

Communication

Identification of a C₂H₂ Transcription Factor (PsCZF3) Associated with RxLR Effectors and Carbohydrate-Active Enzymes in *Phytophthora sojae* Based on WGCNA

Yanhong Hu ^{1,†}, Zhihua He ^{1,†}, Yebin Kang ¹, Wenwu Ye ^{2,*}  and Linkai Cui ^{1,*} ¹ College of Horticulture and Plant Protection, Henan University of Science and Technology, Luoyang 471000, China² Department of Plant Pathology, Nanjing Agricultural University, Nanjing 210095, China

* Correspondence: yeww@njau.edu.cn (W.Y.); cuilk@haust.edu.cn (L.C.)

† These authors contributed equally to this work.

Abstract: *Phytophthora sojae* is a destructive soybean pathogen that orchestrates various secreted proteins (effectors) to modulate plant immunity and facilitate infection. Although a number of effectors have been identified and functionally studied in *P. sojae*, the way these molecules are regulated is marginally known. In this study, we performed a weighted gene correlation network analysis (WGCNA) based on digital RNA-seq, which enabled the identification of a transcription factor (PsCZF3) in *P. sojae*. This transcription factor is a C₂H₂-type zinc finger protein that regulates the transcription of 35 RxLR effectors during the early infection stage. Phylogenetic analysis revealed that PsCZF3 is a highly conserved protein across oomycetes, suggesting that this regulation mechanism may broadly exist in oomycete species. In addition, by building a subnetwork of PsCZF3 and correlated genes, we also found that PsCZF3 contributed to the transcriptional regulation of carbohydrate-active enzymes. Our findings suggest that the activation of PsCZF3 facilitates *P. sojae* infection by up-regulating RxLR effectors and carbohydrate-active enzymes.

Keywords: digital RNA-seq; *Phytophthora sojae*; WGCNA; transcription factor; RxLR effector

Citation: Hu, Y.; He, Z.; Kang, Y.; Ye, W.; Cui, L. Identification of a C₂H₂ Transcription Factor (PsCZF3) Associated with RxLR Effectors and Carbohydrate-Active Enzymes in *Phytophthora sojae* Based on WGCNA. *J. Fungi* **2022**, *8*, 998. <https://doi.org/10.3390/jof8100998>

Academic Editor: Katrina Maria Ramonell

Received: 13 August 2022

Accepted: 17 September 2022

Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Phytophthora sojae is a devastating oomycete pathogen that causes the stem and root rot of soybean crop, resulting in approximately USD1–2 billion losses globally per year [1]. Generally, oomycete pathogens use different mechanisms, such as the production of extracellular toxins, hydrolytic enzymes and inhibitors, and effector proteins, for their virulence [2]. Particularly, the proteaceous effectors, which are one of the most important virulence factors, target numerous host cellular processes, commonly leading to the suppression of PAMP-triggered immunity (PTI), effector-triggered immunity (ETI), and programmed cell death (PCD) [2]. Thus far, the best-studied effector class in oomycetes is the RxLR effectors, which are named for a conserved N-terminal amino acid sequence motif (arginine, any amino acid, leucine, arginine) [2]. Genomic and transcriptomic studies have advanced our understanding of the biology and pathology of *P. sojae* RxLR effectors. With genomic analysis, the *P. sojae* reference genome was found to have approximately 400 genes encoding RxLR effectors [3,4]. Transcriptomic analyses further indicated that these effectors were expressed in a timely manner. The immediate-early effectors were strongly expressed before infection and moderately induced upon infection (2- to 10-fold), while the early effectors were very weakly expressed prior to infection, but strongly induced (10- to 120-fold) during the first 12 h of infection patterns [4,5]. To date, however, the way these effectors are regulated is unknown.

In eukaryotes, the regulation of gene expression generally depends on the activity of trans acting factors such as transcription factors, which subsequently regulate the transcriptional initiation of genes. Transcription factors can be involved in the transcriptional

regulation of different (multiple) genes [6]. In oomycetes, transcription factors have been reported to contribute to pathogen development and virulence. For instance, in both *P. sojae* and *P. infestans*, the MADS-box transcription factors played significant roles in sporulation [7,8], while the Myb transcription factors regulated both vegetative growth and sporulation [9,10]. Other transcription factors have been shown to be involved in the resistance of plant defense response [11–13]. Despite the association of transcription factors with pathogenesis that have been found, no transcription factor has been reported to directly regulate RxLR effectors.

Weighted gene co-expression network analysis (WGCNA) is a powerful network analysis tool that describes the correlation of gene expression based on the microarray or RNA-seq data. It takes advantage of a graph theoretical approach to understanding correlations amongst genes and grouping genes into modules that typically have coordinated biological functions and regulatory mechanisms [14]. Currently, WGCNA has been extensively used to construct gene co-expression networks and identify centrally connected hub genes in humans, animals and plants [14–16]. However, this method is rarely applied to plant pathogens.

The objective of this study was to identify transcription factors associated with RxLR effectors by WGCNA. In the present study, the WGCNA was constructed based on the digital RNA-seq data during the early stage of the *P. sojae* infection. Key gene modules including a large number of RxLR effectors were identified. In one key module, we found a transcription factor closely associated with RxLR effectors and carbohydrate-active enzymes. The structure, DNA-binding sites and conservation of the protein were analyzed in detail. Based on the bioinformatic analyses, our study is expected to provide novel insights into the transcriptional regulation mechanisms of RxLR effectors.

2. Materials and Methods

2.1. Strain and Culture Conditions

P. sojae strain P6497 was used in this study. The strain was grown in 20% V8 liquid medium at 25 °C under darkness for 72 h and was then washed with sterile distilled water in order to collect the culture. The soybean cultivar Williams, which is susceptible to P6497, was grown in a greenhouse at 25 to 30 °C and was used at the second-leaf stage. A mycelial mat (d = 3 cm) was laid on and sandwiched between the upper surfaces of two leaves at 25 °C, respectively, for 0, 6, 12, 24 and 48 h post-inoculation (hpi) [5]. Three biological replicates were applied for each time point. The regions of the leaves in contact with the mycelia were excised together with the mycelia and were preserved in liquid nitrogen.

2.2. RNA Isolation, Library Preparation, and Sequencing

Total RNA was isolated from a mixture of leaf and mycelium using TRIzol reagent (Thermo Scientific, Waltham, MA, USA) according to the manufacturer's recommendation. The quality of RNA was estimated by using 1% agarose gel electrophoresis, and its concentration and purity were assessed by using the NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA). The integrity of total RNA were assessed on the Agilent Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA) using the RNA Nano 6000 Assay Kit. The sequencing library was constructed with KC-Digital Stranded mRNA-seq Library Prep Kit for Illumina (Seqhealth, Wuhan, China), according to the manufacturer's instructions. Oligo (DT) magnetic beads were used to enrich the mRNA with poly-A structure in the total RNA. The RNA was fragmented into ~300-bp fragments by ion interruption. Purified RNA was reversely transcribed into cDNA using random primers and reverse transcriptase, and the unique identifier (UID; a short sequence of 8–10 nt) was added. After the construction of the library, PCR was conducted to enrich the library fragments. Finally, the fifteen libraries were sequenced using the Illumina HiSeq X Ten System.

2.3. Digital RNA-Seq and Data Analysis

Using the Illumina HiSeq X Ten System, transcriptome sequencing generated 150-bp paired-end (PE) raw reads. The raw reads were subjected to quality control. High-quality reads were filtered using Trimmomatic v0.36 (Anthony M. Bolger, Berlin, Germany) [17], and data quality scores were sorted using the same software. The cleaned reads were further processed with the kcUID to remove duplication bias introduced during amplification and sequencing [18]. The consequent reads were grouped based on their UIDs, in which reads carrying the same UID were clustered into the same cluster. In the same cluster, the reads with a minimum of 95% sequence identity were extracted to generate a new sub-cluster through pairwise alignment. After all sub-clusters were generated, multiple sequence alignment was conducted to obtain a consensus sequence for every sub-cluster. The kcUID reads were mapped onto the *P. sojae* genome V3.0 using Hisat2 v2.0.1 (Daehwan Kim, Baltimore, MD, USA) [19]. The read count values on each gene as the original expression of the gene were compared using RSEM v1.2.31 (<http://deweylab.github.io/RSEM/>, accessed on 4 June 2016), and the expressions were standardized by RPKM (reads per kilobase per million mapped reads).

The gene expression profiles of all 15 samples were compared using the hierarchical clustering analysis (HCA). HCA was implemented according to a distance matrix of the Spearman correlation [20], and samples with a similar expression profile were clustered together. HCA was performed using the *gplots* functions in the R package [21]. The genes with a difference of at least two-fold change with *p*-value < 0.05 were considered as significantly differentially expressed genes. The “DESeq2” R language package was used to identify differentially expressed genes (DEGs) between non-infection (mycelia) stages and infection stages [22].

2.4. Weighted Gene Co-Expression Network Analysis

Co-expression analysis was conducted using WGCNA [23]. A matrix of pairwise Pearson correlation coefficients between all pairs of genes was created, and the matrix was converted into an adjacency matrix using the soft threshold power β , which proposes covariant similarity, and the topological overlap measure (TOM) similarity algorithm was used to transform the adjacency matrix into a topological overlap matrix to reduce noise and false correlation. Successively, hierarchical clustering was performed to identify modules. To obtain moderately sized modules, the minimum number of genes was set at 30 and merging modules with highly correlated ($r > 0.75$) eigengenes (defined as the first principal component of a given module and may be considered as a representative of the gene expression profiles in that module). All modules depicted in the hierarchical clustering dendrogram are henceforth referred to by their color labels. The grey module is used to hold all genes that do not clearly belong to any other module. To identify the relationship between modules and various infection stages, module–trait associations were estimated using the correlation between the module eigengene and the phenotype at the infection stages, which enabled the easy identification of the expression set (module) highly correlated with the phenotype. In addition, from the above analysis, we gained the weighted values, which represent the relationships between genes in pairs. The weighted values, which are >0.15 , were used to perform network analysis using Cytoscape v3.7.2 (Paul Shannon, Seattle, WA, USA) [24]. Weighted values > 0.15 were considered highly correlated and strongly regulatory.

2.5. Gene Ontology (GO) Enrichment Analysis

The ClusterProfiler package of R was used to identify the GO categories significantly enriched with the differentially expressed genes that were highly correlated with PsCZF3 [25]. $p < 0.05$ was considered as a threshold of GO enrichment analysis.

2.6. DNA-Binding Site Prediction

In this study, two transcription factors (PsCZF3 and PsCZF4) were identified. Because PsCZF3 had stronger associations with RxLR effectors than PsCZF4, we focused on PsCZF3. The ODN-binding site prediction for the Cys2-His2 zinc finger transcription factor PsCZF3 was performed based on the webtools “ZFModels” (<http://stormo.wustl.edu/ZFModels/>, accessed on 1 April 2014) and “DNA-binding Specificities of Cys2His2 Zinc Finger Proteins” (<http://zf.princeton.edu>, accessed on 9 September 2010). The predicted DNA-binding sequence was represented as a sequence logo [26].

2.7. Identification and Phylogenetic Analysis of PsCZF3 Orthologs

PsCZF3 orthologs were selected based on Blastp searches against NCBI (threshold values were set to $\geq 80\%$ coverage, $\geq 50\%$ identity, and ≥ 200 Max score). The full-length protein sequences of PsCZF3 orthologs were selected for phylogenetic analysis, and the tree was constructed by MEGA 6.0 [27], employing the neighbor-joining (NJ) algorithm. Bootstrap analysis with 1000 replicates was used to evaluate the significance of the nodes. Conserved motifs in the PsCZF3 homologous proteins were analyzed by MEME (<https://meme-suite.org/meme/tools/meme>, accessed on 5 February 2021) [28].

2.8. Validation by Reverse Transcription-Quantitative PCR (RT-qPCR)

The remaining total RNA samples were reversely transcribed to cDNA after sequencing libraries were constructed. Twelve genes from different modules were randomly selected for RT-qPCR analysis. Gene-specific primer pairs of the selected genes (Table S1) were designed using the primer design website (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>, accessed on 18 June 2012) and were commercially synthesized by Sangon Biotech Co., Ltd. (Shanghai, China). RT-qPCR was performed using a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). The reaction mixture with a total volume of 20 μ L contained 10 μ L of 2XWiz Universal SYBR qPCR Master Mix, 2 μ L of cDNA mix, 0.4 μ L of each primer (10 μ M), and 7.2 μ L of RNase-free double-distilled H₂O. The qRT-PCR was programmed at 95 °C for 60 s, followed by 40 cycles of 95 °C for 10 s and 60 °C for 20 s. A melting curve analysis was performed at the end of each PCR reaction at 95 °C for 15 s, 60 °C for 60 s, and 95 °C for 15 s. The *P. sojae* actin gene (ACT, accession number: XM_009530461.1) was used as a reference. Relative expression level was measured by normalizing the expression of targets against ACT, based on the $2^{-\Delta\Delta CT}$ method [29]. RT-qPCR assay was repeated three times and three technical replicates were used for each RT-qPCR.

3. Results

To create a weighted gene co-expression network and explore the transcription factors involved in the regulation of RxLR effectors, we performed the digital RNA-seq of the reference *P. sojae* strain P6497 during the early infection stage. The mycelium was sampled for RNA extraction at 0, 6, 12, 24, and 48 h after inoculation onto susceptible soybean leaf tissues. An average of 82,385,841 raw reads for a sample were generated (Table S2). To analyze the digital RNA-seq data, the raw sequencing reads were first filtered, which yielded 1112 million clean reads (Table S2). For each of the fifteen libraries, at least 98.40% of the clean reads had a quality score of Q30. The cleaned reads were further processed with the kcUID. Eventually, an average of 59 million of the UID reads were selected, and 89.79–97.00% of the reads were uniquely mapped to the *P. sojae* reference genome V3.0 (Table S3). In total, there were 28,142 genes, including both protein and non-protein coding genes. After excluding 4335 genes that showed an RPKM value = 0 in all 15 samples, 23,807 (84.6%) genes were present in at least one library. To test the correlation of the gene expression level among the samples, we conducted a hierarchical clustering analysis (HCA). Fifteen samples were clearly clustered into five groups on the basis of the infection time points, suggesting that there was a high correlation coefficient within the groups (Figure 1A).

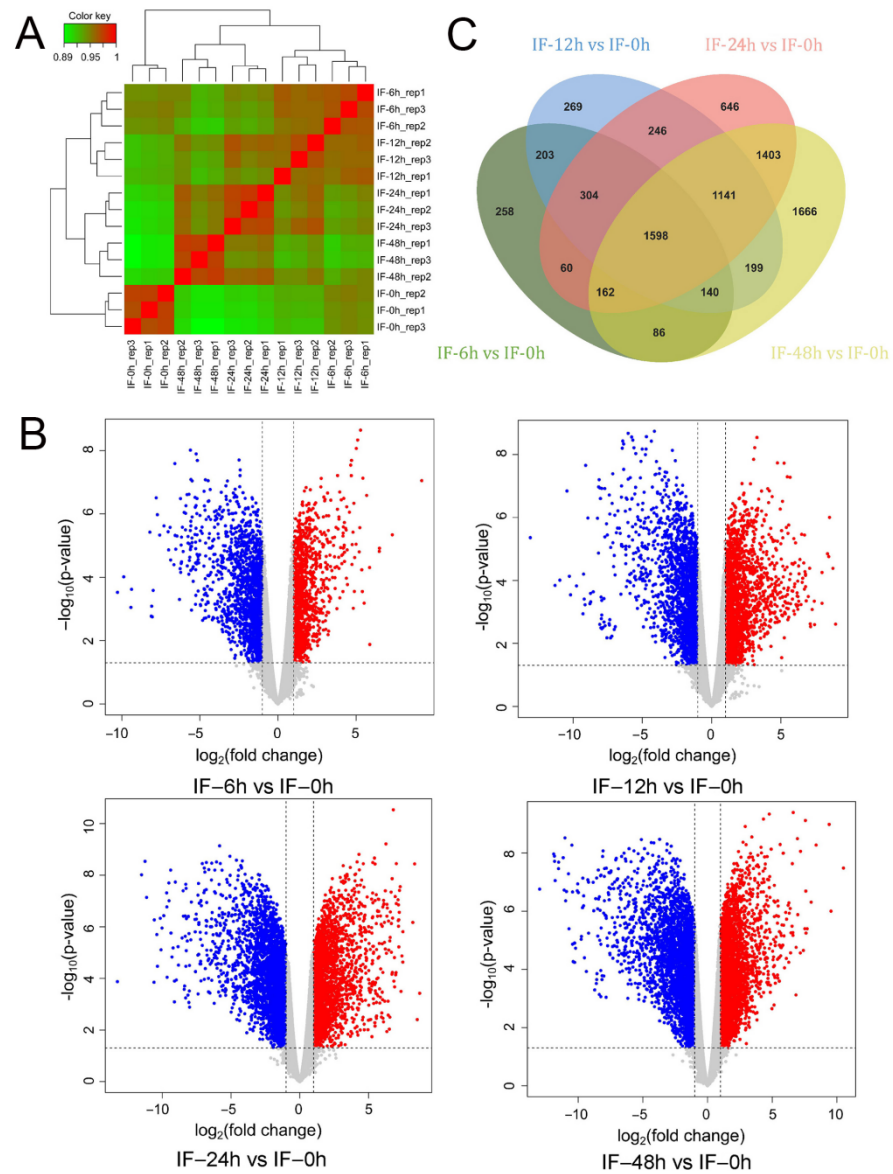


Figure 1. Comparison of gene expression and differentially expressed genes (DEGs) during the early infection stage of *P. sojae* to soybean leaves. (A) Hierarchical clustering analysis (HCA) of transcriptome data from 15 samples during the early infection. Color scale ranging from green to red indicates the inter-sample correlation from low to high. (B) Volcano plots illustrating the DEGs at different sampling time points after the inoculation of *P. sojae* to leaves. (C) Venn diagram showing the DEGs identified from comparison of 6, 12, 24 and 48 hpi with 0 hpi.

To examine the DEGs, we compared the gene expression levels at 6, 12, 24, and 48 hpi to that at 0 hpi. Some 8381 DEGs were observed from the four different expression groups (Figure 1B,C). To validate the DEGs that were identified, we randomly picked 12 genes based on their expression patterns at the five infection time points, and conducted RT-qPCR analysis. The RT-qPCR results show a consistency with those of RNA-seq (Supplementary Figure S1), confirming the gene expression differentiation that we obtained from the DESeq2 analysis.

To determine the average expression pattern of RxLR effectors at each time point of early infection, we measured the expression level values (RPKM) for each infection group by pooling all RxLR effectors. A total of 316 RxLR effector genes were detected by the digital RNA-seq. The expression of these RxLR effectors was markedly up-regulated and reached the peak value at 12 hpi, and then gradually decreased over infection time

(Supplementary Figure S2). Among the detected RxLR effector genes, 93 genes were differentially expressed. Further analysis of the differentially expressed RxLR effectors revealed that the expression of these genes changed more significantly at each time point during the early infection (Supplementary Figure S2), compared to the average expression of all RxLR effectors. Remarkably, the average expression level of the RxLR effectors at 12 hpi was significantly higher than those at other time points, implying that these genes may play a very important role in 12 hpi.

To explore genes that regulated the expression of RxLR effectors during the early stage of infection, we performed a WGCNA by using the 8381 DEGs. WGCNA was designed to identify highly correlated gene clusters and relate them to biological traits [23]. It modularly investigates the co-expressed genes and extracts intramodular hub genes from system networks, increasing the sensitivity to recognize the potential of worthwhile targets for biological regulations [23]. A total of 10 modules were obtained using a soft threshold of 22 (Figure 2A,B). The expression level of RxLR effectors at 12 hpi was the highest during early infection, so we focused on the module which was closely related with 12 hpi. The module eigengenes showed that the red module had the highest correlation with 12 hpi ($R^2 = 0.62$, $p < 0.01$) (Figure 2C). Interestingly, we found that 41 (44%) RxLR effector genes were clustered in the red module, which is much more abundant than those in other modules (Figure 2D). Collectively, these findings suggest that the red module is closely related to RxLR effectors.

In the network of gene interactions, the key regulators, such as transcription factors, are usually at the center of the network regulation and are the 'core' of the study. Transcription factor genes with high connectivity represent essential genes that are the center of the network and might be involved in very important functions in co-expression networks. Therefore, the visualization of the co-expression network for the red module was constructed to explore high connectivity transcription factor genes by Cytoscape v3.7.2. Two putative transcription factor genes PHYSODRAFT_353615 and PHYSODRAFT_286006 were identified from the red module according to the Joint Genome Institute website (JGI) transcription factor database (<https://mycocosm.jgi.doe.gov/mycocosm/proteins-browser/browse;p3YEuM?p=Physo3>, accessed on 12 November 2013)). Pfam v32.0 (<http://pfam.xfam.org/>, accessed on 11 June 2020) analysis revealed that PHYSODRAFT_353615 contained C₂H₂-type zinc finger protein (ZFP) domains, while PHYSODRAFT_286006 possessed a C3HC4-type zinc finger (RING finger) domain. We named PHYSODRAFT_353615 as *PsCZF3* (Genbank accession number: XP_009516505.1), and PHYSODRAFT_286006 as *PsCZF4* (Genbank accession number: XP_009527011.1). In order to identify the genes regulated by the two transcription factors, we constructed a co-expression sub-network employing the two transcription factors as guide genes. We found that 197 genes were highly correlated with *PsCZF3*, and 94 genes were highly correlated with *PsCZF4*. To our surprise, the 94 genes belonged to the members of the 197 genes, and *PsCZF4* was highly associated with *PsCZF3* (Figure 3). Closer examination indicated that among the 197 genes, 35 RxLR effector genes were closely correlated with *PsCZF3*, including 20 that were correlated with *PsCZF4*. These findings suggest that *PsCZF3* and *PsCZF4* can be two important transcription factors related to the RxLR effector regulation during the early stage of infection.

As the connectivity of *PsCZF3* is higher compared to *PsCZF4* in the red module, we would like to focus on *PsCZF3*. Given that *PsCZF3* is a C₂H₂-ZFP transcription factor, we further dissected its protein structure, particularly its DNA-binding sites. *PsCZF3* is composed of 254 amino acids, and contains four conserved C₂H₂-type zinc finger domains (Figure 4A). The four zinc finger domains are the DNA binding domain of *PsCZF3* and able to recognize 12-nucleotide DNA sequence (Figure 4B). We also examined the conservation of the *PsCZF3* protein across other oomycete species, and found that the protein is highly conserved among 16 species in six genera of oomycetes (Figure 4C; Supplementary Table S4), indicating *PsCZF3* may be a regulator that widely controls the expression of RxLR effector genes in different oomycete species.

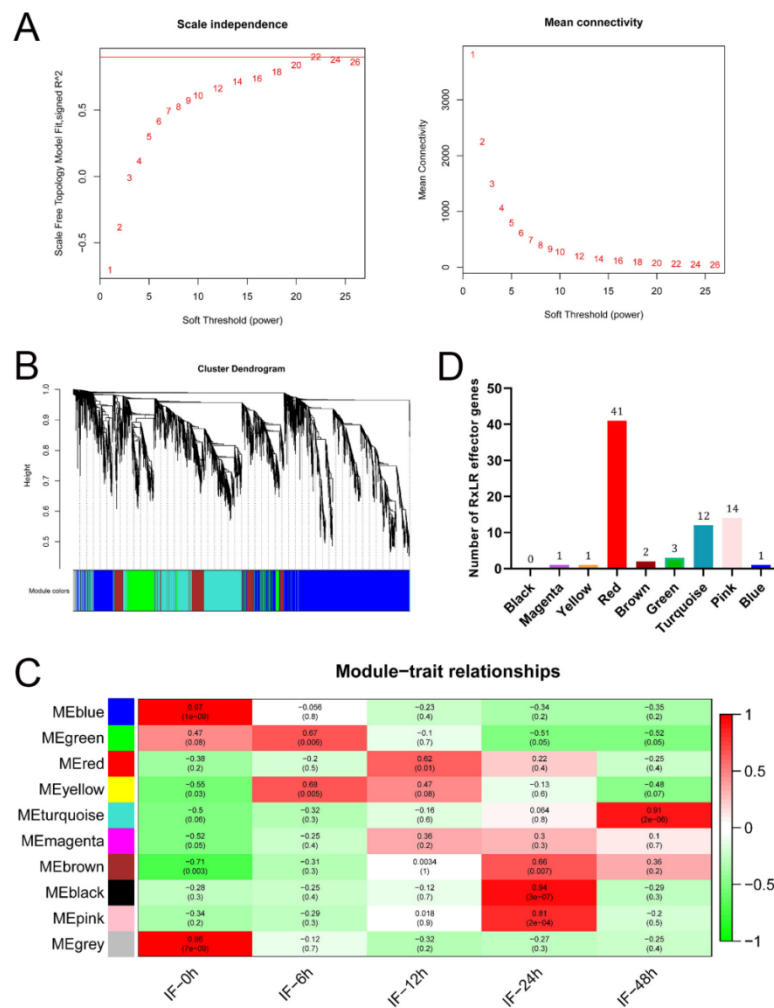


Figure 2. Construction of weighted gene co-expression network analyses (WGCNA) and module detection. **(A)** Analysis of the scale-free topology model fit index for soft threshold powers (β) and the mean connectivity for soft threshold powers. Left, the scale of the free fit index (y axis) as a function of the soft thresholding power (x axis). Right, the mean connectivity (degree, y axis) as a function of the soft thresholding power (x axis). The β value was set to ensure the high-scale independence (near 0.9) and low mean connectivity (near 0). **(B)** Module assignment in hierarchical clustered genes. Genes within different modules are labeled with different colors according to the WGCNA’s conventions. Ten co-expression modules were constructed and shown with distinct colors. The grey module contains all genes that were not involved in clustering and thus were not applied to subsequent analysis. **(C)** Matrix showing a module–trait relationship. Each row corresponds to a module. Each column corresponds to a time result. The module–trait relationship is colored based on their correlation, and red and green colors indicate strong positive and negative correlations, respectively. **(D)** Distribution of differentially expressed RxLR effectors in different modules. The number of RxLR effectors was counted after all modules filtered by the criteria that have weight value greater than 0.15.

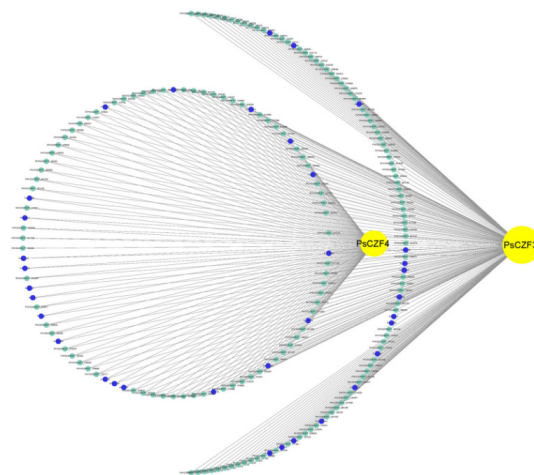


Figure 3. Network of transcription factors–effector interactions in the red modules. Yellow dots represent transcription factors, green dots indicate the correlated genes, and blue dots denote RxLR effector genes. The network was constructed by Cytoscape v3.7.2.

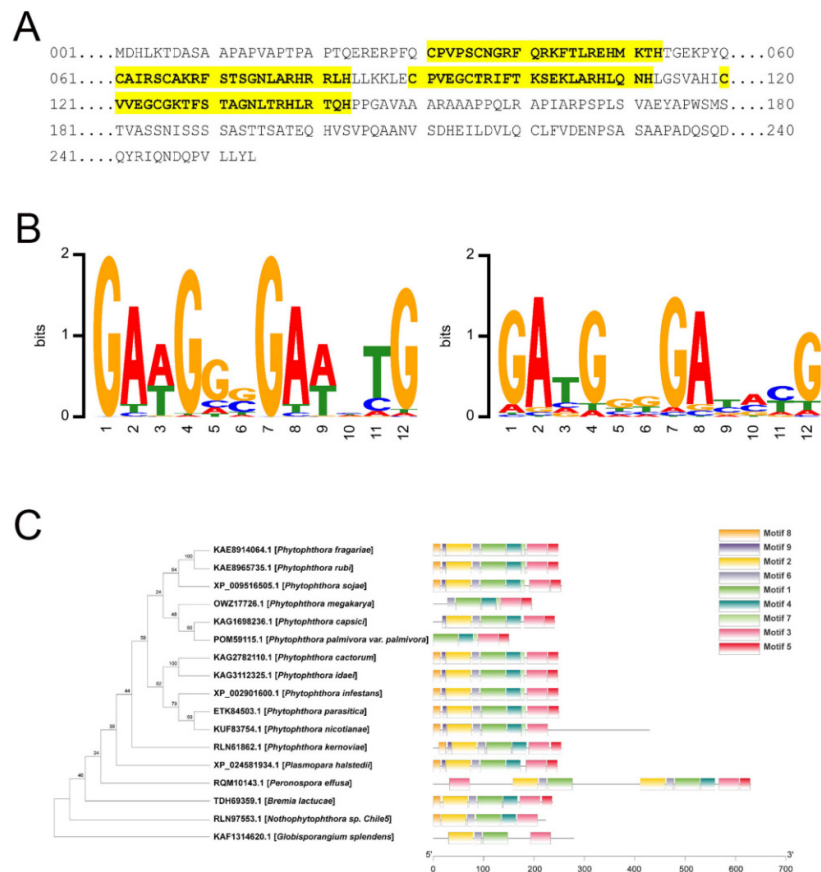


Figure 4. PsCzf3 protein domain analyses. (A) Amino acid sequence and putative C₂H₂-type zinc finger domains of PsCZF3. Amino acid sequences highlighted in yellow indicate predicted C₂H₂-type zinc finger domains. (B) Sequence logos for the predicted transcription factor binding site (TFBS) sequences. Left, TFBS analysis based on the webtool “ZFModels” (<http://stormo.wustl.edu/ZFModels/>, accessed on 1 April 2014); Right, TFBS prediction based on the webtool “DNA-binding Specificities of Cys2His2 Zinc Finger Proteins” (<http://zf.princeton.edu>, accessed on 9 September 2010). (C) Motif analysis of selected PsCZF3 orthologs across oomycetes. Each motif is represented by a numbered colored box. The length of the protein can be estimated while using the scale at the bottom.

In addition to the RxLR effector family, we also noticed that other effector families are present in the aforementioned 197 gene pool, such as N-terminal YxSL[ARK] (Tyr-Xaa-Ser-Leu [Arg/Lys]) motif containing secreted protein (YxSL) [30], protease, necrosis- and ethylene-inducing-like protein (NLP) [31], small cysteine-rich protein (SCP), elicitor, crinkler (CRN) and carbohydrate-active enzymes (CAZymes) (Supplementary Figure S3; Table S5). The gene ontology analysis of the genes that are highly correlated with *PsCZF3* revealed that CAZymes included the “carbohydrate metabolic process”, “hydrolase activity”, “hydrolyzing O-glycosyl compounds”, “pectate lyase activity” and “carbohydrate binding” (Figure 5). These enzymes are also secreted proteins that are mainly involved in the degradation of plant cell walls and promote adhesion, invasion, colonization and the nutrient absorption of pathogens [32,33]. These findings indicate that *PsCZF3* may not only regulate RxLR effectors, but also regulate some carbohydrate-active enzymes during the early stage of infection.

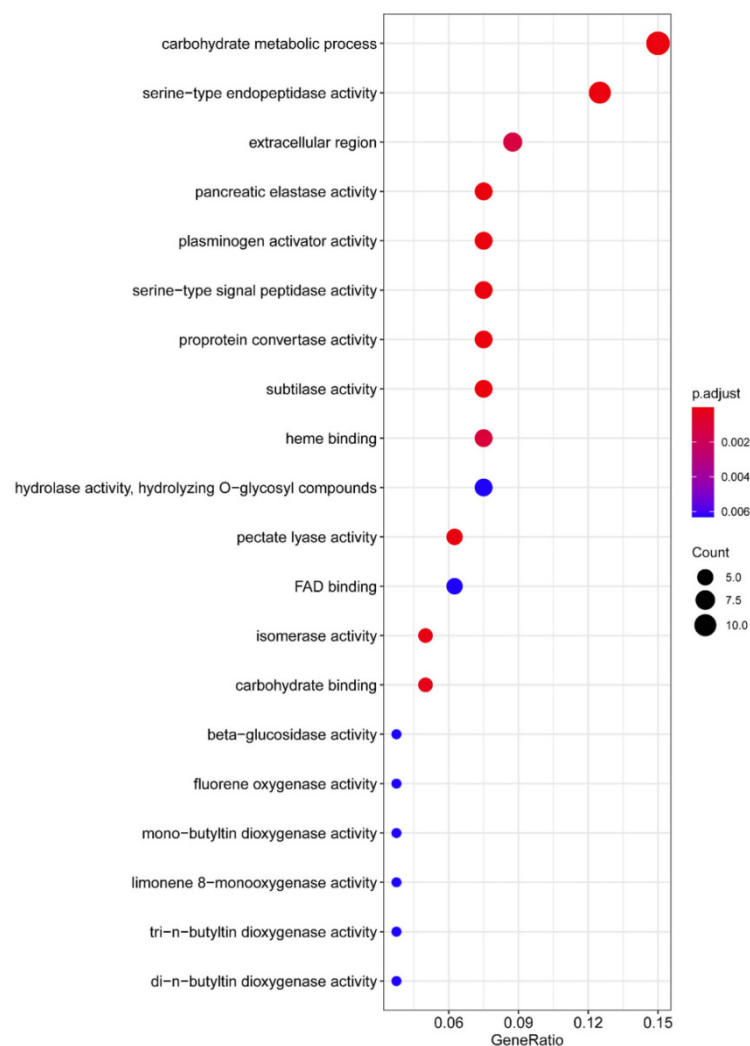


Figure 5. Gene Ontology (GO) enrichment analysis of the genes correlated with *PsCZF3*. Color scale represents the Benjamini–Hochberg adjusted p -values, and the size of circles indicates gene numbers. X-axis, GeneRatio, amount of DEGs enriched in the pathway/amount of all genes in annotation gene set; Y-axis, GO term ($p < 0.05$).

4. Discussion

In this study, we demonstrated that WGCNA in combination with digital RNA-seq to illustrate the regulation network in oomycetes. WGCNA has been widely used for various genomic applications [34,35]. Unlike ordinary clustering methods, WGCNA uses

correlation coefficient weighted values in its analysis, making the connections between genes in the network obey a scale-free network distribution, an algorithm that is more biologically meaningful. One of the biggest advantages of WGCNA is its ability to divide thousands of genes from multiple samples into several modules based on their expression patterns. By analyzing them on a module-by-module basis, the computational workload is decreased while the accuracy is increased. We discovered a C₂H₂-type zinc finger protein transcription factor that is associated with 35 RxLR effector genes and other virulence factors such as CAZymes. This study advances our understanding of transcriptional regulation in oomycetes.

In comparison to digital RNA-seq, the traditional RNA-seq method has some constraints, mainly due to the library construction step that relies on PCR amplification. Systemic errors could be magnified by increasing PCR cycles, which is more prominent in sequencing with the small initial samples. As a result, the sequencing outcomes sometimes do not truly reflect the expression levels of genes or transcripts in different samples. UID-mRNA-Seq is an absolute quantification of the initial molecule by counting UID, and therefore, more accurate gene expression results can be observed. In this study, 23,807 genes of *P. sojae* were detected in at least one library using digital RNA-seq. In the previous study, 14,969 genes of *P. sojae* were detected in at least one library using traditional RNA-seq [5]. The number of detected genes increased from 14,969 to 23,807, indicating that the transcriptome profiling by digital RNA-seq is not only more accurate but also more abundant. In the present study, a total of 316 RxLR effector genes were detected during the early infection stage of *P. sojae* to soybean leaves. The expression of these RxLR effectors was markedly up-regulated and reached the peak value at 12 hpi. The finding is consistent with the prior studies reported by Ye et al. [5], suggesting that 12 hpi is an important time point for RxLR effectors.

C₂H₂-type ZFP is the largest transcription factor superfamily in eukaryotes [36]. The only remarkable feature of the C₂H₂-type ZFPs is its DNA binding domain that contains zinc fingers arranged in tandem [37]. Each zinc finger is able to recognize 3-nucleotide DNA sequences [38], and with a combination of adjacent zinc fingers, C₂H₂-type ZFPs recognize long and complex DNA patterns. C₂H₂ transcription factors were shown to be important for pathogenic fungi, such as *Aspergillus fumigatus* [39,40], *Botrytis cinerea* [41], *Candida albicans* [42], *Cryptococcus neoformans* [43], *Ustilago maydis* [44] and *Magnaporthe oryzae* [45]. The disruption of C₂H₂ transcription factor in these fungi attenuated fungal virulence. In *P. sojae*, the C₂H₂ transcription factor *PsCZF1* was required for growth, development and pathogenesis. *PsCZF1* silenced transformants displayed a deficiency in terms of hyphal growth and sporulation, and loss of virulence on host soybean cultivars [46]. With the emerging CRISPR/Cas9 technology [47], it will be of interest to study the function of *PsCZF3* in the future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jof8100998/s1>, Figure S1: Confirmation of expression of 12 randomly selected DEGs by qRT-PCR; Figure S2: Average gene expression of the RxLR effector family during the early stage of infection; Figure S3: Gene categories correlated with *PsCZF3*; Table S1: Primers used in this study; Table S2: Statistical analysis of digital RNA-seq reads; Table S3: Summary of digital RNA-seq reads mapped to the reference genome; Table S4: *PsCZF3* ortholog genes in the oomycetes; Table S5: Protein-coding genes correlated with *PsCZF3*.

Author Contributions: Conceptualization, L.C. and W.Y.; methodology, Y.H. and Z.H.; software, Y.H. and Z.H.; writing—original draft preparation, all authors; writing—review and editing, L.C. and W.Y.; funding acquisition, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Science Foundation of China (project no. U1804104 and U1404318), and the Key Scientific and Technological Project of Henan Province (project no. 212102110141).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The RNA-Seq data are available in the NCBI BioProject database at <https://www.ncbi.nlm.nih.gov/bioproject/with> accession code PRJNA832871 (accessed on 27 April 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Erwin, D.C.; Ribeiro, O.K. *Phytophthora Diseases Worldwide*; American Phytopathological Society Press: St. Paul, MN, USA, 1996.
- Jiang, R.H.; Tyler, B.M. Mechanisms and evolution of virulence in oomycetes. *Annu. Rev. Phytopathol.* **2012**, *50*, 295–318. [[CrossRef](#)] [[PubMed](#)]
- Tyler, B.M.; Tripathy, S.; Zhang, X.; Dehal, P.; Jiang, R.H.Y.; Aerts, A.; Arredondo, F.D.; Baxter, L.; Bensasson, D.; Beynon, J.L.; et al. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **2006**, *313*, 1261–1266. [[CrossRef](#)] [[PubMed](#)]
- Wang, Q.; Han, C.; Ferreira, A.O.; Yu, X.; Ye, W.; Tripathy, S.; Kale, S.D.; Gu, B.; Sheng, Y.; Sui, Y.; et al. Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. *Plant Cell.* **2011**, *23*, 2064–2086. [[CrossRef](#)]
- Ye, W.; Wang, X.; Tao, K.; Lu, Y.; Dai, T.; Dong, S.; Dou, D.; Gijzen, M.; Wang, Y. Digital gene expression profiling of the *Phytophthora sojae* transcriptome. *Mol. Plant-Microbe Interac.* **2011**, *24*, 1530–1539. [[CrossRef](#)] [[PubMed](#)]
- Sassone-Corsi, P.; Borrelli, E. Transcriptional regulation by trans-acting factors. *Trends Genet.* **1986**, *2*, 215–219. [[CrossRef](#)]
- Leesutthiphonchai, W.; Judelson, H.S. A MADS-box transcription factor regulates a central step in sporulation of the oomycete *Phytophthora infestans*. *Mol. Microbiol.* **2018**, *110*, 562–575. [[CrossRef](#)]
- Lin, L.; Ye, W.; Wu, J.; Xuan, M.; Li, Y.; Gao, J.; Wang, Y.; Wang, Y.; Dong, S.; Wang, Y. The MADS-box transcription factor PsMAD1 is involved in zoosporogenesis and pathogenesis of *Phytophthora sojae*. *Front. Microbiol.* **2018**, *9*, 2259. [[CrossRef](#)]
- Xiang, Q.; Judelson, H.S. Myb transcription factors and light regulate sporulation in the oomycete *Phytophthora infestans*. *PLoS ONE* **2014**, *9*, e92086. [[CrossRef](#)]
- Zhang, M.; Lu, J.; Tao, K.; Ye, W.; Li, A.; Liu, X.; Kong, L.; Dong, S.; Zheng, X.; Wang, Y. A Myb transcription factor of *Phytophthora sojae*, regulated by MAP kinase PsSAK1, is required for zoospore development. *PLoS ONE* **2012**, *7*, e40246. [[CrossRef](#)]
- Gamboa-Melendez, H.; Huerta, A.I.; Judelson, H.S. bZIP Transcription factors in the oomycete *Phytophthora infestans* with novel DNA-binding domains are involved in defense against oxidative stress. *Eukaryot. Cell* **2013**, *12*, 1403–1412. [[CrossRef](#)]
- Sheng, Y.; Wang, Y.; Meijer, H.J.G.; Yang, X.; Hua, C.; Ye, W.; Tao, K.; Liu, X.; Govers, F.; Wang, Y. The heat shock transcription factor PsHSF1 of *Phytophthora sojae* is required for oxidative stress tolerance and detoxifying the plant oxidative burst. *Environ. Microbiol.* **2015**, *17*, 1351–1364. [[CrossRef](#)] [[PubMed](#)]
- Sheng, Y.; Lin, L.; Chen, H.; Pu, T.; Liu, X.; Dong, S.; Ye, W.; Wang, Y. The bZIP transcription factor PsBZP32 is involved in cyst germination, oxidative stress response, and pathogenicity of *Phytophthora sojae*. *Phytopathol. Res.* **2021**, *3*, 1. [[CrossRef](#)]
- Maertens, A.; Tran, V.; Kleensang, A.; Hartung, T. Weighted Gene Correlation Network Analysis (WGCNA) Reveals Novel Transcription Factors Associated with Bisphenol A Dose-Response. *Front. Genet.* **2018**, *9*, 508. [[CrossRef](#)]
- Fan, Y.; Arbab, A.A.I.; Zhang, H.; Yang, Y.; Yang, Z. Lactation Associated Genes Revealed in Holstein Dairy Cows by Weighted Gene Co-Expression Network Analysis (WGCNA). *Animal* **2021**, *11*, 314. [[CrossRef](#)] [[PubMed](#)]
- Zhu, Q.; Gao, S.; Zhang, W. Identification of Key Transcription factors related to bacterial spot resistance in pepper through regulatory network analyses. *Genes* **2021**, *12*, 1351. [[CrossRef](#)] [[PubMed](#)]
- Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
- Shugay, M.; Britanova, O.V.; Merzlyak, E.M.; Turchaninova, M.A.; Mamedov, I.Z.; Tuganbaev, T.R.; Bolotin, D.A.; Staroverov, D.B.; Putintseva, E.V.; Plevova, K.; et al. Towards error-free profiling of immune repertoires. *Nat. Methods* **2014**, *11*, 653–655. [[CrossRef](#)]
- Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)]
- Schober, P.; Boer, C.; Schwarte, L.A. Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [[CrossRef](#)]
- Warnes, G.R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.; Liaw, A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; et al. Gplots: Various R programming tools for plotting data. *R Package Version* **2009**, *2*, 1.
- Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
- Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)] [[PubMed](#)]
- Su, G.; Morris, J.H.; Demchak, B.; Bader, G.D. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinform.* **2014**, *47*, 11–24. [[CrossRef](#)] [[PubMed](#)]
- Yu, G.; Wang, L.; Han, Y.; He, Q. cluster Profiler: An R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)]

26. Pratt, H.; Weng, Z. LogoJS: A Java script package for creating sequence logos and embedding them in web applications. *Bioinformatics* **2020**, *36*, 3573–3575. [[CrossRef](#)] [[PubMed](#)]
27. Tamura, K.; Stecher, G.; Peterson, D.; Filipinski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [[CrossRef](#)] [[PubMed](#)]
28. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [[CrossRef](#)]
29. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)]
30. Levesque, C.A.; Brouwer, H.; Cano, L.; Hamilton, J.P.; Holt, C.; Huitema, E.; Raffaele, S.; Robideau, G.P.; Thines, M.; Win, J.; et al. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* **2010**, *11*, R73. [[CrossRef](#)]
31. Dong, S.; Kong, G.; Qutob, D.; Yu, X.; Tang, J.; Kang, J.; Dai, T.; Wang, H.; Gijzen, M.; Wang, Y. The NLP toxin family in *Phytophthora sojae* includes rapidly evolving groups that Lack necrosis-inducing activity. *Mol. Plant-Microbe Interact.* **2012**, *25*, 896–909. [[CrossRef](#)]
32. Schäfer, W. Molecular mechanisms of fungal pathogenicity to plants. *Annu. Rev. Phytopathol.* **1994**, *32*, 461–477. [[CrossRef](#)]
33. Zerillo, M.M.; Adhikari, B.N.; Hamilton, J.P.; Buell, C.R.; Lévesque, C.A.; Tisserat, N. Carbohydrate-active enzymes in *Pythium* and their role in plant cell wall and storage polysaccharide degradation. *PLoS ONE* **2013**, *8*, e72572. [[CrossRef](#)] [[PubMed](#)]
34. Massonnet, M.; Morales-Cruz, A.; Figueroa-Balderas, R.; Lawrence, D.P.; Baumgartner, K.; Cantu, D. Condition-dependent co-regulation of genomic clusters of virulence factors in the grapevine trunk pathogen *Neofusicoccum parvum*. *Mol. Plant Pathol.* **2018**, *19*, 21–34. [[CrossRef](#)] [[PubMed](#)]
35. Li, Z.; Zhang, W.; Jue, D.; Liu, X.; Jiang, Y.; Tang, J. Transcriptome changes induced by *Botrytis cinerea* stress and weighted gene co-expression network analysis (WGCNA) in *Actinidia chinensis*. *Plant Mol. Biol. Rep.* **2022**, *40*, 389–401. [[CrossRef](#)]
36. Vaquerizas, J.M.; Kummerfeld, S.K.; Teichmann, S.A.; Luscombe, N.M. A census of human transcription factors: Function, expression and evolution. *Nat. Rev. Genet.* **2009**, *10*, 252–263. [[CrossRef](#)]
37. Wolfe, S.A.; Nekudova, L.; Pabo, C.O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183–212. [[CrossRef](#)]
38. Pavletich, N.P.; Pabo, C.O. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **1991**, *252*, 809–817. [[CrossRef](#)]
39. Soriani, F.M.; Malavazi, I.; da Silva Ferreira, M.E.; Savoldi, M.; Von Zeska Kress, M.R.; de Souza Goldman, M.H.; Loss, O.; Bignell, E.; Goldman, G.H. Functional characterization of the *Aspergillus fumigatus* CRZ1 homologue, CrzA. *Mol. Microbiol.* **2008**, *67*, 1274–1291. [[CrossRef](#)]
40. Cramer, R.A.; Perfect, B.Z.; Pinchai, N.; Park, S.; Perlin, D.S.; Asfaw, Y.G.; Heitman, J.; Perfect, J.R.; Steinbach, W.J. Calcineurin target CrzA regulates conidial germination, hyphal growth, and pathogenesis of *Aspergillus fumigatus*. *Eukaryot. Cell* **2008**, *7*, 1085–1097. [[CrossRef](#)]
41. Schumacher, J.; de Larrinoa, I.F.; Tudzynski, B. Calcineurin-responsive zinc finger transcription factor CRZ1 of *Botrytis cinerea* is required for growth, development, and full virulence on bean plants. *Eukaryot. Cell* **2008**, *7*, 584–601. [[CrossRef](#)]
42. Karababa, M.; Valentino, E.; Pardini, G.; Coste, A.T.; Bille, J.; Sanglard, D. CRZ1, a target of the calcineurin pathway in *Candida albicans*. *Mol. Microbiol.* **2006**, *59*, 1429–1451. [[CrossRef](#)] [[PubMed](#)]
43. Chang, Y.C.; Wright, L.C.; Tschärke, R.L.; Sorrell, T.C.; Wilson, C.F.; Kwon-Chung, K.J. Regulatory roles for the homeodomain and C₂H₂ zinc finger regions of *Cryptococcus neoformans* Ste12 α . *Mol. Microbiol.* **2004**, *53*, 1385–1396. [[CrossRef](#)]
44. Flor-Parra, I.; Vranes, M.; Kamper, J.; Perez-Martin, J. Biz1, a zinc finger protein required for plant invasion by *Ustilago maydis*, regulates the levels of a mitotic cyclin. *Plant Cell* **2006**, *18*, 2369–2387. [[CrossRef](#)] [[PubMed](#)]
45. Cao, H.; Huang, P.; Zhang, L.; Shi, Y.; Sun, D.; Yan, Y.; Liu, X.; Dong, B.; Chen, G.; Snyder, J.H.; et al. Characterization of 47 Cys(2)-His(2) zinc finger proteins required for the development and pathogenicity of the rice blast fungus *Magnaporthe oryzae*. *New Phytol.* **2016**, *211*, 1035–1051. [[CrossRef](#)]
46. Wang, Y.; Dou, D.; Wang, X.; Li, A.; Sheng, Y.; Hua, C.; Cheng, B.; Chen, X.; Zheng, X.; Wang, Y. The PsCZF1 gene encoding a C₂H₂ zinc finger protein is required for growth, development and pathogenesis in *Phytophthora sojae*. *Microb. Pathog.* **2009**, *47*, 78–86. [[CrossRef](#)] [[PubMed](#)]
47. Fang, Y.; Tyler, B.M. Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9. *Mol. Plant Pathol.* **2016**, *17*, 127–139. [[CrossRef](#)] [[PubMed](#)]