 doi:10.20944/preprints202001.0048.v1

The NBS-LRR gene class is a small family in *Cucurbita pepo*

Belén Román[1], Pedro Gómez[2], Belén Picó[3], José V. Die[4*]

[1] Department of Plant Breeding and Biotechnology, IFAPA, Córdoba, Spain
[2] Department of Plant Breeding and Biotechnology, IFAPA, Almería, Spain
[3] Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València Valencia, Spain
[4] Department of Genetics – ETSIAM, University of Cordoba, Córdoba, Spain

*Correspondence:
jose.die@uco.es

Abstract

Although *Cucurbita pepo* is one of the most variable species of the plant kingdom, Zucchini morphotype has undergone intensive breeding that has led to a narrow genetic base making the crop vulnerable to pest and diseases. This vulnerability makes the knowledge of resistance genes of utmost importance. In this study, a data mining search of Zucchini summer squash genome database was conducted to identify and annotate members of the NBS-encoding gene family. In order to characterize the retrieved genes in detail, they have been studied in the bases of phylogenetic relationships, structural diversity, conserved protein motifs, gene duplications and promoter region analysis. Our study shows that the NBS-encoding gene family is relatively small in Zucchini (34 members, which are separated into non-TIR- and TIR-NBS-LRR subfamilies) with a significantly lower number of R-genes than in other species. Duplications have not played a major role in the expansion of this type of genes in *C. pepo*. Among the *cis*-regulatory elements presented in these sequences, six motifs are over-represented. These elements were reported to be involved in pathogens or plant stress induced responses. These results will contribute to the identification, isolation and characterization of candidate R-genes, thereby providing insight into NBS gene family evolution in the species.

Keywords: *cis*-regulatory element; data mining; NBS-LRR resistance genes; Zucchini

 doi:10.20944/preprints202001.0048.v1

Introduction

Summer squash are the edible immature fruits of *Cucurbita pepo* L. Most of the widely grown commercial types belong to the elongated forms of three *C. pepo* ssp. *pepo* horticultural groups, Cocozelle, Vegetable marrow, and Zucchini. Selection for long-fruitedness in the subspecies *pepo* was first reported in Italy in the sixteenth century [1]. New cultivars were then developed, beginning with the Vegetable marrow, and followed by the Cocozelle group. Cultivars for the Zucchini group were probably developed in the late nineteenth century, which were mainly selected for their long, uniformly cylindrical fruits. This latter group has undergone intensive breeding in the United States and Europe [2] and nowadays its cultivars dominate both the squash market and the breeding efforts of seed companies. All this breeding effort, while providing a consistent product, also leads to a narrow genetic base. Although *C. pepo* is one of the most variable species of the plant kingdom, Zucchini is the least variable morphotype. Recently, Formisano et al. (2012) [3] using EST-derived SSR, reported higher genetic variability in Cocozelle and Vegetable marrow than Zucchini, which has a more recent origin. This narrow genetic base makes the crop vulnerable to existing pests and diseases, as well as those that could develop in the future. This vulnerability, along with the increasing interest in developing new, environmentally friendly resistant cultivars, makes the knowledge of the structural and functional mechanisms of resistance genes of outmost importance.

NBS-LRR proteins encoded by resistance (R) genes play an important role in the plant responses to pathogens. These proteins contain nucleotide binding sites (NBS) and leucine-rich repeat (LRR) domains. They can be broadly divided in two subclasses, the TIR subclass with a Toll/Interleukine-1 receptor domain in their N terminus and the non-TIR subclass with a coiled-coil (CC) domain instead [4,5]. The majority of the cloned plant genes conferring resistance to pathogens belong to this family (NBS-LRR). The number of R-genes in different plant genomes varies dramatically. Some genomes such as those of apple and wheat contain approximately 1000 R-genes [6,7]. However, there are relatively few R-gene lineages in case of Cucurbitaceae. The vast majority of them are present in the sequenced genomes of cucumber, melon and watermelon. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

Cucurbitaceae family also shows lower number of R-genes per lineage than other species [8]. According to these authors, the frequent loss of R-gene lineages and the deficient duplications in extant lineages have been proposed as the main causes of the scarcity of R-genes in Cucurbitaceae.

Among Cucurbitaceae species, it has been reported different association between NBS-LRR analogue genes and effective resistance to aphid, fungus or virus in melon, cucumber and Luffa [9–11]. Although the characterization of R-genes sequences and phylogenetic analyses have been carried out in some of the most important species of the Cucurbitaceae, this information is still lacking in Zucchini. Fortunately, in the last years a set of genomic tools has become available for Cucurbita genus breeders and researchers, including a transcriptome, a high quality saturated map and a TILLING platform and genome resequencing [12–16]. In addition to these tools, researchers at COMAV Bioinformatics & Genomics and Cucurbits Breeding groups have been driven to release the first draft of the Zucchini genome [15]. This achievement provides an opportunity to conduct the first study of the NBS-encoding gene superfamily at the genome level.

Results

NBS-encoding genes in *Cucurbita pepo*

NBS-encoding genes were identified in the Zucchini genome using a data mining approach through a reiterative process. In the first step, more than 200 and 500 NBS proteins from *Arabidopsis* and rice respectively [17–19] were used to retrieve potential NBS genes from Zucchini. Then, flanking regions of the blast hits were expanded 2,500-5,000bp. The resulting sequences were annotated and classified based on the features of their N-terminal and C-terminal protein structures. Finally, all sequences were used as query against the entire Zucchini genome to check for additional genes that had not been previously identified. In this way we could annotate 29, 3, 1 and 1 genes in each search. At each search step, we also identified a number of nonfunctional genes or pseudogenes based on frameshifts mutations or premature stop codons. These sequences were not considered for further analysis. Thus, the overall analysis revealed that the NBS-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

encoding gene family is composed of 34 members (Table 1). A total of six categories (CNL, TNL, TN, N, NL and RPW8-NL) were identified. Fifteen proteins were predicted to encode CNL domains, while eleven were identified as TNL. Three proteins show a domain similar to the RPW8 *Arabidopsis* powdery mildew resistance gene family [20]. Not all the genes showed a complete structure with both N-terminal domain (TIR, CC, or RPW8) and LRR domain. A total of 2 open reading frames lacked both N-terminal and LRR domains (2N) and 1 NBS gene did not encode an LRR domain (1TN).

Table1. Number genes encoding domains similar to NBS genes in *Cucurbita pepo*

Predicted protein domain	Letter code	Number of genes
TIR-NBS-LRR	TNL	11
CC-NBS-LRR	CNL	15
TIR-NBS	TN	1
NBS-LRR	NL	2
NBS	N	2
RPW8-NBS-LRR	RPW8-NL	3
Total NBS genes		**34**

The NBS genes belonging to the non-TNL subclass outnumbered those in the TNL subclass (1.45:1). A similar ratio has been found in the genome sequences of Cucurbitaceae members cucumber and melon [21,22] as well as in some legumes such as *M. truncatula* and *Glycine max* [23]. This distribution is contrary to what has been found in *Arabidopsis* where TNL sequences are more numerous (1:1.89) [19].

Moreover, blastp searches revealed that the 32 NBS-LRR proteins had the highest degree of identity with known R-proteins from other Cucurbitaceae crops (Supplemental Table S1). A number of proteins matched to cucumber (*Cucumis sativus* L.) and melon (*Cucumis melo* L.), the two species with a sequenced genome most closely related to summer squash [22,24,25]. Most of the TNL proteins showed homology to *C. melo* with the TMV resistance N-protein (XP_008459550) being the most frequent hit (5 TNL proteins). Conversely, the majority of CNL proteins showed homology to *C. sativus* sequences. The most common hit (5 CNL proteins) was the disease-resistance RGA4-like protein (XP_004169509).

 doi:10.20944/preprints202001.0048.v1

Phylogenetic analysis

In order to study the relationships among the Zucchini NBS-encoding genes, a ML phylogenetic tree was constructed based on the region between the P-loop and GLPL motifs. The NBS genes retrieved in this study were separated into non-TIR and TIR-NBS-LRR subfamilies (Figure 1). These sequences were distributed as follows: 20 showed amino-acid sequence similarity to the non-TIR type NBS and 11 to the TIR-type. The Zucchini NBS-LRR family consists of at least 5 distinct subfamilies. The non-TIR type sequences were grouped into 3 classes and the TIR-type sequences into 2 classes. These results are consistent with previous studies in other Cucurbitaceae crops such as cucumber [8,21], melon and watermelon [8,26]. The non-TIR group showed higher number of members than the TIR one, as reported in cucumber, melon, watermelon or pepper [8,21,27]. Each of these subgroups included different numbers of *C. pepo* resistance genes: 16, 2 and 2 in the case of non-TIR family and 5 and 6 in the case of TIR family.

Moreover, a total of 49 NBS sequences were used for phylogenetic tree construction: 31 sequences from Zucchini *C. pepo* retrieved in this study and 18 known NBS genes from three different Cucurbitaceae crops (*Cucumis melo, Cucumis sativus* and *Citrullus lanatus*) [8]. Considering this phylogenetic tree, a clear differentiation between CNL and TNL sequences was also found (Figure 2). The distribution of sequences was 29 in the non-TIR group and 20 in the TIR one, grouped in 5 distinct subgroups: 3 in the CNL family and 2 in the TNL family. Each of these subgroups included different numbers of sequences: 20 in the group with the higher number of sequences of CNL and 8 and 12 in the subgroups found in the TNL subfamily.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

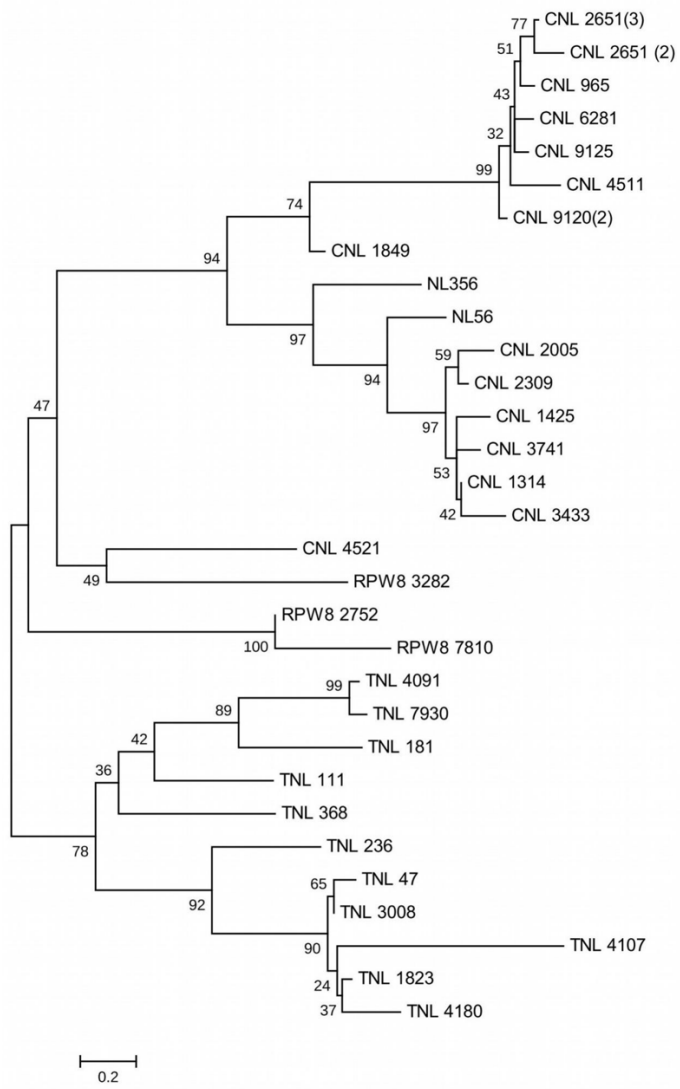


Figure 1. ML phylogenetic tree of retrieved sequences of Zucchini. The phylogenetic tree was constructed based on the translated protein region between the P-loop and GLPL motifs.

Duplication of NBS-LRR genes

To test if gene duplications have occurred in the summer squash genome during evolution, we blasted all predicted proteins against each other. Eight of the 34 NBS genes were identified as duplications. All 8 were subsequently divided into 2 families (Table 2). The maximum number of family members was five and the average number of family members was four. Then, we compared our data to those available in *Arabidopsis*, rice, *Brachypodium distachyon*, as well as other Cucurbitaceae. The results revealed that the average number of NBS members per multigene family (two or more members per family) was higher in summer squash than that of *Arabidopsis* (3.24). However, *C. pepo*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

shows a lower number of families than the other species. The percentage of multigene families is also lower (for example, the percentage of multigene families in *C. pepo* is half that of *Arabidopsis* or rice). However, it is more similar to the values shown by the other Cucurbitaceae members (Figure 3). This may suggest that duplications of NBS genes have not played a major role in the expansion of NBS-encoding genes in *C. pepo*.

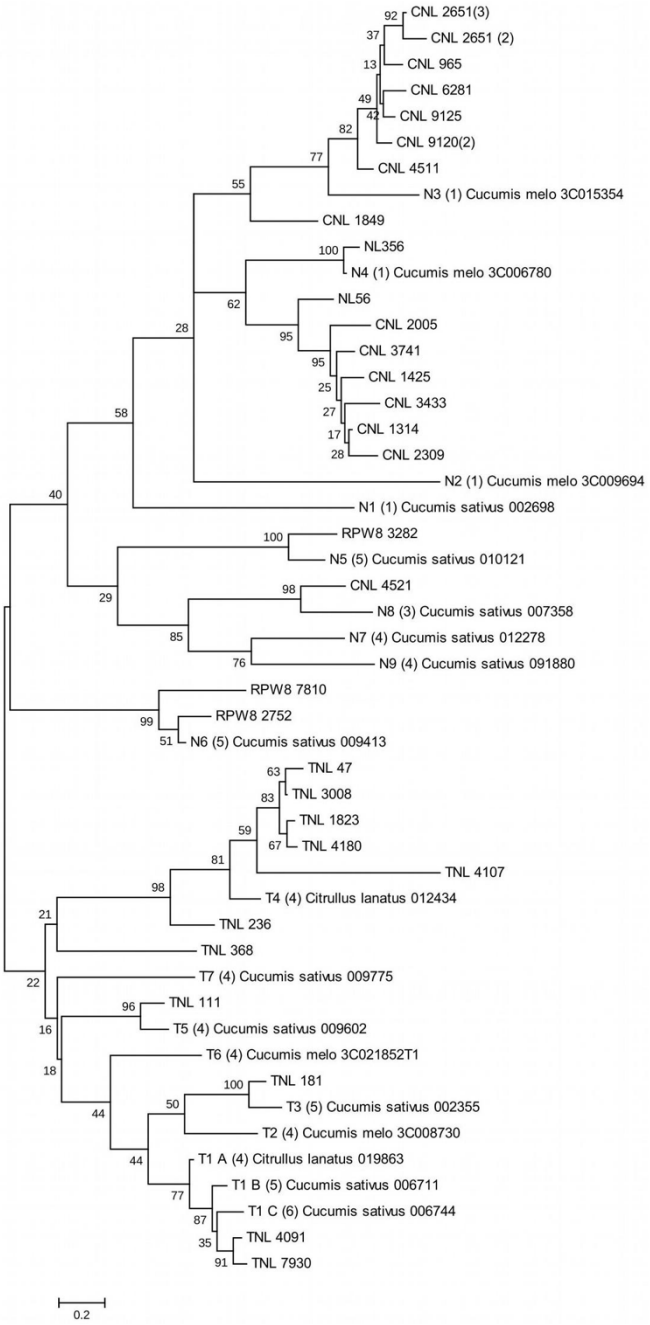


Figure 2: ML phylogenetic tree from the 31 Zucchini *C. pepo* sequences from Zucchini retrieved in this study and 18 known NBS genes from three different Cucurbitaceae crops.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

Table 2. Duplication of NBS-LRR genes in different species

Organization	*C. pepo*	*C. sativus*[1]	*C. melo*[1]	*Arabidopsis*[2]	Rice[2]	*B. distachyon*[2]
Single-genes	26	42	67	93	216	77
Multigenes	8	24	12	81	248	49
Number of families	2	10	6	25	93	20
Maximal family members	5	6	3	7	10	7
Average members per family	4.0	2.4	2	3.24	2.67	2.45
Multigenes/single-gene families	0.31	0.57	0.18	0.87	1.14	0.64
Percentage of multigene families	23.5%	36.4%	15.2%	46.6%	53.4%	38.9%

[1] Data from Lin et al. (2013) [8]
[2] Data from Tan and Wu (2012) [28]

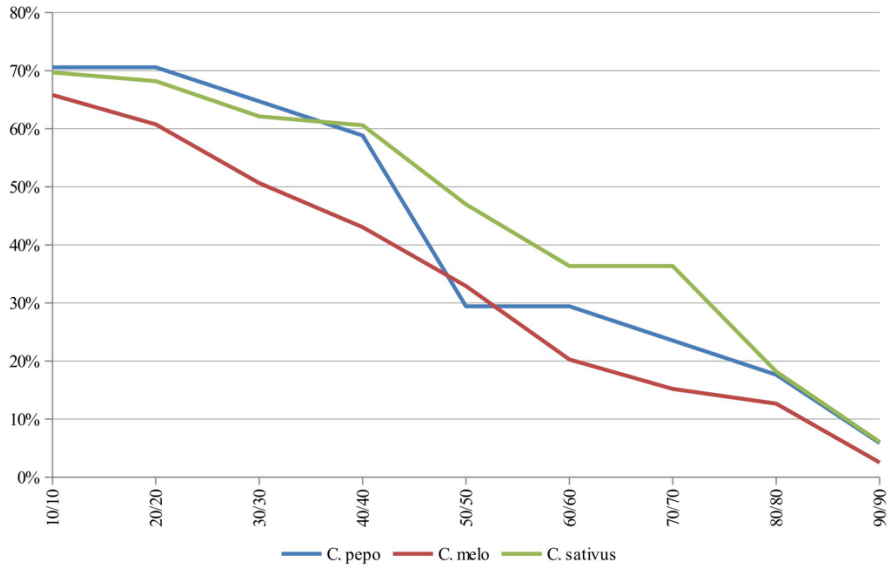


Figure 3: Multigene families across different similarity/coverage thresholds and across other Cucurbitaceae species. The range 60-80 shows intermediate values for *C. pepo* between *C. sativus* and *C. melo*.

Gene structure

The exon/intron positions and phases of the CNL and TNL genes were further analyzed with both coding sequences and genomic sequences. The first structural analysis revealed some likely inaccuracies in automated gene predictions and annotations, thereby affecting the number and size of anomalous exons/introns. Introns occurring in non-conserved locations were reanalyzed by blastx comparisons. Misannotations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

differing from the automated output were assigned to 9 genes. Their annotations were corrected manually. Thus, this step was considered necessary and relevant to determine the final composition of our dataset. Following the corrections, it was found that eleven NBS-LRR genes had zero introns, all of which are CNL genes. More than half of the NBS-LRR genes (15 sequences) had zero to two introns. The TNL genes showed two to four introns. The mean exon number of TNL class (4.27) was higher than that of the CNL class (1.60; P=2.219e^{-07}). The number of exons predicted in all these genes ranged from one to five, with a mean of 2.73 (Supplemental Figure S1). Therefore, the number of introns in R-genes of Zucchini was lower than the number reported in *M. domestica* [7], *B. rapa* [29], *A.thaliana* [19], *V. vinifera* [30,31] and *O. sativa* [18,32] but higher than *P. trichocarpa* [33].

Architectural diversity and motif characterization of *C. pepo* NBS-LRR genes

In order to understand the fine structure of the NBS-LRR-encoding genes in Zucchini, both the TIR- and CC-NBS subfamilies were analyzed separately. Each of these two subfamilies were separated into three parts (N-terminal region, NBS domain, and C-terminal region) and subjected to protein domain and motif analysis that revealed additional features. Conserved motif structures of predicted R-genes were analyzed using the Multiple Expectation Maximization for Motif elicitation suite tool.

CC-NBS-LRR family

On average, the N-terminal region of the CNL proteins spanned ~170 amino acids from the start of the coding region to the beginning of the P-loop of the NBS domain. The presence of an N-terminal CC domain has been identified as a characteristic motif in the N terminus of the CNL proteins. There was strong evidence for a CC motif in 15 proteins. The predicted CC motif was positioned at ~35-55 amino acids from the N terminus corresponding to the motive sequence C3 that was present in all sequences (Supplemental Figure S2). We were able to identify ten distinct motive sequences (C1-C10) in the N-terminal domain from the 15 CNL proteins using MEME search tool. Up to eight motives were present in at least 13 proteins. Although these motives showed less conservation than the NBS domain, it was possible to identify several patterns

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

associated with the subgroups defined by the clades identified in the phylogenetic and intron/exon analyses. For example, motive C8 was not detected in sequences with more than 2 exons whereas motives C3, C6, C7, C9 and C10 were present in almost all members of the CNL family. Moreover, these motives were located on almost identical positions within each sequence. Motive C5 was only detected in the first CNL subclade that started with motive C1 (positions 1-15 aa). The second CNL subclade started with motive C2 (positions 1-15 aa; Supplemental Table S2 and Supplemental Figure S3). The eight major conserved motifs (P-loop, RNBS-A-non-TIR, kinase-2, RNBS-B, RNBS-C, GLPL, RNBS-D-non-TIR and MHDV) that determine the structural character of the NBS domain were found in the CC-NBS-LRR members (Figure 4). The MHDV motif is considered to represent the C-terminal end of the NBS [19]. In total, the eight NBS motifs from P-loop to MHDV spanned ~315 amino acids in the CNL proteins. The RNBS-D-non-TIR (*E*-value: $3.5e^{-146}$) showed the highest conservation, followed by P-loop (E-value: $1.6e^{-120}$) and kinase-2 (*E*-value: $1.3e^{-117}$). With respect to the C-terminal region, we analyzed all amino acids encoded immediately 3' to the encoded MHDV motif. The LRR and C-terminal region was variable in sequence and size, but it registered 500-600 amino acids for most of the predicted proteins. The repeats of the genes were imperfects showing high variability (data not shown).

TIR-NBS-LRR family

The TIR motives of the TNL proteins included ~181 to 186 amino acids and the profile score analysis indicated the presence of ten conserved motives (the least statistically significant motive present in the 11 sequences is TIR4, *E*-value=$9.3e^{-30}$). TIR1 and TIR7 motives were highly conserved, whereas TIR10 motive was present only in five TNL proteins. In general, the TIR motives of the TNL showed same sequence diversity as the N-terminal region of CNL proteins (Supplemental Table S2). The NBS domain spanned ~290 amino acids. The P-loop with the signature sequence GMGGIGKTTxxxxxY showed the highest conservation (*E*-value: $7.7e^{-96}$) followed by RNBS-B (*E*-value: $1.2e^{-72}$) and motive 2 (*E*-value: $3.0e^{-57}$). The eight conserved motifs were present in the TNL subfamily members with the exception of scaffold4107 that has one truncated NBS domain (Figure 5). The RNBS-C showed the highest level of diversity among the eight conserved motifs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

The C-terminal region was the least conserved domain and, as in the CNL sequences, that region showed high variability. Motif L4 with the signature sequence HMC[HY]SNLKQ[FL]W[HQ]GEK was present only in one of the TNL subclades. Three motives (L1-L3) appear between ~20-90 amino acids distance from the 3′ MHDV domain in all the TNL sequences except for scaffold4107. These motives might be essential to maintaining the function of the LRR domain (Supplemental Table S3).

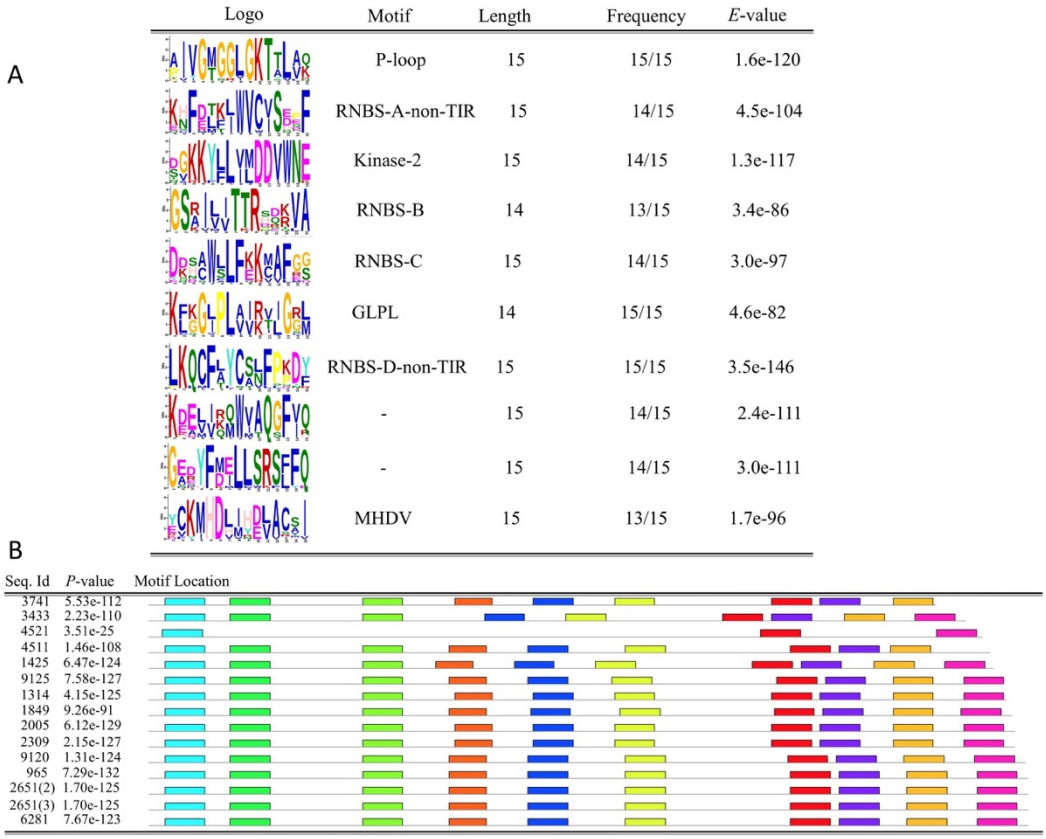


Figure 4. Motif sequences identified through MEME analyses within the NBS domain of 15 CNL genes. **A.** Logo of conserved motifs in the NBS region along their *E*-values. **B.** Schematic diagram of conserved amino acid motifs for each sequence. Colored boxes represent the conserved motifs from P-loop to MHDV. The combined *P*-value to all the motifs for a given sequence is shown.

In this study, the domain analysis of *C. pepo* NBS genes, showed the majority of the three typical conserved domains reported by Cannon et al. (2002) [34] : a variable N-terminal domain CC/TIR, an NBS domain, and a more variable tandem array of approximately 10 to 40 short LRR (leucine-rich-repeat motifs). Nevertheless, a group of 8 proteins related to the TNL or CNL subfamilies had less typical arrangements and lacked the full complements of domains. These included 1 TN (TIR-NBS), 2 NL (NBS-LRR), 2 N (NBS)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

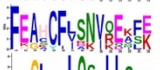
and 3 RPW8-NL proteins that presented a RPW8 motif in the N-terminal domain (lacking the CC or TIR ones). These same atypical families have been found in cucumber [21]. The function of these proteins is not known, but they seem to have the potential to act as adaptors or regulators of TNL and CNL proteins [35]. Two of these domains of plant R proteins, NBS and LRR, seem to be the most crucial in the pathogen recognition process and the activation of signal transduction in the response to pathogen attack. The two domains of the R-genes have different roles and are therefore, subject to different selection pressures. The NBS domain has a role in signal transduction and is relatively conserved. The LRR domain has a role in recognition of the pathogen, so it is expected to be under diversifying selection to adapt to newly emerged pathogen races [36]. The results of this study regarding the NBS domain are consistent with previous reports that have used the last residue of the kinase-2 motif to predict whether they belong to the TNL (D, aspartate) or CNL subfamilies (W, tryptophan; [4,5]. We found in the summer squash NBS proteins the "W" residue for all the CNL members, whereas most of the kinase-2 motifs of the TNL proteins showed a "N" (asparagine) followed by the D residue. There are also other domains that have been reported as conserved in the majority of the NBS region of R-genes [19] that are also present in both types of Zucchini proteins, CNL and TNL: the hydrophobic domain RNBS (with three motifs RNBS-A, B and C), kinase 2, P-loop, GLPL and MHDV.

Analysis of *cis*-elements of the promoters

To investigate whether we could identify potentially conserved transcription factor binding sites in summer squash, we collected the promoter genomic sequences in 2kb windows upstream of 31 predicted NBS-LRR genes. In a first attempt to analyze the dataset, we screened the sequences against the collection of 27,457 *Arabidopsis* promoters. Six CREs, implicated in either response to pathogens or plant stress, were overrepresented. The characterized CREs included: WBOX, DRE, ERE, ATHB5, Gbox and RAV1-B (Table 3). Further statistical analysis indicated an enriched content of the analyzed CREs in NBS-LRR gene promoters of summer squash. None of the promoters contained all six CREs. However, two promoters contained 5 of the six CREs. Moreover, eleven promoters contained at least 3 of the six CREs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

A

Logo	Motif	Length	Frequency	*E*-value
	P-loop	15	11/11	7.7e-96
	-	15	11/11	3.0e-57
	RNBS-A-TIR	15	11/11	1.2e-40
	-	15	5/11	2.8e-27
	Kinase-2	15	11/11	9.0e-47
	RNBS-B	15	11/11	1.2e-72
	RNBS-C	15	10/11	3.6 e-37
	GLPL	15	10/11	5.8e-51
	RNBS-D	15	10/11	1.1e-51
	MHDV	15	10/11	5.6e-52

B

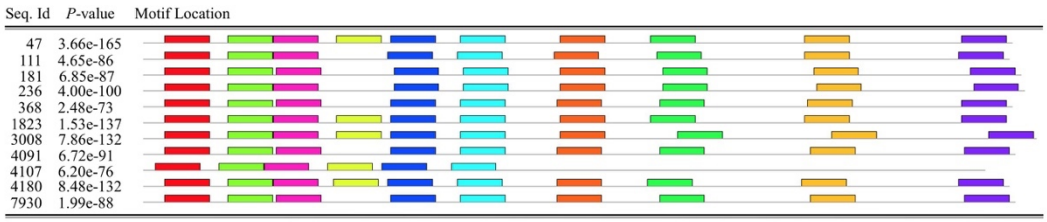


Figure 5. Motif sequences identified through MEME analyses within the NBS domain of 11 TNL genes. **A.** Logo of conserved motifs in the NBS region along their *E*-values. **B.** Schematic diagram of conserved amino acid motifs for each sequence. Colored boxes represent the conserved motifs from P-loop to MHDV. The combined *P*-value to all the motifs for a given sequence is shown.

WBOX elements are the most numerous, existing in 30 of the 31 NBS encoding genes tested, and averaging 3.58 motifs per gene; 10 promoters contained at least 4 predicted WBOX elements. Specifically, the average is 4.9 for the TNL and 2.6 for the CNL subfamilies. In contrast, the average of other element types is 0.19 (ATHB5) and 0.13 (ERE). Although only six and four promoters contained the regulatory elements ATHB5 and ERE, respectively, these two CREs were clearly overrepresented in our query set. The promoter sequences from two CNL genes (scaffold2309 and scaffold 9125) contained both, ATHB5 and ERE element.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

Table 3. Conserved cis-regulatory elements in summer squash NBS-LRR promoters

CRE	Motif	Query[1]	Promoters observed[2]	Motifs observed[3]	Motifs Expected[4]	Enrichment Factor[5]	*P*-value[6]
WBOX	TGACY	31	30	111	102.18	1.41	0.029
DRE	RCCGAC	31	16	22	10.03	2.19	0.014
ERE	AGCCGCC	31	4	4	0.60	6.65	0.005
ATHB5	CAATNATTG	31	4	6	2.23	2.69	0.012
Gbox	CACATG	31	13	17	11.54	1.47	0.022
RAV1-B	CACCTG	31	9	12	6.49	1.85	0.017

[1] Total number of promoters in query set.
[2] Number of promoters observed containing at least 1 copy of the motif.
[3] Total number of motifs in query set.
[4] Total number of motifs expected to occur by chance/2.0kb promoter based on nucleotide frequency in 31 summer squash promoter regions.
[5] Number of motifs observed divided by the number of motifs expected to occur by chance.
[6] Probabilities based on 2,000 Monte Carlo simulations.

WBOX elements are the most numerous, existing in 30 of the 31 NBS encoding genes tested, and averaging 3.58 motifs per gene; 10 promoters contained at least 4 predicted WBOX elements. Specifically, the average is 4.9 for the TNL and 2.6 for the CNL subfamilies. In contrast, the average of other element types is 0.19 (ATHB5) and 0.13 (ERE). Although only six and four promoters contained the regulatory elements ATHB5 and ERE, respectively, these two CREs were clearly overrepresented in our query set. The promoter sequences from two CNL genes (scaffold2309 and scaffold 9125) contained both, ATHB5 and ERE element.

Promoter sequences of the analyzed genes were divided into 200 nucleotide fragments and the content of the six analyzed CREs was calculated. An increase in density of the fragments was identified at a distance of ~1500 and ~900 bp from the start codon (Supplemental Figure S4).

Gene mapping and annotation

Physical chromosomal positions were established for 28 (~90%) of the 31 NBS-LRR identified genes (Supplemental Table S4). These genes were located in 11 chromosomes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

(of the 20 chromosomes in *C. pepo*): chr2, 3, 4, 6, 7, 8, 10, 11, 12, 16 and 18 (Supplemental Figure S5). The remaining three genes mapped in chromosome 00 (consisting of unanchored scaffolds). The genes of the non-TIR family were present on 10 zucchini chromosomes, with at least one representative, being chr8 the one with more genes, although they were not clustered, but distributed in three separated regions of the chromosome. The distribution of the genes of the TIR-NBS-LRR family genes was more limited, with genes on 5 chromosomes. Chr6 had the highest number of NBS genes (8, ~28% of mapped genes) with most of them arranged in a cluster of 7 genes.

The class 1 of non-TIR genes were grouped in Figure 1 in two main clusters, one grouping 8 CNL genes and the second 2 NL and 6 CNL genes. The first group of CNL genes had significant blastn hits with six *C. pepo* genes annotated at http://cucurbitgenomics.org/ (Supplemental Table S4), five as NBS-LRR type disease resistance protein, and one as NB-ARC and LRR domains containing disease resistance protein, most of them showed also significant blastx hits at the NCBI non-redundant protein database with *C. pepo* genes annotated as putative disease resistance gene analogs, RGA1, RGA3 or RGA4, a family known to be involved in plants resistance to pathogens [37]. The highest query cover and identity (99 to 100%) was found for CNL 265 (XP_023538754) and CNL9125 (XP_023539563), both located in chr8. Most of the second group of class 1 non-TIR genes were not annotated at http://cucurbitgenomics.org/, but they have also significant blastx hits at NCBI nr protein database with *C. pepo* putative RGA3 and RGA2 proteins, having NL 56 (XP_023527716) and CNL3741 (XP_023535482), located in chr 3 and 6 respectively, the best hits (99-100%). The class 2 and 3 of non-TIR genes, including one CNL and three RPW8-NL genes, dispersed in chr2, 6, 10, and 16, which had significant blastn hits with three *C. pepo* genes annotated as CC-NBS-LRR and NBS-LRR resistance proteins. The blastx against the nr protein database provided significant blast hit (97-100% query cover and 94-100% identity) with *C. pepo* proteins similar to At4g27220, At5g66900 and At4g33300 genes. The NBS-LRR gene At4g27220 is the orthologue of the cotton gene GbaNA1 that confers resistance to Verticillium wilt caused by *Verticillium dahliae* [38]. The At5g66900 is the N REQUIREMENT GENE 1.1. (NRG1), a key component to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

mediate various TNL signaling pathways, including resistance to *Xanthomonas* and *Pseudomonas* in Nicotiana [39], and the At4g33300 gene encodes a member of the activated disease resistance 1, ADR1 gene family, known to be involved in resistance against *Erysiphe cichoracearum*, the causal agent of Powdery Mildew [40].

The TIR-NBS-LRR genes were grouped into two classes in Figure 1. Class 1 included genes mapped on chr2, 3, 6 and 7, similar to *C. pepo* genes annotated as TIR-NBS-LRR disease resistance proteins, MRGH21 and receptor lectin protein kinase. Most of them showed significant blastx hit (97 to 100% query cover and 91 to 100% identity) with *C. pepo* genes annotated as Tobacco mosaic virus resistance protein N. The gene TNL_4091, mapped at chr7, which had 100% cover and identity with XP_023537669 (disease resistance protein RPS6 like isoform). The TIR-NBS-LRR genes of class 2, mostly clustered at chr6, however, did not show blastn hits with *C. pepo* genes annotated at http://cucurbitgenomics.org/, but have also the best blastx hit with TMV resistance protein N-like at NCBI.

DISCUSSION

Low number of NBS sequences

The data mining approach using the draft version of the genome through a reiterative process has revealed that the NBS-encoding gene family in Zucchini is composed of 34 members. The data obtained indicate that the number of NBS genes in summer squash is significantly lower than in other species: 1015, 400, 147, 126 or 68 in case of *Malus domestica*, *Populus trichocarpa*, *Linum usitatissimum*, *Brachypodium distachyon*, or *Asparagus officinalis*, respectively [28,33,41–43]. Although in other Cucurbitacea species a low number of NBS genes has been also described (70, 75 and 55 in cucumber, melon and water melon, respectively [8], the number found in *C. pepo* in this study has been still lower. Wan et al. (2013) [21], when comparing six species (including two Cucurbitacea) found that the number of NBS-encoding genes does not increase or decrease proportionally to the genome size, although a trend between genome size and total number of predicted R-proteins exists. Moreover, another study in grasses showed that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

the species with larger genomes do not necessarily have more gene members for NBS-LRR gene family [44]. Variation by multiple fold size of the NBS family has been found not only among congeneric species but also among conspecific cultivars or lines suggesting that the gene family size variation is a common phenomenon in the plant genera [45]. The variation of NBS-LRR family size between species could be attributed to gene duplication, deletion, pseudogenization and/or functional diversification [45]. The last case is supported by the necessity of a species to adapt to rapidly changing pathogen populations [7].

There are several explanations for the pronounced low number of NBS-LRR coding genes identified in *C. pepo*, compared to other species: (i) It seems likely that a high copy number of R-genes in plants should be advantageous. However, the limited number of them in plant species suggests that R-genes must have biological cost to balance their expansion in a genome. These costs include energy for transcription or translation and toxic effects, ie. high expression of R-genes may be lethal to plant cells [8]. A study in *Arabidopsis thaliana* has shown the fitness cost of the RPM1 resistance locus, providing evidence that costs prevent fixation by natural selection [46]; (ii) Regarding domestication, it is expected that wild species have more NBS genes than cultivated ones due to the genetic bottleneck that this process implies. In this sense, we could expect a low number of NBS members in a cultivated Zucchini genotype such as Murcia. In support of that expectation, African wild rice (*O. barthii*) and wild soybean have shown higher numbers of NBS-genes when compared to cultivated ones, even though the number of genes in the cultivated rice is larger than the Asian wild rice (*O. rufipogon*; [45]); (iii) Since the NBS-LRR genes are reported to be multi-copy and highly diversified ones [47,48], the genome sequences assembly from Illumina short reads could be underestimating the number of this type of gene detected in the Zucchini genome. In other studies, misannotation is considered to decrease the number of initially identified genes, with about 36% errors in automated annotations in the case of *Arabidopsis* [19]. However, in this study we do not expect a decreased number since the process has included a manual curation from the original automated annotation; (iv) Finally, we found a low number of duplications of NBS-LRR genes in the Zucchini genome. The majority of conserved genes are single copy and/or located as a singletons as reported in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

the cucumber genome [21]. The low number of duplications in Cucurbitaceae genomes has been proposed as a plausible cause to support the low number of R-genes within the Cucurbitaceae family.

Future studies will be needed to determine the functions of the identified sequences. NBS-LRR genes have been reported to undergo alternative splicing [49] and this mechanism might produce non-functional proteins [50]. Moreover, the alternative transcript isoforms may be subjected to nonsense-mediated decay [51] that prevents inappropriate NBS-gene activation and its adverse effects on growth and fitness [52].

Phylogenetic analysis

Phylogenetic analysis of *C. pepo* NBS-LRR proteins showed two distinct families, the TIR-NBS-LRR (TNL) and the (non-TIR) CC-NBS-LRR (CNL). This differentiation of NBS-LRR gene into two main families is both ancient and well known [19,41]. Furthermore, when grouping together these sequences from *C. pepo* with similar ones from other Cucurbitaceae spp., the main groups observed are based on differences TIR/non-TIR sequences instead of species. Thus, CNL sequences from monocotyledons and dicotyledons tend to cluster together, suggesting that the CNL group originated before the divergence of the monocotyledons and dicotyledons (reviewed in [53]. Although TNL and CNL proteins are both involved in pathogen recognition, the two subfamilies are distinct both in sequence and signaling pathways [35]. In *C. pepo*, we have found a slightly higher number of CNL sequences compared to TNLs. These results are similar to those described for potato [54,55], pepper [27] and *Medicago truncatula* [56]. Other genomes like those from *A. thaliana*, *A. lyrata* and soybean contain from two-fold to six-fold more TNL than CNL genes [57,58]. In case of cereal species, TNLs are completely absent, which suggests that the early angiosperm ancestors had few TNLs and that these were lost in the cereal lineage [35].

When considering the 31 NBS-LRR sequences from Zucchini described in this study, a total of 5 subfamilies has been observed. The number of subfamilies is consistent with that observed in other Cucurbitaceae spp., such as *Cucumis sativus* [21]. It is possible, however, that additional subfamilies may be found if additional NBS sequences are added in the future, when updated versions of the genome are available. The number of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

sequences representing each class ranged from 2 to 16 and 5 to 6 for CC-NBS and TIR-NBS subfamilies, respectively. The finding of multiple NBS-LRR subgroups in many plant species, suggests the occurrence of sequence diversity and probable functional variation among different resistance genes with high evolutionary significance [59]. R-gene evolution seems to be facilitated by cluster formation thereby causing recombination and sequence exchange [60]. When grouping our sequences with the 18 gene models for R-genes reported from other Cucurbitaceae spp. [8] , the five subfamilies clustered with members of cucumber, melon and watermelon. Of the three species considered, none of the members lacked clade. Subsequently, there is no evidence to support formation of divergent subfamilies. In both the non-TIR and TIR group, most of the NBS genes within the same clade were from at least two species. This fact indicated that they had ancient origins which preceded a split between species. This characteristic has also been described in watermelon [26].

The ratio of nonsynonymous (K_a) to synonymous (K_s) nucleotide substitutions among triplets encoding the amino acids of the NBS domain was below 1 for each of the five main groups detected in the phylogenetic tree. This shows purifying selection for all the groups that refers to selection against nonsynonymous substitutions at the DNA level. Heterogeneity of impact of selection on NBS-encoding genes has been previously reported in plants [35]. Although, diversifying selection has been shown in the NBS domain of resistance genes in faba bean [61], the NBS domain is reported to be more commonly subjected to purifying selection, such as the dominant selection reported in pepper [27] or rice [62]. By contrast, the LRR domain seems to be more variable and experiences diversifying selection [63,64]. This reflects the role of the LRR domain in recognizing constantly evolving pathogen ligands [6].

An examination of the proportion of multi-gene families across similarity/coverage thresholds was made using NBS sequences from other Cucurbitaceae. Following this examination, a temporal difference in NBS-encoding gene expansion among species was estimated. Previously, strict standards were used as the proxy for the detection of recent duplication events (>80%; [65,66]. When using these standards, a considerably larger proportion of NBS-encoding genes with identities higher than 80% were found in Zucchini and cucumber compared to melon. This may indicate that duplications have

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

not only played a minor role in summer squash, but also recent duplications have likely dominated the NBS-encoding genes. Similar findings have been found in the case of maize with more recent duplications than sorghum and rice [65].

Gene and protein structure

In this study, the mean exon number of Zucchini TNL class was higher than that of CNL class. A percentage of 73% CNL were encoded by a single exon while all TNL proteins were encoded by genes with at least 3 exons. Thus, *C. pepo* TNL sequences contained more introns than CNLs as *Arabidopsis* [19], linseed [41], poplar [33], apple [7] or cereals [6]. The existence of introns in eukaryotic genes is believed to provide an evolutionary advantage by increasing protein diversity through exon shuffling and alternative splicing [67]. However, this eukaryotic feature needs the exclusion of intronic sequences, which requires considerable energy expenditure and can lead to splicing errors. According to Carmel et al. (2007) [68], the probability of intronic gains is positively correlated with the level of evolutionary conservation of the gene. Moreover, Gorlova et al. (2014) [69] have analyzed the relationship between intronic burden and the level of evolutionary conservation of genes. It was determined that genes with a greater intronic burden had lower density of missense and nonsense mutations in the coding regions of the gene. This suggests that they are under a stronger pressure from purifying selection. Our results regarding the mode of selection of the clade containing the TNL proteins are in agreement with these studies and show the lowest value of the K_a/K_s ratio. This indicates purifying selection supported by a higher amount of synonymous substitutions. The introns in resistance genes are vulnerable to changes such as insertions and deletions when exposed to various evolutionary forces. Therefore, they have been repository for developing genetic markers based on intron length polymorphism as reported in the case of NBS-LRR genes of maize [27].

Promoters

Analysis of promoter regions of the *C. pepo* NBS-LRR genes did not reveal uniformity in the numbers of six overrepresented *cis* elements (WBOX, DRE, ERE, ATHB5, Gbox and RAV1-B). Among the six *cis* regulatory elements, the WBOX was the most abundant. It

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

averaged 3.58 motifs per gene, specifically 4.9 for the TNL and 2.6 for the CNL subfamilies. Promoter regions of CNL genes in soybean also showed an average of 2.77 WBOX cassettes per gene [70]. Similarly, in linseed, all the R-genes contained WBOX domains [41]. Nevertheless, this ratio is lower than those found for WBOX elements in *Brachypodium* [28] or *Medicago truncatula* [56]. The finding that from the 31 upstream regions of *C. pepo* NBS genes analysed, 30 had at least one WBOX motif indicates that this element is important for regulation of most if not all NBS-LRR family genes. WBOXs have been often statistically overrepresented in promoters of defense-related genes [71,72]. In addition, mutational analyses of WBOX in several genes have clearly demonstrated their relevance to immune regulation [73–75]. WBOX motifs have been described upstream in the NPR1 gene, which is a positive regulator of inducible plant disease resistance [74] and upstream of most *Arabidopsis* pathogen response genes [76,77]. Furthermore, it has also been shown that they activate NBS-LRR genes in grape [78]. Moreover, Mohr et al. (2010) [79], demonstrated that one or more of the 3 WBOX identified in the promoter of RPP8 (an *Arabidopsis* downy mildew R gene) is necessary for this gene expression and function. The widespread presence of WBOX motifs upstream of so many NBS genes does not necessarily imply that all the NBS genes of the species are under identical regulatory mechanism and other less conserved factors might be involved in fine regulation of these genes [56].

Gene mapping

Genetic studies aimed to identify genes involved in resistance to pathogens in *C. pepo* are now advancing due to the availability of genomic resources for this and other closely-related species. Two of the two most important pathogens of *C. pepo* are powdery mildew caused by *Podosphaera xanthii* and the Potyvirus zucchini yellow mosaic virus (ZYMV). Two recent studies provide candidates for resistance genes to these pathogens. Holdsworth et al. (2016) [80], mapped a candidate region of 76.4 kb including Pm-0, a major resistance gene from the wild species *Cucurbita okeechobeensis* subsp. *martinezii* introgressed into *C. pepo* and *C. moschata*. The most interesting candidate that maps in this region is the homolog of At5g66900, corresponding with the RPW8_3282 gene of our study (Cp4.1LG10g02750.1). In fact, the CAPS markers that cosegregate with

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

PM resistance in the analysis by Holdsworth et al. (2016) is in this gene, so Cp4.1LG10g02750 mapped in chr10 is likely be involved in PM response. The MU-CU-16 *C. pepo* variety from which the reference genome was obtained is highly susceptible to PM. Resistance to this pathogen comes from *C. moschata* and wild *Cucurbita*. The *C. moschata* homolog of Cp4.1LG10g02750.1 is CmoCh03G010060.1 (Supplemental Table S4). Sequences from sets of resistant/susceptible *C. moschata* accessions could be cloned, analyzed and compared to that one of susceptible *C. pepo* accessions. Regarding ZYMV, Capuozzo et al. (2017) [81], identified SNPs linked to virus resistance in *C. pepo*. The most linked SNPs is located in gene Cp4.1LG08g08000 (a dead box ATP-dependent RNA helicase, mapped at chr8 6315834-6317495) which is located very close, less than 8 kb, to CNL2651 (Cp4.1LG08g07980 chr8 6324046-6342276). Therefore, this gene is a good candidate to be involved in the ZYMV resistance trait.

Conclusions

Since production of summer squash can be severely limited by pathogens, developing disease resistance cultivars is one of the most important objectives in plant breeding. In this study, we report the first analysis of the NBS-encoding genes for *Cucurbita pepo* by using the draft genome sequence of the species. These results will contribute to the identification, isolation and characterization of candidate R-genes, as well as gaining insight into the NBS gene family evolution in the species. These genes can be also used as candidates in the recently reported TILLING platform of the species [14] developed with the genotype MU, which is the same genotype that has been used to generate the genome sequences. Thus, these regions could provide an essential tool for the manipulation of important summer squash genes conferring resistance for molecular breeding and genetic improvement.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

Materials and Methods

Retrieval and identification of Zucchini NBS-encoding R-genes

A set of NBS gene sequences was identified in the draft genome of Zucchini (*C. pepo* L.) using a reiterative process. Zucchini (Spanish cultivar MU16, belonging to *C. pepo* ssp. pepo) assembly and annotation genome v0.1 was downloaded from http://www.cucurbigene.net/genome-0.1/. First, a set of candidate NBS genes was selected from the protein coding sequences from *A. thaliana* and rice [18,19] as the query against the CucurbiGene database (COMAV Bioinformatics & Genomics and Cucurbits Breeding groups) using a tblastn search and a threshold of E=1e^{-5}. In the second step, all blast hits were expanded to 2,500-5,000 bp from both ends of the hits, and then the expanded nucleotide fragments were annotated using the gene-finding programs FGENESH (http://www.softberry.com) and GENSCAN (http://genes.mit.edu/GENSCAN.html) to obtain information on complete ORFs and on intergenic regions. Proteins lacking NBS motifs (Pfam: PF00931) or with no match to R-gene proteins (blastp search against the nr database at NCBI) were excluded from further analysis. Each of these annotated sequences was surveyed to determine whether they encoded TIR, CC, or LRR motifs using the Pfam database (http://pfam.xfam.org/), SMART protein motif analysis (http://smart.embl-heidelberg.de/; [82], and COILS program to detect CC domains with a threshold = 0.9 (http://toolkit.tuebingen.mpg.de/pcoils; [83]), in order to obtain detailed information on protein motifs, domains and families and classify these NBS-encoding genes. The third step aimed at a complete search of candidate NBS genes in the *C. pepo* genome. All sequences classified as TNL or CNL were used as query, so that we could search the entire Zucchini genome to check whether there were additional related genes that had not been identified by the foregoing work. Finally, all the amino sequences selected above were compared (if possible) to the transcriptome database at the CucurbiGene database and manually modified to correct missannotation. Predicted proteins were aligned using CLUSTALW [84]. After alignment, all identical sequences were checked manually and overlapping sequences were discarded.

Phylogenetic analysis

Multiple alignment of translated protein sequences was carried out using MUSCLE as implemented in the Molecular Evolutionary Genetics Analysis software-MEGA ver. 5.1. [85]. First, we studied 31 out of the 34 NBS encoding sequences identified from *C. pepo*. Three sequences (1 TN and 2 N) that did not show a complete structure with both N-terminal and LRR domains, were discarded for the phylogenetic study. Then, these Zucchini sequences, were compared with eighteen R-gene models previously reported in three Cucurbitaceae species: cucumber, melon and watermelon [8]. In both cases, once the amino acid sequence alignments between the P-loop and GLPL motifs were carried out and followed by manual adjustment, the phylogenetic tree was constructed by the Maximum likelihood (ML) analysis using the evolutionary model Jones-Taylor-Thornton (JTT) with the bootstrap support of 100 replicates implemented in MEGA version 5. We calculated the ratio of non-synonymous K_a to synonymous K_s nucleotide substitutions (K_a/K_s ratio) among triplets encoding the amino acids of Zucchini NBS-LRR genes for each class to determine either diversifying (K_a/K_s>1) or purifying (K_a/K_s<1)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2020 doi:10.20944/preprints202001.0048.v1

selection, using the method described by [86], with MEGA 5.10 program. This ratio was estimated for each of the Zucchini R-genes groups determined by the phylogenetic tree.

Gene duplication, gene structure and prediction of conserved motifs

NBS gene duplication events of the NBS-LRR genes were also investigated. NBS domain was defined as the region extending from the P-loop to the MHDV motif. By gene duplication we mean the criteria defined by [28] : (1) the alignment covered >70% of the longer gene; (2) the aligned region had an identity >70%. Exon/intron structures were obtained using the online Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) with both coding sequences and genomic sequences [87]. To investigate the diversity and structure of NBS-encoding genes in Zucchini, their predicted amino acid sequences were subjected to domain and motif analysis. NBS-encoding genes were divided into three components, namely the N-terminal, NBS domain and LRR C-terminal regions. They were then analyzed individually using the Multiple Expectation Maximization for Motif elicitation (MEME) system (http://meme.nbcr.net/meme/) on members on TIR-NBS-LRR and CC-NBS-LRR subfamilies with the following conditions: (1) optimum motif width was set to 10 and 15; (2) maximum number of motifs was designed to identify a number of motifs with an *E*-value<$1e^{-20}$.

Identification and analysis of the promoter regions

The promoter sequence (~ 2,000 bp from the start ATG codon) was obtained for each predicted NBS-LRR R-gene. The extracted sequences were screened against a collection of 27,457 *Arabidopsis* promoters (Arabidopsis Promoter Element Discovery Tools, http://stan.cropsci.uiuc.edu/tools.php). *Cis*-regulatory elements (CREs) overrepresented in the dataset and known to be involved in regulation during the resistance response and under stressed conditions were selected for further analyzed. Occurrence and distribution of CREs over a given promoter sequence analysis were performed using standard Python scripts. The expected frequency of each motif was calculated using the average G+C content of 35% observed in these promoters. Probabilities were estimated based on 2,000 Monte Carlo simulations. The characterized CREs included: WBOX sequence associated with the WRKY transcription factor ([TGAC(C/T)]; [88]), DRE boxes associated with multiple abiotic stresses ([(G/A)CCGAC]; [89]), ethylene responsive element (ERE/GCCbox- [AGCCGCC]; [90]), ATHB5, a positive regulator of ABA responsiveness ([CAATNATTG]; [91]), Gbox involved in dehydration-responsive expression ([CACATG]; [92]) and RAV1-B element involved in defense mechanism signaling cascade ([CACCTG]; [93]).

Mapping of NBS-LRR genes

The NBS-LRR candidate genes were mapped to their physical position using Blast [94] against the *C. pepo* mRNA collection v 4.1 (annotated and mapped to the last version of *C. pepo* genome), available at http://cucurbitgenomics.org/, and described in [95]. Only the best hit was considered (full coverage of both query and subject). Those NBS-LRR sequences with no hit in the mRNA collection were mapped using Blast against the *C. pepo* reference genome v4.1. Mapchart [96] was used for visualization. Since some NBS-LRR candidate genes had no hit with *C. pepo* genes annotated

at http://cucurbitgenomics.org/, all sequences were blasted against the mRNA collection of the close species *C. moschata* (also annotated and mapped against the reference *C. moschata* genome v1, available at http://cucurbitgenomics.org/). Also, we used Blastx to find the best protein hit of *C. pepo* and other *Cucurbita*, mainly *C. moschata* and *C. maxima* annotated at NCBI non-redundant protein sequences database. The two best hits were selected.

References

1. Paris, H.S. Summer squash, *in* Handbook of plant breeding, vegetables I. ed. J. Prohens, and F. Nuez (Springer, New York), **2008**, 351-379.
2. Paris, H. S. Summer squash: history, diversity, and distribution. *Horttechnology* **1996**, *6*, 6–13, doi:10.21273/HORTTECH.6.1.6.
3. Formisano, G.; Roig, C.; Esteras, C.; Ercolano, M. R.; Nuez, F.; Monforte, A. J.; Picó, M. B. Genetic diversity of Spanish *Cucurbita pepo* landraces: an unexploited resource for summer squash breeding. *Genet. Resour. Crop Evol.* **2012**, *59*, 1169–1184, doi:10.1007/s10722-011-9753-y.
4. Meyers, B. C.; Dickerman, A. W.; Michelmore, R. W.; Sivaramakrishnan, S.; Sobral, B. W.; Young, N. D. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* **1999**, *20*, 317–332, doi:10.1046/j.1365-313x.1999.t01-1-00606.x.
5. Pan, Q.; Wendel, J.; Fluhr, R. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* **2000**, *50*, 203–213, doi:10.1007/s002399910023.
6. 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.; Hulbert, S. H. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* **2002**, *12*, 1871–1884, doi:10.1101/gr.454902.
7. Perazzolli, M.; Malacarne, G.; Baldo, A.; Righetti, L.; Bailey, A.; Fontana, P.; Velasco, R.; Malnoy, M. Characterization of resistance gene analogues (RGAs) in apple (Malus × domestica Borkh.) and their evolutionary history of the Rosaceae family. *PLoS ONE* **2014**, *9*, e83844, doi:10.1371/journal.pone.0083844.
8. Lin, X.; Zhang, Y.; Kuang, H.; Chen, J. Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC Genomics* **2013**, *14*, 335, doi:10.1186/1471-2164-14-335.
9. Narusaka, M.; Kubo, Y.; Hatakeyama, K.; Imamura, J.; Ezura, H.; Nanasato, Y.; Tabei, Y.; Takano, Y.; Shirasu, K.; Narusaka, Y. Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. *PLoS ONE* **2013**, *8*, e55954, doi:10.1371/journal.pone.0055954.
10. Saha, D.; Rana, R. S.; Sureja, A. K.; Verma, M.; Arya, L.; Munshi, A. D. Cloning and characterization of NBS-LRR encoding resistance gene candidates from Tomato Leaf Curl New Delhi Virus resistant genotype of Luffa cylindrica Roem. *Physiological and Molecular Plant Pathology* **2013**, *81*, 107–117, doi:10.1016/j.pmpp.2012.11.007.
11. Dogimont, C.; Chovelon, V.; Pauquet, J.; Boualem, A.; Bendahmane, A. The Vat locus encodes for a CC-NBS-LRR protein that confers resistance to Aphis gossypii infestation and A. gossypii-mediated virus resistance. *Plant J.* **2014**, *80*, 993–1004, doi:10.1111/tpj.12690.
12. Blanca, J.; Cañizares, J.; Roig, C.; Ziarsolo, P.; Nuez, F.; Picó, B. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* **2011**, *12*, 104, doi:10.1186/1471-2164-12-104.
13. Esteras, C.; Gómez, P.; Monforte, A. J.; Blanca, J.; Vicente-Dólera, N.; Roig, C.; Nuez, F.; Picó, B. High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. *BMC Genomics* **2012**, *13*, 80, doi:10.1186/1471-2164-13-80.

14. Vicente-Dólera, N.; Troadec, C.; Moya, M.; del Río-Celestino, M.; Pomares-Viciana, T.; Bendahmane, A.; Picó, B.; Román, B.; Gómez, P. First TILLING platform in Cucurbita pepo: a new mutant resource for gene function and crop improvement. *PLoS ONE* **2014**, *9*, e112743, doi:10.1371/journal.pone.0112743.
15. Montero-Pau, J.; Blanca, J.; Bombarely, A.; Ziarsolo, P.; Esteras, C.; Martí-Gómez, C.; Ferriol, M.; Gómez, P.; Jamilena, M.; Mueller, L.; Picó, B.; Cañizares, J. De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the Cucurbita genus. *Plant Biotechnol. J.* **2018**, *16*, 1161–1171, doi:10.1111/pbi.12860.
16. Xanthopoulou, A.; Montero-Pau, J.; Mellidou, I.; Kissoudis, C.; Blanca, J.; Picó, B.; Tsaballa, A.; Tsaliki, E.; Dalakouras, A.; Paris, H. S.; Ganopoulou, M.; Moysiadis, T.; Osathanunkul, M.; Tsaftaris, A.; Madesis, P.; Kalivas, A.; Ganopoulos, I. Whole-genome resequencing of Cucurbita pepo morphotypes to discover genomic variants associated with morphology and horticulturally valuable traits. *Hortic. Res.* **2019**, *6*, 94, doi:10.1038/s41438-019-0176-9.
17. Meyers, B. C. Genome-Wide Analysis of NBS-LRR-Encoding Genes in *Arabidopsis*. *THE PLANT CELL ONLINE* **2003**, *15*, 809–834, doi:10.1105/tpc.009308.
18. 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. Genomics* **2004**, *271*, 402–415, doi:10.1007/s00438-004-0990-z.
19. 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, doi:10.1105/tpc.009308.
20. Xiao, S.; Ellwood, S.; Calis, O.; Patrick, E.; Li, T.; Coleman, M.; Turner, J. G. Broad-spectrum mildew resistance in Arabidopsis thaliana mediated by RPW8. *Science* **2001**, *291*, 118–120, doi:10.1126/science.291.5501.118.
21. Wan, H.; Yuan, W.; Bo, K.; Shen, J.; Pang, X.; Chen, J. Genome-wide analysis of NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *BMC Genomics* **2013**, *14*, 109, doi:10.1186/1471-2164-14-109.
22. Garcia-Mas, J.; Benjak, A.; Sanseverino, W.; Bourgeois, M.; Mir, G.; González, V. M.; Hénaff, E.; Câmara, F.; Cozzuto, L.; Lowy, E.; Alioto, T.; Capella-Gutiérrez, S.; Blanca, J.; Cañizares, J.; Ziarsolo, P.; Gonzalez-Ibeas, D.; Rodríguez-Moreno, L.; Droege, M.; Du, L.; Alvarez-Tejado, M.; Puigdomènech, P. The genome of melon (Cucumis melo L.). *Proc Natl Acad Sci USA* **2012**, *109*, 11872–11877, doi:10.1073/pnas.1205415109.
23. Shao, Z.-Q.; Zhang, Y.-M.; Hang, Y.-Y.; Xue, J.-Y.; Zhou, G.-C.; Wu, P.; Wu, X.-Y.; Wu, X.-Z.; Wang, Q.; Wang, B.; Chen, J.-Q. Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant Physiol.* **2014**, *166*, 217–234, doi:10.1104/pp.114.243626.
24. Huang, S.; Li, R.; Zhang, Z.; Li, L.; Gu, X.; Fan, W.; Lucas, W. J.; Wang, X.; Xie, B.; Ni, P.; Ren, Y.; Zhu, H.; Li, J.; Lin, K.; Jin, W.; Fei, Z.; Li, G.; Staub, J.; Kilian, A.; van der Vossen, E. A. G.; Li, S. The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* **2009**, *41*, 1275–1281, doi:10.1038/ng.475.
25. Wóycicki, R.; Witkowicz, J.; Gawroński, P.; Dąbrowska, J.; Lomsadze, A.; Pawełkowicz, M.; Siedlecka, E.; Yagi, K.; Plader, W.; Seroczyńska, A.; Śmiech, M.; Gutman, W.; Niemirowicz-Szczytt, K.; Bartoszewski, G.; Tagashira, N.; Hoshi, Y.; Borodovsky, M.; Karpiński, S.; Malepszy, S.; Przybecki, Z. The genome sequence of the North-European cucumber (Cucumis sativus L.) unravels evolutionary adaptation mechanisms in plants. *PLoS ONE* **2011**, *6*, e22728, doi:10.1371/journal.pone.0022728.
26. Harris, K. R.; Wechter, W. P.; Levi, A. Isolation, sequence analysis, and linkage mapping of nucleotide binding Site-Leucine-rich repeat disease resistance gene analogs in watermelon. *J. Amer. Soc. Hort. Sci.* **2009**, *134*, 649–657.
27. Wan, H.; Yuan, W.; Ye, Q.; Wang, R.; Ruan, M.; Li, Z.; Zhou, G.; Yao, Z.; Zhao, J.; Liu, S.; Yang, Y. Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper:

characterization, genetic variation, functional divergence and expression patterns. *BMC Genomics* **2012**, *13*, 502, doi:10.1186/1471-2164-13-502.

28. Tan, S.; Wu, S. Genome Wide Analysis of Nucleotide-Binding Site Disease Resistance Genes in *Brachypodium distachyon*. *Comp. Funct. Genomics* **2012**, *2012*, 418208, doi:10.1155/2012/418208.
29. Mun, J.-H.; Yu, H.-J.; Park, S.; Park, B.-S. Genome-wide identification of NBS-encoding resistance genes in Brassica rapa. *Mol. Genet. Genomics* **2009**, *282*, 617–631, doi:10.1007/s00438-009-0492-0.
30. Malacarne, G.; Perazzolli, M.; Cestaro, A.; Sterck, L.; Fontana, P.; Van de Peer, Y.; Viola, R.; Velasco, R.; Salamini, F. Deconstruction of the (paleo)polyploid grapevine genome based on the analysis of transposition events involving NBS resistance genes. *PLoS ONE* **2012**, *7*, e29762, doi:10.1371/journal.pone.0029762.
31. Velasco, R.; Zharkikh, A.; Troggio, M.; Cartwright, D. A.; Cestaro, A.; Pruss, D.; Pindo, M.; Fitzgerald, L. M.; Vezzulli, S.; Reid, J.; Malacarne, G.; Iliev, D.; Coppola, G.; Wardell, B.; Micheletti, D.; Macalma, T.; Facci, M.; Mitchell, J. T.; Perazzolli, M.; Eldredge, G.; Viola, R. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* **2007**, *2*, e1326, doi:10.1371/journal.pone.0001326.
32. Monosi, B.; Wisser, R. J.; Pennill, L.; Hulbert, S. H. Full-genome analysis of resistance gene homologues in rice. *Theor. Appl. Genet.* **2004**, *109*, 1434–1447, doi:10.1007/s00122-004-1758-x.
33. Kohler, A.; Rinaldi, C.; Duplessis, S.; Baucher, M.; Geelen, D.; Duchaussoy, F.; Meyers, B. C.; Boerjan, W.; Martin, F. Genome-wide identification of NBS resistance genes in Populus trichocarpa. *Plant Mol. Biol.* **2008**, *66*, 619–636, doi:10.1007/s11103-008-9293-9.
34. Cannon, S. B.; Zhu, H.; Baumgarten, A. M.; Spangler, R.; May, G.; Cook, D. R.; Young, N. D. Diversity, Distribution, and Ancient Taxonomic Relationships Within the TIR and Non-TIR NBS-LRR Resistance Gene Subfamilies. *J. Mol. Evol.* **2002**, *54*, 548–562, doi:10.1007/s0023901-0057-2.
35. McHale, L.; Tan, X.; Koehl, P.; Michelmore, R. W. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* **2006**, *7*, 212, doi:10.1186/gb-2006-7-4-212.
36. Jiang, H.; Wang, C.; Ping, L.; Tian, D.; Yang, S. Pattern of LRR nucleotide variation in plant resistance genes. *Plant Sci.* **2007**, *173*, 253–261, doi:10.1016/j.plantsci.2007.05.010.
37. Sekhwal, M. K.; Li, P.; Lam, I.; Wang, X.; Cloutier, S.; You, F. M. Disease resistance gene analogs (rgas) in plants. *Int. J. Mol. Sci.* **2015**, *16*, 19248–19290, doi:10.3390/ijms160819248.
38. Li, N.-Y.; Zhou, L.; Zhang, D.-D.; Klosterman, S. J.; Li, T.-G.; Gui, Y.-J.; Kong, Z.-Q.; Ma, X.-F.; Short, D. P. G.; Zhang, W.-Q.; Li, J.-J.; Subbarao, K. V.; Chen, J.-Y.; Dai, X.-F. Heterologous Expression of the Cotton NBS-LRR Gene GbaNA1 Enhances Verticillium Wilt Resistance in Arabidopsis. *Front. Plant Sci.* **2018**, *9*, 119, doi:10.3389/fpls.2018.00119.
39. Qi, T.; Seong, K.; Thomazella, D. P. T.; Kim, J. R.; Pham, J.; Seo, E.; Cho, M.-J.; Schultink, A.; Staskawicz, B. J. NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in Nicotiana benthamiana. *Proc Natl Acad Sci USA* **2018**, *115*, E10979–E10987, doi:10.1073/pnas.1814856115.
40. Andolfo, G.; Di Donato, A.; Darrudi, R.; Errico, A.; Aiese Cigliano, R.; Ercolano, M. R. Draft of Zucchini (Cucurbita pepo L.) Proteome: A Resource for Genetic and Genomic Studies. *Front. Genet.* **2017**, *8*, 181, doi:10.3389/fgene.2017.00181.
41. Kale, S. M.; Pardeshi, V. C.; Barvkar, V. T.; Gupta, V. S.; Kadoo, N. Y. Genome-wide identification and characterization of nucleotide binding site leucine-rich repeat genes in linseed reveal distinct patterns of gene structure. *Genome* **2013**, *56*, 91–99, doi:10.1139/gen-2012-0135.
42. Arya, P.; Kumar, G.; Acharya, V.; Singh, A. K. Genome-wide identification and expression analysis of NBS-encoding genes in *Malus* x *domestica* and expansion of NBS genes family in Rosaceae. *PLoS ONE* **2014**, *9*, e107987, doi:10.1371/journal.pone.0107987.
43. Die, J. V.; Castro, P.; Millán, T.; Gil, J. Segmental and Tandem Duplications Driving the Recent NBS-LRR Gene Expansion in the Asparagus Genome. *Genes (Basel)* **2018**, *9*, doi:10.3390/genes9120568.

44. Li, J.; Ding, J.; Zhang, W.; Zhang, Y.; Tang, P.; Chen, J.-Q.; Tian, D.; Yang, S. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol. Genet. Genomics* **2010**, *283*, 427–438, doi:10.1007/s00438-010-0527-6.
45. Zhang, M.; Wu, Y.-H.; Lee, M.-K.; Liu, Y.-H.; Rong, Y.; Santos, T. S.; Wu, C.; Xie, F.; Nelson, R. L.; Zhang, H.-B. Numbers of genes in the NBS and RLK families vary by more than four-fold within a plant species and are regulated by multiple factors. *Nucleic Acids Res.* **2010**, *38*, 6513–6525, doi:10.1093/nar/gkq524.
46. Tian, D.; Traw, M. B.; Chen, J. Q.; Kreitman, M.; Bergelson, J. Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. *Nature* **2003**, *423*, 74–77, doi:10.1038/nature01588.
47. Yang, S.; Zhang, X.; Yue, J.-X.; Tian, D.; Chen, J.-Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* **2008**, *280*, 187–198, doi:10.1007/s00438-008-0355-0.
48. 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, doi:10.1105/tpc.104.025502.
49. Ferrier-Cana, E.; Macadré, C.; Sévignac, M.; David, P.; Langin, T.; Geffroy, V. Distinct post-transcriptional modifications result into seven alternative transcripts of the CC-NBS-LRR gene *JA1tr* of *Phaseolus vulgaris*. *Theor. Appl. Genet.* **2005**, *110*, 895–905, doi:10.1007/s00122-004-1908-1.
50. Simpson, C. G.; Fuller, J.; Maronova, M.; Kalyna, M.; Davidson, D.; McNicol, J.; Barta, A.; Brown, J. W. S. Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J.* **2008**, *53*, 1035–1048, doi:10.1111/j.1365-313X.2007.03392.x.
51. Gassmann, W. Alternative splicing in plant defense. *Curr. Top. Microbiol. Immunol.* **2008**, *326*, 219–233, doi:10.1007/978-3-540-76776-3_12.
52. Gloggnitzer, J.; Akimcheva, S.; Srinivasan, A.; Kusenda, B.; Riehs, N.; Stampfl, H.; Bautor, J.; Dekrout, B.; Jonak, C.; Jiménez-Gómez, J. M.; Parker, J. E.; Riha, K. Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant antibacterial defense. *Cell Host Microbe* **2014**, *16*, 376–390, doi:10.1016/j.chom.2014.08.010.
53. Marone, D.; Russo, M. A.; Laidò, G.; De Leonardis, A. M.; Mastrangelo, A. M. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *Int. J. Mol. Sci.* **2013**, *14*, 7302–7326, doi:10.3390/ijms14047302.
54. Jupe, F.; Pritchard, L.; Etherington, G. J.; Mackenzie, K.; Cock, P. J. A.; Wright, F.; Sharma, S. K.; Bolser, D.; Bryan, G. J.; Jones, J. D. G.; Hein, I. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics* **2012**, *13*, 75, doi:10.1186/1471-2164-13-75.
55. Lozano, R.; Ponce, O.; Ramirez, M.; Mostajo, N.; Orjeda, G. Genome-wide identification and mapping of NBS-encoding resistance genes in Solanum tuberosum group phureja. *PLoS ONE* **2012**, *7*, e34775, doi:10.1371/journal.pone.0034775.
56. Ameline-Torregrosa, C.; Wang, B.-B.; O'Bleness, M. S.; Deshpande, S.; Zhu, H.; Roe, B.; Young, N. D.; Cannon, S. B. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* **2008**, *146*, 5–21, doi:10.1104/pp.107.104588.
57. Guo, Y.-L.; Fitz, J.; Schneeberger, K.; Ossowski, S.; Cao, J.; Weigel, D. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant Physiol.* **2011**, *157*, 757–769, doi:10.1104/pp.111.181990.
58. Kang, Y. J.; Kim, K. H.; Shim, S.; Yoon, M. Y.; Sun, S.; Kim, M. Y.; Van, K.; Lee, S.-H. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol.* **2012**, *12*, 139, doi:10.1186/1471-2229-12-139.
59. Rout, E.; Nanda, S.; Nayak, S.; Joshi, R. K. Molecular characterization of NBS encoding resistance genes and induction analysis of a putative candidate gene linked to *Fusarium* basal rot resistance in *Allium sativum*. *Physiological and Molecular Plant Pathology* **2014**, *85*, 15–24, doi:10.1016/j.pmpp.2013.11.003.

60. Joshi, R. K.; Nayak, S. Perspectives of genomic diversification and molecular recombination towards R-gene evolution in plants. *Physiol. Mol. Biol. Plants* **2013**, *19*, 1–9, doi:10.1007/s12298-012-0138-2.
61. Palomino, C.; Satovic, Z.; Cubero, J. I.; Torres, A. M. Identification and characterization of NBS-LRR class resistance gene analogs in faba bean (Vicia faba L.) and chickpea (Cicer arietinum L.). *Genome* **2006**, *49*, 1227–1237, doi:10.1139/g06-071.
62. Yang, S.; Feng, Z.; Zhang, X.; Jiang, K.; Jin, X.; Hang, Y.; Chen, J.-Q.; Tian, D. Genome-wide investigation on the genetic variations of rice disease resistance genes. *Plant Mol. Biol.* **2006**, *62*, 181–193, doi:10.1007/s11103-006-9012-3.
63. Ellis, J.; Dodds, P.; Pryor, T. Structure, function and evolution of plant disease resistance genes. *Curr. Opin. Plant Biol.* **2000**, *3*, 278–284, doi:10.1016/S1369-5266(00)00080-7.
64. Jones, J. D. G.; Dangl, J. L. The plant immune system. *Nature* **2006**, *444*, 323–329, doi:10.1038/nature05286.
65. Mace, E.; Tai, S.; Innes, D.; Godwin, I.; Hu, W.; Campbell, B.; Gilding, E.; Cruickshank, A.; Prentis, P.; Wang, J.; Jordan, D. The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC Plant Biol.* **2014**, *14*, 253, doi:10.1186/s12870-014-0253-z.
66. Zhong, Y.; Yin, H.; Sargent, D. J.; Malnoy, M.; Cheng, Z.-M. M. Species-specific duplications driving the recent expansion of NBS-LRR genes in five Rosaceae species. *BMC Genomics* **2015**, *16*, 77, doi:10.1186/s12864-015-1291-0.
67. Kriventseva, E. V.; Koch, I.; Apweiler, R.; Vingron, M.; Bork, P.; Gelfand, M. S.; Sunyaev, S. Increase of functional diversity by alternative splicing. *Trends Genet.* **2003**, *19*, 124–128, doi:10.1016/S0168-9525(03)00023-4.
68. Carmel, L.; Wolf, Y. I.; Rogozin, I. B.; Koonin, E. V. Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res.* **2007**, *17*, 1034–1044, doi:10.1101/gr.6438607.
69. Gorlova, O.; Fedorov, A.; Logothetis, C.; Amos, C.; Gorlov, I. Genes with a large intronic burden show greater evolutionary conservation on the protein level. *BMC Evol. Biol.* **2014**, *14*, 50, doi:10.1186/1471-2148-14-50.
70. Nepal, M. P.; Benson, B. V. CNL disease resistance genes in soybean and their evolutionary divergence. *Evol. Bioinform. Online* **2015**, *11*, 49–63, doi:10.4137/EBO.S21782.
71. Eulgem, T.; Weigman, V. J.; Chang, H.-S.; McDowell, J. M.; Holub, E. B.; Glazebrook, J.; Zhu, T.; Dangl, J. L. Gene expression signatures from three genetically separable resistance gene signaling pathways for downy mildew resistance. *Plant Physiol.* **2004**, *135*, 1129–1144, doi:10.1104/pp.104.040444.
72. Maleck, K.; Levine, A.; Eulgem, T.; Morgan, A.; Schmid, J.; Lawton, K. A.; Dangl, J. L.; Dietrich, R. A. The transcriptome of Arabidopsis thaliana during systemic acquired resistance. *Nat. Genet.* **2000**, *26*, 403–410, doi:10.1038/82521.
73. Lebel, E.; Heifetz, P.; Thorne, L.; Uknes, S.; Ryals, J.; Ward, E. Functional analysis of regulatory sequences controlling PR-1 gene expression in Arabidopsis. *Plant J.* **1998**, *16*, 223–233, doi:10.1046/j.1365-313x.1998.00288.x.
74. Yu, D.; Chen, C.; Chen, Z. Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. *Plant Cell* **2001**, *13*, 1527–1540, doi:10.1105/tpc.010115.
75. Turck, F.; Zhou, A.; Somssich, I. E. Stimulus-dependent, promoter-specific binding of transcription factor WRKY1 to Its native promoter and the defense-related gene PcPR1-1 in Parsley. *Plant Cell* **2004**, *16*, 2573–2585, doi:10.1105/tpc.104.024810.
76. Zheng, Z.; Mosher, S. L.; Fan, B.; Klessig, D. F.; Chen, Z. Functional analysis of Arabidopsis WRKY25 transcription factor in plant defense against Pseudomonas syringae. *BMC Plant Biol.* **2007**, *7*, 2, doi:10.1186/1471-2229-7-2.
77. Li, W.; Li, M.; Zhang, W.; Welti, R.; Wang, X. The plasma membrane-bound phospholipase Ddelta enhances freezing tolerance in Arabidopsis thaliana. *Nat. Biotechnol.* **2004**, *22*, 427–433, doi:10.1038/nbt949.

78. Marchive, C.; Mzid, R.; Deluc, L.; Barrieu, F.; Pirrello, J.; Gauthier, A.; Corio-Costet, M.-F.; Regad, F.; Cailleteau, B.; Hamdi, S.; Lauvergeat, V. Isolation and characterization of a Vitis vinifera transcription factor, VvWRKY1, and its effect on responses to fungal pathogens in transgenic tobacco plants. *J. Exp. Bot.* **2007**, *58*, 1999–2010, doi:10.1093/jxb/erm062.
79. Mohr, T. J.; Mammarella, N. D.; Hoff, T.; Woffenden, B. J.; Jelesko, J. G.; McDowell, J. M. The Arabidopsis downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements. *Mol. Plant Microbe Interact.* **2010**, *23*, 1303–1315, doi:10.1094/MPMI-01-10-0022.
80. Holdsworth, W. L.; LaPlant, K. E.; Bell, D. C.; Jahn, M. M.; Mazourek, M. Cultivar-Based Introgression Mapping Reveals Wild Species-Derived Pm-0, the Major Powdery Mildew Resistance Locus in Squash. *PLoS ONE* **2016**, *11*, e0167715, doi:10.1371/journal.pone.0167715.
81. Capuozzo, C.; Formisano, G.; Iovieno, P.; Andolfo, G.; Tomassoli, L.; Barbella, M. M.; Pico, B.; Paris, H. S.; Ercolano, M. R. Inheritance analysis and identification of SNP markers associated with ZYMV resistance in Cucurbita pepo. *Mol. Breeding* **2017**, *37*, 99, doi:10.1007/s11032-017-0698-5.
82. Letunic, I.; Doerks, T.; Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **2012**, *40*, D302-5, doi:10.1093/nar/gkr931.
83. Lupas, A. Prediction and analysis of coiled-coil structures. *Meth. Enzymol.* **1996**, *266*, 513–525.
84. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680, doi:10.1093/nar/22.22.4673.
85. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739, doi:10.1093/molbev/msr121.
86. Nei, M.; Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **1986**, *3*, 418–426, doi:10.1093/oxfordjournals.molbev.a040410.
87. Hu, B.; Jin, J.; Guo, A.-Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297, doi:10.1093/bioinformatics/btu817.
88. Dong, J.; Chen, C.; Chen, Z. Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response. *Plant Mol. Biol.* **2003**, *51*, 21–37, doi:10.1023/a:1020780022549.
89. Sakuma, Y.; Maruyama, K.; Qin, F.; Osakabe, Y.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc Natl Acad Sci USA* **2006**, *103*, 18822–18827, doi:10.1073/pnas.0605639103.
90. Ohme-Takagi, M.; Suzuki, K.; Shinshi, H. Regulation of ethylene-induced transcription of defense genes. *Plant Cell Physiol.* **2000**, *41*, 1187–1192, doi:10.1093/pcp/pcd057.
91. Johannesson, H.; Wang, Y.; Hanson, J.; Engström, P. The *Arabidopsis thaliana* homeobox gene *ATHB5* is a potential regulator of abscisic acid responsiveness in developing seedlings. *Plant Mol. Biol.* **2003**, *51*, 719–729.
92. Abe, H.; Yamaguchi-Shinozaki, K.; Urao, T.; Iwasaki, T.; Hosokawa, D.; Shinozaki, K. Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* **1997**, *9*, 1859–1868, doi:10.1105/tpc.9.10.1859.
93. Kagaya, Y.; Ohmiya, K.; Hattori, T. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* **1999**, *27*, 470–478, doi:10.1093/nar/27.2.470.
94. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
95. Montero-Pau, J.; Blanca, J.; Esteras, C.; Martínez-Pérez, E. M.; Gómez, P.; Monforte, A. J.; Cañizares, J.; Picó, B. An SNP-based saturated genetic map and QTL analysis of fruit-related traits

 doi:10.20944/preprints202001.0048.v1

in Zucchini using Genotyping-by-sequencing. *BMC Genomics* **2017**, *18*, 94, doi:10.1186/s12864-016-3439-y.

96. Voorrips, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78, doi:10.1093/jhered/93.1.77.