agronomy

MDPI

Article

Phylogenetic Analyses and Transcriptional Survey Reveal the Characteristics, Evolution, and Expression Profile of NBS-Type Resistance Genes in Papaya

Qian Jiang [1,2], Yu Wang [1], Aisheng Xiong [3], Hui Zhao [1], Ruizong Jia [1], Mengyao Li [4], Huaming An [2], Changmian Ji [1,*] and Anping Guo [1,*]

[1] Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Institute of Tropical Bioscience and Biotechnology & Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China
[2] Agricultural College, Guizhou University, Guiyang 550025, China
[3] College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China
[4] College of Horticulture, Sichuan Agricultural University, Chengdu 611130, China
* Correspondence: jichangmian@itbb.org.cn (C.J.); guoanping@itbb.org.cn (A.G.)

Abstract: *Carica papaya* maintains an abnormally small but complete NLR family while showing weak disease resistance. To better understand their origin, evolution, and biological function, we identified 59 NLR genes via a customized RGAugury and investigated their characteristics, evolutionary history, and expression profiles based on the improved papaya genome and large-scale RNA-seq data. The results indicated that duplication is a major evolutionary force driving the formation of the papaya NLR family. Synteny analyses of papaya and other angiosperms showed that both insertion and inheritance-derived NLRs are present in papaya. Transcriptome-based expression and network analyses revealed that NLRs are actively involved in biotic stress responses. For example, a papaya-specific inserted TNL was up-regulated strongly by the fungal infection. Both transcriptome and qRT-PCR analyses confirmed the expression divergence of an RNL and an RCNL, a pair of tandem duplication genes involved in different co-expression modules. Furthermore, we observed an inserted gene cluster composed of five duplicated CNLs, showing dosage effects and functional differentiation of disease-resistance genes during evolution. This research will enhance our knowledge of the special NLR family in papaya, which may serve as a model plant for disease-resistance genetic studies.

Keywords: *Carica papaya*; NLRs; insertion-derived; biotic stress; expression divergence; transcriptomic network

Citation: Jiang, Q.; Wang, Y.; Xiong, A.; Zhao, H.; Jia, R.; Li, M.; An, H.; Ji, C.; Guo, A. Phylogenetic Analyses, and Transcriptional Survey Reveal the Characteristics, Evolution and Expression Profile of NBS-Type Resistance Genes in Papaya. *Agronomy* **2023**, *13*, 970. https://doi.org/10.3390/agronomy13040970

Academic Editors: Zhiyong Li, Chaolei Liu and Jiezheng Ying

Received: 3 February 2023
Revised: 22 March 2023
Accepted: 23 March 2023
Published: 25 March 2023

1. Introduction

Papaya (*Carica papaya* L.) is a tropical fruit that originated in Central America and benefits tropical and subtropical regions for its high commercial values [1,2]. Papaya fruits have a high nutritional value with abundant vitamin C, vitamin A, and carotenoids [3,4]. Papaya latex contains two important proteolytic enzymes: papain and chymopapain, which are widely used in food, chemicals, and pharmaceutical industries [5–7]. The draft genome of papaya was first reported in 2008, which revealed that the papaya genome contains fewer genes than most angiosperms. Specifically, significantly fewer disease-resistance genes were identified from the papaya genome [1]. In 2022, two updated genomes of the transgenic SunUp and its progenitor Sunset papaya were released [2]. The genome quality was substantially improved, and the lack of disease-resistance genes was reconfirmed in the updated genome release [2]. Among major diseases in papaya-growing regions, PRSV (papaya ringspot virus) is the most destructive and widespread virus to papaya, but no PRSV resistance has been found in papaya germplasm so far [2]. Although genetic engineering made papaya resistant to PRSV infestation [8], the rapid evolution of PRSV and the diversity in lineages led to the loss of transgenic papaya resistance in some regions [9].

The disease resistance (R) gene is a crucial breakthrough in plant disease resistance breeding [10,11]. Intracellular R proteins are often members of the NLR (nucleotide-binding site leucine-rich repeat receptor) family, which played important roles in effector-triggered immunity (ETI) [12–15]. Full-length NLR proteins usually contain a specific N-terminal domain, a conserved nucleotide-binding site (NBS) domain, and an extremely variable C-terminal leucine-rich repeat (LRR) domain, and could be divided into three subclasses based on the N-terminal domain: Toll/Interleukin-1 receptor (TIR)-NBS-LRR (TNL), resistance to powdery mildew8 (RPW8)-NBS-LRR (RNL), and coiled-coil (CC)-NBS-LRR (CNL) [11,13,15,16]. NLR gene copy number and subclass composition often varied greatly in angiosperms for the rapid gene loss and gain [17]. Liu et al. [17] surveyed 305 angiosperms and found that 93 species have no TNL, 23 species have no RNL, and only 1 species has no CNL. Moreover, TNL gene loss events occur frequently throughout dicots and magnoliid diversification, and the TNL is even absent in all monocots [17–19]. It is certain that papaya contains all three subclasses of NLRs and the number is comparatively lower than major angiosperms, although previous studies have reported slightly different numbers of NLR genes (50, 54, and 55) [1,17,20,21]. In terms of disease resistance of papaya, there is resistance against *Colletotrichum brevisporum* [22,23], but no resistance against PRSV in papaya germplasm [2]. The unique NLR gene family composition of papaya attracts us to study their characteristics, evolution, and functional roles.

Here, we identified 59 NLR genes from the improved papaya genome using a customized RGAugury pipeline and performed comprehensive analysis on them, including structural composition, sequence diversity, chromosomal distribution, and phylogenetic analysis. Gene duplication inspection and the syntenic analysis helped us to trace the evolutionary origin of papaya NLRs. Additionally, the expression profile and divergence of NLRs among transcriptome experiments were carefully investigated to elucidate their potential biological roles and functional differentiation under different biotic and abiotic stresses.

2. Materials and Methods

2.1. Identification and Similarity Comparison of NLRs

For the identification of NLRs, we customized the RGAugury pipeline [24] with two major changes: (1) The identification of powdery mildew8 (RPW8) domain was added. (2) NLRs in the original "others" group were assigned to classes, which are named by abbreviated terms of domains. To sum up, the customized pipeline mainly included the following steps: First, potential RGAs (Resistance gene analogs) were obtained by blast to the RGAdb [24]. Second, NBS, TIR, and RPW8 domains were identified by pfam_scan [25]. Third, LRR and CC domains were identified by InterProScan [26] and nCoils [27]. Finally, all identified results were summarized. In the study, TIR, RPW8, CC, NBS, and LRR domains were represented by their initials T, R, C, N, and L, respectively. The customized RGAugury is available in a git repository at https://github.com/reductase4/papaya_NLRs (accessed on 31 January 2023).

In this study, we collected genomes of three basal angiosperms and six eudicots to compare the NLR family of papaya with other angiosperms (for genome data source, see Table S1). The three basal angiosperms (*Nymphaea colorata*, *Nymphaea thermarum*, and *Euryale ferox*) and six eudicots (*Vitis vinifera*, *Prunus persica*, *Citrus sinensis*, *Theobroma cacao*, *Tarenaya hassleriana*, and *Arabidopsis thaliana*) keep syntenic associations with papaya, and they are gradually close to papaya in taxon phylogeny (Figure 1). Then we performed the customized RGAugury pipeline on each genome with the command line: `RGAugury_modified.pl -p longest_transcript.pep.fa -d SUPERFAMILY, SMART, gene3d, Pfam`. Gene IDs and classes of 59 NLRs from papaya are listed in Table S2. Regarding similarity comparison, we applied EMBOSS Needle with the Needleman–Wunsch algorithm [28] to generate the optimal global alignment of two sequences and calculate the similarity of pairwise sequences.

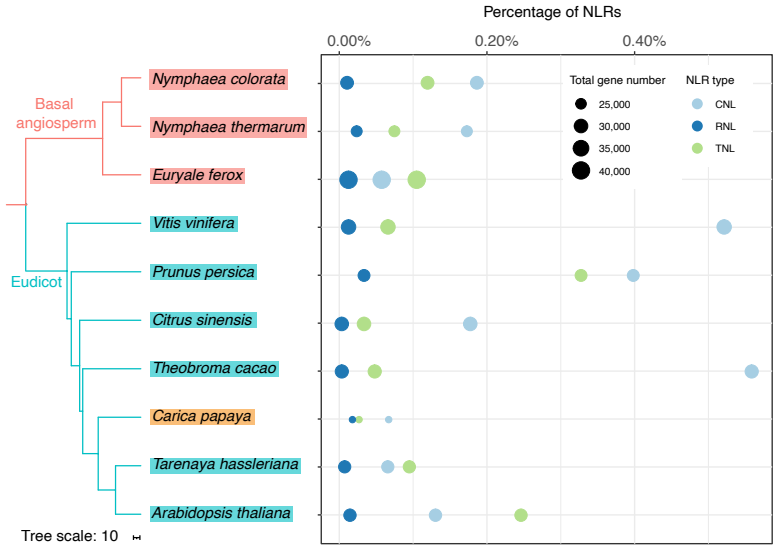

Figure 1. Phylogeny of papaya and selected species with the genome-wide proportion of NLRs (CNL%, RNL%, TNL%). Phylogenomic relationships of species refer to http://www.timetree.org (accessed on 15 November 2022) [29] and research of water lily genome [30].

2.2. Duplication Modes Identification and Chromosomal Localization of Papaya NLRs

DupGen_finder was used to identify genome-wide duplication modes of papaya, including whole-genome, tandem, proximal, transposed, and dispersed duplication (WGD, TD, PD, TRD, DSD) [31]. According to DupGen_finder's description [31], three basal angiosperms (*N. colorata*, *N. thermarum*, and *E. ferox*) were used as the outgroup for the transposed duplication detection. We wrote python scripts to extract gene pairs involving at least one NLR gene from genome-wide duplication identification, in which genes are allowed to be reused in duplication gene pairs. Gene density of chromosomes was calculated by a window size of 100 kb. Finally, papaya chromosomes with NLR loci and duplication associations were visualized by TBtools [32].

2.3. Phylogenetic Analysis of NLRs in Papaya

Referring to the phylogenetic analysis of NLR proteins performed by Gao et al. [33] and Shao et al. [16], we used the NBS domain to construct the phylogenetic tree and NBS genes from Rhodophyta were used as an outgroup in this study. All NLR proteins shared the conserved domain NBS, and Rhodophyta is outside the green plants. A tree of all the full-length NLRs found in papaya and nine other selected species was built in order to illustrate the phylogeny of NLRs in angiosperms. To fully investigate the phylogeny of NLRs in papaya, a tree of full-length NLRs was also constructed.

The NBS domain sequences of NLRs were aligned using the UPP algorithm [34], and the aligned sequences were trimmed using trimAl with the "-gappyout" option [35]. FastTree [36,37] was used to construct the approximately-maximum-likelihood phylogenetic tree with the "-pseudo" and "-gamma" options. The reliability of the tree was estimated by the Shimodaira–Hasegawa test [38] with 1000 resamples.

2.4. Analysis of Conserved Motifs, Domains, and Gene Structures of NLRs in Papaya

Conserved motifs were discovered using MEME Suite 5.5.0 [39]. Conserved domains were re-confirmed in NCBI Conserved Domain Database (CDD) by Batch CD-Search interface with an expected value threshold of 0.05 [40]. We visualized the phylogenetic tree of NLRs with their conserved motifs, domains, and gene structures using TBtools [32].

2.5. Estimation of Divergent Times

We constructed multiple alignments for protein-coding DNA sequences of duplication gene pairs by ParaAT [41] and formatted them into AXT formats. Ks values were calculated by KaKs_Calculator 2.0 [42] with the YN model [43]. According to previous research [44,45], the formula $T = K/2r$ was used to calculate the divergence times, where T represents the divergence time, K means the divergence distance (using Ks values here), r indicates the mutation rate of substitution per site per year. We used a mutation rate of 12×10^{-9}, which we collected from *A. thaliana* under high temperature (29 °C) [46]. Considering that papaya is long-lived in high-temperature environments and R genes evolve rapidly, the mutation rate should be higher, and actual divergent times should be over-estimated.

2.6. Micro-Synteny Analysis of NLRs among Species

Pairwise synteny of papaya and other selected genomes was obtained from the following command line: `python -m jcvi.compara.catalog ortholog --full`. We observed and compared synteny near the loci of all NLRs, and visualized the local synteny around typical NLRs by command line: `python -m jcvi.graphics.synteny`. The analysis was performed following instructions in https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version) (accessed on 23 November 2022).

2.7. Expression Pattern and Co-Expression Analysis of NLRs in Papaya

Transcriptome sequencing data of five projects were downloaded from NCBI (PRJNA560275, PRJNA591254, PRJNA692338, PRJNA352643, and PRJNA470602), including 52 papaya samples of different tissues under biotic and abiotic stresses, such as three viruses, fungi, ethylene, and drought stresses (Table S6) [3,22,23,47,48]. We performed the quality control of the downloaded RNA-seq data by fastp [49], then aligned clean reads to the sunset genome by HISAT2 and assemble transcripts in StringTie [50]. Principal Component Analysis (PCA) was used to evaluate the biological reproducibility of treatments. The gene expression levels were calculated with Fragments Per kb per Million reads (FPKM) values. Differentially expressed genes (DEGs) were detected by DESeq2 with cutoffs of FDR < 0.05 and abs(fold change) > 2 [51,52]. Heatmaps of gene expression level and fold change values of NLRs were generated by R scripts. Weighted gene co-expression network analysis (WGCNA) was applied to cluster highly correlated genes into modules [53–55]. The co-expression network was illustrated by Cytoscape [56].

2.8. Real-Time Quantitative PCR Analysis of NLRs in PRSV Infected Papaya

Three-month seedlings of 'Tainong No. 2' papaya were prepared for inoculation. A 10% *w*/*v* inoculum was made by grinding the PRSV-infected leaves with 0.05 M phosphate buffer (pH 7.0). PBS buffer was used as a control inoculum. Inoculum was applied to the rubbed areas and rinsed away with distilled water after 10 min. After 7 days, fresh leaves of the same part were sampled and stored at −80 °C.

Total RNA was isolated from the control and PRSV-infected leaves using Spin Column Plant Total RNA Purification Kit (Sangon Biotech, Shanghai, China). Extracted RNA was reverse transcribed into cDNA using PrimeScriptTM II 1st Strand cDNA Synthesis Kit (TaKaRa, Beijing, China) according to the manufacturer's instructions. qRT-PCR was performed on the Applied Biosystems 7500 Real-Time PCR System (Thermo Fisher Scientific Inc.) using 2 × Q3 SYBR qPCR Master Mix (Universal). The qRT-PCR system was 20 µL, including 0.8 µL of forward and reverse primers (10 µM), 1 µL of cDNA (50 ng/µL), 10 µL of 2 × Q3 SYBR qPCR Master Mix, and 9.2 µL of ddH_2O. The relative expression was analyzed using the $2^{-\Delta\Delta C_T}$ method [57] with papain as the housekeeping gene. Specific primers of genes were listed in Table S7. Three biological replicates and three technical replicates were used in this study.

3. Results

3.1. A Set of Complete and Simplified NLRs in Papaya

To make the identification results appropriate to the analysis in this study, we modified the RGAugury pipeline [24]. A total of 59 NLRs were identified from the improved papaya genome [2], including 15 CNLs, 8 T(C)NLs, 5 R(C)NLs, and 31 partial (single or two domains) NLRs (Tables 1 and S2). The results showed papaya NLRs in this study than in previous studies [1,17,20,21], which should attribute to the recent optimization of identification methods and the substantial improvement in genome quality.

To further compare NLR genes in papaya with those in other angiosperms, we surveyed the NLR genes of nine angiosperms (*N. colorata*, *N. thermarum*, *E. ferox*, *V. vinifera*, *P. persica*, *C. sinensis*, *T. cacao*, *T. hassleriana*, and *A. thaliana*), which are gradually close to papaya in taxon phylogeny (Figure 1 and Table S1). Consistent with the previous report, papaya has fewer NLR genes than other angiosperms [1]. Furthermore, we compared the genome-wide proportion of three subclasses of NLR genes (TNL%, CNL%, and RNL%) in papaya with those in other species, and found that papaya has the lowest TNL% (papaya: 0.03%, other species: 0.03∼0.33%), a relative low CNL% (papaya: 0.07%, other species: 0.06∼0.56%) and a medium RNL% (papaya: 0.02%, other species: 0.00∼0.03%) (Figure 1 and Table 1). The proportion of RNL was the lowest among the three subclasses of NLRs in all surveyed species (Figure 1). The proportion of CNL was usually higher than that of TNL in most species, except for one basal angiosperm (*E. ferox*) and two brassicales (*A. thaliana* and *T. hassleriana*) (Figure 1). These results are consistent with the previous study of 305 angiosperms, in which CNL is the predominant NLR subclass in 268 species, and TNL predominates 36 species [17]. Moreover, Liu et al. [17] also observed that RNL is a minority subclass, which accounts for no more than 10% of all NLRs.

Similar to most angiosperms, papaya retains members of all three NLR subclasses, and CNL predominates in three NLR subclasses (Figure 1). However, each NLR subclass of papaya is simplified, containing only very few genes but maintaining an acceptable genome-wide proportion (Figure 1 and Table 1). In summary, papaya contains a set of complete and simplified NLRs, which is unique among angiosperms and makes papaya a suitable model plant for studying basic disease-resistance genes.

Table 1. The number of NLRs identified in papaya and selected species.

Species	Total	NLR	NBS	CN	TN	RN	NL	TCN	RCN	CNL	TNL	RNL	TCNL	RCNL
Nymphaea colorata	28,438	255	63	17	12	2	57	0	1	53	34	3	9	4
Nymphaea thermarum	25,461	157	21	15	10	2	31	1	1	44	19	6	3	4
Euryale ferox	40,049	134	8	9	8	0	13	1	0	23	42	5	16	9
Vitis vinifera	31,845	439	52	41	8	2	126	0	1	166	21	4	7	11
Prunus persica	26,873	374	24	7	17	0	105	3	1	107	88	9	11	2
Citrus sinensis	24,456	112	7	12	2	0	25	0	1	53	10	1	0	1
Theobroma cacao	29,181	275	15	13	3	0	63	0	0	163	14	1	0	3
Carica papaya	**22,416**	**59**	**10**	**3**	**1**	**0**	**16**	**1**	**0**	**15**	**6**	**4**	**2**	**1**
Tarenaya hassleriana	27,396	76	7	1	4	0	15	0	0	18	26	2	3	0
Arabidopsis thaliana	27,628	166	4	1	14	1	25	3	1	36	68	4	9	0

3.2. Different Similarity Distribution of Three NLR Subclasses

To explore the difference in genetic diversity of NLRs in angiosperms, we conducted a meticulous evaluation of the pairwise similarity of amino acid sequences for three NLR subclasses (Figure 2). As aforementioned, we observed that papaya has significantly fewer CNLs and TNLs than other species. However, what is interesting is that the similarity distribution of CNLs in papaya is similar to other species, while the similarity distribution of TNLs in papaya is different to other species, which lacks a similarity distribution higher than 78% and lower than 38%. Furthermore, the median of similarity among CNLs of papaya (26.4%) is lower than that of other species (31.3∼40%), while the median of similarity among TNLs of papaya (48.1%) is higher than that of other species (27.2∼44.6%) (Figure 2 and Table S3). RNLs are rare in all species, and even absent in *C. sinensis* and *T. cacao*.

However, there are four RNLs in papaya, and their median value of similarity in papaya (26.15%) is much lower than that in other eudicots (50.3∼70.9%) (Figure 2 and Table S3).

The lack of high similarity distribution of NLRs is consistent with the previous study of the papaya genome [1]. This was proposed to be a consequence of the lack of recent genome duplication in the papaya genome, which is atypical of other angiosperms; see the study of Ming et al. [1]. The absence of TNL gene pairs with low similarity in papaya, which is not common in angiosperms, suggests that the TNL-lost event had occurred in papaya. Additionally, the small size of RNLs and irregular distribution of RNL similarity indicated that the loss of RNL is a continuous process in angiosperm genomes.

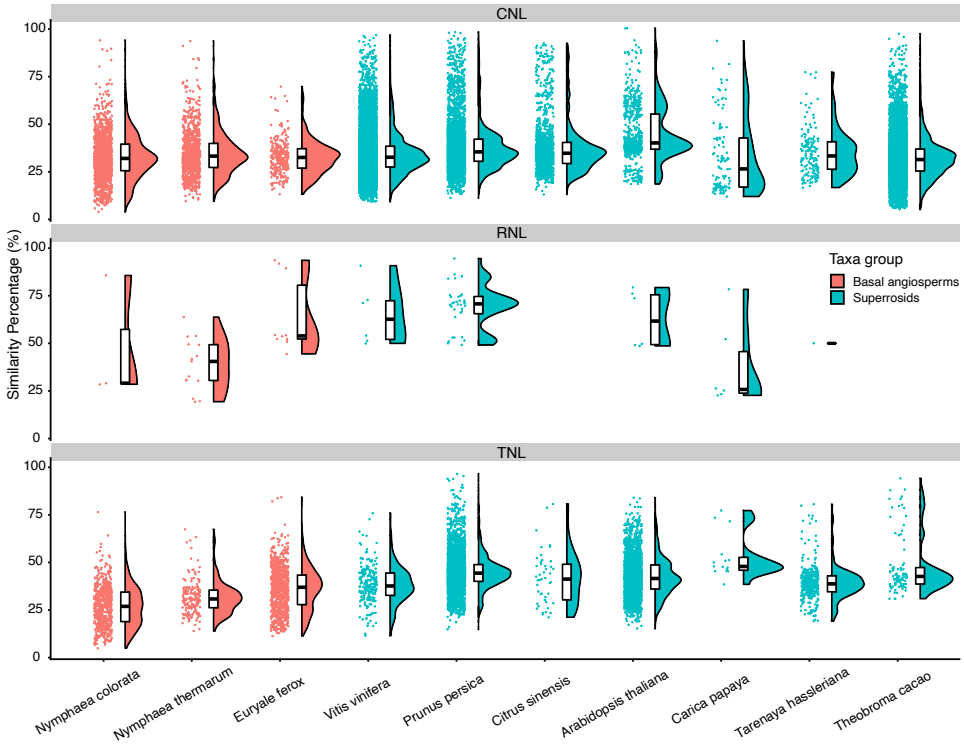

Figure 2. Similarity comparison of NLRs (CNL, RNL, TNL) in papaya and selected species.

3.3. Chromosome Distribution and Duplication Modes of NLRs in Papaya

Figure 3 exhibits locations of NLRs on nine chromosomes of papaya, with the least numbers of NLRs on Chr01 (one RNL) and Chr09 (one NL and one TNL). Chr02 contains five NLRs, Chr03 contains seven NLRs, Chr08 contains eight NLRs, while Chr04, Chr05, Chr06, and Chr07 contains nine NLRs, respectively (Figure 3). Most NLR genes are clustered at loci with relatively high gene density on chromosomes (Figure 3). Duplication gene pairs involving at least one NLR gene were extracted from the identification of genome-wide duplication. We identified 79 duplication gene pairs involving at least 1 NLR gene (51 DSD, 18 TD, 8 PD, 1 WGD, and 1 TRD) and observed that gene duplication events frequently occur at NLR-clustered loci (Figure 3 and Table S4). For example, five CNLs clustered on Chr06 (*sunset06G0024840*, *sunset06G0024850*, *sunset06G0024860*, *sunset06G0024870*, and *sunset06G0024880*) were identified as a segment of consecutive tandem duplication genes (Table S4). One TNL and two TCNLs clustered on Chr02 (*sunset02G0020440*, *sunset02G0020450*, and *sunset02G0020470*) were identified as proximal duplication genes (Table S4). The potential functional differentiation and expression divergence of these duplicated genes still remains unclear.

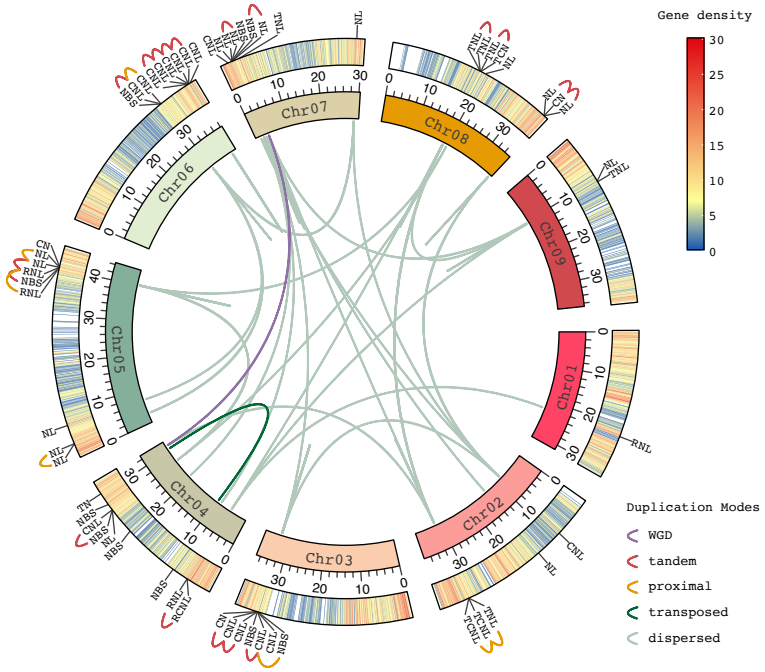

Figure 3. Physical locations and duplication modes of NLRs in papaya genome.

3.4. Conserved Sequences, Gene Structures, and Phylogeny of NLRs in Papaya

Three subclasses of full-length NLRs were found clustered in distinct clades of the tree after we studied the phylogeny of these genes in papaya and other selected species (Figure A1). To clearly show the phylogeny and structures of papaya NLRs, we constructed a maximum likelihood phylogenetic tree using 28 full-length papaya NLRs and 2 NBS genes of Rhodophyta as the outgroup (Figure 4A). The two NBS genes, *BWQ96_09803* from *Gracilariopsis chorda* and *CHC_T00003089001* from *Chondrus crispus*, were chosen as the outgroup of the papaya NLR tree because they are the closest to NLR genes among the outgroup genes in Figure A1. Meanwhile, we analyzed the conserved motifs, conserved domains, and gene structures for NLRs on the tree (Figure 4B–D). The phylogeny, conserved sequences, and gene structures showed that the RCNL and TCNL are highly related to RNL and TNL, respectively (Figure 4A–D). Additionally, we noticed that two RNL genes (*sunset05G0021430* and *sunset05G0021370*) fall inside the CNL clade (Figures 4A and A1). The classification of these two genes deserves more discussion because a segment of their N-terminal sequence could match both the CC and RPW8 domains simultaneously in the results of pfam_scan.

Overall, members of the same NLR subclass generally shared similar compositions of protein and gene structures, suggesting similar functions among them (Figure 4A–D). For conserved motifs, motif 1 was present in all NBS domains while motif 9 was present in all TIR domains on the tree (Figure 4B,C). The CC and LRR domains were more variable, with motif 6 typically present in CC domains and motif 5 regularly repeated in LRR domains (Figure 4B,C). The majority of the domains identified by RGAugury, which combines several databases and domain-specific identification tools [24], could be confirmed in the CDD database, with the exception of a small number of CC and PRW8 domains (Figure 4C). Conserved domain analysis showed that all TNLs in papaya contain the PPP1R42 (protein phosphatase 1 regulatory subunit 42) domain architecture, which influences centrosome separation, interacts with PP1 (protein phosphatase-1), and promotes its activity [58,59] (Figure 4C). Moreover, most papaya TNLs had the C-JID domain (C-terminal jelly roll/Ig-like), which has been demonstrated to be identified by pathogen effectors specifically [60] (Figure 4C). Gene structure analysis indicated that most NLRs clustered in the same clade

displayed similar gene structures, except for a few CNLs with abnormally large intron insertions (Figure 4D).

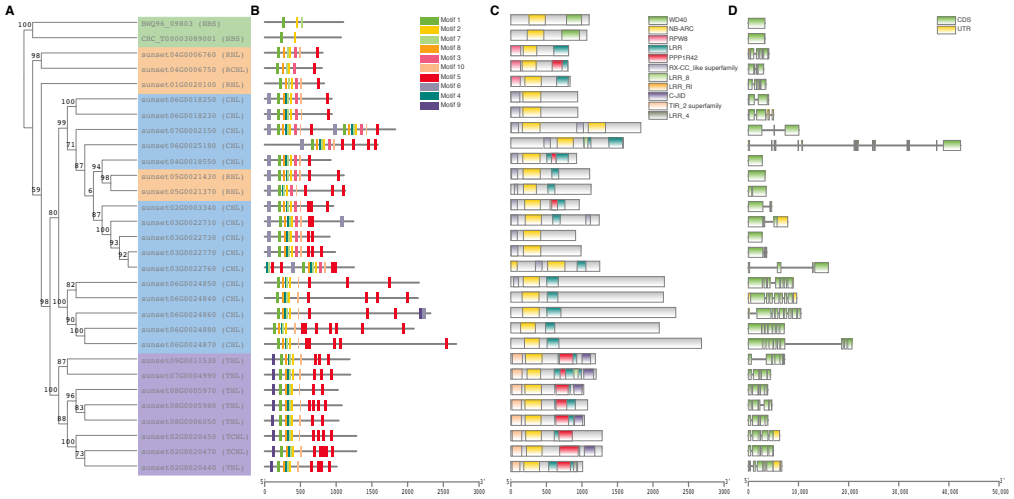

Figure 4. Phylogeny, conserved motifs, domains, and gene structures of papaya NLRs (CNL, RNL, TNL). (**A**) Phylogeny of papaya NLRs. NBS genes from Rhodophyta were used as the outgroup. The numbers at the nodes mean branch support values, which are evaluated by the Shimodaira–Hasegawa test. Green indicates the NBS subclass. Orange indicates the RNL subclass. Blue indicates the CNL subclass. Purple indicates the TNL subclass. (**B**) Conserved motifs of papaya NLRs. (**C**) Conserved domains of papaya NLRs. (**D**) Gene structures of papaya NLRs.

3.5. Inheritance- and Insertion-Origin of NLRs in Papaya Genome

To explore the evolutionary origin of papaya NLRs, local syntenic relationships of full-length NLR genes were investigated among selected angiosperms, covering basal angiosperms to eudicots (Table S5). By analysis, syntenic orthologs of RNL (*sunset01G0020100*) were discovered in genomes of *V. vinifera*, *P. persica*, *C. sinensis*, and *T. cacao*, demonstrating that this gene is conservative in dicots (Figure 5A). Additionally, papaya and basal angiosperms (*N. colorata* and *N. thermarum*) retained the syntenic relationship of RNL (*sunset04G0006760*) and RCNL (*sunset04G0006750*), suggesting that these two genes are derived from basal angiosperms (Figure 5B). Table S5 showed that papaya is the only eudicot that inherited both genes from basal angiosperms in this study. Moreover, we observed an inversion in the conserved syntenic region of Chr04: 5.01–5.14 Mb in papaya, which comprises the RNL and RCNL genes (Figure 5B). These findings revealed that papaya RNL genes are usually acquired via inheritance, which is the most typical origin of functional genes in plants.

In contrast to the dominance of inheritance in RNL genes, the local synteny analysis revealed that all TNL and most CNL genes are insertion-derived in papaya (Table S5). Compared to adjacent orthologous genes among selected species, a papaya TNL (*sunset07G0004990*) was found to be an insertion-derived gene located at Chr07: 3.78-3.99 Mb of papaya (Figure 6A). Additionally, a cluster of duplicated CNL genes (*sunset06G0024840*, *sunset06G0024850*, *sunset06G0024860*, *sunset06G0024870*, and *sunset06G0024880*) were found to be inserted into Chr06: 37.15–36.91 Mb of the papaya genome (Figure 6C). The upstream of these five CNLs on the papaya genome matched best with Chr06: 1.54–1.71 Mb region of *V. vinifera*, while the downstream matched best with Chr14: 4.72–5.08 Mb region of *V. vinifera* (Figure 6C). This syntenic relationship around the inserted loci was also kept in other dicots, demonstrating the confident insertion-origin of papaya-specific CNLs. Notably, the inserted loci were reshaped by a chromosome rearrangement during the chromosome evolution of papaya (Figure 6C). We calculated synonymous substitution rates (Ks) values of the five duplicated CNLs and estimated their divergent times based

on a mutation rate of 12×10^{-9} base substitutions per site per generation [46] (Table 2). Since previous research estimated that Caricaceae diverge from its sister clade around 35.5–39.4 Mya and papaya diverge from its sister clade around 24.2–25.1 Mya [61,62], we deduced that *sunset06G0024860* diverge around the divergent time of Caricaceae, while clade I and clade II diverge around the divergent time of papaya (Figure 6B). Moreover, *sunset06G0024870* and *sunset06G0024880* were generated by the most recent duplication event (Figure 6B).

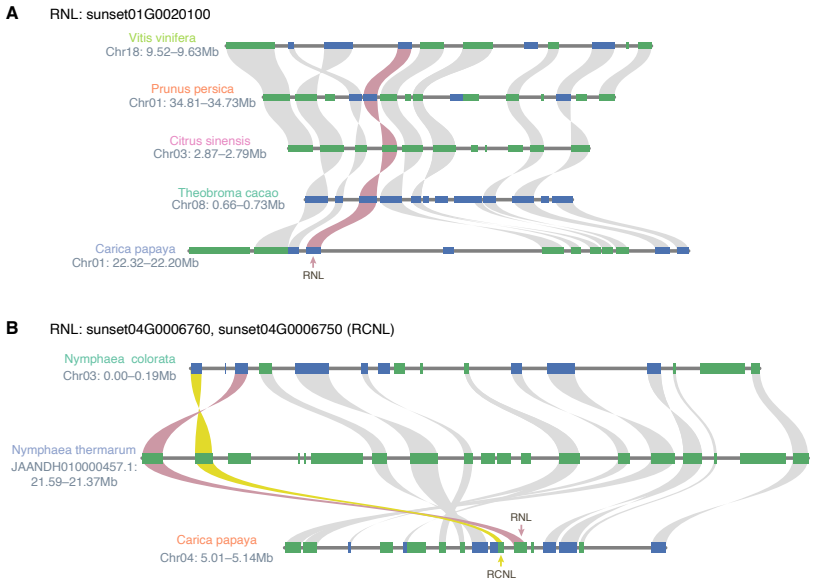

Figure 5. Inheritance-derived NLRs in papaya. (**A**) Conservation of a single RNL (*sunset01G0020100*). (**B**) Conservation of RNL (*sunset04G0006760*) and RCNL (*sunset04G0006750*).

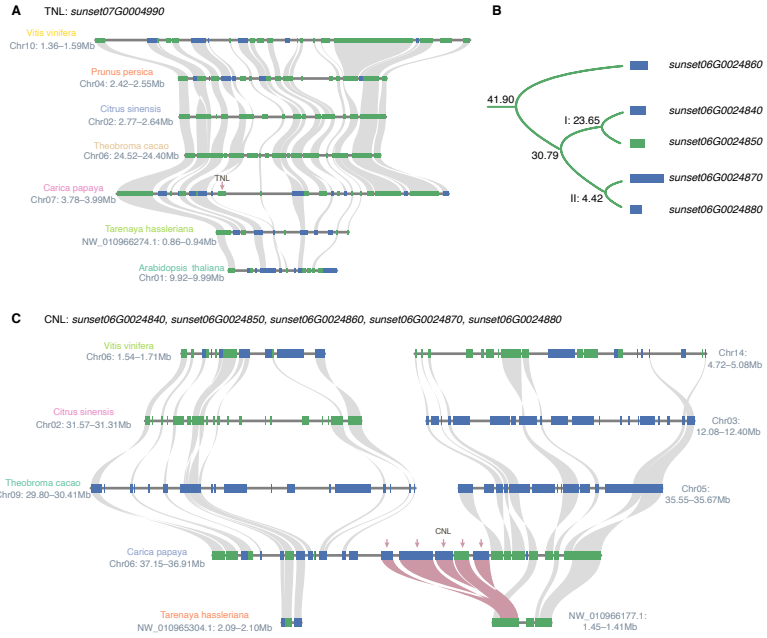

Figure 6. Insertion of NLRs in papaya. (**A**) Insertion of a single TNL (*sunset07G0004990*). (**B**) Evolution model of five duplicated CNLs. The numbers at the nodes indicated divergent times (Mya) estimated in Table 2. (**C**) Insertion of five CNLs in the papaya genome (*sunset06G0024840*, *sunset06G0024850*, *sunset06G0024860*, *sunset06G0024870*, *sunset06G0024880*).

Table 2. Duplication modes and divergent times of a five-CNL-cluster in papaya.

Duplicate 1	Type 1	Duplicate 2	Type 2	E-Value	Duplication Mode	Ks	Divergent Time (Mya)
sunset06G0024860	CNL	*sunset06G0024870*	CNL	8.56×10^{-82}	tandem	1.01	41.90
sunset06G0024840	CNL	*sunset06G0024860*	CNL	0	dispersed	0.93	38.81
sunset06G0024860	CNL	*sunset06G0024880*	CNL	2.62×10^{-82}	dispersed	0.87	36.19
sunset06G0024850	CNL	*sunset06G0024860*	CNL	3.64×10^{-84}	tandem	0.87	36.06
sunset06G0024850	CNL	*sunset06G0024870*	CNL	0	dispersed	0.74	30.79
sunset06G0024840	CNL	*sunset06G0024850*	CNL	2.30×10^{-83}	tandem	0.57	23.65
sunset06G0024870	CNL	*sunset06G0024880*	CNL	0	tandem	0.11	4.42

3.6. Expression of NLR Genes in Papaya under Biotic and Abiotic Stresses

We investigated the expression levels of NLR genes under biotic and abiotic stresses to mine their functions. A total of 52 papaya samples' RNA-seq data were collected from 5 NCBI projects, including treatments of PRSV, papaya mosaic virus (PAPMV), papaya leaf-distortion mosaic virus (PLDMV), fungi (*C. brevisporum*), ethylene, drought, tissues of leaf, root, and sap (Figure A2 and Table S6). We found that more NLR genes are up-regulated by biotic stresses and more NLR genes are down-regulated by abiotic stresses (Figure 7). This was consistent with the recognition that NLRs are involved in the process of pathogen resistance [63,64]. Additionally, more NLRs responded to fungal infection than to PRSV, PAPMV, and PLDMV infection (Figure 7). For fungal infection, more NLRs responded to stress in resistant papaya than that in susceptible papaya (Figure 7). The papaya-specific inserted TNL (*sunset07G0004990*) was strongly up-regulated by the infection of fungi and viruses, and especially highly expressed in fungi-infected papaya (Figures 7 and A2).

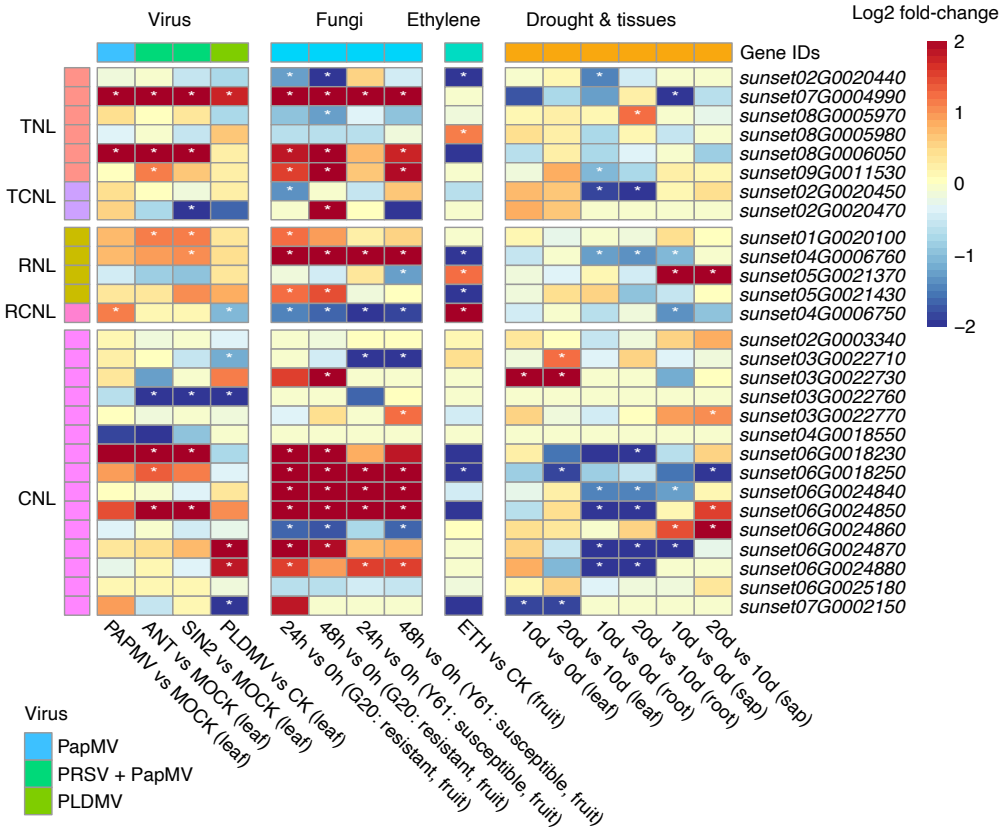

Figure 7. Log2 fold-change expression of NLRs in papaya under biotic and abiotic stresses. DEGs are marked by white starts with Log2 fold-change > 1 and $P_{FDR} < 0.05$. ANT indicates 'Antagonism' samples, which are inoculated with simultaneous PapMV and PRSV [47]. SIN2 indicates 'Sinergism2' samples, which are inoculated with stepwise PapMV and PRSV [47].

As shown in Figure 5B, the tandem duplication genes RNL(*sunset04G0006760*) and RCNL (*sunset04G0006750*) were both conserved from basal angiosperms. However, the RNL was up-regulated, whereas the RCNL was down-regulated by fungal infection (Figure 7). Meanwhile, the RNL is highly expressed in virus-infected papaya samples, and the RCNL is highly expressed in papaya sap (Figure A2). Among five CNL genes of the papaya-specific CNL cluster, which derived from tandem duplication and insertion, four (*sunset06G0024840*, *sunset06G0024850*, *sunset06G0024870*, and *sunset06G0024880*) were up-regulated by fungal infection to varying degrees, whereas one (*sunset06G0024860*) was down-regulated (Figure 7). These findings showed the dosage effect and expression divergence of duplicated genes, both of which contribute to the functional diversity of disease-resistance genes in papaya.

The expression levels of duplication genes discussed above were analyzed in PRSV-infected papaya by qRT-PCR technology. The results showed that the relative expression level of RNL (*sunset04G0006760*) is strongly up-regulated by PRSV infection, whereas the relative expression levels of RCNL (*sunset04G0006750*) are not significantly different between control and PRSV infected papaya (Figure 8A). Among five CNL genes of the papaya-specific CNL cluster, the relative expression levels of three (*sunset06G0024840*, *sunset06G0024870*, and *sunset06G0024880*) in PRSV infected papaya were significantly higher than in control (Figure 8B). The relative expression level of *sunset06G0024850* in PRSV infected papaya is higher than in the control, but not significantly (Figure 8B). Moreover, the relative expression levels of *sunset06G0024860* were not significantly different between control and treatment (Figure 8B). Although these results are slightly inconsistent with those of transcriptome analysis of virus-infected papaya, they also showed the dosage effect and expression divergence of duplicated genes as above. The 'Antagonism' and 'Sinergism2' samples of PRJNA560275 were papaya inoculated with simultaneous or stepwise PapMV and PRSV, and samples were collected from systemic leaves at 60 dpi (day post infection) [47]. In this study, we inoculated papaya seedlings with single PRSV and collected samples at 7 dpi. Therefore, both the virus inoculation and sampling time point contributed to the inconsistency.

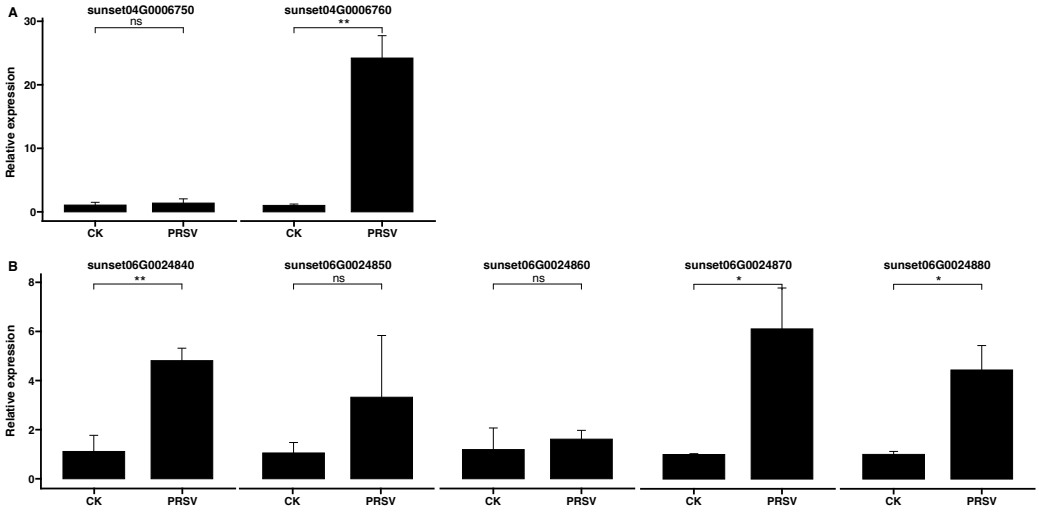

Figure 8. Real-time quantitative PCR analysis of NLR genes in PRSV infected papaya. (**A**) Relative expression levels of RNL duplication gene pairs. (**B**) Relative expression levels of CNL duplication gene pairs. Error bars indicate the standard deviation. Statistical significance is evaluated by T-test and represented by symbols (ns: $p > 0.05$, *: $p <= 0.05$, and **: $p <= 0.01$).

3.7. Co-Expression Network Analysis of NLRs in Papaya Infected by Fungi

To investigate the potential interactions involving NLR genes, WGCNA was performed based on transcription data from papaya samples infected by fungi. As demonstrated in Figure 9A, the gene set in the lightcyan module was significantly correlated

with the papaya anthracnose resistant/susceptible traits (r = 0.88). Gene set in the midnightblue module showed a strong positive correlation (r = 0.8) with fungal infection or not, while the brown module showed a strong positive correlation (r = 0.9) with infection time (0/24/48 h). In addition, it is also worth noting that the turquoise module is strongly negatively correlated with fungal infection (r = −0.89) and infection time (r = −0.88). We found 11 NLRs involved in these 8 modules, including 3 (NBS: *sunset03G0022700*, CNL: *sunset03G0022710*, and RCNL: *sunset04G0006750*) in black, 3 (RNL: *sunset04G0006760*, CNL: *sunset06G0018250*, and CNL: *sunset06G0024840*) in midnightblue, 4 (CNL: *sunset06G0024850*, CNL: *sunset06G0024860*, CNL: *sunset06G0025180*, and TNL: *sunset08G0006050*) in turquoise, and 1 (TNL: *sunset07G0004990*) in brown. The gene and eigengene expressions of turquoise, brown, and midnightblue modules are presented in Figure 9B–D. Eigengenes of the brown and midnightblue module were up-regulated by fungal infection (Figure 9C,D), while that of the turquoise module was down-regulated (Figure 9B). The hub-network of the midnightblue module was constructed, and two CNLs (*sunset06G0018250* and *sunset06G0024840*) and one RNL (*sunset04G0006760*) were found to be core hub-genes for fungal infection (Figure 9E and Table S8). The hub-network and hub-genes of the co-expression genes would be beneficial to improve the fungal resistance for papaya breeding.

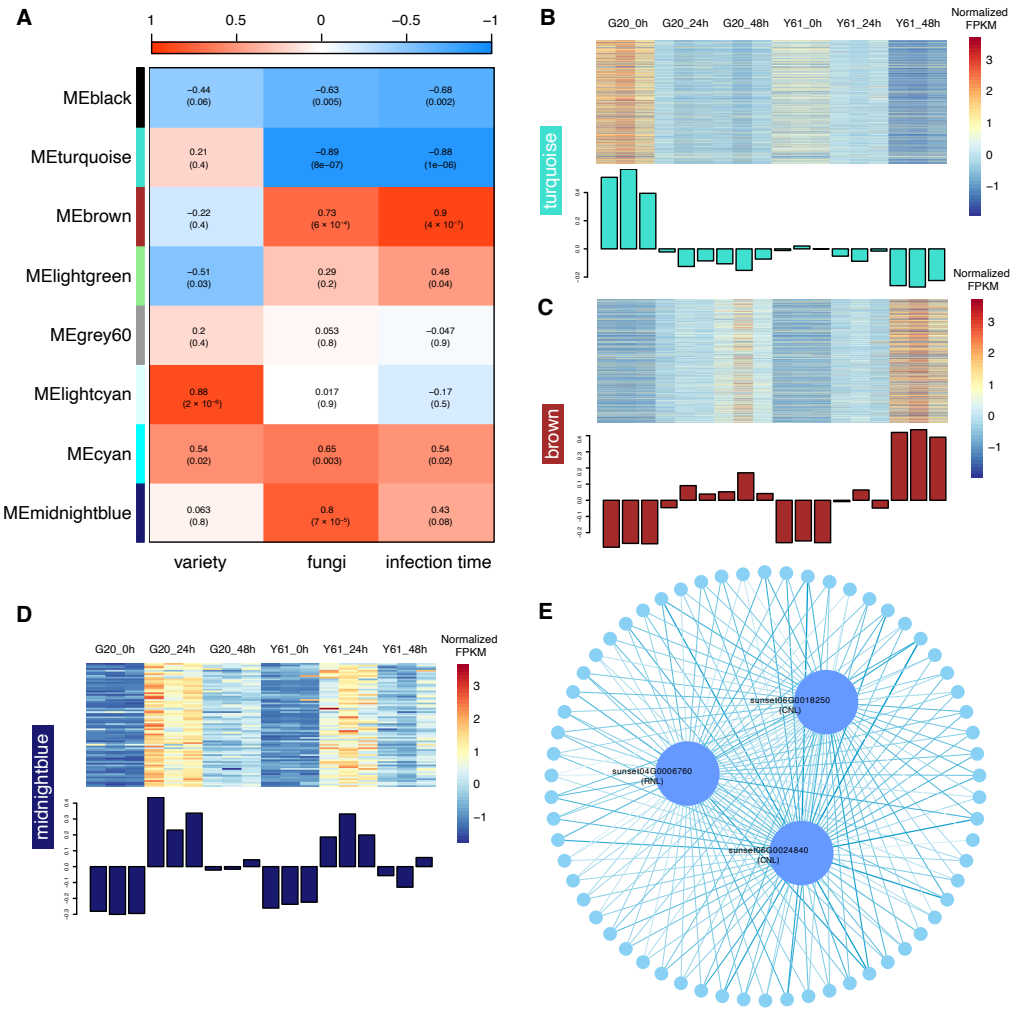

Figure 9. Gene co-expression network analysis of NLRs in papaya. (**A**) Correlation of gene modules and traits. (**B**–**D**) Eigengene and total gene expression of turquoise, brown, and midnightblue modules. (**E**) Gene co-expression network of three NLRs in the midnightblue module.

4. Discussion

In this study, we have identified and classified NLR family members more accurately than the previous research on papaya NLRs [20] due to the increasing awareness of crucial domain architectures in NLR and the development of NLR identification tools. In addition, excellent tools (e.g. NLR-Annotator [65], NLGenomeSweeper [66], and Homology-based R-gene Prediction (HRP) [11]) have also been recently developed to facilitate the NLR identification. For example, NLGenomeSweeper complements R genes that have not been annotated in the genome [66], and HRP has a high performance in full-length R-gene discovery [11]. We believe that these tools will help us further improve the accuracy of gene recognition and the quality of subsequent bioinformatic mining.

Based on the distribution of sequence similarity, we propose that both the loss of NLRs and the lack of recent genome duplication lead to the smaller NLR family of papaya compared to typical angiosperms. The similar subclasses composition and genome-wide proportion of papaya and typical eudicots revealed that the papaya NLR family is a simplified set of NLRs in typical eudicots. The study of papaya NLRs will help us better understand the complex and diverse disease-resistance genes in eudicots, which comprise the largest species diversity of living angiosperms [67,68].

Gene duplication is an important mechanism driving gene diversification and genetic innovation, and its roles in species evolution and adaptation have been well recognized [69,70]. Genes involved in signaling, transport, and metabolism tend to retain copies after duplication more than those involved in the maintenance of genome stability and organelle function [69]. As we know, NLR is a large and diverse family that activates ETI in response to pathogen effectors and mediates immune signaling subsequently [14,64]. In this study, we found that DSD, TD, and PD are the main duplication modes in papaya that contribute to the expansion of the NLR family. DSD is prevalent among plant genomes, which generate gene copies with unclear mechanisms, neither colinear nor neighboring [31]. For the remaining duplication modes with known mechanisms, TD and PD dominated papaya NLRs, while WGD and TRD contributed little. Among 50 large gene families investigated in *A. thaliana*, the NLR family was classified into the family group containing moderate TD and low WGD [71]. In most angiosperms, WGD accompanied by massive TRD usually contributed more than TD and PD [31,72]. Despite the absence of recent genome duplication in papaya, gene pairs produced from WGD (1698 pairs) and TRD (4374 pairs) were significantly more numerous than those from TD (1192 pairs) and PD (696 pairs). This implies that various duplication modes have contributed differently to the papaya NLR gene set than the papaya genome. In addition, the research on plant gene duplication found that more than 63% of gene pairs derived from duplication events showed expression divergence [31]. The expression divergence of duplication pairs was well confirmed in three TD-derived NLR pairs of RNL (*sunset04G0006760*) and RCNL (*sunset04G0006750*), CNL (*sunset06G0024860*) and CNL (*sunset06G0024850*), and CNL (*sunset06G0024860*) and CNL (*sunset06G0024870*).

Natural evolution involved the fission and fusion of chromosomes, which resulted in genome rearrangement and assembly [73]. Gene deletion and insertion occurred frequently in recombination regions [74]. The papaya genome suffered chromosome fission and fusion at the region of Chr06: 37.15–36.91 Mb, because this region matched with two collinear segments on *V. vinifera*, *C. sinensis*, *T. cacao*, and *T. hassleriana* genomes. A cluster of tandem duplicated CNLs (*sunset06G0024840*, *sunset06G0024850*, *sunset06G0024860*, *sunset06G0024870*, and *sunset06G0024880*) were discovered to be inserted into the two segments of this recombination region. Three genes in this cluster were involved in two co-expression modules, which correlated strongly to the fungal infection. Dosage effect and expression divergence were both observed in this tandem duplication CNL-cluster.

Plant NLRs are typically classified into two functional roles in pathogen-induced immune responses: sensor NLRs that recognize pathogen effectors and helper NLRs that assist other NLRs in triggering the immune response [64,75]. To catch up with the evolution of pathogens, it is more urgent for sensor NLRs than for helper NLRs to accelerate gene diversity and genetic novelty. RNL is an ancient and conserved subclass of plant NLRs and

is preferred as a helper [16,75]. We speculate that the simplified NLR family maintains the essential ETI in papaya by preserving a few conserved RNLs and CNLs as 'helpers' and inserting relatively abundant TNLs and CNLs for recognizing variable pathogen effectors. In general, papaya lacks germplasm, which is disease resistant [2]. As a result, it has evolved into other resistant pathways, such as the excretion of latex-containing proteins and enzymes that have demonstrated toxicity [76].

5. Conclusions

In this study, 59 NLR genes were identified from the improved papaya genome, including 15 CNLs, 8 T(C)NLs, 5 R(C)NLs, and 31 partial NLRs. Compared to typical angiosperms, papaya retains complete and simplified subclasses of NLR genes, making papaya a suitable plant model for studying basic disease-resistance genes. The absence of NLR gene pairs with low similarity indicates the loss of ancestral NLR genes. Moreover, members of the same subclass are close on the phylogenetic tree, and they have similar conserved sequences, gene structures, and functional recognition sites, suggesting their similar biological functions. Pairwise synteny analysis revealed the inheritance- and insertion-derived NLR genes. One tandem duplication gene pair (RNL *sunset04G0006760* and RCNL *sunset04G0006750*) is found to be inheritance-derived. Meanwhile, we observed an inserted gene cluster composed of five duplicated CNL genes. The insertion loci of the CNL cluster are found to have been reshaped by a chromosome rearrangement during the chromosome evolution of papaya. We further estimated their divergent time and established an evolutionary model. Expression and transcriptomic network analysis revealed that papaya NLR genes actively respond to biotic stress. Both the public RNA-seq data and qRT-PCR experiment confirmed the dosage effect and functional differentiation of the papaya NLR genes. This study gives new perspectives on the evolution of NLR genes and provides a basis for the disease-resistant breeding of papaya.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy13040970/s1, Table S1: Data source of plant genomes; Table S2: NLRs identified in papaya genome; Table S3: Summary of pairwise similarity (%) of NLRs; Table S4: Duplication modes of NLRs in papaya; Table S5: Syntenic paralogs of NLRs in papaya; Table S6: Data sources of transcriptome; Table S7: Gene-specific primers for qRT-PCR; Table S8: Network edges of midnightblue module.

Author Contributions: Conceptualization, A.G., C.J. and Q.J.; methodology, A.X.; software, Q.J. and Y.W.; formal analysis, Q.J., Y.W. and H.Z.; investigation, A.X., M.L. and H.A.; resources, R.J. and M.L.; data curation, Q.J., Y.W. and C.J.; writing—original draft preparation, Q.J.; writing—review and editing, A.G., C.J. and Q.J.; supervision, A.G. and H.A.; funding acquisition, Q.J. and C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Hainan Provincial Natural Science Foundation of China (No. 321QN289), Hainan Major Science and Technology Project (No. ZDKJ202002), Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (No. 1630052020007 and No. 1630052021017) the innovation platform for Academinicians of Hainan Province and Hainan Postdoctoral Research Funding Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Transcriptional data were downloaded from NCBI. Sunset genome and annotation files were downloaded from Genome Warehouse of China National Center for Bioinformation. Other genome and annotation files were downloaded from public databases listed in Table S1. Scripts were available in a git repository at: https://github.com/reductase4/papaya_NLRs (accessed on 31 January 2023).

Conflicts of Interest: The authors declare no conflict of interest. The funding body had no role in the design of the study, collection, analysis, interpretation of data, and writing the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

NLR	Nucleotide-binding site (NBS) leucine-rich repeat (LRR) receptor
ETI	Effector-triggered immunity
TIR	Toll/Interleukin-1 receptor
RPW8	Resistance to Powdery Mildew Locus 8
CC	Coiled-coil
Mya	Millions of years ago
Ks	Synonymous substitution rates
PRSV	Papaya ringspot virus
PAPMV	Papaya mosaic virus
PLDMV	Papaya leaf-distortion mosaic virus
WGCNA	Weighted gene co-expression network analysis
WGD	Whole-genome duplication
TD	Tandem duplication
PD	Proximal duplication
TRD	Transposed duplication
DSD	Dispersed duplication
PCA	Principal Component Analysis
FPKM	Fragments Per kb per Million reads
DEG	Differentially expressed genes

Appendix A

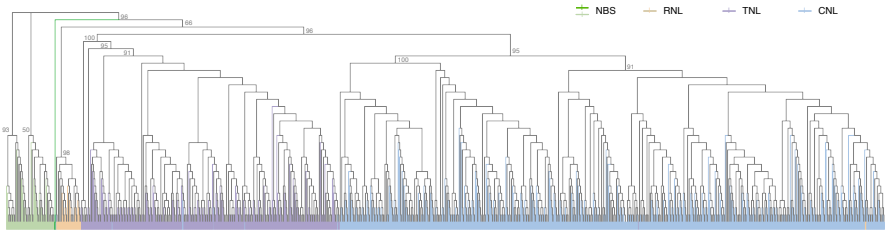

Figure A1. Phylogeny of NBS domains of NLRs. NLRs of three subclasses (RNL, TNL, CNL) of papaya and nine selected species were used to construct the NBS domain tree, in which NBS genes from Rhodophyta were used as the outgroup. The dark green indicates the two NBS genes used as the outgroup in Figure 4A. The numbers at the nodes mean branch support values, which are evaluated by the Shimodaira–Hasegawa test.

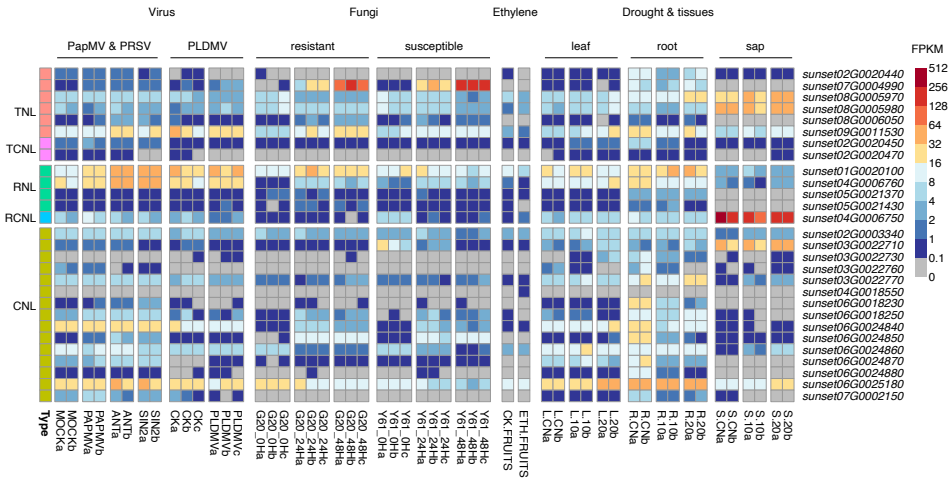

Figure A2. FPKM values of NLRs in papaya under biotic and abiotic stresses. ANT indicates 'Antagonism' samples, which are inoculated with simultaneous PapMV and PRSV [47]. SIN2 indicates 'Sinergism2' samples, which are inoculated with stepwise PapMV and PRSV [47].

References

1. Ming, R.; Hou, S.; Feng, Y.; Yu, Q.; Dionne-Laporte, A.; Saw, J.H.; Senin, P.; Wang, W.; Ly, B.V.; Lewis, K.L.T.; et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **2008**, *452*, 991–996. [CrossRef]
2. Yue, J.; VanBuren, R.; Liu, J.; Fang, J.; Zhang, X.; Liao, Z.; Wai, C.M.; Xu, X.; Chen, S.; Zhang, S.; et al. SunUp and Sunset genomes revealed impact of particle bombardment mediated transformation and domestication history in papaya. *Nat. Genet.* **2022**, *54*, 715–724. [CrossRef]
3. Shen, Y.H.; Lu, B.G.; Feng, L.; Yang, F.Y.; Geng, J.J.; Ming, R.; Chen, X.J. Isolation of ripening-related genes from ethylene/1-MCP treated papaya through RNA-seq. *BMC Genom.* **2017**, *18*, 671. [CrossRef]
4. Chan-León, A.C.; Estrella-Maldonado, H.; Dubé, P.; Ortiz, G.F.; Espadas-Gil, F.; May, C.T.; Prado, J.R.; Desjardins, Y.; Santamaría, J.M. The high content of β-carotene present in orange-pulp fruits of *Carica papaya* L. is not correlated with a high expression of the CpLCY-β2 gene. *Food Res. Int.* **2017**, *100*, 45–56. [CrossRef]
5. Liu, J.; Sharma, A.; Niewiara, M.J.; Singh, R.; Ming, R.; Yu, Q. Papain-like cysteine proteases in *Carica papaya*: Lineage-specific gene duplication and expansion. *BMC Genom.* **2018**, *19*, 26. [CrossRef] [PubMed]
6. Liu, Y.; Huang, C.C.; Wang, Y.; Xu, J.; Wang, G.; Bai, X. Biological evaluations of decellularized extracellular matrix collagen microparticles prepared based on plant enzymes and aqueous two-phase method. *Regen. Biomater.* **2021**, *8*, rbab002. [CrossRef]
7. Hafid, K.; John, J.; Sayah, T.M.; Domínguez, R.; Becila, S.; Lamri, M.; Dib, A.L.; Lorenzo, J.M.; Gagaoua, M. One-step recovery of latex papain from *Carica Papaya* using three phase partitioning and its use as milk-clotting and meat-tenderizing agent. *Int. J. Biol. Macromol.* **2020**, *146*, 798–810. [CrossRef]
8. Lobato-Gómez, M.; Hewitt, S.; Capell, T.; Christou, P.; Dhingra, A.; Girón-Calva, P.S. Transgenic and genome-edited fruits: Background, constraints, benefits, and commercial opportunities. *Hortic. Res.* **2021**, *8*, 166. [CrossRef] [PubMed]
9. Wu, Z.; Mo, C.; Zhang, S.; Li, H. Characterization of papaya ringspot virus isolates infecting transgenic papaya 'Huanong No. 1' in South China. *Sci. Rep.* **2018**, *8*, 8206. [CrossRef] [PubMed]
10. Wang, W.; Feng, B.; Zhou, J.M.; Tang, D. Plant immune signaling: Advancing on two frontiers. *J. Integr. Plant Biol.* **2020**, *62*, 2–24. [CrossRef]
11. Andolfo, G.; Dohm, J.C.; Himmelbauer, H. Prediction of NB-LRR resistance genes based on full-length sequence homology. *Plant J. Cell Mol. Biol.* **2022**, *110*, 1592–1602. [CrossRef]
12. Jones, J.D.G.; Dangl, J.L. The plant immune system. *Nature* **2006**, *444*, 323–329. [CrossRef]
13. Dangl, J.L.; Horvath, D.M.; Staskawicz, B.J. Pivoting the plant immune system from dissection to deployment. *Science* **2013**, *341*, 746–751. [CrossRef] [PubMed]
14. Lolle, S.; Stevens, D.; Coaker, G. Plant NLR-triggered immunity: From receptor activation to downstream signaling. *Curr. Opin. Immunol.* **2020**, *62*, 99–105. [CrossRef] [PubMed]
15. Delaux, P.M.; Schornack, S. Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science* **2021**, *371*, eaba6605. [CrossRef]
16. 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. [CrossRef]
17. Liu, Y.; Zeng, Z.; Zhang, Y.M.; Li, Q.; Jiang, X.M.; Jiang, Z.; Tang, J.H.; Chen, D.; Wang, Q.; Chen, J.Q.; et al. An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. *Mol. Plant* **2021**, *14*, 2015–2031. [CrossRef] [PubMed]
18. Shao, Z.Q.; Xue, J.Y.; Wu, P.; Zhang, Y.M.; Wu, Y.; Hang, Y.Y.; Wang, B.; Chen, J.Q. Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant Physiol.* **2016**, *170*, 2095–2109. [CrossRef]
19. Li, Q.; Jiang, X.M.; Shao, Z.Q. Genome-Wide Analysis of NLR Disease Resistance Genes in an Updated Reference Genome of Barley. *Front. Genet.* **2021**, *12*, 694682. [CrossRef]
20. Porter, B.W.; Paidi, M.; Ming, R.; Alam, M.; Nishijima, W.T.; Zhu, Y.J. Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Mol. Genet. Genom. MGG* **2009**, *281*, 609–626. [CrossRef]
21. Ngou, B.P.M.; Heal, R.; Wyler, M.; Schmid, M.W.; Jones, J.D.G. Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *Nat. Plants* **2022**, *8*, 1146–1152. [CrossRef]
22. Yang, M.; Zhou, C.; Yang, H.; Kuang, R.; Huang, B.; Wei, Y. Identification of WRKY transcription factor genes in papaya and response of their expresion to *Colletotrichum brevisporum* infection. *J. Northwest A F Univ. (Nat. Sci. Ed.)* **2022**, *50*, 127–138+154. [CrossRef]
23. Yang, M.; Zhou, C.; Yang, H.; Kuang, R.; Liu, K.; Huang, B.; Wei, Y. Comparative transcriptomics and genomic analyses reveal differential gene expression related to *Colletotrichum brevisporum* resistance in papaya (*Carica papaya* L.). *Front. Plant Sci.* **2022**, *13*, 1038598. [CrossRef] [PubMed]
24. Li, P.; Quan, X.; Jia, G.; Xiao, J.; Cloutier, S.; You, F.M. RGAugury: A pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genom.* **2016**, *17*, 852. [CrossRef]
25. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [CrossRef]
26. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef] [PubMed]

27. Lupas, A.; Van Dyke, M.; Stock, J. Predicting coiled coils from protein sequences. *Science* **1991**, *252*, 1162–1164. [CrossRef] [PubMed]
28. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [CrossRef]
29. Kumar, S.; Suleski, M.; Craig, J.M.; Kasprowicz, A.E.; Sanderford, M.; Li, M.; Stecher, G.; Hedges, S.B. TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol. Biol. Evol.* **2022**, *39*, msac174. [CrossRef]
30. Zhang, L.; Chen, F.; Zhang, X.; Li, Z.; Zhao, Y.; Lohaus, R.; Chang, X.; Dong, W.; Ho, S.Y.W.; Liu, X.; et al. The water lily genome and the early evolution of flowering plants. *Nature* **2020**, *577*, 79–84. [CrossRef]
31. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* **2019**, *20*, 38. [CrossRef]
32. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef] [PubMed]
33. Gao, Y.; Wang, W.; Zhang, T.; Gong, Z.; Zhao, H.; Han, G.Z. Out of Water: The Origin and Early Diversification of Plant R-Genes. *Plant Physiol.* **2018**, *177*, 82–89. [CrossRef]
34. Nguyen, N.P.D.; Mirarab, S.; Kumar, K.; Warnow, T. Ultra-large alignments using phylogeny-aware profiles. *Genome Biol.* **2015**, *16*, 124. [CrossRef]
35. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
36. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641–1650. [CrossRef] [PubMed]
37. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS ONE* **2010**, *5*, e9490. [CrossRef] [PubMed]
38. Shimodaira, H.; Hasegawa, M. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Mol. Biol. Evol.* **1999**, *16*, 1114. [CrossRef]
39. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [CrossRef]
40. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [CrossRef]
41. Zhang, Z.; Xiao, J.; Wu, J.; Zhang, H.; Liu, G.; Wang, X.; Dai, L. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **2012**, *419*, 779–781. [CrossRef] [PubMed]
42. Wang, D.; Zhang, Y.; Zhang, Z.; Zhu, J.; Yu, J. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* **2010**, *8*, 77–80. [CrossRef]
43. Yang, Z.; Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [CrossRef]
44. Nystedt, B.; Street, N.R.; Wetterbom, A.; Zuccolo, A.; Lin, Y.C.; Scofield, D.G.; Vezzi, F.; Delhomme, N.; Giacomello, S.; Alexeyenko, A.; et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **2013**, *497*, 579–584. [CrossRef] [PubMed]
45. Liu, Y.; Wang, S.; Li, L.; Yang, T.; Dong, S.; Wei, T.; Wu, S.; Liu, Y.; Gong, Y.; Feng, X.; et al. The Cycas genome and the early evolution of seed plants. *Nat. Plants* **2022**, *8*, 389–401. [CrossRef]
46. Belfield, E.J.; Brown, C.; Ding, Z.J.; Chapman, L.; Luo, M.; Hinde, E.; Van Es, S.W.; Johnson, S.; Ning, Y.; Zheng, S.J.; et al. Thermal stress accelerates *Arabidopsis thaliana* mutation rate. *Genome Res.* **2021**, *31*, 40–50. [CrossRef]
47. Vargas-Mejía, P.; Vega-Arreguín, J.; Chávez-Calvillo, G.; Ibarra-Laclette, E.; Silva-Rosales, L. Differential Accumulation of Innate- and Adaptive-Immune-Response-Derived Transcripts during Antagonism between Papaya Ringspot Virus and Papaya Mosaic Virus. *Viruses* **2020**, *12*, 230. [CrossRef]
48. Gamboa-Tuz, S.D.; Pereira-Santana, A.; Zamora-Briseño, J.A.; Castano, E.; Espadas-Gil, F.; Ayala-Sumuano, J.T.; Keb-Llanes, M.Á.; Sanchez-Teyer, F.; Rodríguez-Zapata, L.C. Transcriptomics and co-expression networks reveal tissue-specific responses and regulatory hubs under mild and severe drought in papaya (*Carica papaya* L.). *Sci. Rep.* **2018**, *8*, 14539. [CrossRef] [PubMed]
49. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef] [PubMed]
50. Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*, 1650–1667. [CrossRef] [PubMed]
51. Anders, S.; McCarthy, D.J.; Chen, Y.; Okoniewski, M.; Smyth, G.K.; Huber, W.; Robinson, M.D. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **2013**, *8*, 1765–1786. [CrossRef]
52. Liu, S.; Wang, Z.; Zhu, R.; Wang, F.; Cheng, Y.; Liu, Y. Three Differential Expression Analysis Methods for RNA Sequencing: Limma, EdgeR, DESeq2. *J. Vis. Exp.* **2021**, *175*, e62528. [CrossRef]
53. Zhang, B.; Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*, 17. [CrossRef] [PubMed]
54. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef] [PubMed]
55. Langfelder, P.; Luo, R.; Oldham, M.C.; Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **2011**, *7*, e1001057. [CrossRef] [PubMed]

56. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
57. 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]
58. Wang, R.; Sperry, A.O. Identification of a novel Leucine-rich repeat protein and candidate PP1 regulatory subunit expressed in developing spermatids. *BMC Cell Biol.* **2008**, *9*, 9. [CrossRef]
59. DeVaul, N.; Wang, R.; Sperry, A.O. PPP1R42, a PP1 binding protein, regulates centrosome dynamics in ARPE-19 cells. *Biol. Cell* **2013**, *105*, 359–371. [CrossRef]
60. Ma, S.; Lapin, D.; Liu, L.; Sun, Y.; Song, W.; Zhang, X.; Logemann, E.; Yu, D.; Wang, J.; Jirschitzka, J.; et al. Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* **2020**, *370*, eabe3069. [CrossRef]
61. Antunes Carvalho, F.; Renner, S.S. A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. *Mol. Phylogenetics Evol.* **2012**, *65*, 46–53. [CrossRef] [PubMed]
62. Rockinger, A.; Sousa, A.; Carvalho, F.A.; Renner, S.S. Chromosome number reduction in the sister clade of *Carica Papaya* with concomitant genome size doubling. *Am. J. Bot.* **2016**, *103*, 1082–1088. [CrossRef] [PubMed]
63. Cui, H.; Tsuda, K.; Parker, J.E. Effector-triggered immunity: From pathogen perception to robust defense. *Annu. Rev. Plant Biol.* **2015**, *66*, 487–511. [CrossRef] [PubMed]
64. Wang, J.; Song, W.; Chai, J. Structure, biochemical function, and signaling mechanism of plant NLRs. *Mol. Plant* **2023**, *16*, 75–95. [CrossRef] [PubMed]
65. Steuernagel, B.; Witek, K.; Krattinger, S.G.; Ramirez-Gonzalez, R.H.; Schoonbeek, H.J.; Yu, G.; Baggs, E.; Witek, A.I.; Yadav, I.; Krasileva, K.V.; et al. The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol.* **2020**, *183*, 468–482. [CrossRef]
66. Toda, N.; Rustenholz, C.; Baud, A.; Le Paslier, M.C.; Amselem, J.; Merdinoglu, D.; Faivre-Rampant, P. NLGenomeSweeper: A Tool for Genome-Wide NBS-LRR Resistance Gene Identification. *Genes* **2020**, *11*, 333. [CrossRef]
67. Friis, E.M.; Pedersen, K.R.; Crane, P.R. The emergence of core eudicots: New floral evidence from the earliest Late Cretaceous. *Proc. Biol. Sci.* **2016**, *283*, 20161325. [CrossRef]
68. Guo, X.; Fang, D.; Sahu, S.K.; Yang, S.; Guang, X.; Folk, R.; Smith, S.A.; Chanderbali, A.S.; Chen, S.; Liu, M.; et al. Chloranthus genome provides insights into the early diversification of angiosperms. *Nat. Commun.* **2021**, *12*, 6930. [CrossRef]
69. Li, Z.; Defoort, J.; Tasdighian, S.; Maere, S.; Van de Peer, Y.; De Smet, R. Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms. *Plant Cell* **2016**, *28*, 326–344. [CrossRef]
70. Glenfield, C.; Innan, H. Gene Duplication and Gene Fusion Are Important Drivers of Tumourigenesis during Cancer Evolution. *Genes* **2021**, *12*, 1376. [CrossRef]
71. Cannon, S.B.; Mitra, A.; Baumgarten, A.; Young, N.D.; May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* **2004**, *4*, 10. [CrossRef]
72. Yang, F.S.; Nie, S.; Liu, H.; Shi, T.L.; Tian, X.C.; Zhou, S.S.; Bao, Y.T.; Jia, K.H.; Guo, J.F.; Zhao, W.; et al. Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat. Commun.* **2020**, *11*, 5269. . [CrossRef] [PubMed]
73. Wang, K.; de la Torre, D.; Robertson, W.E.; Chin, J.W. Programmed chromosome fission and fusion enable precise large-scale genome rearrangement and assembly. *Science* **2019**, *365*, 922–926. [CrossRef] [PubMed]
74. Ling, H.Q.; Ma, B.; Shi, X.; Liu, H.; Dong, L.; Sun, H.; Cao, Y.; Gao, Q.; Zheng, S.; Li, Y.; et al. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* **2018**, *557*, 424–428. [CrossRef] [PubMed]
75. Tamborski, J.; Krasileva, K.V. Evolution of Plant NLRs: From Natural History to Precise Modifications. *Annu. Rev. Plant Biol.* **2020**, *71*, 355–378. [CrossRef]
76. Konno, K.; Hirayama, C.; Nakamura, M.; Tateishi, K.; Tamura, Y.; Hattori, M.; Kohno, K. Papain protects papaya trees from herbivorous insects: Role of cysteine proteases in latex. *Plant J. Cell Mol. Biol.* **2004**, *37*, 370–378. [CrossRef] [PubMed]