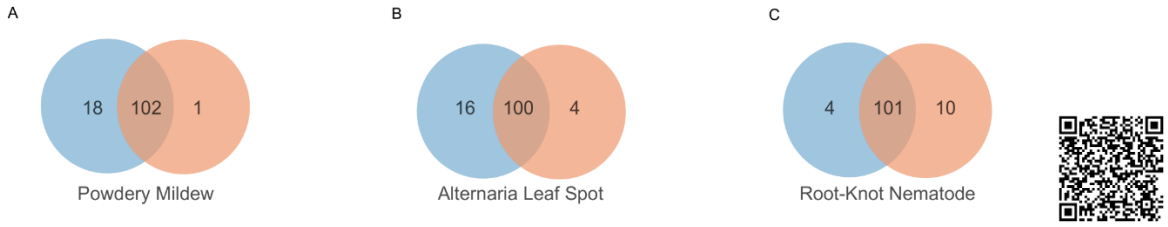

Fig. 4 Cis-Regulatory Elements (CREs) found in the promoter region of differentially expressed kinase genes in three diseases affecting cucumber (*Cucumis sativus*): **(A)** Powdery Mildew (*Podosphaera xanthii*), **(B)** Alternaria Leaf Spot (*Alternaria cucumerina*), **(C)** Root-Knot Nematode (*Meloidogyne incognita*). Genes were considered differentially expressed by log2 fold change > 1.5 and p-value (corrected by FDR – false Discovery Rate) < 0.05. Blue represents downregulated genes, and red represents upregulated genes. The CREs present in the intersection region were discarded for further analyses. Blue: downregulation, Red: upregulation. QR code for a link with a high-quality image.

The PPRs of PK-encoding genes potentially associated with resistance (R genes) or susceptibility (S genes) of plants in each of the pathosystems studied in this work were analyzed for the types and categories of CREs found in these regions. In PM, it was possible to notice that among the 115 different CREs identified, 12 types of four categories (Development, Hormone, Light, and Unknown) were found exclusively in the PPRs of the R genes, while 3 CREs, of two categories (Light and Unknown) were exclusive in the PPRs of the S genes. One hundred CREs were found in common and classified into seven possible categories (Core, Development, Hormone, Light, Biotic Stress, Abiotic Stress, and Unknown) (Figure 5A). In ALS, 112 types of CREs were identified, 88 of which were common among the R and S genes. Uniquely, 18 types of CREs, from four different categories (Developmental, Hormone, Light and Unknown) were detected in the S genes, and 6 CREs from three categories (Developmental, Hormone and Unknown) were observed exclusively in the R genes (Figure 5B). Finally, for

RKN, a total of 103 CREs were identified, of which 5 CREs from four categories (Abiotic stress, Development, Light and Unknown) were exclusive to the S genes, 16 were CREs from four categories (Abiotic stress, Core, Development and Light), exclusive to the R genes, and 82 CREs from the seven categories found in the PPRs of the R and S genes (Figure 5C).

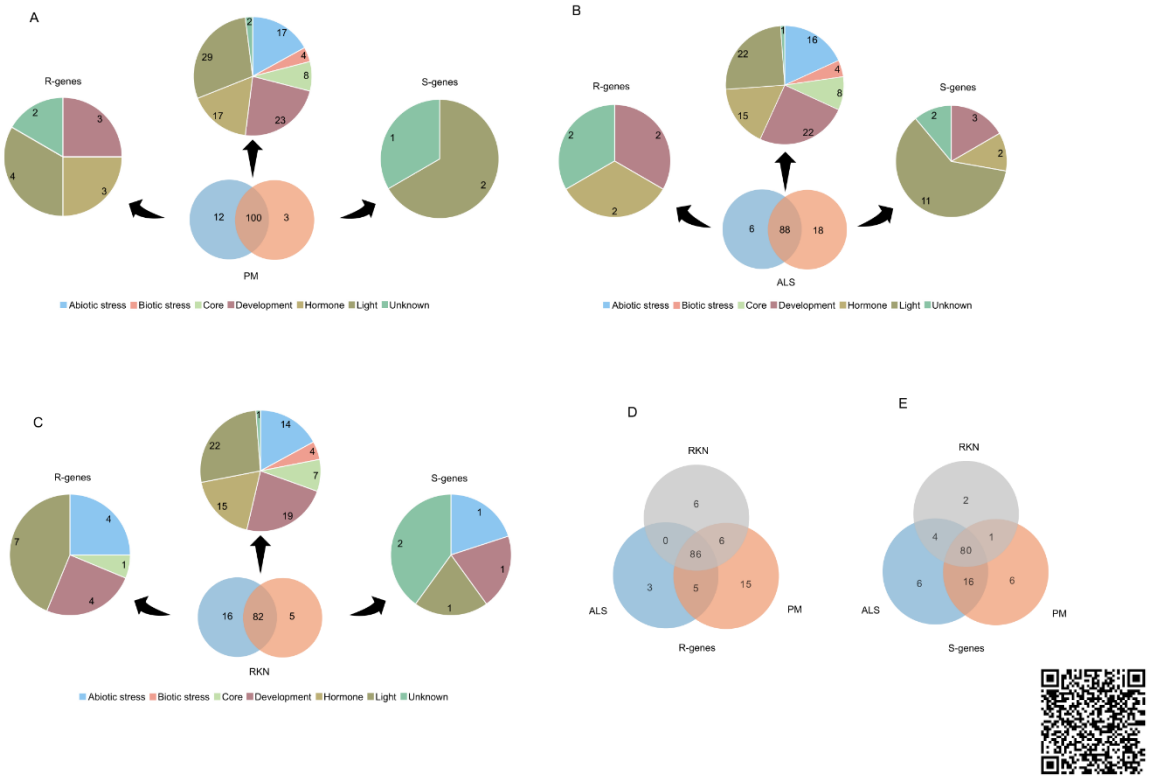

Fig. 5 Categories of Cis-Regulatory Elements (CREs) observed in the Potencially Promoter Region of kinase genes in context of biotic stresses caused by **(A)** Powdery Mildew (*Podophaera xanthii*), **(B)** Alternaria Leaf Spot (*Alternaria cucumerina*), and **(C)** Root-Knot Nematode (*Meloidogyne incognita*). **(A-C)** Venn's diagrams represent the number of CREs of resistance (R genes, blue) and susceptibility (S genes, red), and pie charts represent the seven categories (Abiotic stress, Biotic stress, Core, Development, Hormone, Light, and Unknown) of CREs according to the consensus classification obtained after the literature review. **(D)** Number of different CREs found in R genes, and **(E)** S genes. QR code for a link with a high-quality image.

An overview of these CREs can be seen in figures 5D and 5E, depicting genes R and S, respectively. Comparing the CREs present in the PPR of the R genes, it is observed that 86 CREs were found in common among the three pathosystems (Figure 5D), 5 CREs in common between ALS and PM, and 6 CREs in common between PM and RKN. Exclusive CREs were

found in each pathosystem, with 3 CREs in ALS, 15 CREs in PM, and 6 CREs in RKN. Upon analyzing the CREs of the S genes, it is observed that 80 CREs are common among the three pathosystems (Figure 5-E), 16 CREs in common between ALS and PM, 1 CRE in common between PM and RKN, and 4 CREs in common between ALS and RKN. Additionally, 6 exclusive CREs were found in ALS, 6 in PM, and 2 in RKN. The full description of the CREs can be found in Supplementary Tables S4-S6.

Through the graph theory approach (Andolfo et al., 2019; Qi et al., 2011), the similarity between the PPRs of the entire set of genes (312 DEGs) was analyzed. This analysis was carried out with the aim of investigating the pattern of similarity of the promoter sequences (Figure 6, Supplementary Table S7). From a multiple alignment of sequences of the PPRs of the DEGs, a similarity matrix was obtained. Similarity values ranged from 0.26 to 0.77, with a median of 0.36. Based on these values, a weighted network was built to deepen the understanding of the relationships between the PPRs. The classification of the nodes of this network was carried out considering the number of connections of each node, dividing it into three categories: 1-30, 31-60 and 61-100 connections. The number of connections per node ranged from 1 to 92, reflecting the diversity in the promoter regions. As one of the highlights, 12 PPRs had the highest number of connections, which is an indication that their promoter regions share sequence similarity. The genes of the other classes showed intermediate and lower numbers of connections, respectively, and 35 connections showed higher similarity values ranging from 0.51 to 0.77. The five gene families with the highest promoter region similarity values were: CsGy1G006700 (RLK-Pelle_LRK10L-2) / CsGy1G006720 (RLK-Pelle_LRK10L-2), CsGy1G011515 (RLK-Pelle_DLSV) / CsGy5G030060 (RLK-Pelle_L-LEC), CsGy1G011515 (RLK-Pelle_DLSV) / CsGy6G017830 (RLK-Pelle_WAK), CsGy3G029220 (RLK-Pelle_LRR-XII-1) / CsGy3G029230 (RLK-Pelle_LRR-XII-1), CsGy5G030060 (RLK-Pelle_L-LEC) / CsGy6G017830 (RLK-Pelle_WAK) (Supplementary Table S8).

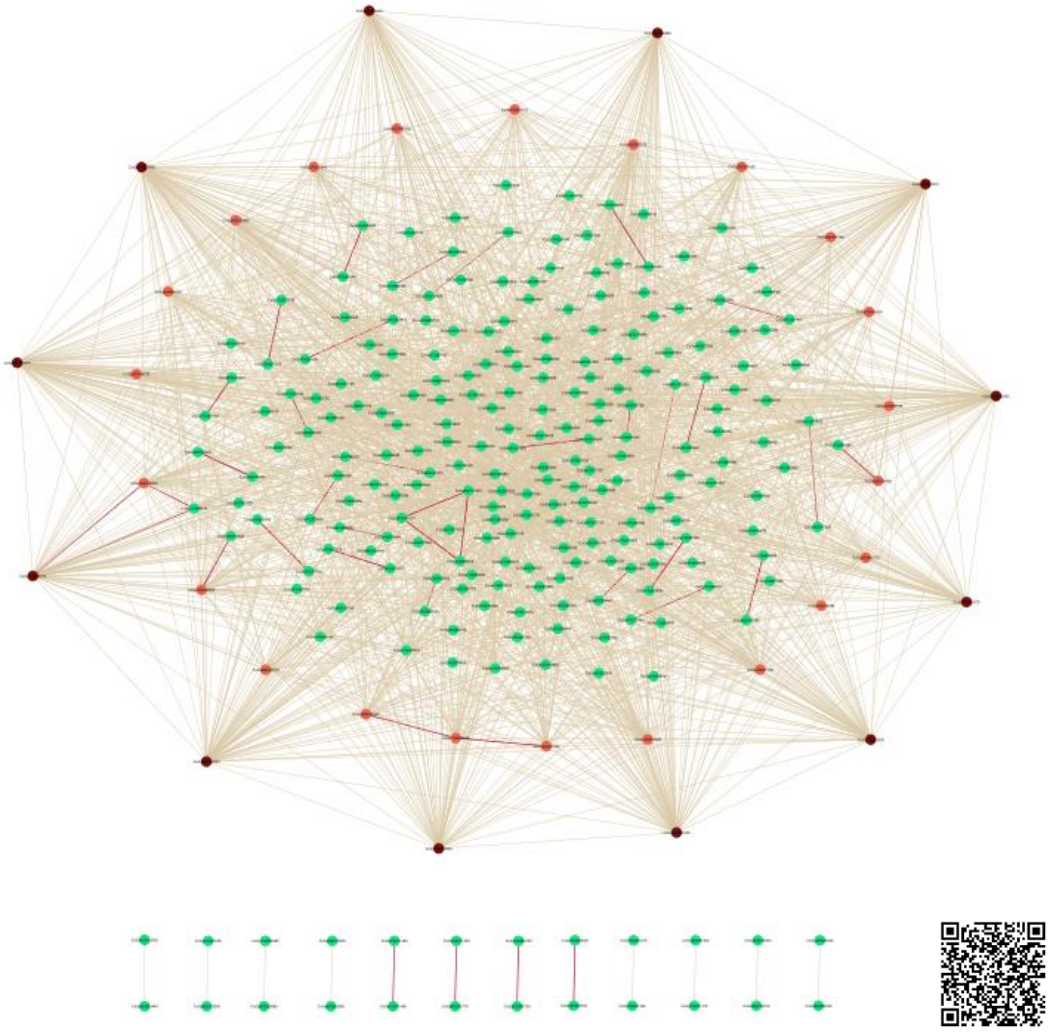

Fig. 6 Weighted Undirected Network of promoter region with 2 kpb upstream of Transcription Start Site (TSS) of differentially expressed kinase genes in response to biotic stress. Nodes represent promoter region, edges represent similarity. Green nodes: 0-30 connections, Orange nodes: 31-60 connections, Red nodes: 61-92 connections. Lion edges: similarity (weight) < 0.5, Red edges: similarity ranging from 0.51 to 0.77. Among the nodes, 24 nodes with only one connection can be observed in the bottom of the image. The similarity (weights) > 0.4 was considered to build the network. QR code for a link with a high-quality image.

3.3 Integration of Coexpression Network Analysis and Cis-Regulatory elements in PK Genes

To better understand the coexpression patterns of these differentially expressed kinase genes in the three pathosystems studied, a comparative analysis was performed. Two networks with 7413 connections were modeled, the first considering control conditions and the second under stress conditions (Figure 7). The presence of different communities within the networks was observed, and the control and stress conditions presented, respectively, four and five communities. In the network that represents the control condition (Figure 7-A, Supplementary Table S8), a cohesive structure can be observed, which evidences the three pathosystems studied, as well as a community represented by the single gene CsGy3G042300 (Family RLK-Pelle_LRR-III). In the network that represents the stress condition (Figure 7-B, Supplementary Table S8), a sparser behavior was observed between the connections, with three communities, in addition to two other communities formed by a single gene each, CsGy3G031785 (RLK-Pelle_WAK_LRK10L-1) and CsGy4G007910 (RLK-Pelle_DLSV).

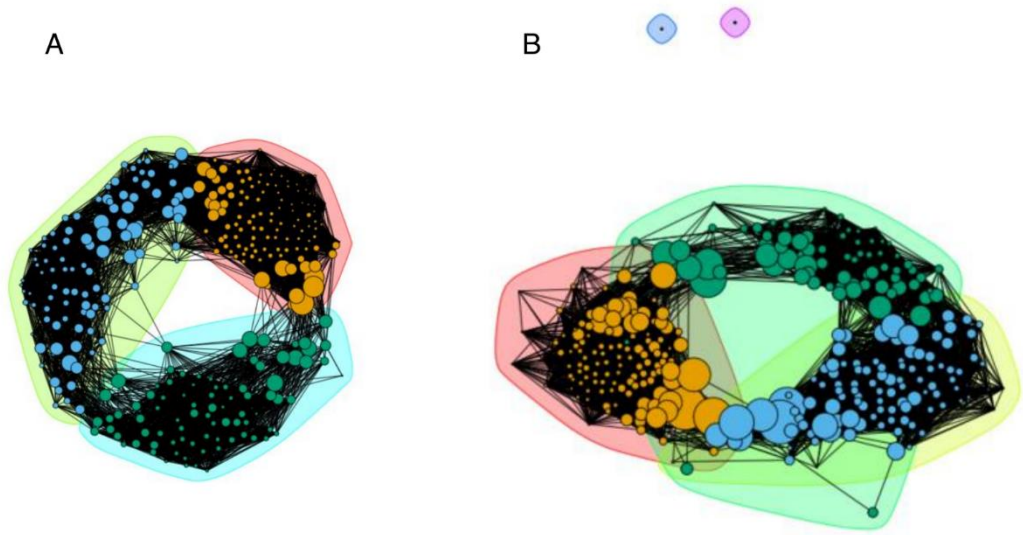

Fig. 7 Coexpression networks for *Cucumis sativus* protein kinase genes in context stresses caused by Powdery Mildew (*Podophaera xanthii*), Alternaria Leaf Spot (*Alternaria cucumerina*), and Root-Knot Nematode (*Meloidogyne incognita*). **(A)** Control conditions. **(B)** Stress conditions. The nodes represent 312 kinase genes coexpression in different treatment conditions. These genes were considered differentially expressed based on the criteria of log2 fold change > 1.5 and p-value < 0.05 with FDR correction. The edges represent the Pearson Correlation Coefficient > 0.6. Colors represent the number hubs observed. The full description

of the graphs can be observed in Supplementary Table S8. QR code for a link with a high-quality image.

The hub scores within the control network revealed that the highest values were associated with the following genes: CsGy7G021418 (RLK-Pelle_LRR-I-1), CsGy4G018630 (CAMK_OST1L), CsGy5G009180 (CMGC_CDK-CRK7-CDK9), CsGy1G032760 (RLK-Pelle_LRR-V), and CsGy2G027270 (RLK-Pelle_L-LEC) (Supplementary Table S8). In contrast, hub scores in the stress condition network exhibited peak values for the genes CsGy7G004490 (BUB), CsGy5G005320 (ULK_Fused), CsGy6G028880 (RLK-Pelle_LRR-XI-1), CsGy1G013210 (RLK-Pelle_Extensin), and CsGy1G001260 (RLK-Pelle_LRR-XI-1), that represent the sparser values (Supplementary Table S8).

The betweeness centrality was estimated. Nodes with high betweenness centrality act as a central bridge across many paths that connect different parts of the network. This node is important for maintaining the cohesion of the network, considering the shortest pathways. The genes exhibiting the highest betweenness centrality values under the control condition were CsGy5G012770 (RLK-Pelle_LRR-II), CsGy6G033410 (RLK-Pelle_DLSV), CsGy3G039460 (RLK-Pelle_CrRLK1L-1), CsGy4G021850 (PEK_PEK), and CsGy2G000730 (STE_STE7) (Supplementary Table S8). In the stress condition, genes with the highest betweenness centrality values included CsGy4G021100 (RLK-Pelle_DLSV), CsGy1G009470 (RLK-Pelle_LRR-VI-1), CsGy2G000730 (STE_STE7), CsGy1G008330 (TKL-Pl-4), and CsGy3G042020 (RLK-Pelle_RLCK-Os).

The Kleinberg's Hub Centrality was calculated to assess nodes with direct connectivity to other nodes exhibiting high betweenness centrality, considering both intermediation and direct connections to other essential nodes in the network. In the control condition, the gene pairs with the highest Kleinberg's Hub Centrality scores included CsGy1G018010 (RLK-Pelle_LRR-XII-1) / CsGy1G027550 (AGC_NDR), CsGy2G026380 (RLK-Pelle_CrRLK1L-1) / CsGy3G043350 (RLK-Pelle_LRR-XI-1), CsGy2G022380 (RLK-Pelle_LRR-III) / CsGy7G021440 (RLK-Pelle_LRR-I-1), CsGy2G027250 (RLK-Pelle_L-LEC) / CsGy3G026760 (RLK-Pelle_LRR-XI-1), CsGy2G006030 (RLK-Pelle_RLCK-VI) / CsGy3G000410 (TKL-Pl-4). Under stress conditions, gene pairs with the highest Kleinberg's Hub Centrality scores were CsGy1G019680 (RLK-Pelle_RLCK-XV) / CsGy6G033470 (RLK-Pelle_DLSV), CsGy2G027290 (RLK-Pelle_L-LEC) / CsGy6G027400 (CAMK_CDPK), CsGy2G026380 (RLK-Pelle_CrRLK1L-1) / CsGy6G013300 (TKL-Pl-5), CsGy3G000205

(RLK-Pelle_RLCK-VIIa-2) / CsGy5G013950 (CMGC_MAPK), CsGy2G014270 (CAMK_OST1L) / CsGy4G021120 (RLK-Pelle_DLSV).

After the identification of the communities, the differentially expressed kinase genes were separated according to the community and their PPRs were analyzed and categorized according to the consensus classification for the control and stress groups (Figure 8). Considering the control condition, it is possible to observe a certain uniformity in the distribution of the categories of CREs by the communities (hubs) 1, 2 and 3; and in community 4, a reduction in the number of CREs involved in immune responses. In the stress condition, similar behavior can be observed in communities 1, 2 and 3, but the stress caused by the diseases caused a disturbance in the network leading to the formation of two communities with unique genes, reflecting a reduction in the number of CREs in communities 4 and 5.

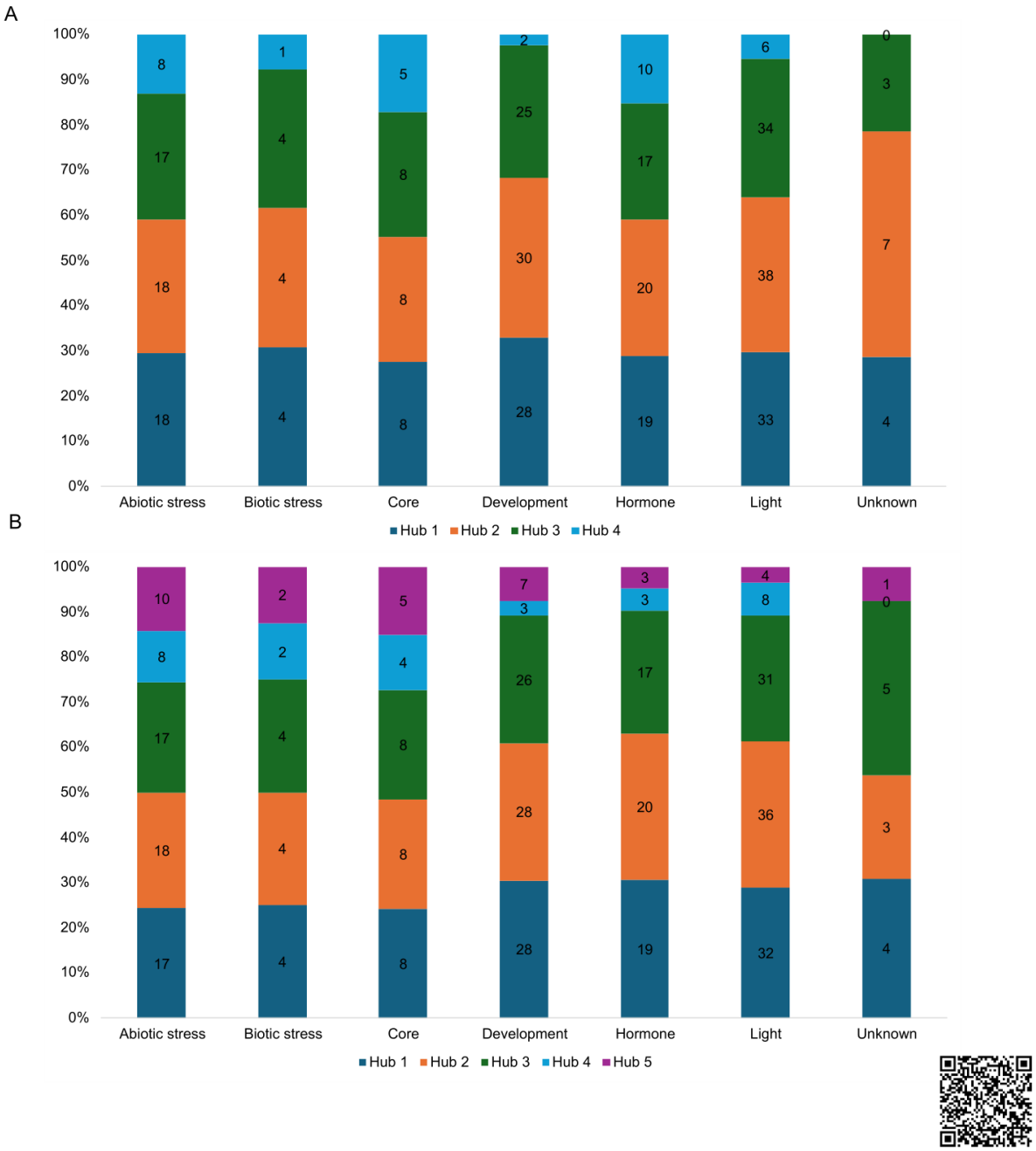

Fig. 8 Distribution of Cis-Regulatory Elements (CREs) according to the consensus classification to the **(A)** control and **(B)** stress conditions. This analysis considered the three pathosystems Powdery Mildew (*Podophaera xanthii*), Alternaria Leaf Spot (*Alternaria cucumerina*), and Root-Knot Nematode (*Meloidogyne incognita*) representing biotic stress for *Cucumis sativus*. Hub 4 e 5 contains CREs from single genes each after the perturbation caused by biotic stress. QR code for a link with a high-quality image.

After identifying the communities, we proceeded to segregate the differentially expressed kinase genes based on their community affiliations. Subsequently, their PPRs were meticulously analyzed and categorized according to the consensus classification, delineating

between the control and stress groups (Figure 8). Notably, under the control condition (Figure 8-A), a notable uniformity in the distribution of CRE categories across communities (hubs) 1, 2, and 3 was discernible. Conversely, in community 4, a discernible decrease in the number of CREs associated with immune responses was observed. Transitioning to the stress condition (Figure 8-B), a parallel trend was observed in communities 1, 2, and 3, albeit with variations. The stress instigated by the diseases, however, induced perturbations in the network dynamics, resulting in the emergence of two distinct communities harboring unique genes. This restructuring was accompanied by a notable reduction in the number of CREs within communities 4 and 5.

It was observed, in the control condition in relation to stress, a reduction in CREs of hub 1 of the Abiotic stress category and Light, in hub 2 there was a reduction of CREs in the Development, Light and Unknown categories. In hub 3, the same behavior was observed in the Development, Light and Unknown categories, and similar behavior in hub 4 Biotic stress, Core, Development, Hormone and Light. These reductions in the amount of CREs per hub in several categories contributed to the formation of hub 5 in the stress condition.

4 DISCUSSION

4.1 The Promoter Region of the Cucumber Kinase Genes

In this research, it was possible to identify a specific set of 125 CREs present in the PPR of genes encoding protein kinases (PKs) differentially expressed in relation to three diseases of great relevance to cucumber culture. PKs are fundamental in the processes of phosphorylation and, consequently, regulation of the activity of a series of proteins that are demanded in response to environmental stresses (Tena et al., 2011; Lin et al., 2021). The identification of genes responsive to eliciting stimuli is critical to understanding the intricate regulatory mechanisms that govern plant responses to biotic stress. In the case of PK-coding genes, this approach is even more important, considering its association with the regulatory role.

This knowledge is essential because it broadens the understanding of the role of these genes in the activation and suppression of protein expression and activity, a regulation that begins at the level of promoter sequences through their CREs (Zou et al., 2011). In addition, it is necessary to consider the need to investigate other factors associated with cucumber biotic stress, especially the specific TFs induced by pathogen elicitors that trigger the activation of

these CREs. The complex interaction between TFs and CREs results in regulated transcription activation.

The regulation of gene expression in eukaryotic organisms is a process composed of several distinct steps, and transcriptional control is considered the crucial point of this mechanism, as it is presented at the beginning of the expression process. Transcriptional control determines gene expression and influences the number of messenger RNA (mRNA) molecules synthesized (Lodish et al., 2005). In this context, TFs together with CREs play a decisive role for the initial phase of regulation. In other words, CREs have the function of allowing TFs to recognize specific DNA sequences in the promoter region of the gene and, after this binding, promote the formation of regulatory complexes, which have the ability to modulate RNA polymerase II activity and enable gene transcription (Schier & Taatjes, 2020).

An effective approach in constructing transcriptional regulation models is to consider the presence or absence of CREs, along with their combinatorial relationships, number of copies, and/or location in relation to the TSS (Zou et al., 2011). These models, known as cis-regulatory codes, have been successful in explaining gene expression in a variety of organisms, ranging from yeast (Pilpel et al., 2001; Beer et al., 2004), humans (Das et al., 2006), fruit flies (Segal et al., 2008), and Arabidopsis (Zou et al., 2011). Thus, the study of the PPR of genes is the initial step to study the modulation of gene expression.

4.2 Prediction of Cis-Regulatory Elements

The growing volume of information on gene expression and its regulation in various biological contexts demands efficient organization and interpretation of these data (Pope & Medzhitov, 2018). A significant challenge in this scenario is the absence of a clear functional classification of CREs. Despite the existence of several categories and types of CREs, the degeneration or polymorphism of the cis-regulatory code in the specificity of binding to TFs contributes to the complexity of gene expression regulation (Haberer et al., 2004; Kim & Wysocka, 2023; Ohno et al., 2018; Rockman et al., 2002). It is assumed that the diversity of categories and types of CREs may be associated with different types of signals emitted by the various types of CREs (Shrestha et al., 2018).

In this study, we sought to detect the CREs of PK-encoding genes that have been responsive to biotic stresses, represented in this study by PM, ALS and RKN, diseases that severely affect cucumber crops. To overcome the lack of a definitive classification of the categories and types of CREs, an approach based on a classificatory consensus derived from a

literature review of recent work was adopted (Ablazov et al., 2016; ain-Ali et al., 2021; Chen et al., 2021; Chowdhury et al., 2023; Deshmukh et al., 2017; Howlader et al., 2017; Mongkolsiriwatana et al., 2009; Pépin et al., 2021; Shariatipour et al., 2020; Sun et al., 2018; Tripathi et al., 2018; Wang et al., 2023). The resultant categories were Core, Abiotic stress, Biotic stress, Hormone, Development, Light and Unknown. From this consensus classification, hormone-responsive elements were listed as a type of CRE within the general Hormone category.

This classification method provides a basis for interpreting the functional implications of the CREs identified in response to biotic stress, as well as in the other categories. In this context, TFs can be categorized into four groups depending on their mode of regulation: constitutive, signal-dependent, inducible, and cell-type-specific expression (Pope & Medzhitov, 2018). These categories, in turn, delineate four fundamental patterns of gene expression: ubiquitous and constitutive, ubiquitous and inducible, cell-type-specific, and cell-type and inducible (Pope & Medzhitov, 2018). However, there was a disparity in classifications between CREs and TFs. While these discrete categories represent the extremes of a continuous spectrum of gene expression patterns, intermediate patterns can often be derived by adjusting one of the parameters of gene expression (Pope & Medzhitov, 2018).

Several strategies aimed at understanding the molecular basis of plant responses to different types of biotic stresses on a genomic scale have been implemented, seeking to identify new target genes involved in defense mechanisms (Berg et al., 2015). Cucumber, due to its versatility in diverse applications, has been the subject of intensive study in various genomic approaches, especially in stress-related contexts. Among these investigations, transcriptome analyses have been widely employed to clarify the repertoire and function of disease-responsive genes and their role in environmental challenges (Dong et al., 2020; Sá et al., 2020; Xu et al., 2017; Wang et al., 2018).

In the expectation of highlighting the factors responsible for the positive and negative regulation of genes during biotic stress, despite the considerable progress of the analyses, there is still a significant gap in the literature regarding the understanding of the modulation of these responses at the transcriptional level. This includes the need to clarify why certain genes are upregulated while others are downregulated. A high number of CREs could be observed in downregulated genes, with differences only in a small number of CREs in upregulated genes (Figure 4), showing a restricted number of CREs that may be responsible for triggering differential expression.

Through the analysis of the genes designated as R and S, we were able to discern a preponderant effect of the elements of the categories Development, Hormone and Light on the PPRs of the R genes in the pathosystems. This suggests that these categories of CREs may play crucial roles in regulating resistance responses in plants. The exclusive presence of types of CREs from the developmental, hormone and light categories in the PPRs of the R genes, in the pathosystems studied, points to a possible association of these elements with resistance, encompassing other metabolic processes in addition to those related to biotic stress. This observation directs research to investigate the molecular mechanisms activated by light and those associated with developmental processes at different stages of the plant.

On the other hand, we observed that CREs activated by light-responsive factors, development, and elements of unknown category also participate in the responses of S genes. This shows that, in addition to kinase genes and their CREs, other elements may be involved in the specific response of these genotypes, since they share up to a certain level CREs of common categories. However, it is important to emphasize that the analysis of the specific role of CREs in the PPRs of the R and S genes did not provide definitive results, highlighting the intricate complexity of the mechanisms underlying resistance and susceptibility, which vary according to the pathogen in question. For a more comprehensive understanding of these mechanisms, it is imperative to investigate and identify the transcription factors involved in the activation of each CRE belonging to the previously mentioned categories. This more detailed approach will certainly contribute significantly to unraveling the intricate mechanisms employed by plants in regulating their responses to biotic stress.

The definition of gene promoters enables the creation of uniform gene expression systems with predictable results, through the application of synthetic promoters and artificial expression, resulting in a significant improvement in the quality of the products obtained (Bilas et al., 2016). A detailed understanding of CREs is critical to unraveling the functioning of these promoters, especially in the context of transcriptional control exerted by TFs or regulatory proteins in response to specific environmental stimuli (Venter, 2007). This becomes particularly relevant when considering the critical role played by TFs and other regulatory proteins in the precise control of transcriptional processes. The quantitative analysis of core promoters allowed to perform precise genetic engineering of regulated gene expression, expanding the possibilities of manipulation of synthetic biological circuits (Ede et al., 2016). Thus, by improving knowledge about promoters and their associated CREs, the ability to propose more efficient gene expression systems is improved and possibilities for biotechnological applications are

revealed, such as increasing the quality of the products generated (Kumari et al., 2020). Such understanding, when detailed, becomes a fundamental step for optimizing genetic engineering strategies and maximizing the benefits associated with the precise control of gene expression in contexts of environmental stresses.

The approach employing DNA sequence similarity matrix by graph theory has been widely used in several researches, such as in the elucidation of evolutionary models (Melián et al., 2010), in the reduction of methodological complexity providing results comparable to or superior to those of the phylogenetic approach, which have computational limitations depending on the size of the sequence (Das et al., 2020); there is research in which it has been used to assess the evolutionary conservation of the MLO gene promoter region (Andolfo et al., 2019). Using this approach, we identified a set of 12 genes with the highest number of similar promoters, as well as a set of 35 connections (70 promoters) with the highest levels of similarity.

The similarity values of the set of 312 promoters (ranging from 0.26 to 0.77) indicate diversity of promoter sequences, representing possible differences in gene regulation patterns. The construction of the weighted network based on these similarity values provided a visual representation of the relationships between the promoters, facilitating a broader understanding of the transcriptional regulatory landscape of the genes encoding differentially expressed PKs in the face of the different types of biotic stresses studied, while the classification of nodes into categories according to the number of connections favored an understanding of the relative importance of each promoter in the network.

The identification of 12 promoters with the highest number of connections highlights the presence of highly connected subnetworks, suggesting a possible cooperation between these genes in gene regulation. On the other hand, the promoters of the genes of the other classes with intermediate (31-60 connections) and low (0-30 connections) numbers of connections may play more specific or isolated roles. This analysis may point to the convergence of different signaling pathways or to the coordinated regulation of specific biological processes, given the similarity of PPRs (Chowdhury & Koyutürk, 2010; Ge, 2011). The identification of highly connected subnetworks contributes to the characterization of the mechanisms underlying differential gene expression and the manifestation of specific phenotypes.

Understanding the real function of CREs is essential for understanding the evolution of the regulatory complexity of genes and their promoters and is a fascinating and fundamental aspect to understand the adaptation of species to different environments and stress conditions,

including biotic stress. Shared ancestry among members of gene families and their similar functions suggests a common origin (Walsh & Stephan, 2001), but variations in their sequences may result in different levels of activity or in proteins with slightly different characteristics and functions, as polymorphism or degeneration of CREs may induce the manifestation of different molecular phenotypes or cis-regulatory splicing (Lu et al., 2012; Walsh & Stephan, 2001). This genetic diversity within a family can confer evolutionary advantages, allowing different members to act more effectively in specific contexts through transcriptomic and proteomic diversity (Lu et al., 2012).

The differential regulation of copies of a gene, whether positive or negative, highlights the complexity and plasticity of the regulatory system. In response to different situations or stresses, one copy of a gene may be more beneficial to the cell, and specific factors may increase its need in a particular cellular context. In addition, spatial or temporal variations in gene expression can result in different levels of activity in specific tissues or at different times in development. This modulation in gene expression offers a significant adaptive advantage, allowing species to adjust their responses to diverse environmental stimuli. Differential regulation also plays an important role in tissue specialization and the coordination of complex biological processes.

4.3 The Biotic Stress Coexpression Networks and their Cis-Regulatory Elements

Gene coexpression networks are different from regulatory networks. The network built in this approach involves only coexpression of kinase genes. This approach used gene coexpression networks to identify communities or sub-networks allowing the dissection of gene clusters that work together in related biological processes (immune response, development, light-responsive processes), potentially due to the similarity of PPRs (Ge, 2011). The analysis of CREs emerges as an effective approach to describe the complex regulatory networks involved in the response to biotic stress. The modeling of these expression patterns of different PK-encoding genes that act in plant defense helps to understand their biological role in disease defense systems. In this modeling, the nodes represent PK genes, and the connections indicate a correlation of 0.6 or more, suggesting that these genes participate in similar pathways of responding to pathogens.

Another type of gene regulatory networks (GRN) arises from the physical and regulatory interaction between TFs and CREs, which are essential and can provide important information about the cis-regulatory code. These networks play a key role in controlling

transcription and translation processes, especially in response to biotic stress. TFs, non-coding RNAs, and RNA-binding proteins collaborate in translation, while transcriptional regulation is performed by GRNs (Bulky and Walhout, 2013). The main mechanism for regulating gene expression occurs through the modification of gene transcription levels (Bulky & Walhout, 2013). Each gene has a specific set of regulatory TFs, allowing for spatiotemporal expression patterns necessary in both development and response to disease (Bulky & Walhout, 2013).

CREs are identified as Transcription Factors Binding Sites (TFBSs) that play a direct role in altering transcription levels by acting as transcriptional activators or repressors. The interaction between TFs and CREs is a fundamental stage in the modulation of gene expression, especially in the face of biotic stimuli. The hypothesis is that coexpressed genes can have similar TFs that activate conserved CREs. Given that identifying specific TFs for each kinase gene is challenging, the strategy adopted in this study involved identifying communities under control and stress conditions, as well as identifying CREs associated with these communities. This approach enriches the information, moving closer to the understanding of the "cis-regulatory code". For example, the WRKY family of transcription factors has had its ability to bind to specific CREs, such as the element W-box, performing several functions in the response of plants to biotic stresses (Chen, 2019).

For the analysis of network attributes, "betweenness centrality" is an attribute of the nodes of a network, which are attributes of the genes in the networks in this study and measures the frequency with which this node appears on the shortest path between all pairs of nodes in the network (Freeman, 1977; Girvan & Newman, 2002). In the control condition, the CsGy1G001260 (RLK-Pelle_LRR-II) gene showed the highest value of "betweeness centrality", indicating that a high flow of information passes through its connections when there are no disturbances in the network. In the stress condition, the CsGy1G012600 (RLK-Pelle_DLSV) gene stood out. In fact, these genes play fundamental roles in the respective networks, representing communication bottlenecks (Yu et al., 2007), since it is common after the activation of the immune response of plants to occur the formation of communities and bottlenecks (Dietz et al., 2010). The protein "Enhanced Disease Susceptibility 1" (EDS1) is an important node in signal transmission after pathogen attack as well as in certain abiotic stresses (Wiermer et al., 2005).

The edge betweenness measures the number of times a specific connection is traversed by the shortest paths between pairs of nodes in the network (Teixeira et al., 2016). This attribute together with betweeness centrality are essential to maintain the connectivity of the networks,

and the genes with high values presented by the connections of the studied networks inform the high level of information that crosses these structures, indicating that this gene is implicated in defense against disorders (Crossley, 2014), or in our case, biotic stress. Connections with a high degree of betweeness indicates a network with high vulnerability, which in turn indicates that there is a high degree of communication between elements that share such connections.

Communities (core elements) can be identified as regulators of biological mechanisms. The highest network hub score in the control condition was presented by the CsGy4G025570 (RLK-Pelle_LRR-I-1) gene, indicating that it is the main regulator of gene expression in the network. The stress condition caused disturbance in the network, given that the highest hub score in the stress condition was presented by the CsGy7G012830 (BUB) gene, indicating that this gene is a central element regulating biotic stress. The importance of communities emerges from their representation of highly interconnected multiprotein complexes, playing key roles in intricate processes such as RNA splicing and proteolysis. These features are universal and constitute essential elements in complex global biological networks, including those related to signaling pathways (Dietz et al., 2010).

The formation of communities with individual genes was noted in both conditions examined. In stressful situations, it is common to observe the creation of communities with specific genes, suggesting that the disturbance caused by stress breaks connections, resulting in the loss of efficiency of these genes and their disconnection from the network (Arnatkevičiūtė, 2019; Dietz, 2010). On the other hand, in the control condition, this pattern may indicate that these genes may participate in other pathways or interactions with other genes that were not addressed in the present study (Oldham, 2018). If a gene has been identified as differentially expressed, this may suggest its involvement in the stress response or its suppression in the control conditions. When a gene disconnects from the network, it contributes to an inhomogeneous topology, indicating that its position in the network is essential. Relating this to the centrality-lethality rule, it can be considered that genes that disconnect from the network, especially under stress conditions, may represent critical points whose removal or disruption has a more significant impact on overall functionality, thus supporting the correlation between centrality and lethality in biological networks (He & Zhang, 2006; Zotenko et al., 2008).

In the context of coexpression network perturbation, a notable phenomenon emerges: the reassignment of certain functional roles of key regulatory elements (CREs) in response to stress, subsequently resulting in the emergence of new hub nodes. Specifically, hubs 1 to 4 in control conditions undergo a transformation, where the activities of their associated CREs are

redirected towards alternative metabolic pathways following stress-induced disturbances. This transformative process culminates in the establishment of a novel hub, hub 5, expanding the network from its original configuration of four hubs observed under control conditions.

It's crucial to emphasize that the CREs identified within hubs 4 and 5 originate from genes distinct to each respective hub. Furthermore, these genes exhibit CRE distributions across seven distinct categories, except for hub 4, which notably lacks CREs categorized as Unknown. This delineation underscores the specificity and diversity of regulatory mechanisms operating within the system. The significance of these findings extends beyond mere network dynamics, offering profound insights into the intricate regulatory landscapes governing plant resilience against disease. The study unveils the dynamic interplay between CREs and their target genes, illuminating the precise regulatory mechanisms underpinning plant defense mechanisms. Notably, the identification of context-dependent effects among CREs underscores the adaptive nature of plant responses to a myriad of pathogenic challenges.

These analyses afforded a comprehensive understanding of community architectures within each network, enabling insightful comparisons between control and stress conditions. Through this, we could discern which CREs potentially contributed to the modulation of responses in the face of network perturbations. Moreover, the elucidation of interplay between different signaling pathways mediated by these CREs reveals a sophisticated regulatory framework orchestrating plant defense response. This intricate network coordinates responses to individual stressors and highlights the interconnectedness of various signaling cascades in mounting an effective defense strategy. In summary, this study sheds light on the dynamic nature of regulatory networks in response to stress and underscores the complexity and adaptability of plant defense mechanisms. By unraveling these intricate interactions, it paves the way for targeted strategies aimed at enhancing plant resilience against diverse pathogens.

5 CONCLUSION

We conducted a comprehensive analysis of the promoter regions associated with differentially expressed kinase genes under the influence of various pathogens, including Powdery Mildew, Alternaria Leaf Spot, and Root-Knot Nematode. This investigation yielded a total of 51,339 cis-regulatory elements (CREs) within the potential promoter region (PPR), spanning 2 kbp upstream. These CREs were classified into 125 distinct types and further categorized into seven functional classes based on their responsiveness: Abiotic Stress, Biotic Stress, Core, Development, Hormone, Light, and Unknown. Notably, the CREs categorized

under Development, Hormone, and Light categories exhibited enhanced effectiveness in response to biotic stress conditions. This insight underscores the significance of these regulatory elements in mediating plant responses to pathogenic challenges.

REFERENCES

Ablazov, A., Tombuloglu, H., 2016. Genome-wide identification of the mildew resistance locus O (MLO) gene family in novel cereal model species Brachypodium distachyon. Eur J Plant Pathol 145, 239–253. https://doi.org/10.1007/s10658-015-0833-2

ain-Ali, Q., Mushtaq, N., Amir, R., Gul, A., Tahir, M., Munir, F., 2021. Genome-wide promoter analysis, homology modeling and protein interaction network of Dehydration Responsive Element Binding (DREB) gene family in Solanum tuberosum. PLoS One 16, e0261215. https://doi.org/10.1371/journal.pone.0261215

Andolfo, G., Iovieno, P., Ricciardi, L., Lotti, C., Filippone, E., Pavan, S., Ercolano, M.R., 2019. Evolutionary conservation of MLO gene promoter signatures. BMC Plant Biol 19, 150. https://doi.org/10.1186/s12870-019-1749-3

Arnatkevičiūtė, A., Fulcher, B.D., Fornito, A., 2019. Uncovering the Transcriptional Correlates of Hub Connectivity in Neural Networks. Front Neural Circuits 13. https://doi.org/10.3389/fncir.2019.00047

Beer, M.A., Tavazoie, S., 2004. Predicting Gene Expression from Sequence. Cell 117, 185–198. https://doi.org/10.1016/S0092-8674(04)00304-6

Berg, J.A., Appiano, M., Santillán Martínez, M., Hermans, F.W.K., Vriezen, W.H., Visser, R.G.F., Bai, Y., Schouten, H.J., 2015. A transposable element insertion in the susceptibility gene CsaMLO8 results in hypocotyl resistance to powdery mildew in cucumber. BMC Plant Biol 15, 1–17. https://doi.org/10.1186/S12870-015-0635-X/FIGURES/5

Berg, J.A., Hermans, F.W.K., Beenders, F., Lou, L., Vriezen, W.H., Visser, R.G.F., Bai, Y., Schouten, H.J., 2020. Analysis of QTL DM4.1 for Downy Mildew Resistance in Cucumber Reveals Multiple subQTL: A Novel RLK as Candidate Gene for the Most Important subQTL. Front Plant Sci 11. https://doi.org/10.3389/FPLS.2020.569876/PDF

Biłas, R., Szafran, K., Hnatuszko-Konka, K., Kononowicz, A.K., 2016. Cis-regulatory elements used to control gene expression in plants. Plant Cell, Tissue and Organ Culture (PCTOC) 127, 269–287. https://doi.org/10.1007/s11240-016-1057-7

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Brandes, U., 2001. A faster algorithm for betweenness centrality. J Math Sociol 25, 163–177. https://doi.org/10.1080/0022250X.2001.9990249

Bulky, M.L., Walhout, M., 2013. Handbook of Systems Biology, in: Walhout, A.J., Vidal, M., Dekker, J. (Eds.), Handbook of Systems Biology Concepts and Insights. Academic Press, Amsterdan, pp. 65–88. https://doi.org/10.1016/C2010-0-67190-2

Charoensawan, V., Wilson, D., Teichmann, S.A., 2010. Lineage-specific expansion of DNA-binding transcription factor families. Trends in Genetics 26, 388–393. https://doi.org/10.1016/j.tig.2010.06.004

Chen, L., Liu, J., Liu, Z., Xia, P., Yuan, X., Ning, Y., 2021. Genome-wide identification and expression analysis of the MLO gene family reveal a candidate gene associated with powdery mildew susceptibility in bitter gourd (Momordica charantia). Eur J Plant Pathol 159, 163–178. https://doi.org/10.1007/s10658-020-02152-0

Chen, X., Li, C., Wang, H., Guo, Z., 2019. WRKY transcription factors: evolution, binding, and action. Phytopathology Research 1. https://doi.org/10.1186/s42483-019-0022-x

Chowdhury, A.T., Hasan, Md.N., Bhuiyan, F.H., Islam, Md.Q., Nayon, Md.R.W., Rahaman, Md.M., Hoque, H., Jewel, N.A., Ashrafuzzaman, Md., Prodhan, S.H., 2023. Identification, characterization of Apyrase (APY) gene family in rice (Oryza sativa) and analysis of the expression pattern under various stress conditions. PLoS One 18, e0273592. https://doi.org/10.1371/journal.pone.0273592

CHOWDHURY, S.A., KOYUTÜRK, M., 2009. IDENTIFICATION OF COORDINATELY DYSREGULATED SUBNETWORKS IN COMPLEX PHENOTYPES, in: Biocomputing 2010. WORLD SCIENTIFIC, pp. 133–144. https://doi.org/10.1142/9789814295291_0016

Chuong, E.B., Elde, N.C., Feschotte, C., 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18, 71–86. https://doi.org/10.1038/nrg.2016.139

Cosby, R.L., Chang, N.-C., Feschotte, C., 2019. Host–transposon interactions: conflict, cooperation, and cooption. Genes Dev 33, 1098–1116. https://doi.org/10.1101/gad.327312.119

Costa, F., Aono, A., Silva, E., Pereira, W., 2023. The Cucumber Kinome: Uncovering and Analyzing the Expression of Cucumis sativus Protein Kinases in Response to Biotic Stress.

Crossley, N.A., Mechelli, A., Scott, J., Carletti, F., Fox, P.T., McGuire, P., Bullmore, E.T., 2014. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain 137, 2382–2395. https://doi.org/10.1093/brain/awu132

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1–9.

Csárdi G, Nepusz T, Traag V, Horvát S, Zanini F, Noom D, Müller K, 2024. igraph: Network Analysis and Visualization in R. R package version 2.0.2.

Das, D., Nahlé, Z., Zhang, M.Q., 2006. Adaptively inferring human transcriptional subnetworks. Mol Syst Biol 2. https://doi.org/10.1038/msb4100067

Das, S., Das, A., Bhattacharya, D.K., Tibarewala, D.N., 2020. A new graph-theoretic approach to determine the similarity of genome sequences based on nucleotide triplets. Genomics 112, 4701–4714. https://doi.org/10.1016/j.ygeno.2020.08.023

Deshmukh, R., Singh, V.K., Singh, B.D., 2017. Mining the Cicer arietinum genome for the mildew locus O (Mlo) gene family and comparative evolutionary analysis of the Mlo genes from Medicago truncatula and some other plant species. J Plant Res 130, 239–253. https://doi.org/10.1007/s10265-016-0868-2

Diehl, A.G., Ouyang, N., Boyle, A.P., 2020. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. Nat Commun 11, 1796. https://doi.org/10.1038/s41467-020-15520-5

Dietz, K., Jacquot, J., Harris, G., 2010. Hubs and bottlenecks in plant molecular signalling networks. New Phytologist 188, 919–938. https://doi.org/10.1111/j.1469-8137.2010.03502.x

Dong, J., Wang, Y., Xian, Q., Chen, X., Xu, J., 2020. Transcriptome analysis reveals ethylene-mediated defense responses to Fusarium oxysporum f. Sp. Cucumerinum infection in Cucumis sativus L. BMC Plant Biol 20, 1–10. https://doi.org/10.1186/S12870-020-02537-7/FIGURES/6

Ede, C., Chen, X., Lin, M.-Y., Chen, Y.Y., 2016. Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. ACS Synth Biol 5, 395–404. https://doi.org/10.1021/acssynbio.5b00266

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

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

FitzPatrick, D.R., 2017. Developmental biology of the eye, in: Taylor and Hoyt's Pediatric Ophthalmology and Strabismus. Elsevier, pp. 25–31. https://doi.org/10.1016/B978-0-7020-6616-0.00004-9

Freeman, L.C., 1977. A Set of Measures of Centrality Based on Betweenness. Sociometry 40, 35. https://doi.org/10.2307/3033543

Ge, S.X., 2011. Large-scale analysis of expression signatures reveals hidden links among diverse cellular processes. BMC Syst Biol 5, 87. https://doi.org/10.1186/1752-0509-5-87

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99, 7821–7826. https://doi.org/10.1073/pnas.122653799

Haberer, G., Hindemitt, T., Meyers, B.C., Mayer, K.E.X., 2004a. Transcriptional Similarities, Dissimilarities, and Conservation of cis-Elements in Duplicated Genes of Arabidopsis. Plant Physiol 136, 3009–3022. https://doi.org/10.1104/PP.104.046466

Haberer, G., Hindemitt, T., Meyers, B.C., Mayer, K.F.X., 2004b. Transcriptional Similarities, Dissimilarities, and Conservation of cis-Elements in Duplicated Genes of Arabidopsis. Plant Physiol 136, 3009–3022. https://doi.org/10.1104/pp.104.046466

He, X., Zhang, J., 2006. Why Do Hubs Tend to Be Essential in Protein Networks? PLoS Genet 2, e88. https://doi.org/10.1371/journal.pgen.0020088

Howlader, J., Park, J.-I., Kim, H.-T., Ahmed, N.U., Robin, A.H.K., Sumi, K.R., Natarajan, S., Nou, I.-S., 2017. Differential Expression under Podosphaera xanthii and Abiotic Stresses Reveals Candidate MLO Family Genes in Cucumis melo L. Trop Plant Biol 10, 151–168. https://doi.org/10.1007/s12042-017-9194-7

Jacinto Junior, S.G., Lucena, E.M.P. de, 2022. Alterações morfofisiológicas e metabólitos secundários produzidos por feijoeiros submetidos ao estresse hídrico: uma revisão integrativa.

Research, Society and Development 11, e445111436506. https://doi.org/10.33448/rsd-v11i14.36506

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

Kim, S., Wysocka, J., 2023. Deciphering the multi-scale, quantitative cis-regulatory code. Mol Cell 83, 373–392. https://doi.org/10.1016/j.molcel.2022.12.032

Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM 46, 604–632. https://doi.org/10.1145/324133.324140

Kummari, D., Palakolanu, S.R., Kishor, P.B.K., Bhatnagar-Mathur, P., Singam, P., Vadez, V., Sharma, K.K., 2020. An update and perspectives on the use of promoters in plant genetic engineering. J Biosci 45, 119. https://doi.org/10.1007/s12038-020-00087-6

Latchman, D.S., 2013. Eukaryotic Transcription Factors, in: Brenner's Encyclopedia of Genetics. Elsevier, pp. 537–539. https://doi.org/10.1016/B978-0-12-374984-0.01548-5

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

Lin, L., Wu, J., Jiang, M., Wang, Y., 2021. Plant Mitogen-Activated Protein Kinase Cascades in Environmental Stresses. Int J Mol Sci 22, 1543. https://doi.org/10.3390/ijms22041543

Lodish, H., Berk, A., Kaiser, C., 2004. Molecular Cell Biology. W. H. Freeman.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, Z., Jiang, P., Xing, Y., 2012. Genetic variation of pre-mRNA alternative splicing in human populations. WIREs RNA 3, 581–592. https://doi.org/10.1002/wrna.120

Melián, C.J., Alonso, D., Vázquez, D.P., Regetz, J., Allesina, S., 2010. Frequency-Dependent Selection Predicts Patterns of Radiations and Biodiversity. PLoS Comput Biol 6, e1000892. https://doi.org/10.1371/journal.pcbi.1000892

Mongkolsiriwatana C, Pongtongkam P, Peyachoknagul S, 2009. In silico Promoter Analysis of Photoperiod-Responsive Genes Identified by DNA Microarray in Rice (Oryza sativa L.). Agriculture and Natural Resources 43.

Ohno, K., Takeda, J., Masuda, A., 2018. Rules and tools to predict the splicing effects of exonic and intronic mutations. WIREs RNA 9. https://doi.org/10.1002/wrna.1451

Oldham, S., Fornito, A., 2019. The development of brain network hubs. Dev Cogn Neurosci 36, 100607. https://doi.org/10.1016/j.dcn.2018.12.005

Pépin, N., Hebert, F.O., Joly, D.L., 2021. Genome-Wide Characterization of the MLO Gene Family in Cannabis sativa Reveals Two Genes as Strong Candidates for Powdery Mildew Susceptibility. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.729261

Peremarti, A., Twyman, R.M., Gómez-Galera, S., Naqvi, S., Farré, G., Sabalza, M., Miralpeix, B., Dashevskaya, S., Yuan, D., Ramessar, K., Christou, P., Zhu, C., Bassie, L., Capell, T., 2010.

Promoter diversity in multigene transformation. Plant Mol Biol 73, 363–378. https://doi.org/10.1007/s11103-010-9628-1

PHYLIP, n.d. PHYLIP (Phylogeny Inference Package) Version 3.697 [WWW Document]. URL https://phylipweb.github.io/phylip (accessed 2.19.24).

Pilpel, Y., Sudarsanam, P., Church, G.M., 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29, 153–159. https://doi.org/10.1038/ng724

Pope, S.D., Medzhitov, R., 2018. Emerging Principles of Gene Expression Programs and Their Regulation. Mol Cell 71, 389–397. https://doi.org/10.1016/j.molcel.2018.07.017

Porto, M.S., Pinheiro, M.P.N., Batista, V.G.L., dos Santos, R.C., de Albuquerque Melo Filho, P., de Lima, L.M., 2014. Plant Promoters: An Approach of Structure and Function. Mol Biotechnol 56, 38–49. https://doi.org/10.1007/s12033-013-9713-1

Priest, H.D., Filichkin, S.A., Mockler, T.C., 2009. cis-Regulatory elements in plant cell signaling. Curr Opin Plant Biol 12, 643–649. https://doi.org/10.1016/j.pbi.2009.07.016

Purugganan, M.D., 1998. The molecular evolution of development. BioEssays 20, 700–711. https://doi.org/10.1002/(SICI)1521-1878(199809)20:9<700::AID-BIES3>3.0.CO;2-K

Qi, X., Wu, Q., Zhang, Y., Fuller, E., Zhang, C.-Q., 2011. A Novel Model for DNA Sequence Similarity Analysis Based on Graph Theory. Evolutionary Bioinformatics 7, EBO.S7364. https://doi.org/10.4137/EBO.S7364

Rockman, M. V., Wray, G.A., 2002. Abundant Raw Material for Cis-Regulatory Evolution in Humans. Mol Biol Evol 19, 1991–2004. https://doi.org/10.1093/oxfordjournals.molbev.a004023

Russell, J., Zomerdijk, J.C.B.M., 2006. The RNA polymerase I transcription machinery. Biochemical Society Symposia 73, 203–216. https://doi.org/10.1042/bss0730203

Sa, R., Liu, D., Chen, L., Liu, L., Zhang, Y., Zhang, X., Xu, C., 2020. 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

Schier, A.C., Taatjes, D.J., 2020. Structure and mechanism of the RNA polymerase II transcription machinery. Genes Dev 34, 465–488. https://doi.org/10.1101/gad.335679.119

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., Gaul, U., 2008. Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature 451, 535–540. https://doi.org/10.1038/nature06496

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 13, 2498–2504. https://doi.org/10.1101/gr.1239303

Shariatipour, N., Heidari, B., 2020. Meta-Analysis of Expression of the Stress Tolerance Associated Genes and Uncover their Cis-Regulatory Elements in Rice (Oryza sativa L.). Open Bioinforma J 13, 39–49. https://doi.org/10.2174/1875036202013010039

Shrestha, A., Khan, A., Dey, N., 2018. cis–trans Engineering: Advances and Perspectives on Customized Transcriptional Regulation in Plants. Mol Plant 11, 886–898. https://doi.org/10.1016/j.molp.2018.05.008

Sun, R., Wang, S., Ma, D., Liu, C., 2018. Genome-Wide Analysis of LRR-RLK Gene Family in Four Gossypium Species and Expression Analysis during Cotton Development and Stress Responses. Genes (Basel) 9, 592. https://doi.org/10.3390/genes9120592

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., Wang, T., 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res 24, 1963–1976. https://doi.org/10.1101/gr.168872.113

Sundaram, V., Wysocka, J., 2020. Transposable elements as a potent source of diverse *cis* -regulatory sequences in mammalian genomes. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190347. https://doi.org/10.1098/rstb.2019.0347

Teixeira, A.S., Santos, F.C., Francisco, A.P., 2016. Spanning Edge Betweenness in Practice. pp. 3–10. https://doi.org/10.1007/978-3-319-30569-1_1

Tena, G., Boudsocq, M., Sheen, J., 2011. Protein kinase signaling networks in plant innate immunity. Curr Opin Plant Biol 14, 519–529. https://doi.org/10.1016/j.pbi.2011.05.006

Tripathi, R.K., Bregitzer, P., Singh, J., 2018. Genome-wide analysis of the SPL/miR156 module and its interaction with the AP2/miR172 unit in barley. Sci Rep 8, 7085. https://doi.org/10.1038/s41598-018-25349-0

Venter, M., 2007. Synthetic promoters: genetic control through cis engineering. Trends Plant Sci 12, 118–124. https://doi.org/10.1016/j.tplants.2007.01.002

Walsh, J.B., Stephan, W., 2001. Multigene Families: Evolution, in: Encyclopedia of Life Sciences. Wiley. https://doi.org/10.1038/npg.els.0001702

Wang, J., Wu, Xiaohua, Wang, Y., Wu, Xinyi, Wang, B., Lu, Z., Li, G., 2023. Genome-wide characterization and expression analysis of the MLO gene family sheds light on powdery mildew resistance in Lagenaria siceraria. Heliyon 9, e14624. https://doi.org/10.1016/j.heliyon.2023.e14624

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

Wang, X., Haberer, G., Mayer, K.F., 2009. Discovery of cis-elements between sorghum and rice using co-expression and evolutionary conservation. BMC Genomics 10, 284. https://doi.org/10.1186/1471-2164-10-284

Wiermer, M., Feys, B.J., Parker, J.E., 2005. Plant immunity: the EDS1 regulatory node. Curr Opin Plant Biol 8, 383–389. https://doi.org/10.1016/j.pbi.2005.05.010

Windram, O., Penfold, C.A., Denby, K.J., 2014. Network Modeling to Understand Plant Immunity. Annu Rev Phytopathol 52, 93–111. https://doi.org/10.1146/annurev-phyto-102313-050103

Xu, Q., Xu, X., Shi, Y., Qi, X., Chen, X., 2017. 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

Young, R.A., 1991. RNA POLYMERASE II. Annu Rev Biochem 60, 689–715. https://doi.org/10.1146/annurev.bi.60.070191.003353

Yu, D., Chen, C., Chen, Z., 2001. Evidence for an Important Role of WRKY DNA Binding Proteins in the Regulation of *NPR1* Gene Expression. Plant Cell 13, 1527–1540. https://doi.org/10.1105/TPC.010115

Yu, H., Kim, P.M., Sprecher, E., Trifonov, V., Gerstein, M., 2007. The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. PLoS Comput Biol 3, e59. https://doi.org/10.1371/journal.pcbi.0030059

Yu, J., Wu, S., Sun, H., Wang, X., Tang, X., Guo, S., Zhang, Z., Huang, S., Xu, Y., Weng, Y., Mazourek, M., McGregor, C., Renner, S.S., Branham, S., Kousik, C., Wechter, W.P., Levi, A., Grumet, R., Zheng, Y., Fei, Z., 2023. CuGenDBv2: an updated database for cucurbit genomics. Nucleic Acids Res 51, D1457–D1464. https://doi.org/10.1093/nar/gkac921

Zhao, Z., Dong, Y., Wang, J., Zhang, G., Zhang, Z., Zhang, A., Wang, Z., Ma, P., Li, Y., Zhang, X., Ye, C., Xie, Z., 2022. Comparative transcriptome analysis of melon (Cucumis melo L.) reveals candidate genes and pathways involved in powdery mildew resistance. Sci Rep 12, 4936. https://doi.org/10.1038/s41598-022-08763-3

Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., Zhao, P.X., Rhee, S.Y., Fei, Z., 2016. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. Mol Plant 9, 1667–1670. https://doi.org/10.1016/j.molp.2016.09.014

Zotenko, E., Mestre, J., O'Leary, D.P., Przytycka, T.M., 2008. Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. PLoS Comput Biol 4, e1000140. https://doi.org/10.1371/journal.pcbi.1000140

Zou, C., Sun, K., Mackaluso, J.D., Seddon, A.E., Jin, R., Thomashow, M.F., Shiu, S.-H., 2011. *Cis* -regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences 108, 14992–14997. https://doi.org/10.1073/pnas.1103202108

ABBREVIATIONS LIST

ALS (Alternaria Leaf Spot)

CRES (Cis-Regulatory Elements)

CuGenDBv2 (Cucurbit Genomic DataBase v2)

DEGs (Differentially Expressed Genes)

FDR (False Discovery Rate)

GRNs (gene regulatory networks)

HMM (Hidden Markov Model)

log2fc (log2 fold change)

ORFs (Open Reading Frames)

PKs (Protein Kinases)

PlantCARE (Plant Cis Acting Regulatory Element)

PM (Powdery Mildew)

RKN (Root-Knot Nematode)

TFBSs (Transcription Factors Binding Sites)

TFs (Transcription Factors)

TPM (transcripts per million)

TSS (Transcription Start Site)

UTR (Untranslated Region)

WGD (Whole Genome Duplication)

PART THREE - FINAL CONSIDERATIONS

The findings of this thesis underscore the crucial importance of protein kinases in immunity and disease resistance in cucumbers. The comprehensive characterization of the 835 protein kinases in cucumbers, distributed in 123 families and 20 groups, provides a solid basis for future investigations into the specific functional roles of these protein kinases. Notably, the identification of 312 protein kinases with significant relevance in the biotic stress response, classified into 10 functional groups, paves the way for the development of targeted strategies to increase disease resistance in cucumbers and other Cucurbit species. In addition, the analysis of the promoter regions of the kinase genes differentially expressed under the influence of various pathogens revealed the presence of 51,339 cis-regulatory elements, classified into 125 distinct types and seven functional classes. These cis-regulatory elements, particularly those related to development, hormones, and light, have been shown to be effective in responding to biotic stress, highlighting their importance in mediating plant responses to pathogenic challenges. These findings are highly relevant to the advancement of agricultural practices, offering significant contributions to strengthening crop resistance against various diseases. In addition to contributing significantly to the understanding of the molecular basis of plant-pathogen interactions, the results of this research can be applied to the genetic engineering of plants that are more resistant to diseases, promoting agricultural sustainability and food security. Future work should investigate the specific functions of the identified protein kinase groups and the most effective cis-regulatory elements under biotic stress conditions. Studies exploring the interaction between these regulatory elements and hormonal signaling mechanisms may provide deeper insights into plant responses to pathogens. In addition, the application of the techniques developed in this study to other plant species can broaden the scope of the findings, benefiting a wider range of agricultural crops. Continued research in this area promises significant advances in the development of more robust and disease-resistant plant varieties.