agronomy

MDPI

Article

Genome-Wide Characterization of NLRs in *Saccharum spontaneum* L. and Their Responses to Leaf Blight in *Saccharum*

Zhoutao Wang, Fu Xu, Hui Ren, Guilong Lu, Youxiong Que and Liping Xu *

Key Laboratory of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture and Rural Affairs, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2170102010@fafu.edu.cn (Z.W.); 2200101003@fafu.edu.cn (F.X.); 3190130030@fafu.edu.cn (H.R.); 2190102004@fafu.edu.cn (G.L.); queyouxiong@fafu.edu.cn (Y.Q.)

* Correspondence: xuliping@fafu.edu.cn

Abstract: Sugarcane is an important sugar and potential energy crop, and the complexity of its genome has led to stagnant progress in genome decipherment and hindered the genome-wide analysis of the nucleotide binding site leucine-rich repeat (NLR) receptor until the genome of *Saccharum spontaneum* was published. From the genome of *S. spontaneum*, 724 allelic and non-allelic NLRs were identified and classified into five types (N, NL, CN, CNL, and P) according to domain architectures and integrity and at least 35 genes encoded non-canonical domains. The phylogenetic analysis indicated NLRs containing the coiled-coil (CC) domain separated from those without CC in six Poaceae species, including *S. spontaneum*. The motif analysis determined the characteristics and potential functions of the 137 representative non-allelic NLRs, especially the core motifs contained in the NBS and LRR domains, which indicated that motifs were regularly distributed among clades. Through transcription factor binding site (TFBS) profiles, we predicted that the most important transcription regulator of NLRs in sugarcane were ERF, MIKC_MADS, and C2H2. In addition, based on three sets of transcriptome data from two sugarcane hybrids and one *S. spontaneum* clone infected by the necrotrophic fungal pathogen *Stagonospora tainanensis* causing sugarcane leaf blight (SLB), the expression dynamics of NLRs responding to the infection in three sugarcane clones were compared. The different genetic background led to the significant difference of NLRs response to SLB in different sugarcane clones, and we got an inference of the potential mechanism of SLB resistance. These results provided a basic reference and new insights to further study and utilize the NLRs.

Keywords: *Saccharum spontaneum*; NLR receptor; genome-wide characterization; sugarcane leaf blight; gene expression

Citation: Wang, Z.; Xu, F.; Ren, H.; Lu, G.; Que, Y.; Xu, L. Genome-Wide Characterization of NLRs in *Saccharum spontaneum* L. and Their Responses to Leaf Blight in *Saccharum*. *Agronomy* **2021**, *11*, 153. https://doi.org/10.3390/agronomy11010153

Received: 14 December 2020
Accepted: 11 January 2021
Published: 15 January 2021

1. Introduction

The immobility of plants causes passivity to accept environmental impacts. In response, plants have evolved into a sophisticated, multi-layered mechanism for responding to pathogenic challenge and triggering immune responses to counteract pathogen attack. Immune responses are mainly induced by plasma membrane pattern recognition receptors (PRRs) and cytoplasmic recognition receptors encoded by resistance (R) genes. PRRs directly perceive relatively conservative small molecules, which were defined as pathogen associated molecular patterns (PAMPs) and were secreted externally to the plant cells by invading pathogens [1]. However, in order to overcome the recognition of PRRs, pathogens can also directly generate effectors into plant cells to disturb cell homeostasis [2]. Fortunately, the effector is also acting as the trigger of the second defense by specific interactions with a class of disease resistance proteins with a nucleotide binding site (NBS) and a leucine-rich repeat (LRR) domain, serve as recognition receptors encoded by NLR or NBS-LRR genes in plants, and strongly revealed the evil invading behavior of pathogens [3,4]. A hallmark of the resistance mediated by R proteins is hypersensitive response (HR), which often facilitates a localized cell programmed death response around the infected tissues to

prevent the further spread of disease symptom and infection of pathogens [5–7]. The two defense modes mentioned above are the so-called PAMP triggered immunity (PTI) and effector triggered immunity (ETI), respectively.

The highly conservative NBS domain is also referred to as the NB-ARC domain since it was also found both in human apoptotic protease-activating factor-1 (Apaf-1) and caenorhabditis elegans death-4 protein (CED-4) [8,9]. Most of the members of the NLR family can be roughly divided into three typical classes in accordance with different amino-terminal domains named toll/interleukin-1 receptor (TIR), coiled-coil (CC), and resistance to powdery mildew8 (RPW8), which are referred to as TIR-NLRs (TNLs), CC-NLRs (CNLs), and RPW8-NLRs (RNLs), respectively [10]. However, the NLR family has a high structural diversity. Except for the three typical classes aforementioned, it also contains many other small classes based on structural diversity. Truncation of NLRs is one way of diversity formation, which contains truncation of a single domain, such as LRR (named CN- or TN-type), RPW8, TIR, or CC (named NL-type). There is only one N domain left, which is named the N-type. The presence of atypical domains, or integrated domains, for instance, LIM, BED, WRKY, MAPKKK, SD, PK, and ZF, can also increase the structural diversity of the NLR family [11].

With the development of plant genomics, the genome-wide analysis of important gene families at different taxonomic levels has become available. At the beginning of the 21st century, pioneering genome-wide studies of NLR started in *Arabidopsis* [12] and rice (*Oryza sativa*) [13], and other species were also subsequently quickly involved. These efforts correspond to the importance of NLR in ecology and breeding and simultaneously revealed that the number of NLR genes varies from less than one hundred to several thousand in different species [14]. Although sugarcane is an important sugar crop, accounting for about 80% of total sugar and cultivated in more than one hundred countries in the world as well as acting as a potential energy crop, the complexity of sugarcane genome has led to stagnant progress in genome decipherment. Gratifyingly, one clone named AP85-441 belongs to *S. spontaneum*, which is one of the two ancestor species of modern sugarcane cultivars (*Saccharum* spp. hybrids, Poaceae), in which the genome was recently published by Zhang et al. [15]. Sugarcane hybrids are derived from interspecific crosses between *S. officinarum* and *S. spontaneum*, and have approximately 10–15% chromosomes derived from *S. spontaneum,* which is the most important contributor to many important biotic resistance and abiotic tolerance [16,17].

The first cloned R gene was *Hm1* from maize in 1992, and subsequently over about 25 years, the number of cloned R genes has steadily increased and some mechanisms of NLR genes recognizing pathogens are already known [18]. New disease-resistant NLR genes and the mechanisms of NLR perception are also constantly being explored and discovered [19–21], which provides valuable resources for people to fully understand the working mechanism of NLR genes and then utilize NLR genes in breeding work. Undoubtedly, the genome-wide identification and analysis of the resistant gene family NLR in *S. spontaneum* also represents an important task and a vital precondition in sugarcane disease-resistant breeding. In the current study, we first used the Hmmer and Blast software to search all NLR genes in the *S. spontaneum* genome, and compared the number and density of NLR genes with five other different species in Poaceae. Then, conservative motif, gene structure, and upstream TFBSs were analyzed for those genes containing both the complete NBS and LRR domain, and the distribution characteristics of all NLR genes on the chromosomes were described. Finally, based on the transcriptome data of sugarcane response to sugarcane diseases leaf blight, we detected some active NLR genes expressed during the sugarcane response to disease infection.

2. Materials and Methods

2.1. Sugarcane S. spontaneum and Other Poaceae Species Genome Resources

The whole genome assembly v4.1 of the haploid *S. spontaneum*, AP85-441, containing eight homologous groups (Chr1–Chr8) of four members (A–D) each (32 chromo-

somes), as well as the protein, gene, CDS, and gff3 annotation files, were downloaded from the following link: http://www.life.illinois.edu/ming/downloads/Spontaneum_genome/ [15]. Genome sequences and annotation files of five other Poaceae species, including millet (*Setaria italica*, Setaria_italica_v2.0), rice (Japonica Group, IRGSP-1.0), *Brachypodium distachyon* (Brachypodium_distachyon_v3.0), sorghum (*Sorghum bicolor*, Sorghum_bicolor_NCBIv3), and maize (*Zea mays*, B73 RefGen_v4) were downloaded from the EnsemblPlants database (http://plants.ensembl.org/species.html). The six species above belong to three subfamilies of Poaceae, the Panicoideae (*S. spontaneum*, *S. bicolor*, *Z. mays,* and *S. italica*), Pooideae (*B. distachyon*), and Ehrhartoideae (*O. sativa*).

2.2. Identification of NLR Genes

In order to search for the total NLR genes as much as possible, two search strategies were used in the current study. One search strategy was based on the software Hmmer v3 using the Hidden Markov Model (HMM) corresponding to the NBS (NB-ARC) family (PF00931, NBS.hmm), which was downloaded from the Pfam database (http://pfam.xfam.org/). This NBS.hmm profile was built by nine NBS domain sequences from other species. The high-quality, non-redundant, and complete NBS domain protein sequences were obtained using the "hmmsearch" command from the Hmmer v3 with the E-value < 1×10^{-15} and CDD tool [22] from NCBI with the E-value < 1×10^{-3}. Then, those NBS sequences were aligned by the Clustalw software and a *S. spontaneum*-specific HMM (Ss-NBS.hmm) was built using the "hmmbuild" command from the Hmmer v3. Again, the "hmmsearch" command was used based on the "Ss-NBS.hmm" profile with the E-value < 1×10^{-2}, and all searched proteins were filtered by the CDD tool with the E-value < 1×10^{-3}.

Another search strategy was based on local Blast [23]. First, 7,125 plant NLR genes in the NCBI protein database were downloaded using the search keyword "NBS-LRR", and a local Blast database specially for the NLR gene search was built by the "makeblastdb" command. Except for those proteins searched by Hmmer, all the remaining proteins from the *S. spontaneum* annotation were searched by "blastp" against the NLR database, and candidate proteins from blastp results were filtered by the CDD tool with the E-value < 1×10^{-3}, similar to the aforementioned strategy. Following the same approaches, the NLR genes in the genomes of five other Poaceae species *S. italica*, *O. sativa*, *B. distachyon*, *S. bicolor,* and *Z. mays* were also subsequently identified, as much as possible. In addition, atypical domains can also be detected by the CDD tool on some sequences. By analyzing the integrality of the NBS domain, these proteins were classified into two types: complete genes (encoding the complete NBS domain) and partial genes (encoding the incomplete NBS domain), by the CDD tool.

2.3. Multiple Sequence Alignment and Phylogenetic Analysis

The phylogenetic analysis was conducted to investigate whether there was an obvious segregation of NLR proteins from different species of Poaceae, as well as from different gene types. In this study, a total of 200 NLR proteins were randomly selected. Among them, 50 NLR proteins were selected from *S. spontaneum* and 150 on average were selected from each of the other Poaceae species, including *S. spontaneum*, *S. italica*, *O. sativa*, *B. distachyon*, *S. bicolor,* and *Z. mays*. As the indispensable and most conservative domains in NLRs, NBSs are ideal sequences for constructing the phylogenetic tree [24]. The complete NBS domain regions for every protein were extracted using an in-house perl script and the multiple alignment were performed by the MEGA X software using the ClustalW algorithm with default parameter settings. After trimming the poorly aligned regions at both ends of aligned proteins, the Maximum Likelihood method based on the WAG with the Freqs. (+F) model was used to conduct a phylogenetic tree and the bootstrap value was set to 1000 (test the node) [24]. Another phylogenetic tree of 137 representative NLR proteins were also constructed using the above methodology.

2.4. Sequence Analysis

The types of CNL and NL containing both NBS and LRR domains were recognized as the dominant class of NLR genes [12,25,26], which are our focus in the sequence analysis. Due to the polyploidy characteristic of modern sugarcane and its ancestor species *S. officinarum* and *S. spontaneum*, alleles are rich among them. Considering that most alleles are highly similar in their sequences and functions, just one representative allele was selected from each allele group. Finally, 137 non-redundant and non-allelic genes were retained from a total of 724 NLRs, which can well represent the NLR family of *S. spontaneum*.

The MEME suit for the protein sequence analysis was used to identify the conserved motifs [27]. The main parameters were set as the following: the distribution of motifs (-mod), anr; the maximum number of motifs to find (-nmotifs), 20; the minimum motif width (-minw), and maximum motif width (-maxw) of each motif, 10 and 50 residues. For the structure analysis of the 137 genes, we first extracted annotations for all genes, and then used the web tool GSDS 2.0 (http://gsds.cbi.pku.edu.cn) to visualize annotation files. The distribution of CDSs, introns, and UTRs were mapped in each gene sequence. The number of exons of each gene was counted through annotation files, too.

To study the distribution of TFBSs in each gene promoter region, the upstream sequences (1500 bp) of the 137 NLR genes were retrieved and then submitted to the web tool PlantRegMap (http://plantregmap.cbi.pku.edu.cn/binding_site_prediction.php) with a threshold p-value $\leq 1 \times 10^{-5}$. The first 12 transcription factors (TFs) with the largest number of predicted TFBSs were extracted and the TFBS distribution was plotted in the promoter region of each NLR gene using GSDS 2.0, as well. Some genes were removed due to the absent TFBSs in their promoter regions based on prediction.

2.5. Chromosomal Distribution of NLR Genes

All NLR genes (407 complete NLR genes encoding the complete NBS domain and 317 partial genes) were mapped to eight homologous groups of four members each of *S. spontaneum* based on the physical location information extract from the gff3 annotation file of *S. spontaneum* genome using the web tool MG2C v2.1 (http://mg2c.iask.in/mg2c_v2.0/). In order to clearly display the distributions of the NLR gene, only the name of NLR genes with the complete NBS domain was tagged along the edge of the chromosome. The gene numbers were tagged at the top of each group of chromosomes.

2.6. Plant Materials and Treatments

Sugarcane leaf blight, one of the most serious biotic constraints for sugarcane, is caused by the necrotrophic fungal pathogen *S. tainanensis*. The expression profiles of NLR genes in response to natural infections of *S. tainanensis* at different stages in *S. spontaneum* and sugarcane hybrids are our interest here. For *S. spontaneum*, more than 50 plants of *S. spontaneum* clone SES208 were cultivated in the field under the same environmental conditions in Fuzhou, China. Two different disease development stages were defined: For the early stage of infection, most lesions or spots were pale yellow; for the medium to late stage of infection, most lesions were reddish-brown, bright red, or brown together with part of the necrotic tissue. Four biological replicates were performed for each infection stage and for the healthy control, and thus total 12 samples were obtained. For each infected replicate, more than 100 lesions clipped from random leaves and random plants were pooled. Similarly, control samples were also randomly collected from different leaves of different plants.

The expression patterns of NLRs in response to the *S. tainanensis* infection in *Saccharum* spp. hybrids are also our focus. One susceptible clone FN12-047 derived from the cross of *S. tainanensis*-resistant variety ROC22 and -susceptible variety YT93-159, together with its resistant male parent ROC22 were selected for the present transcriptional study. Both sugarcane clones were cultivated in the field under the same environmental conditions in Fuzhou, China. After the appearance of SLB symptoms visible to the naked eye, a symptom

observation was performed daily to identify the development of disease symptoms and the infected leaves were collected for pathogen inspection. For each sugarcane clone, two plants with a similar growth vigor and disease severity in the same clone were selected for sampling, and the leaf located at the same leaf position was collected, pooled together for one replicate before RNA extraction, and finally subjected to RNA-sequencing. The healthy leaves at the same leaf position without any disease symptom were selected as the control. The infected leaves with most lesions were reddish-brown, bright red, or brown together with a little part of necrotic tissue were selected as the case. Three biological replicates were performed for each type of resistance reaction types and a total of 12 samples were obtained, six for each clone.

2.7. Total RNA Extraction and mRNA-Sequencing

A total of 12 RNA samples from sugarcane ancestor species *S. spontaneum* and 18 samples from sugarcane hybrids were isolated using the mirVana™ miRNA Isolation Kit (Ambion Life Technologies, Austin, TX, USA) as recommended by the manufacturer. RNA integrity was evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The samples with the RNA integrity number (RIN) ≥ 7 were subjected to the subsequent DNA library construction. The libraries were constructed using the TruSeq Stranded mRNA LTSample Prep Kit (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Then, these libraries were sequenced on the Illumina sequencing platform (HiSeqTM 2500) and 150 bp paired-end reads were generated.

2.8. Identification of Differentially Expressed NLR Genes

As an efficient and time-saving transcriptome reads alignment tool, Hisat2 was used as the aligner in the current study [28]. After the alignment, the SAM files are converted into BAM files, and then Featurecounts was used to quantify gene expression [29]. DESeq2 provided the most accurate differential analysis and has been confirmed by Sahraeian et al. [28]. For sugarcane-*S. tainanensis* transcriptome data, each sample has three biological repeats, and thus we chose DESeq2 as the software for the differentially expressed genes (DEGs) analysis. All differentially expressed genes need to meet the following screening criteria: p-value < 0.05, |log2 (FoldChange)| > 1.

3. Results

3.1. Identification of NLR Genes

We searched 140 complete and non-redundant NBS domain sequences from *S. spontaneum* using raw NBS.hmm downloaded from the Pfam database and built a *S. spontaneum*-specific Ss-NBS.hmm file. Using the Ss-NBS.hmm, we searched for total proteins in the *S. spontaneum* genome again and obtained 721 NLR proteins. Simultaneously, three extra NLR proteins were obtained by the Blast method. Finally, 724 non-redundant NLR proteins encoded by 724 genes were obtained, and their detailed information including protein/gene identity (id), gene name, gene type, length of coding sequence (CDS length), number of amino acids (no. of AA), molecular weight (MW), and protein isoelectric point (pI) were presented in Table S1. The aforementioned proteins were classified into two classes by the CDD tool based on the integrality of the NBS domain. Of which, one class containing 407 proteins with complete NBS was divided into four types according to the composition of main domains (Table 1), including N (120, only NBS existed), CN (66, only CC and NBS existed), NL (105, only NBS and LRR existed), and CNL (120, CC, NBS, and LRR existed). These 407 proteins, encoded by 407 NLR genes, are the focus of our analysis in the current study. In addition, 317 partial proteins without the complete NBS domain were termed as P-type.

Table 1. The classification of nucleotide-binding site-leucine-rich repeat (NLR) proteins/genes and their corresponding numbers in six Poaceae species genomes.

NLR Protein Type	*Saccharum Spontaneum*	*Oryza Sativa*	*Brachypodium Distachyon*	*Zea Mays*	*Setaria Italica*	*Sorghum Bicolor*
N	120	59	73	36	48	66
TN	0	1	1	1	1	1
CN	66	63	92	37	67	81
NL	102	51	37	85	52	61
CNL	119	83	121	63	123	126
P (Partial proteins)	317	110	96	55	140	73
Total	724	367	420	277	431	408
Coding genes of NLR protein	724	344	346	139	407	333
Coding genes in the genome	83,826	35,825	35,125	39,591	35,831	34,118
Percentage of NLR genes	0.88%	0.96%	0.99%	0.35%	1.14%	0.98%
Genome size (Mb)	3133.3	375	271.2	2135.1	405.7	708.7
Average NLR gene density (Mb $^{-1}$)	0.2	0.9	1.3	0.1	1.0	0.5

Moreover, a total of 420, 277, 367, 431, and 407 NLR proteins encoded by 246, 139, 344, 407, and 333 NLR genes were identified in *B. distachyon*, *Z. mays*, *O. sativa*, *S. italica,* and *S. bicolor*, respectively (Table 1). Among the six Poaceae species, *S. italica* has the largest NLR gene percentage (1.14%) in the genome, followed by *B. distachyon* (0.99%). Then, in terms of the average NLR gene density in the genome, the *B. distachyon* is the highest (1.3 Mb $^{-1}$), followed by *S. italica* (1.0 Mb $^{-1}$). Whether the NLR gene percentage or density, *Z. mays* is the smallest (0.35%, 0.1). Although the NLR gene percentage of *S. spontaneum* is high (0.88%), the average NLR gene density in the genome is very low (0.2 Mb $^{-1}$). Interestingly, in the six Poaceae species, there is almost no NLR protein containing TIR domain.

3.2. Diversity of Integrated Domains

As mentioned above, the presence of integrated domains can also increase the structural diversity of the NLR family. A total of 45 distinct Pfam domains were identified in the six Poaceae species involving 85 genes using the Pfam database, which were shown in Table S2. We preliminarily estimated that the NLR genes carrying integrated domains in Poaceae species account for about 3.7% according to the data of the aforementioned six species. Of the 724 NLR genes identified in *S. spontaneum*, 35 (4.8%) encoded at least one (ranked from one to eight) non-canonical NLR domain, or integrated domain, representing 28 distinct Pfam domains (Figure 1). Among these Pfams, protein kinase domain (Pkinase, 8), integrase core domain (rve, 4), RNA-dependent DNA polymerase (RVT_1, 3), and WRKY DNA-binding domain (3) occur more frequently. Interestingly, most of the genes (57.0%) carrying integrated domains are partial NLRs, although partial NLRs account for only 43.8% of all NLR genes. Thirty-six integrated domains were detected in 64 *Arabidopsis* accessions [30], and eight of them, i.e., Pkinase, WRKY, Pkinase_Try, PAH, PP2, B3, Calmodulin_bind, and AAA, were observed coexisting in the aforementioned six Poaceae species and *Arabidopsis*, which are marked with black triangles in Figure 1.

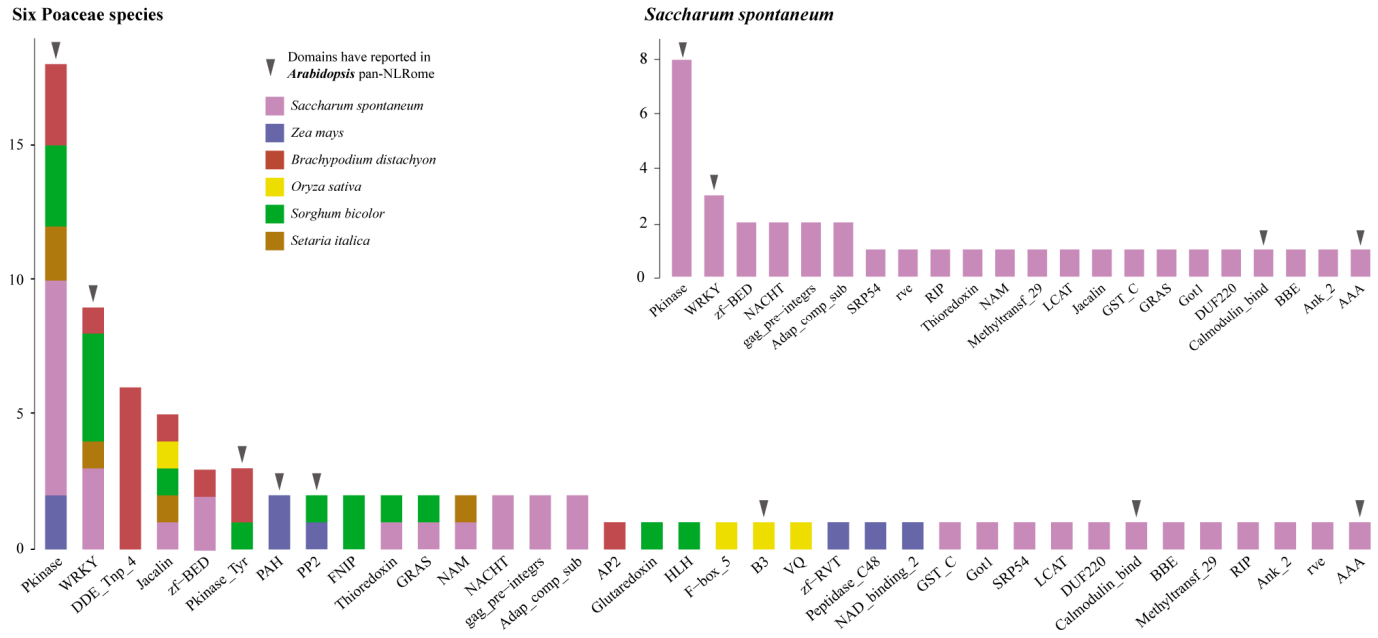

Figure 1. A total of 45 distinct Pfam domains were identified in the six Poaceae species involving 85 NLRs and 28 distinct Pfam domains were identified in *Saccharum spontaneum*.

3.3. Comparative Phylogenetic Analysis of NLR Genes among Six Poaceae Species

The phylogenetic analysis was constructed using complete NBS sequences extracted from 200 NLR proteins randomly selected from *S. spontaneum*, *S. italica*, *O. sativa*, *B. distachyon*, *S. bicolor*, and *Z. mays*. Among the 200 proteins, 50 (12 CNs + 12 NLs + 12 Ns + 14 CNLs) were selected from *S. spontaneum*, another 150 were selected on average from the five other species, each having 30 proteins (7 CNs + 7 NLs + 7 Ns + 9 CNLs, Figure 2). We labelled the proteins with an acronym for its species name and protein name given in this study. Since there are almost no TIR-containing NLRs screened from the six Poaceae species, the separation between the two typical classes cannot be confirmed in the six Poaceae species. However, interestingly, we can clearly see that proteins containing the CC domain separated from those without the CC domain. Based on this phenomenon, we can confidently divide 200 protein sequences into four clades, CN-I and CN-II (containing CC domains) clades, N-I and N-II (without CC domains) clades. Clade N-I was comprised of 10 proteins, of which eight belong to the NL-type. Clade N-II contains 79 proteins, of which 65 (82.3%) were truncated to the CC domain. However, we found that 14 CC-containing proteins in N-II were mainly derived from *O. sativa* (7, 50.0%). Of the 15 proteins in clade CN-I, 13 belong to the CNL protein. Clade CN-II contained the largest number (96) of proteins in four clades and 83.3% (80) of the proteins contained CC domains. However, we found that 16 CC-containing proteins in N-II were mainly derived from *S. italica* (7, 43.8%).

3.4. Gene Structure and Motif Composition of NL and CNL Proteins

For different types of NLR genes in Poaceae species, the genes with CC domains and those without CC domains have undergone an obvious evolutionary segregation. As a member of Panicoideae, *S. spontaneum* also shows this characteristic (Figure 2). We selected two main types CNL and NL (full-length genes) for the detailed sequence analysis, including a complete phylogenetic analysis, motif prediction, and gene structure analysis (Figure 3). In the two types, CNL and NL, for highly similar or identical protein sequences caused by alleles, only one protein sequence was retained. Finally, 137 representative NL and CNL protein sequences were retained for the subsequent analysis. From the phylogenetic tree constructed by the NBS domain of all CNL and NL genes from *S. spontaneum*, the two types of genes did have a distinct segregation in the evolution (Figure 3A). Five clades,

I, II, III, IV, and V were divided. Clearly, in clade II, 39 genes were included, 38 of them were NL-type genes. However, in clade III, all genes (16) belong to the CNL-type, and in clade IV, 22 of the 24 genes belong to the CNL-type. However, in clades I and V, the above separation characteristics are not obvious.

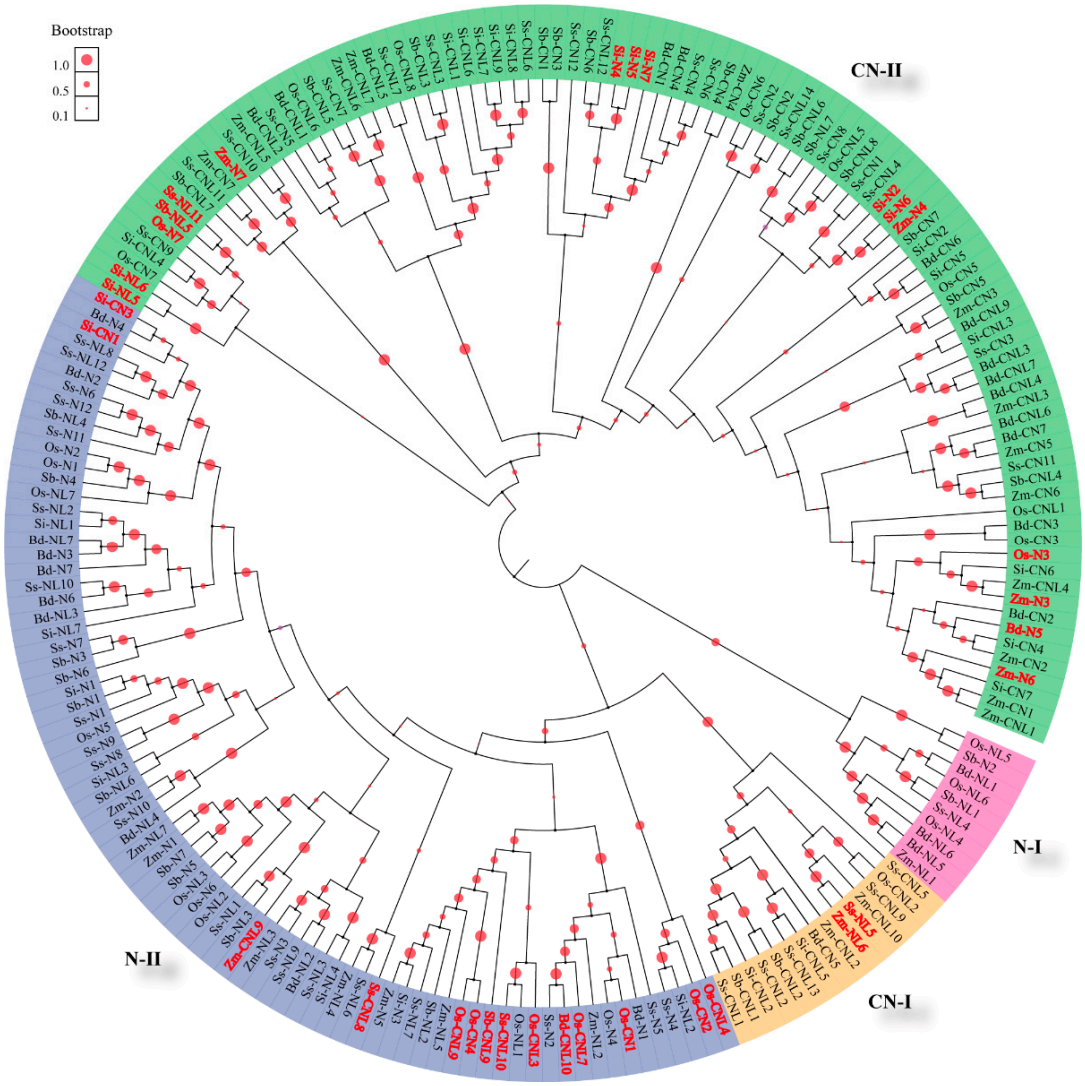

Figure 2. Comparative phylogenetic analysis of NLR genes among six Poaceae plants. A total of 200 NLR proteins randomly selected from *Saccharum spontaneum*, *Setaria italica*, *Oryza sativa*, *Brachypodium distachyon*, *Sorghum bicolor,* and *Zea mays* were constructed into a phylogenetic analysis. For clarity, each protein is encoded as follows: acronym for species name + protein name (protein type N, NL, CN, and CNL + number). Fifty NLR proteins belong to *S. spontaneum* (12 CNs + 12 NLs + 12 Ns + 14 CNLs), and 150 NLR proteins (7 CNs + 7 NL7s + 7 Ns + 9 CNLs) on average were selected from each of the five other species. Bootstrap values (set at 1000 replications) are indicated in every branch with a red circle size.

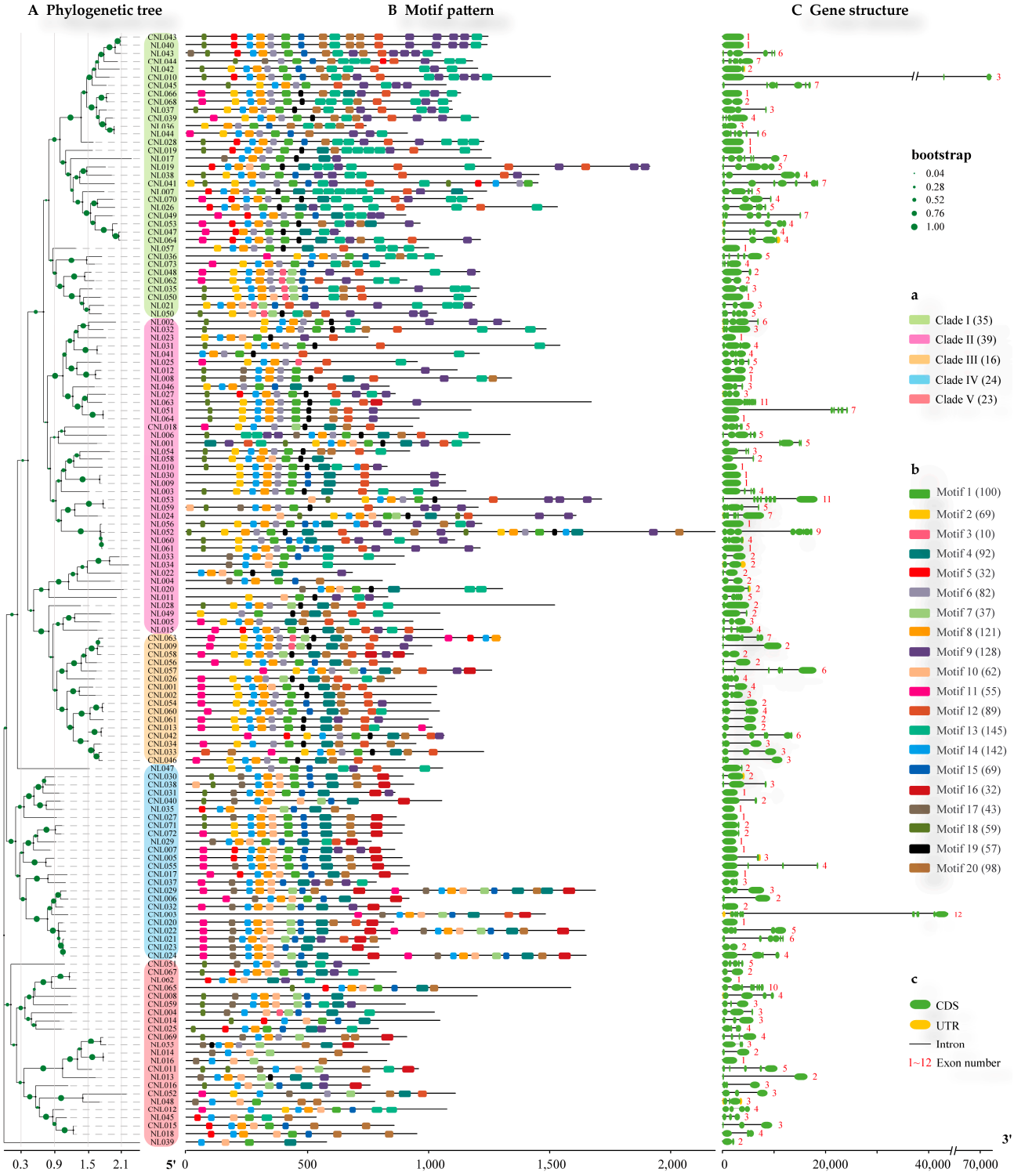

Figure 3. Phylogenetic relationships, gene structure, and architecture of conserved protein motifs in NLR genes from *Saccharum spontaneum*. (**A**) The phylogenetic tree was constructed based on the NBS sequences of 137 CNL and NL-type proteins from *S. spontaneum* using MEGA X. Five clades of I, II, III, IV, and V are marked in different colors. (**B**) The motif composition of *S. spontaneum* NLR proteins. The motifs 1-20 are marked with different colors. (**C**) Structure of *S. spontaneum* NLR genes. Yellow boxes represent untranslated 5′- and 3′-UTRs; green boxes represent CDSs; black lines represent introns. The length of proteins and genes can be measured using the scale at the bottom.

In order to explore the gene characteristics within each clade, we used the MEME to search the conserved motifs. The motif prediction results were arranged behind the

phylogenetic tree (Figure 3B). A total of 20 distinct conserved motifs were found with the E-value < 6.8×10^{-313}. A detailed amino acid composition of each motif is provided in Figure S1. Among the 20 motifs, motif 13 has the largest number. However, most of motif 13 is distributed in clade I (93, 64.1%), followed by clade II (31, 21.4%). Moreover, some proteins have several of the motif 13, for instance, CNL44, CNL68, CNL28, and CNL19 have five and NL07 have six. Motifs 9 and 13 show similar distribution characteristics, while the number of motif 9 (128) is less than that of motif 13. Motif 16 (32) mainly appears in clade IV (24, 75.0%) and motif 1 rarely appears in clade V. In addition, clades III and IV are mainly composed of CNL-type proteins, of which most have one to two of motif 11. Similarly, for clade II, which consists mainly of NL-type proteins, most of the sequences contain motif 18. The distribution of other motifs in five clades is relatively uniform.

The structures (location of CDSs and UTRs) of the gene sequences encoding 137 NLR proteins were shown in Figure 3C. The exon number of each gene was calculated, and then the number was tagged after the gene sequence. The average number of exons contained in these gene sequences was 3.4, while the number of exons in different genes varies greatly. Twenty-three genes contain only one exon, but three genes have more than 10 exons. For instance, CNL65 has only one exon, while NL53 and NL63 have 11 exons, and CNL03 has 12 exons.

3.5. Chromosomal Distribution of NLR Genes

The 407 NLR genes encoding the complete NBS domain and 317 partial genes were mapped to eight homologous groups of four members each (Figure 4). The number of NLR genes in each homologous group showed obvious differences. Respectively, 22, 84, 21, 27, 95, 70, 61, and 28 genes mapped to Chr1–Chr8. Chr5 carried the most NLR genes (95), followed by Chr2 (84), Chr6 (70), and Chr7 (61), which together account for 76.0% of the total. The remaining 24.0% were distributed in Chr1 (22), Chr3 (21), Chr4 (27), and Chr8 (28), and the gene number of each group was less than 30. Whilst, the distribution characteristic of the NLR gene on each chromosome was shown to be higher in gene density at both ends of the chromosome. Moreover, in Chr4 and Chr8, NLR genes were concentrated mainly on one chromosome arm. Interestingly, many NLR genes have been clustered on chromosomes. These gene clusters contain two to seven NLR genes, and most of the clusters were formed by the same type of genes.

3.6. Transcription Factor Binding Sites in NLR Promoters

To further study the potential regulatory mechanisms of NLR genes, 3747 TFBSs involving 34 TFs were predicted in the promoter regions (−1500 bp) of all the 73 CNL and 67 NL genes. According to the number of binding sites, 12 TFs (ERF, MIKC-MADS, C2H2, Dof, BBR-BPC, LBD, MYB, TALE, GRAS, G2-like, WRKY, and HD-ZIP) with the largest number of binding sites were selected, and a total of 3109 TFBSs were tagged in the promoter regions (Figure S2), while promoter regions of CNL09, CNL53, CNL65, CNL66, and CNL67 were absent. Of these 12 TFs, ERF accounts for nearly half (48.5%) of the binding sites. Some NLR genes contained abundant TFBSs at its upstream regions, such as CNL14, CNL19, and CNL38, etc., but some contain only one TFBS, such as NL21 and NL40, etc. Although ERF has the most binding sites, these binding sites do not exist in most of the NLR genes, but only in the promoter regions of 72 genes (Figure S2). However, for MIKC_MADS, and C2H2, although the number of binding sites in both was less than that of ERF, they existed more in NLR promoter regions (79 and 74, respectively) than in ERFs.

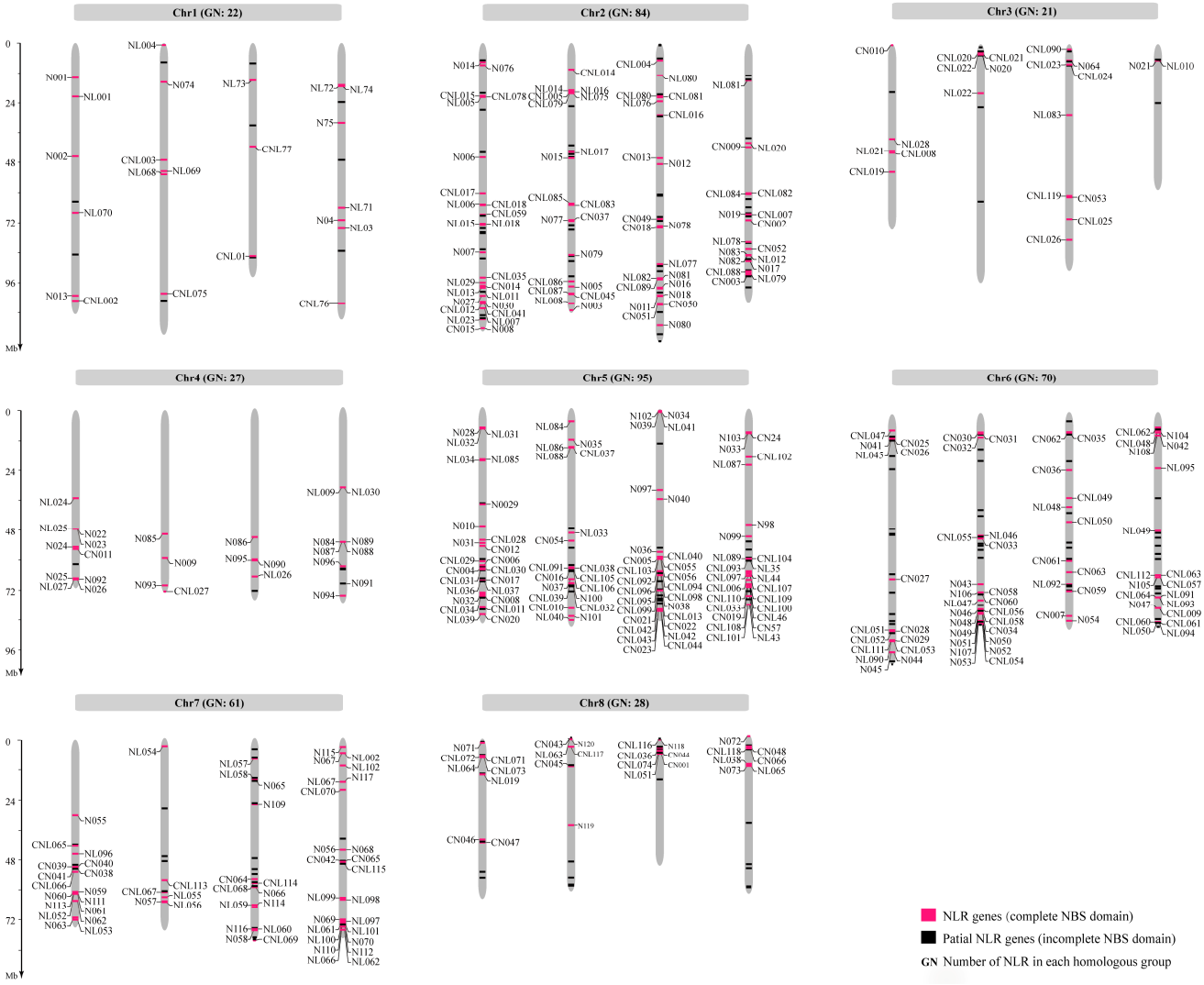

Figure 4. Physical distribution of NLR genes on *Saccharum spontaneum* chromosomes. All NLR genes (407 NLR genes encoding the complete NBS domain, and 317 for partial genes) were mapped to eight homologous groups (Chr1–Chr8) of four members (A–D) each (32 chromosomes). Red lines on the chromosomes represent NLR genes encoding the NBS domains and black lines represent partial NLR genes encoding the incomplete NBS domains.

3.7. The NLR Expression Responses to SLB in One S. spontaneum Clone and Two Sugarcane Hybrids

Three sets of transcriptome data were applied to study the NLR expression dynamics of sugarcane in response to SLB (Figure 5). These data were obtained from one *S. spontaneum* clone named SES208 and two sugarcane hybrids named ROC22 and FN12-047. ROC22 is resistant to SLB, and FN12-047 and SES208 are susceptible to SLB. A total of 20, 19, and 45 differentially expressed NLR genes (DENLRs) were detected in SES208, FN12-047, and ROC22, respectively. In each sugarcane clone, the expression of all five types of NLRs was activated or inhibited, including partial genes (P-type) which were also actively involved in the responses to SLB. In some types of NLRs, the numbers of differentially expressed partial genes were more than half, such as the N-type NLRs in SES208 and FN12-047. In either clone, the number of NLRs responding to SLB in the medium to late disease stage was significantly higher than that in the early disease stage. Continuous DENLRs (marked with an orange triangle in Figure 5) both in the early and medium to late disease stages were infrequent for just 5, 5, and 8 in SES208, FN12-047, and ROC22. Interestingly, the number of DENLRs detected in ROC22 was more than twice that in the susceptible SES208

and FN12-047. However, there were more downregulated (fold change < 0.50) NLRs in ROC22, but a few downregulated NLRs in FN12-047 and SES208. Especially in the early stage in ROC22, nine NLRs (69.2%) were downregulated.

Figure 5. The NLR (differential) expression profiles responses to sugarcane leaf blight (SLB) in one *Saccharum spontaneum* clone SES208 and two sugarcane hybrids of ROC22 and FN12-047. ROC22 was resistant to SLB, and FN12-047 and SES208 were susceptible to SLB. Respectively, 20, 19, and 45 differentially expressed NLR genes (DENLRs) were detected in SES208, FN12-047 and ROC22. Continuous DENLRs both in the early and medium to late disease stages were marked with an orange triangle.

3.8. A Few DENLRs Coexist in Three Sugarcane Clones

The transcriptome data of SES208 and the two hybrids were not from the same project, however, the experimental materials were planted in the same geographical environment, and the disease classification of experimental materials followed the same standard. In addition to comparing the quantitative differences of NLR responses of the three sugarcane clones after infection of *S. tainanensis*, we are more concerned about which NLRs were active in the disease stages of each clone. Based on the Venn diagram (Figure 6A) of the DENLR sets derived from three clones, no DENLR was shared by them. ROC22 shared eight DENLRs (DENLR set f, Figure 6B) with FN12-047 and shared two DENLRs (DENLR set d) with SES208. Similarly, FN12-047 shared more DENLRs with ROC22 than with SES208 (2, DENLR set e). Each clone has its own unique DENLR set, of which ROC22 has the most (DENLR set b). Moreover, the most downregulated NLRs (22, 91.0%) were

included in the unique DENLR set of ROC22. Obviously, the number of DENLRs shared between hybrids was significantly higher than that of each hybrid with SES208.

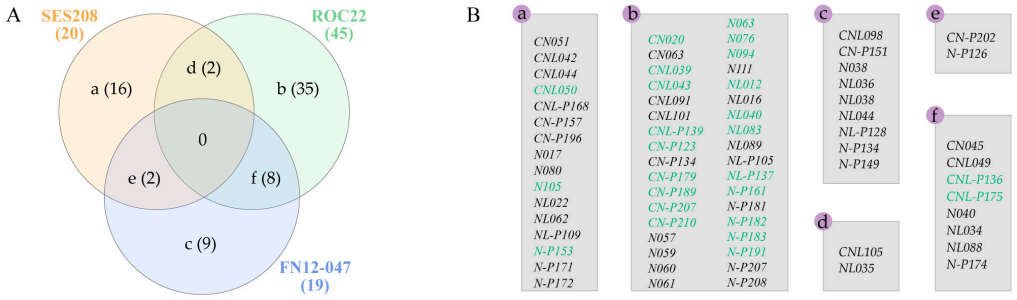

Figure 6. Comparisons of expression of NLR in three sugarcane accessions in response to sugarcane leaf blight (SLB). (**A**) Venn diagram of the differentially expressed NLR gene (DENLR) sets derived from three clones of SES208, ROC22, and FN12-047. (**B**) Specific members of each DENLR set. The genes labeled green were downregulated DENLRs.

4. Discussion

4.1. Mining NLR Family Genes Based on Genome-wide is More Representative

Although there are many transcriptome data available for sugarcane derived from the studies on different breeding traits, such as biotic resistance and abiotic tolerance, NLR genes mining based on transcriptome data are limited. Due to the fact that the high expression of the NLR gene is harmful to plants themselves, the expression of the NLR gene is low, even not expressed under unnecessary circumstances [4]. Therefore, mining NLR family genes based on genome-wide will be more important due to the fact that more complete members were collected. Fully searching for its family members is the first step in the genome-wide gene family analysis. In the NLR gene family, NBS being the most conservative domain was used to build the HMM model for searching NLR genes, in the current study. However, in order to fish out all members of the family as far as possible, we also used Blast to search for sequences containing incomplete NBS domains. The combination of Hmmer and Blast was crucial to find the maximum number of NLR genes.

4.2. The Number and Type of NLRs in Six Poaceae Species

Through this combination of Hmmer and Blast, we searched for genome-wide NLR genes in six Poaceae species *S. spontaneum*, *S. italica*, *O. sativa*, *B. distachyon*, *S. bicolor,* and *Z. mays*. Among the above six species, the percentage of NLR genes from *Z. mays* is lowest (0.35), while the percentage of NLR genes is roughly similar (0.88–1.14%) for the other five. Commonly, the number of NLR genes in plants is consistent with the total number of genes in the genome, but with exceptions. For example, although *Carica papaya*, *Citrullus lanatus,* and *Cucumis sativus* have 27,769, 23,440, and 21,503 genes, respectively, just 46 (0.17%), 42 (0.18%), and 65 (0.30%) NLR genes were discovered in *C. papaya*, *C. lanatus,* and *C. sativus* genomes, respectively [14]. Moreover, *Z. mays* (139, 0.35%) in the current study should be one of the exceptions. These exceptions indicate the existence of species-specific mechanisms of the NLR gene during expansion and evolution. The pathogen *avr* genes are evolving at a very fast rate and encode polymorphic protein effectors [31,32]. Being the receptors of the effector *avr* genes, it is not surprising that there are a large number of NLR members which are present in high diversity in plants, and thus it is the inevitable choice for plants to adapt to multiple adversities caused by pathogens. Indeed, in rice cultivar Tetep, although few cloned NLRs showed resistance to > 6 blast pathogen strains, multiple NLRs are noticeably essential for Tetep's broad-spectrum resistance to blast [33].

In general, CNLs can be found both in monocot and dicot genomes, but TNLs are almost missing in monocots and widely distributed in dicots [34]. With the exception of one TNL present in the *S. spontaneum* genome, they are absent in the aforementioned NLRs we identified, which confirms that this conclusion is due to the fact that the above six

plant species belong to monocot. Although TNL genes exist widely in dicots, there are some exceptions, for instance, both the core eudicot *Beta vulgaris* [35] and Ranunculacea *Aquilegia coerulea* [36], are absent in the TNL gene. The two types of RNL and RN have not been found in Poaceae plants at present, which has been confirmed by this study and the previous study [14]. However, we can find many RNL and RN genes in Brassicaceae and Fabaceae [14]. Due to the missing TNL and RNL genes in the *S. spontaneum* genome, it is impossible to analyze whether the NLR genes can be clearly divided into two or three classes in the evolution according to the RPW8, TIR, and CC domain. Interestingly, in the phylogenetic analysis, the type CNL with the CC domain and the type NL without the CC domain have a distinct trend towards separation (Figure 3). Likewise, this trend also exists in the five other Poaceae species in the current study (Figure 2).

4.3. Diversity of Integrated Domains in S. spontaneum

Some integrated domains carried by NLRs are thought to be required for effector recognition, which has been confirmed by several studies [20,37,38]. For instance, in *Arabidopsis*, a receptor complex consisting of RRS1-R and PRS4 proteins can recognize two bacterial effectors, AvrRps4 or PopP2, and then triggers a defense. However, in this complex, the target of the pathogen effectors directly function to is the WRKY domain contained in the RRS1-R NB-LRR protein [37]. Li et al. [20] found that the Sw-5b NLR immune receptor recognizes the NSm effector not only using a single domain LRR, but also depending on the N-terminal Solanaceae domain. Although 28 integrated domains in *S. spontaneum* and total 45 integrated domains in six Poaceae species are detected, only a few (8) integrated domains coexist in *Arabidopsis* (Figure 1). Even among the six Poaceae species, the shared integrated domains are also scanty. To some extent, this characteristic reflects the high integrated domain diversity among species, especially the distant species. It is noteworthy that both kinase and WRKY domains occur frequently in *S. spontaneum* and *Arabidopsis*.

4.4. Characteristics of Motifs Derived from NBS and LRR Domains

Most known NLR genes are those encoding proteins with a nucleotide-binding site and a C-terminal leucine-rich repeat domain [34,39]. Therefore, this study mainly focused on NL- and CNL-type genes, and the 137 representative NL and CNL proteins were retained for further analysis. In the motif prediction, we find that they exist in different motifs, such as motifs 1, 9, 11, 13, 16, and 18. Different clades are present in the above motifs, which may be highly correlated with gene evolution and these specific motifs may contribute to the functional divergence of NLR genes.

As the most conservative domain of NLR protein, the NBS domain contains four main conserved motifs [12,40]. The first is the P-loop, which binds to ATP and may have a crucial effect on the activity of the R protein [41]. The three other motifs were named Kinase 2, Kinase 3, and GLPL and their structural characteristics have been intensively studied and described in the report of Meyers et al. [12]. P-loop sequences were detected in motifs 2, 5, and 17 (Figure 7A–C) in the current study. According to the conservativeness of amino acid residues at each locus in three P-loop sequences, we can infer that the conserved sequence of P-loop in sugarcane is G[M/L]GG[V/L/I/M]GK[T/S]TL[A/V]. In the phylogenetic analysis (Figure 3), motif 17 mainly appeared in clades IV and V. Motif 14 containing kinase 2 (Figure 7D) and motif 8 containing kinase 3 (Figure 7E) are present in almost all NLR proteins. Based on the predicted results of motifs 14 and 8, the conserved sequences of kinase 2 and kinase 3 are [L/I][L/I]VLDDVW and S[R/K][I/V/F][I/L/V][I/V/L]T[T/S]R. In addition, motif 6 containing GLPL (Figure 7F) is absent in clade IV consisting mainly of CNL-type genes, and the conserved sequence of GLPL is G[L/V]PLA[I/A/V]K. We noted that the conserved sequences of P-loop (G[M/L/F]GG[L/M/I]GK), kinase 2 (L[L/I]VLD[D/N]V[W/D], and GLPL (G[L/I]PLA[L/I][K/E]) of the NLR genes in the common bean (*Phaseolus vulgaris* L.) have also been reported, while no kinase 3 was predicted [42]. Clearly, there are no significant differences in the sequences of P-loop, kinase 2, and GLPL between the two

species *P. vulgaris* and *S. spontaneum* in different families, indicating that they are highly conservative. In addition, Zhang et al. [14] suggested that the most frequent P-loop core was GFGGLGK or GMGGLGK in monocots, which is consistent with our observation, and the second amino acid Phe of P-loop core can be found in motif 5 (Figure 7B).

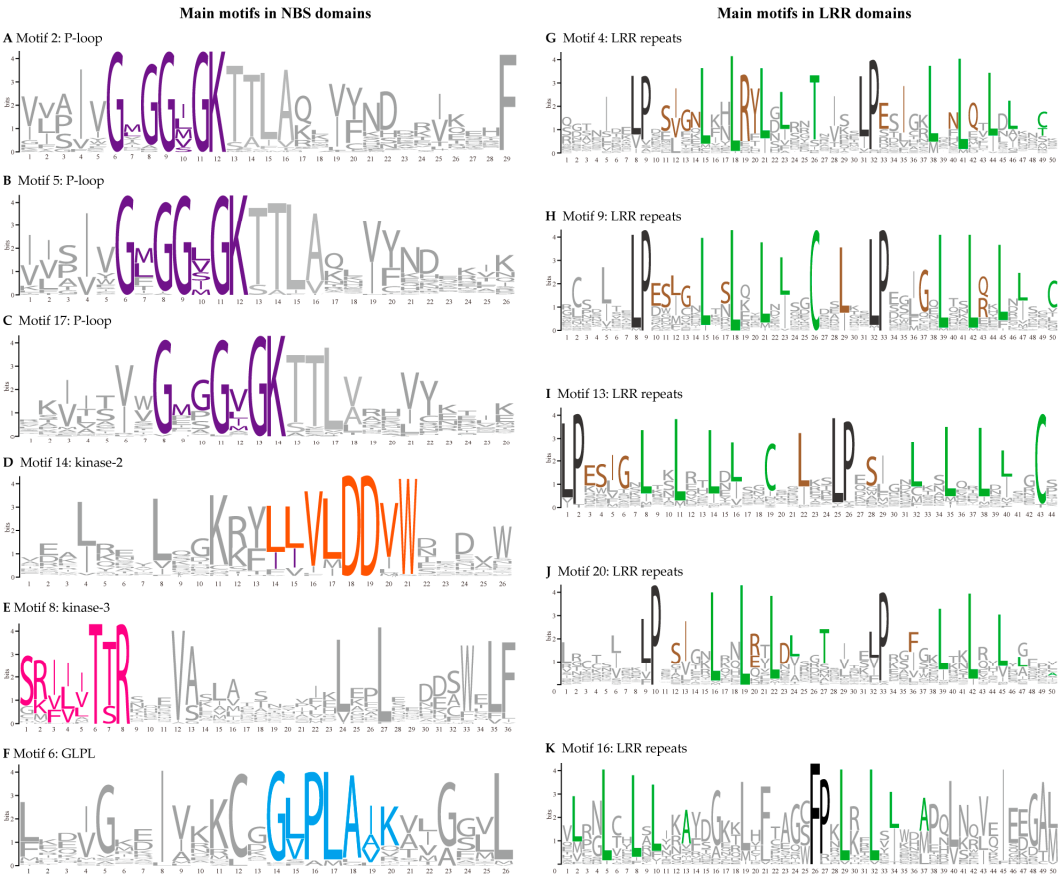

Figure 7. The amino acid sequences of motifs in NLR proteins. (**A**–**C**) Represent the predicted motif sequences of three P-loops. (**D**–**F**) Represent the predicted motif sequences of kinase-2, kinase-3, and GLPL, respectively. (**G**–**K**) Represent 10 LRR repeats with various degrees of differences.

The LRR domain is a common amino-structure in a wide range of species, from viruses to eukaryotes, whose functions were mainly involved in ligand binding and protein-protein interactions. LRR repeats in the LRR domains vary among different species, and the largest, minimal, and the mean of repeats are 8, 25, and 14, respectively, in *Arabidopsis* [43]. However, the largest repeats of LRR may be considerably higher in other species, for instance, in the Dm3 protein, CNL resistance gene candidate 2 (RGC2) proteins of lettuce, as many as 47 LRRs appeared [44]. In our study, 20 motifs with the E-value $< 6.8 \times 10^{-313}$ were predicted and analyzed. It is inevitable that some motifs containing LRR are not identified. These motifs may appear in fewer genes. Therefore, we can only cautiously infer that the numbers of LRR repeats in NLR genes of *S. spontaneum* are less than 20, such as CNL10. In addition, the minimal LRR repeats may be close to two, such as NL39. That is to say, the range of change in the LRR repeats number of different NLR genes in the *S. spontaneum* genome is considerable. Although LRR repeats exist in five motifs with similar structures (Figure 7G–K), there are also obvious differences among them, and the distribution of each clade has certain characteristics. For instance, among the five motifs, motif 13 has the largest number (145) mainly distributed in clade I (93). In motif 16, we did not find the double amino acid residue combination (Leu and Pro) at upstream of LRR, which is an obvious difference between LRRs in motif 16 and the four other motifs with LRR repeats (Figure 7G–K), and motif 16 is mainly distributed in clade IV. In addition, we

found that Thr appears most frequently in the position of the 26th residue of motif 4, while the corresponding position of motif 9 (containing the 26 residues) is Cys.

What is the function of quantity variance and sequence diversity of LRR repeats among NLR genes for plants? In *Arabidopsis*, the NLR recognition of Peronospora parasitica1 (RPP1) can interact with the *Arabidopsis thaliana* recognized1 (ATR1), an effector from the oomycete pathogen *Hyaloperonospora Arabidopsidis* [45]. Moreover, different *RPP1* alleles can recognize the products of different *ATR1* alleles due to the diversity of the LRR domain, thus triggering plant immunity [46]. For *S. spontaneum*, the abundant structural specificities and quantitative variations in the LRR domain among different NLR genes may improve its ability to recognize potential pathogens. Meanwhile, the specificity of the LRR structure is also evident in the same NLR gene.

4.5. Several Transcription Factors May Regulate the Expression of the NLR Gene

To date, no report on the prediction of TFBSs in the upstream promoter region of NLR genes was found. In this study, a total of 3,747 TFBSs of several TFs were predicted. Especially for TFs ERF, MIKC_MADS, C2H2, and Dof, the number of predicted TFBSs is considerable, and their binding sites are present in nearly half of the promoters of the 137 non-allelic full-length NLR genes. It is possible that these TFs are the main regulators of the NLR gene expression in sugarcane, and some NLR genes are regulated by a variety of TFs. Although the predicted TFBS number of ERF was largest in the current study, there is no report on the direct regulation of the NLR gene by ERF until now. Despite this, we still note that Pre et al. [47] illustrated the JA- and ethylene-responsive expression of some resistance-related genes, including *PLANT DEFENSIN1.2*, depending on ORA59, a APETALA2/Ethylene response factor (AP2/ERF) domain TF from *Arabidopsis*. For MIKC_MADS and C2H2, although the number of binding sites in both are less than that of ERF, they exist more in NLR promoter regions (79 and 74, respectively) than ERF (72). MIKC_MADS are exclusive to plants, which play an important role in growth and development regulation and signal transduction [48]. Studies on C2H2 responses to abiotic stress have been reported in several species [49,50]. However, there are a few reports on MIKC_MADS and C2H2 responding to biotic stress, let alone their direct involvement in regulating the expression of the NLR gene. Kim et al. [51] illustrated that when threatened by bacterial *Xanthomonas campestris* and fungal *Colletotrichum coccodes*, the expression of *CAZFP1*, a pepper gene encoding a C2H2-zinc finger protein, was significantly induced at the transcriptional level. Similarly, *StZFP1*, a C2H2-type zinc finger protein gene in potato, was also activated when facing the infection of *Phytophthora infestans* [52]. However, whether *CAZFP1* or *StZFP1*, their specific roles are unknown. This study suggests that ERF, MIKC_MADS, C2H2, and Dof are major TFs regulating the NLR expression, which need to be further studied in the future.

4.6. NLR Expressions May Determine the Resistance Levels to SLB in Sugarcane Clones

The NLR expression patterns of one resistant sugarcane clone ROC22 and two susceptible clones FN12-047 and SES208 responding to SLB were studied. The number of DENLRs in the resistant clone was significantly higher than that in two susceptible clones. Moreover, the number of DENLRs that were shared between any two clones were small. This may be due to the difference in the genetic background. In particular, the number of shared DENLRs between the hybrids and *S. spontaneum* were less than that between two hybrids, which indicated that the genetic background difference between the sugarcane hybrids and *S. spontaneum* were obvious. However, we got a more interesting finding. In susceptible clones of FN14-047 and SES208, a few downregulated NLRs were observed. However, this phenomenon was just the opposite in ROC22 and more than 90% of the downregulated NLRs only existed in ROC22. Why are there many downregulated NLRs in the resistant clone ROC22 at the early or medium to late disease stage?

S. tainanensis belongs to the plant-pathogenic species of *Dothideomycetes*, a necrotrophic fungal pathogen [53,54]. Necrotrophic pathogens can interact with their hosts in an inverse

gene-for-gene manner, in which necrotrophic effectors (NEs) are recognized by specific dominant genes in the host leading to host-mediated programmed cell death in the infection site allowing the pathogen to cause the disease [55]. The specific dominant genes are mainly considered as NLRs [55,56]. In the resistant ROC22, many NLR expressions were inhibited, which may make ROC22 unable to interact with *S. tainanensis* NEs and fail to initiate downstream defense responses, such as programmed cell death. In this way, *S. tainanensis* could not hijack host NLRs to achieve further infection, and ROC22 also obtained resistance. Therefore, only these NLRs downregulated in ROC22 need to be further studied in relation to the sugarcane resistance to SLB.

5. Conclusions

In sum, a total of 724 non-redundant NLRs encoded by 724 genes were identified in *S. spontaneum*. At least 35 (4.8%) NLRs contained at least one integrated domain. Interestingly, we clearly observed that the NLRs containing CC domains separated from those without CC domains. This characteristic was also present in five Poaceae species *B. distachyon*, *Z. mays*, *O. sativa*, *S. italica,* and *S. bicolor*. According to the chromosomal distribution of NLRs, Chr5 carried the most (95), followed by Chr2 (84), Chr6 (70), and Chr7 (61), which together account for 76.0% of the total. The average number of exons contained in 137 representative NLRs was 3.4, while the number of exons in different genes varied greatly (1–12). In the motif analysis, we found that different motifs were distributed regularly in different clades. We discussed in detail the distribution characteristics of some important motifs, especially motifs 2, 5, 17, 14, 8, and 6 in the NBS domain and motifs 4, 9, 13, 20, and 16 in the LRR domain. In the promoter regions (−1,500 bp) of 137 representative NLR genes, 3,747 TFBSs involving 34 TFs were predicted. For ERF, MIKC_MADS, C2H2, and Dof, the number of predicted TFBSs is considerable. In addition, based on three sets of transcriptome data from two sugarcane hybrids and one *S. spontaneum* clone infected by the necrotrophic fungal pathogen *S. tainanensis* causing sugarcane leaf blight (SLB), the expression dynamics of NLRs responding to the infection in different sugarcane clones were compared. The difference of genetic background led to the significant difference of the NLR response to SLB in different sugarcane clones, and we got an inference of the potential mechanism of SLB resistance. These results provided a basic reference and new insights to further study and utilize the NLRs and may aid in sugarcane disease-resistant breeding.

Supplementary Materials: Supplementary materials can be found at https://www.mdpi.com/2073-4395/11/1/153/s1. Figure S1: Twenty distinct conserved motifs of 137 representative non-allelic NLR proteins were found with the E-value < 6.8×10^{-313}. Among the 20 motifs, motif 13 has the largest number. However, most of motif 13 is distributed in clade I (93, 64.1%), followed by clade II (31, 21.4%). Moreover, some proteins have several of motif 13, for instance, CNL44, CNL68, CNL28, and CNL19 have five and NL07 have six. Motifs 9 and 13 show similar distribution characteristics, while the number of motif 9 (128) is less than that of motif 13. Motif 16 (32) mainly appears in clade IV (24, 75.0%) and motif 1 rarely appears in clade V. In addition, clades III and IV are mainly composed of CNL-type proteins, of which most have one to two of motif 11. Similarly, for clade II, which consists mainly of NL-type proteins, most of the sequences contain motif 18. The distribution of other motifs in five clades is relatively uniform; Figure S2: Predicted transcription factor binding site (TFBSs) profile of CNL and NL-type genes in promoters. Promoter sequences (–1,500 bp) of all the 73 CNL and 64 NL genes are analyzed by PlantCARE. The top 12 predicted TFBSs were labeled in promoter regions with different color boxes (promoter regions of CNL9, CNL53, CNL65, CNL66, and CNL67 were absent). The upstream length to the translation initiation sites can be inferred according to the scale at the bottom; Table S1: Detailed information on 724 NLRs including protein/gene identity (id) in genome, gene name, gene type, CDS length, number of amino acids, molecular weight (MW), and isoelectric point (pI) of protein; Table S2: A total of 45 distinct Pfam domains were identified in the six Poaceae species involving 85 NLR genes.

Author Contributions: Conceptualization, Z.W., H.R., F.X., and L.X.; methodology, Z.W.; software, Z.W.; validation, H.R., F.X., and G.L.; writing—original draft preparation, Z.W.; writing—review and

editing, L.X. and Y.Q.; supervision, Y.Q.; project administration, L.X.; funding acquisition, L.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 31571732, the earmarked fund for the Modern Agriculture Technology of China (CARS-17) and the Science and Technology Innovation Project of FAFU (KFA17513A).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

NLRs	Nucleotide binding site leucine-rich repeats
CC	Coiled-coil
TFBSs	Transcription factor binding sites
SLB	Sugarcane leaf blight
PRRs	Pattern recognition receptors
PAMPs	Pathogen associated molecular patterns
NBS	Nucleotide-binding site
LRR	Leucine-rich repeat
TIR	Toll/interleukin-1 receptor
RPW8	Powdery mildew8
TNLs	TIR-NLRs
CNLs	Coiled-coil-NLRs
RNLs	Powdery mildew8-NLRs
HMM	Hidden markov model
TFs	Transcription factors
DENLRs	Differentially expressed NLR genes

References

1. Boller, T.; Felix, G. A renaissance of elicitors: Perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* **2009**, *60*, 379–406. [CrossRef] [PubMed]
2. Rafiqi, M.; Ellis, J.G.; Ludowici, V.A.; Hardham, A.R.; Dodds, P.N. Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Curr. Opin. Plant Biol.* **2012**, *15*, 477–482. [CrossRef] [PubMed]
3. Mace, E.; Tai, S.; Innes, D.; Godwin, I.; Hu, W.; Campbell, B.; Gilding, E.; Cruickshank, A.; Prentis, P.; Wang, J.; et al. The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC Plant Biol.* **2014**, *14*, 253. [CrossRef] [PubMed]
4. Lai, Y.; Eulgem, T. Transcript-level expression control of plant NLR genes. *Mol. Plant Pathol.* **2018**, *19*, 1267–1281. [CrossRef] [PubMed]
5. Coll, N.S.; Vercammen, D.; Smidler, A.; Clover, C.; Van Breusegem, F.; Dangl, J.L.; Epple, P. *Arabidopsis* type I metacaspases control cell death. *Science* **2010**, *330*, 1393–1397. [CrossRef] [PubMed]
6. Jones, J.D.; Dangl, J.L. The plant immune system. *Nature* **2006**, *444*, 323–329. [CrossRef]
7. Zipfel, C. Plant pattern-recognition receptors. *Trends Immunol.* **2014**, *35*, 345–351.
8. Van Ooijen, G.; Mayr, G.; Albrecht, M.; Cornelissen, B.J.C.; Takken, F.L.W. Transcomplementation, but not physical association of the CC-NB-ARC and LRR domains of tomato R protein Mi-1.2 is altered by mutations in the ARC2 subdomain. *Mol. Plant* **2008**, *1*, 401–410. [CrossRef]
9. Van der Biezen, E.A.; Jones, J.D. The NB-ARC domain: A novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.* **1998**, *8*, R226–R227. [CrossRef]
10. Shao, Z.Q.; Xue, J.Y.; Wang, Q.; Wang, B.; Chen, J.Q. Revisiting the origin of plant NBS-LRR genes. *Trends Plant Sci.* **2019**, *24*, 9–12.
11. Li, X.; Kapos, P.; Zhang, Y. NLRs in plants. *Curr. Opin. Immunol.* **2015**, *32*, 114–121. [CrossRef] [PubMed]
12. Meyers, B.C.; Kozik, A.; Griego, A.; Kuang, H.; Michelmore, R.W. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **2003**, *15*, 809–834. [CrossRef] [PubMed]
13. Bai, J.; Pennill, L.A.; Ning, J.; Lee, S.W.; Ramalingam, J.; Webb, C.A.; Zhao, B.; Sun, Q.; Nelson, J.C.; Leach, J.E.; et al. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* **2002**, *12*, 1871–1884. [CrossRef] [PubMed]

14. Zhang, Y.; Xia, R.; Kuang, H.; Meyers, B.C. The diversification of plant NBS-LRR defense genes directs the evolution of microRNAs that target them. *Mol. Biol. Evol.* **2016**, *33*, 2692–2705. [CrossRef] [PubMed]
15. Zhang, J.; Zhang, X.; Tang, H.; Zhang, Q.; Hua, X.; Ma, X.; Zhu, F.; Jones, T.; Zhu, X.; Bowers, J.; et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **2018**, *50*, 1565–1573. [CrossRef]
16. D'Hont, A.; Grivet, L.; Feldmann, P.; Rao, S.; Berding, N.; Glaszmann, J.C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol. Gen. Genet.* **1996**, *250*, 405–413.
17. Wang, L.P.; Jackson, P.A.; Lu, X.; Fan, Y.H.; Foreman, J.W.; Chen, X.K.; Deng, H.H.; Fu, C.; Ma, L.; Aitken, K.S. Evaluation of sugarcane × progeny for biomass composition and yield components. *Crop. Sci.* **2008**, *48*, 951–961. [CrossRef]
18. Kourelis, J.; van der Hoorn, R.A.L. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell* **2018**, *30*, 285–299. [CrossRef]
19. Dracatos, P.M.; Bartoš, J.; Elmansour, H.; Singh, D.; Karafiatova, M.; Zhang, P.; Steuernagel, B.; Svačina, R.; Cobbin, J.C.; Clark, B.; et al. The Coiled-Coil NLR *Rph1*, confers leaf rust resistance in barley cultivar Sudan. *Plant Physiol.* **2019**, *179*, 1362–1372.
20. Li, J.; Huang, H.; Zhu, M.; Huang, S.; Zhang, W.; Dinesh-Kumar, S.P.; Tao, X. A plant immune receptor adopts a two-step recognition mechanism to enhance viral effector perception. *Mol. Plant* **2019**, *12*, 248–262. [CrossRef]
21. Liu, C.; Cui, D.; Zhao, J.; Liu, N.; Wang, B.; Liu, J.; Xu, E.; Hu, Z.; Ren, D.; Tang, D.; et al. Two *Arabidopsis* receptor-like cytoplasmic kinases SZE1 and SZE2 associate with the ZAR1-ZED1 complex and are required for effector-triggered immunity. *Mol. Plant* **2019**, *12*, 967–983. [CrossRef] [PubMed]
22. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [CrossRef] [PubMed]
23. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef] [PubMed]
24. Lozano, R.; Hamblin, M.T.; Prochnik, S.; Jannink, J.L.J.B.G. Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genom.* **2015**, *16*, 360. [CrossRef]
25. Caplan, J.; Padmanabhan, M.; Dinesh-Kumar, S.P. Plant NB-LRR immune receptors: From recognition to transcriptional reprogramming. *Cell Host Microbe* **2008**, *3*, 126–135. [CrossRef] [PubMed]
26. Sanseverino, W.; Hermoso, A.; D'Alessandro, R.; Vlasova, A.; Andolfo, G.; Frusciante, L.; Lowy, E.; Roma, G.; Ercolano, M.R. PRGdb 2.0: Towards a community-based database model for the analysis of R-genes in plants. *Nucleic Acids Res.* **2013**, *41*, D1167–D1171. [CrossRef]
27. 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]
28. Sahraeian, S.M.E.; Mohiyuddin, M.; Sebra, R.; Tilgner, H.; Afshar, P.T.; Au, K.F.; Asadi, N.B.; Gerstein, M.B.; Wong, W.H.; Snyder, M.P.; et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat. Commun.* **2017**, *8*, 59. [CrossRef]
29. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]
30. Van de Weyer, A.L.; Monteiro, F.; Furzer, O.J.; Nishimura, M.T.; Cevik, V.; Witek, K.; Jones, J.D.; Dangl, J.L.; Weigel, D.; Bemm, F. A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* **2019**, *178*, 1260–1272. [CrossRef]
31. Abramovitch, R.B.; Anderson, J.C.; Martin, G.B. Bacterial elicitation and evasion of plant innate immunity. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 601–611. [CrossRef] [PubMed]
32. Chisholm, S.T.; Coaker, G.; Day, B.; Staskawicz, B.J. Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* **2006**, *124*, 803–814. [CrossRef] [PubMed]
33. Wang, L.; Zhao, L.; Zhang, X.; Zhang, Q.; Jia, Y.; Wang, G.; Li, S.; Tian, D.; Li, W.H.; Yang, S. Large-scale identification and functional analysis of NLR genes in blast resistance in the Tetep rice genome sequence. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 18479–18487. [CrossRef] [PubMed]
34. Jacob, F.; Vernaldi, S.; Maekawa, T. Evolution and conservation of plant NLR functions. *Front. Immunol.* **2013**, *4*, 297. [CrossRef] [PubMed]
35. Tian, Y.; Fan, L.; Thurau, T.; Jung, C.; Cai, D. The absence of TIR-type resistance gene analogues in the sugar beet (*Beta vulgaris* L.) genome. *J. Mol. Evol.* **2004**, *58*, 40–53. [CrossRef]
36. Collier, S.M.; Hamel, L.P.; Moffett, P. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant. Microbe Interact.* **2011**, *24*, 918–931. [CrossRef]
37. Sarris, P.F.; Duxbury, Z.; Huh, S.U.; Ma, Y.; Segonzac, C.; Sklenar, J.; Derbyshire, P.; Cevik, V.; Rallapalli, G.; Saucet, S.B.; et al. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **2015**, *161*, 1089–1100. [CrossRef]
38. Cesari, S.; Thilliez, G.; Ribot, C.; Chalvon, V.; Michel, C.; Jauneau, A.; Rivas, S.; Alaux, L.; Kanzaki, H.; Okuyama, Y.; et al. The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell* **2013**, *25*, 1463–1481. [CrossRef]
39. Jones, J.D.; Vance, R.E.; Dangl, J.L. Intracellular innate immune surveillance devices in plants and animals. *Science* **2016**, *354*, aaf6395. [CrossRef]

40. Zhou, T.; Wang, Y.; Chen, J.Q.; Araki, H.; Jing, Z.; Jiang, K.; Shen, J.; Tian, D. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genom.* **2004**, *271*, 402–415. [CrossRef]
41. Wan, H.; Yuan, W.; Ye, Q.; Wang, R.; Ruan, M.; Li, Z.; Zhou, G.; Yao, Z.; Zhao, J.; Liu, S.; et al. Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: Characterization, genetic variation, functional divergence and expression patterns. *BMC Genom.* **2012**, *13*, 502. [CrossRef] [PubMed]
42. Wu, J.; Zhu, J.; Wang, L.; Wang, S. Genome-wide association study identifies NBS-LRR-encoding genes related with anthracnose and common bacterial blight in the common bean. *Front. Plant Sci.* **2017**, *8*, 1398. [CrossRef] [PubMed]
43. McHale, L.; Tan, X.; Koehl, P.; Michelmore, R.W. Plant NBS-LRR proteins: Adaptable guards. *Genome Biol.* **2006**, *7*, 212. [CrossRef] [PubMed]
44. Kuang, H.; Woo, S.S.; Meyers, B.C.; Nevo, E.; Michelmore, R.W. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **2004**, *16*, 2870–2894. [PubMed]
45. Krasileva, K.V.; Dahlbeck, D.; Staskawicz, B.J. Activation of an *Arabidopsis* resistance protein is specified by the *in planta* association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell* **2010**, *22*, 2444–2458.
46. Steinbrenner, A.D.; Goritschnig, S.; Staskawicz, B.J. Recognition and activation domains contribute to allele-specific responses of an *Arabidopsis* NLR receptor to an oomycete effector protein. *PLoS Pathog.* **2015**, *11*, e1004665. [CrossRef]
47. Pre, M.; Atallah, M.; Champion, A.; De Vos, M.; Pieterse, C.M.; Memelink, J. The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiol.* **2008**, *147*, 1347–1357. [CrossRef]
48. Kaufmann, K.; Melzer, R.; Theissen, G. MIKC-type MADS-domain proteins: Structural modularity, protein interactions and network evolution in land plants. *Gene* **2005**, *347*, 183–198. [CrossRef]
49. Huang, J.; Yang, X.; Wang, M.M.; Tang, H.J.; Ding, L.Y.; Shen, Y.; Zhang, H.S. A novel rice C2H2-type zinc finger protein lacking DLN-box/EAR-motif plays a role in salt tolerance. *Biochim. Biophys. Acta* **2007**, *1769*, 220–227. [CrossRef]
50. Yu, G.H.; Jiang, L.L.; Ma, X.F.; Xu, Z.S.; Liu, M.M.; Shan, S.G.; Cheng, X.G. A soybean C2H2-type zinc finger gene *GmZF1* enhanced cold tolerance in transgenic *Arabidopsis*. *PLoS ONE* **2014**, *9*, e109399. [CrossRef]
51. Kim, S.H.; Hong, J.K.; Lee, S.C.; Sohn, K.H.; Jung, H.W.; Hwang, B.K.J.P.M.B. CAZFP1, Cys2/His2-type zinc-finger transcription factor gene functions as a pathogen-induced early-defense gene in *Capsicum annuum*. *Plant Mol. Biol.* **2004**, *55*, 883–904. [CrossRef] [PubMed]
52. Tian, Z.D.; Zhang, Y.; Liu, J.; Xie, C.H. Novel potato C2H2-type zinc finger protein gene, *StZFP1*, which responds to biotic and abiotic stress, plays a role in salt tolerance. *Plant Biol. (Stuttg.)* **2010**, *12*, 689–697. [CrossRef] [PubMed]
53. Haridas, S.; Albert, R.; Binder, M.; Bloem, J.; LaButti, K.; Salamov, A.; Andreopoulos, B.; Baker, S.E.; Barry, K.; Bills, G.; et al. 101 *Dothideomycetes* genomes: A test case for predicting lifestyles and emergence of pathogens. *Stud. Mycol.* **2020**, *96*, 141–153. [CrossRef] [PubMed]
54. Hane, J.K.; Lowe, R.G.; Solomon, P.S.; Tan, K.C.; Schoch, C.L.; Spatafora, J.W.; Crous, P.W.; Kodira, C.; Birren, B.W.; Galagan, J.E.; et al. *Dothideomycete* plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *Plant Cell* **2007**, *19*, 3347–3368. [CrossRef]
55. Faris, J.D.; Friesen, T.L. Plant genes hijacked by necrotrophic fungal pathogens. *Curr. Opin. Plant. Biol.* **2020**, *56*, 74–80.
56. Han, G.Z. Origin and evolution of the plant immune system. *New Phytol.* **2019**, *222*, 70–83. [CrossRef]