

RESEARCH ARTICLE

Open Access

Identification and distribution of the NBS-LRR gene family in the Cassava genome

Roberto Lozano¹, Martha T Hamblin², Simon Prochnik³ and Jean-Luc Jannink^{1,4*}

Abstract

Background: Plant resistance genes (R genes) exist in large families and usually contain both a nucleotide-binding site domain and a leucine-rich repeat domain, denoted NBS-LRR. The genome sequence of cassava (*Manihot esculenta*) is a valuable resource for analysing the genomic organization of resistance genes in this crop.

Results: With searches for Pfam domains and manual curation of the cassava gene annotations, we identified 228 NBS-LRR type genes and 99 partial NBS genes. These represent almost 1% of the total predicted genes and show high sequence similarity to proteins from other plant species. Furthermore, 34 contained an N-terminal toll/interleukin (TIR)-like domain, and 128 contained an N-terminal coiled-coil (CC) domain. 63% of the 327 R genes occurred in 39 clusters on the chromosomes. These clusters are mostly homogeneous, containing NBS-LRRs derived from a recent common ancestor.

Conclusions: This study provides insight into the evolution of NBS-LRR genes in the cassava genome; the phylogenetic and mapping information may aid efforts to further characterize the function of these predicted R genes.

Background

In the tropics, cassava (*Manihot esculenta*) is the third biggest source of carbohydrates after rice and maize, feeding almost a billion people daily (www.fao.org). Most importantly, it is one of the major food crops in sub-Saharan Africa. However, two viral diseases threaten cassava productivity: Cassava Mosaic Disease (CMD) and Cassava Brown Streak Disease (CBSD) [1,2]. While these viruses were previously associated with lowlands, a new variant of the Cassava Brown Streak virus was found recently to infect this crop at altitudes above 1000 m [3]. In Uganda, the disease is pandemic, and its devastating effects, makes this virus a major concern for food security in central and east Africa [4]. Because of these diseases, an understanding of the molecular basis of disease resistance in cassava is a priority. As a first step, we have used the cassava genome sequence to identify and classify members of a major class of disease-resistance genes.

Studies in model plant species have shown that, unlike vertebrates, plants lack a somatic adaptive immune system [5]. To resist pathogens, plants have developed an advanced innate immune system consisting of a multiple layered network of defense proteins. One of these layers, Effector-Triggered Immunity (ETI), acts inside the cell via proteins encoded by a class of defense genes called R genes [6]. The most common disease resistance genes cloned to date are those belonging to the NBS-LRR family, named after the domains they typically contain: the nucleotide binding sites (NBS) and the leucine-rich repeat (LRR). This highly conserved gene family has structural and functional homology to the mammalian nucleotide-binding oligomerization domain (NOD)-LRR protein family, which functions in inflammatory and immune responses [7,8].

The NBS domain is part of the larger ~300 amino acid NB-ARC domain and contains strictly ordered motifs [9]. The NBS region binds and hydrolyzes ATP and GTP and primarily works as a signal transduction switch following pathogen recognition. LRR domains typically consist of 20–30 amino acid repeats that are often implicated in protein-protein interaction and, more precisely, bind to pathogen-derived molecules [10]. The LRR domain is thought to be the primary determinant of

* Correspondence: jj332@cornell.edu

¹Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, USA

⁴United States Department of Agriculture, Agricultural Research Service (USDA-ARS) R.W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA

Full list of author information is available at the end of the article

pathogen recognition specificity [11-13]. NBS-LRR proteins can recognize a wide variety of taxonomically unrelated pathogens, including viruses, bacteria, fungi, and even insects [14]. Activation of these genes results in a hypersensitive response (HR), a localized form of host-programmed cell death [15].

Resistance genes encoding NBS domains can be further classified into two major groups according to the presence or absence of different domains in the N-terminal region. The first group is comprised of proteins carrying the TOLL/interleukin-1 receptor (TIR) and are named TNL proteins (for TIR-NBS-LRR). The second, non-TIR-NBS-LRR group is usually known as CNL (for CC-NBS-LRR), because most of its members encode a coiled-coil (CC) N-terminal domain. Despite the name, we can find members of this group with Zinc finger or RPW8 domains instead of a coiled-coil [16-18]. This division is reflected in both phylogenetic analysis and their signaling pathways [19]. Both TIR and CC domains are involved in downstream specificity and signaling regulation [20].

While molecular techniques can be used to analyse NBS-LRR genes in plants lacking a genome sequence [21], the increasing number of sequenced plant genomes has facilitated the study of the NBS-LRR family in dicots and monocots, including *Arabidopsis thaliana* [22], *Arabidopsis lyrata* [23], *Oriza sativa* [24,25], *Vitis vinifera* [26], *Glycine max* [27], *Malus domestica* [28], *Solanum tuberosum* [29,30], and *Solanum lycopersicum* [31]. In most of these studies, the NBS-LRR genes exist in large, diverse families that are clustered on the genome [32,33].

The genomic clustering of R genes is thought to facilitate rapid R gene evolution in plant genomes via recombination. These clusters vary in size and complexity and fall into two types based on the phylogenetic relationship of their members. Commonly, clusters contain closely-related genes (same recent ancestor) of the same type, but they can also be heterogeneous, with NBS-LRR genes that are phylogenetically distant from each other (i.e., clusters can contain both TNL and CNL genes) [14,34].

In a recent effort to accelerate functional R gene discovery in cassava, several Resistant Gene Analogs (RGA) were identified using molecular techniques [35]. The *Manihot esculenta* genome comprises 12,977 scaffolds (L50 = 258,147 bp) [36] and together with gene annotations, and the genetic map [37], represent powerful tools for identifying and mapping resistance genes.

Among the 30,666 annotated protein-coding genes, we identified 228 belonging to the NBS-LRR family. Annotation of functional domains, physical position, as well as expression profiling and phylogenetic analysis was performed on these genes. Our results provide significant insights into the evolution of this gene family in the

cassava genome, and the results also generated an extensive R gene database that will accelerate future efforts for disease resistance breeding in this crop.

Methods

Cassava genome resources

The whole v4.1 genome assembly of the AM560-2 genotype comprising 12,977 scaffolds, as well as the whole genome annotation (30,666 genes), were downloaded from Phytozome [38] (<http://www.phytozome.net/> accessed on 01/24/2014). Subsequently, a genetic map was used to anchor scaffolds from v4.1 onto the genetic map, creating 18 pseudomolecules (*M. esculenta* v5.0, <http://phytozome.jgi.doe.gov>).

Identification of NBS-LRR genes

Predicted proteins from the cassava genome were scanned using HMMER v3 [39] using the Hidden Markov Model (HMM) corresponding to the Pfam [40] NBS (NB-ARC) family (PF00931; <http://pfam.sanger.ac.uk/>). From the proteins obtained using the raw NBS HMM, a high-quality protein set (E-value < 1×10^{-20} and manual verification of an intact NBS domain) was aligned and used to construct a cassava-specific NBS HMM using hmmbuild from the HMMER v3 suite. This new cassava-specific HMM was used, and all proteins with an E-value lower than 0.01 were selected. NBS-LRR genes were further filtered based on manual curation and functional annotation against both the closest homolog from *Arabidopsis* and the UNIREF100 sequence database. Most of the proteins that were removed had at least a partial kinase domain, but no relationship to NBS-LRR genes; this result was expected because the NBS domain has smaller kinase subdomains (Additional file 1).

NBS-associated conserved domains

NBS-encoding resistance genes usually have additional domains such as TIR, CC, or RPW8 in the N-terminal domain and a variable number of LRR domains in the carboxy-terminal region [5]. Conserved, associated domains were identified using a hmmpfam comparison to Pfam v27 [40]. The raw TIR HMMs (PF01582), RPW8 (PF05659), and LRR (PF00560, PF07723, PF07725, and PF12799) were downloaded (<http://pfam.xfam.org>) and used to mine the previous NBS-encoding gene candidates to identify distinct domains. Results were confirmed using both the NCBI Conserved Domains Tool [41] and Multiple Expectation for Motif Elicitation (MEME) [42]. Paircoil2 was used [43] with a P score cut-off of 0.03, because coiled-coil domains cannot be identified through conventional Pfam searches (Additional file 2).

Identification of partial NBS-LRR genes

Due to the rapid evolution of the NBS-LRR family, our pipeline might not identify some genes that belong to

the NBS-LRR cluster, but which have lost the NBS domain, or a large part of it. To try to identify all of these genes, we used an in-house script to download all the proteins from NCBI that included an “NBS-LRR” tag in their names. Later these proteins were formatted as a BLAST database. The remaining proteins from the cassava annotation were searched with BLAST [44] against this database. We kept high similarity genes as partial genes that could be pseudogenes caused by deletion, insertion, or frameshift mutation.

Alignment and phylogenetic tree estimation

We conducted this analysis to confirm the separation between the two main NBS-LRR groups in cassava and to learn about the phylogenetic history of the genes within each main branch. The NB-ARC domain region for every protein that carried a full-length NBS, as revealed by MEME [42], was extracted (counting 250 aa after the p-loop). Sequences with less than 90% of the full-length NB-ARC domain were excluded from posterior analysis. The multiple alignment was performed using clustalW [45] on 157 full NBS-domain cassava genes under default parameters. The resulting alignment was manually curated using Jalview [46], and poorly aligned regions at both ends were trimmed. A phylogenetic tree was then inferred in MEGA6 [47] by using the Maximum Likelihood method based on the Whelan and Goldman + freq. Model [48]. The tree with the highest log-likelihood was selected. Initial trees for the heuristic search were obtained by applying the Neighbour-Joining method to the matrix of pairwise distances estimated using a JTT model [49]. The nodes were tested by bootstrap analysis with 1000 replicates. Two additional trees were constructed using the same methodology, but which included reference resistance NBS-LRR genes from other species (Additional file 3). All trees were rooted using the NBS domain of the Human apoptotic protease-activating factor-1 (APAF-1).

Anchoring NBS-LRR genes to the cassava pseudomolecules

The NBS-LRR candidate genes were mapped to their physical position in the cassava genome using the cassava pseudomolecule assembly v5 and the genes from annotation v4.1 (Phytozome, <http://phytozome.jgi.doe.gov>). Genes were mapped to their position in the pseudomolecule file using Blast+ [50]. Only the top hit was considered (full coverage of both query and subject). CIRCOS [51] and Mapchart [52] were used for visualization.

Genes were arranged in different clusters. As described previously [29], an NBS-LRR cluster is defined as two or more NBS-LRR genes that are closer than 200 kb and separated by no more than eight non-NBS-LRR genes.

To test the statistical significance of this definition, we compared the cluster frequencies of NBS-LRR genes with the mean cluster frequencies obtained from 1000 iterations of a random sample of genes. Each random sample consisted of 205 genes, which is the same number of NBS-LRR genes that is anchored to its chromosomal positions.

Expression analysis of NBS genes under biotic stresses

RNAseq data were obtained from two experiments. The first measured changes in the transcriptome after infecting plants with Cassava Brown Streak Virus (CBSV) [53]. This study was focused on detecting genes involved in the steady state defence response by carrying out a transcriptome analysis 12 months after graft inoculation of CBSV. In the experiment, leaf samples were collected from three CBSV-inoculated and control plants, and two cassava genotypes were used, Kaleso (resistant to CBSV) and Albert (susceptible). RNA was sequenced using the Illumina HiSeq 2000 platform to generate 50 bp single end reads. BWA aligner was used to map the reads against the cassava genome. FPKM (Fragments per Kilobase of exon per Million fragments mapped) were calculated for each gene, but only transcripts showing an FPKM > 1 were kept for further analysis. Differential expression was calculated using the R package DEGseq [54].

The second RNAseq experiment measured the transcriptome response of a susceptible cassava plant (cultivar MCOL1522) infected with both a pathogenic and a non-pathogenic strain of *Xanthomonas axonopodis* pv. Manihotis (causal agent of Cassava Bacterial Blight) [55]. Two biological replicates were performed and RNA samples were collected at 0, 5, and 7 days post inoculation. RNAseq was run using Illumina technology to give 100 bp pair-end reads. FPKM values for all annotated cassava genes were obtained using cufflinks v2.0.2 [56]. Differential expression was calculated using NOISeqBIO v2.6.0 [57].

Differentially expressed genes from both experiments were scanned for NBS-LRR genes. It was expected that these genes were overexpressed during infection if they were contributing to the response against the pathogens.

Availability of supporting data

Phylogenetic raw data are available through the Data Dryad digital repository, doi:10.5061/dryad.tp030. Nucleic acid and protein sequences for every gene presented in this article are available in the Phytozome v10.1 repository, <http://phytozome.jgi.doe.gov> (Manihot esculenta v4.1).

Results

Identification of NBS-LRR genes

The cassava-specific HMM for the NBS-LRR domain identified 490 gene candidates. This initial dataset was

filtered based on several criteria (Additional file 1). Finally, a total of 228 non-redundant NBS-encoding R gene candidates were identified in the v4.1 release of *Manihot esculenta* genome, as well as 99 partial genes (without the NBS domain) (Table 1, Additional file 4). Analysing each NBS-LRR candidate allowed us to classify them into the TNL or CNL families (Table 1). Proteins belonging to the CNL group include 117 with full-length domains (CC, NBS and LRR). However, 64 proteins from this group lacked a domain and were classified as follows; N_{CC} (10, only NBS domain from the CC type), CN (11, N terminal domain and NBS, but lacking the LRR), NL_{CC} (43, NBS and LRR from the CC type, but lacking the N terminal domain). The remaining 47 genes belonged to the TNL group and were distributed as follows: TNL (29), N_{TIR} (4), TN (5), NL_{TIR} (9).

The average number of exons among the full-length NBS-encoding genes (CNL and TNL) in the cassava genome was 3.35, a value that is approximately half the average number of exons among all predicted cassava genes (6.17). As expected from previous studies, the average number of exons from the TNL family was higher (5.48) than those in CNL genes (2.82). Moreover, 35% of all the CNL genes were encoded by a single exon. This result is consistent with *Arabidopsis thaliana*, *Malus domestica*, or *Brassica rapa*, where CNLs and TNLs have 2.2, 2.3, 3.4, and 5.3, 5.2, 6.4 exons per gene on average, respectively [18,22,28].

Phylogenetic analysis

To study the evolutionary relationships among the newly discovered NBS-LRR genes, we built a phylogenetic tree using the conserved NB-ARC domain. Predicted NBS-LRR genes that contained no or partial NB-ARC domain were excluded. Alignment of the amino acid sequences revealed the NBS-subdomains, including the p-loop, kinase-2, kinase-3, and GLPLA. This alignment also showed a previously reported diagnostic site [58] that can differentiate CNL and TNL proteins right after the kinase-2 sub-domain (Additional file 5). As expected, the phylogenetic tree separated TNL and CNL genes into two different clades (Figure 1). For clarity, we labelled the genes with their type and chromosome position. The TNL clade is comprised of 33 genes, including several incomplete genes (NL_{TIR}, N_{TIR}). These genes are distributed among nine chromosomes (Figures 1 and 2), with a relatively high density on chromosome 17. On the other hand, the CNL clade has three main groups: CC(I), CC(II), and a separate clade that includes those proteins encoding an RPW8 (Resistant to powdery mildew in *A. thaliana*) domain. This strong separation has not been observed in previous studies where the RPW8s genes grouped together with the CC(II) group [29].

For comparative purposes, we included well characterized and manually curated resistance genes from *Arabidopsis thaliana*, *Cucumis melo*, *Hordeum vulgare*, *Solanum tuberosum*, and *Zea mays*, among others

Table 1 NBS-LRR genes and their classification in different genomes

Predicted domains	Letter code	<i>Manihot esculenta</i>	<i>Oriza sativa</i>	<i>Populus trichocarpa</i>	<i>Arabidopsis thaliana</i>
CNL Type					
NB	N _{CC}	10	25	27	1
CC NB	CN	11	7	19	5
CC NB LRR	CNL	117	160	119	51
NB LRR	NL _{CC}	43	16	71	4
Other	-	N/A*	264	43	1
TNL Type					
NB	N _{TIR}	4	-	19	-
TIR NB	TN	5	3	13	21
TIR NB LRR	TNL	29	-	64	83
NB LRR	NL _{TIR}	9	-	12	2
Other	-	N/A*	N/A*	15	39
Partial	P	99	60	N/A*	N/A*
Total		327	535	402	207
# of genes in the genome		30666	37544	45550	25498
Genome size		~760 Mb	389 Mb	485 Mb	125 Mb

Source: This article IRGSP (2005); (Yang et al. [26]); (Meyers et al. [22]); (Zhou et al., [25]).

*Not Analyzed.

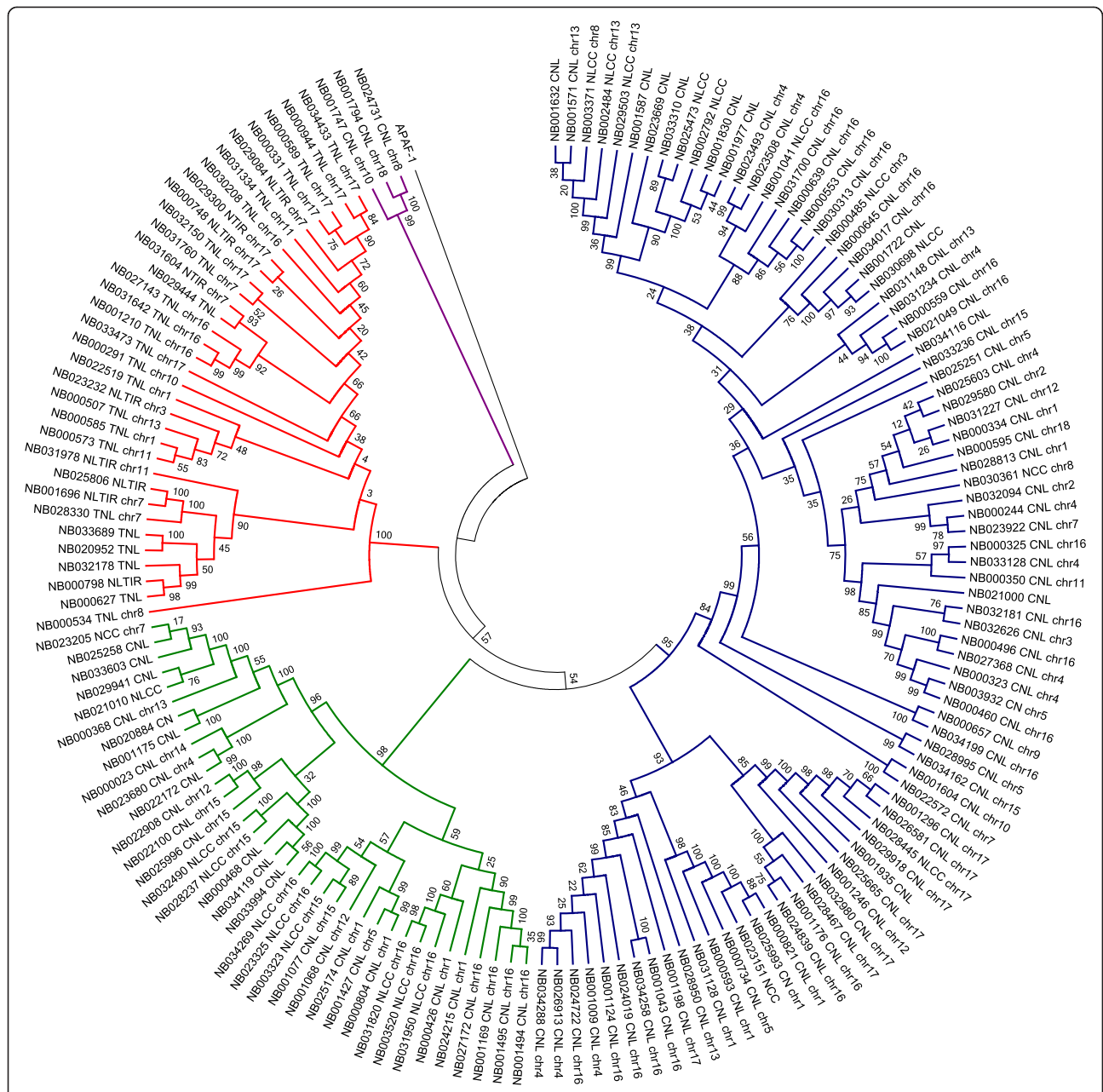
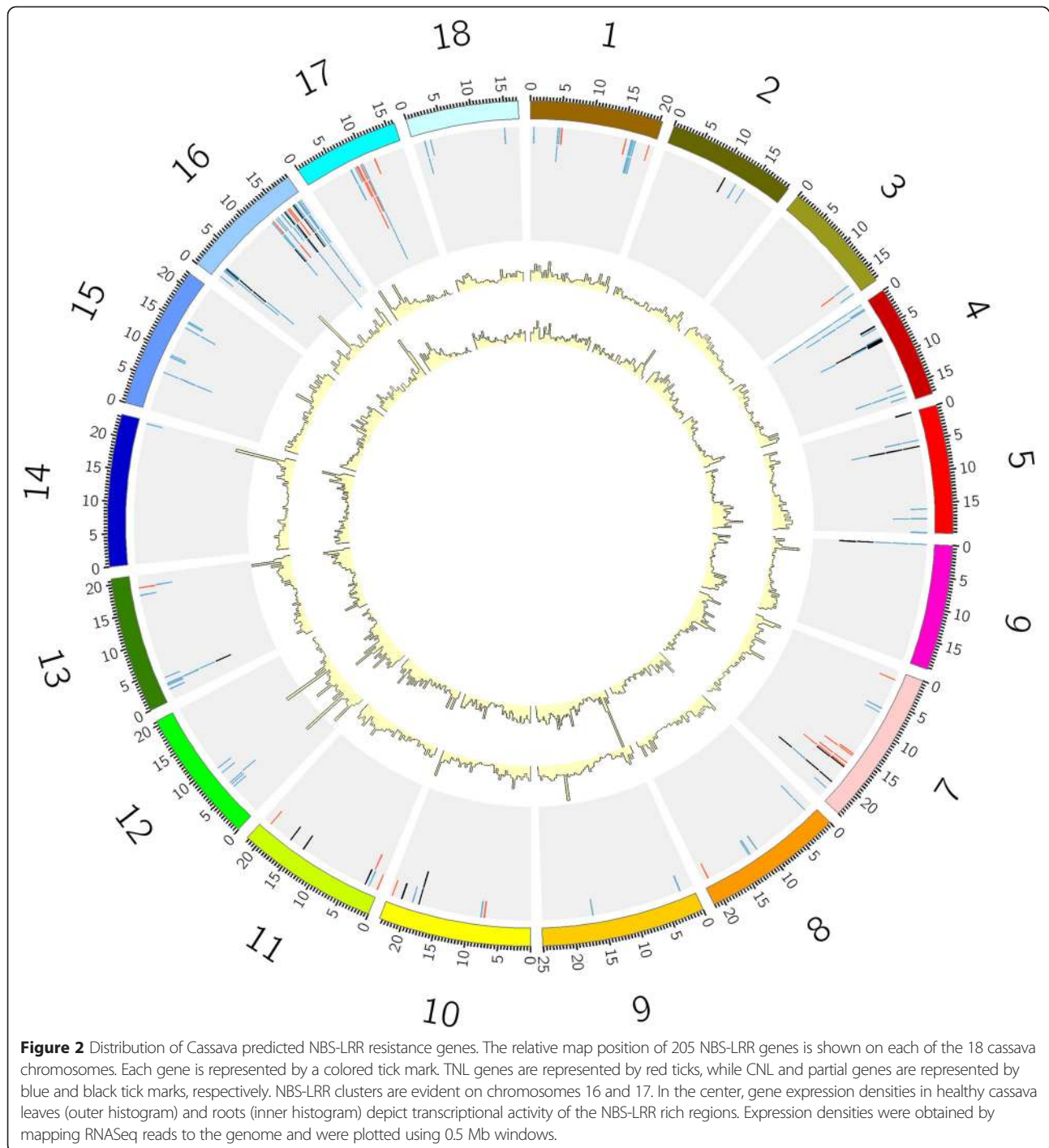


Figure 1 Phylogenetic reconstruction of the NBS-LRR proteins in *Manihot esculenta*. A maximum likelihood tree was constructed using 157 NBS domains. Percent bootstrap values (1000 iterations) are indicated in every branch. Each protein is encoded as follows: NB + ID number (same as phytozome ID) + Domains present (TNL, TN, NLTIR, NLTLR, CNL, CN, NCC, NLCC) + chromosome assignment (if available). Red, green, purple, and blue correspond to TIR, CNL-2, RPW8, and CNL-1 clades, respectively. APAF-1 was used as an outgroup.

(Additional file 3) into a second phylogenetic tree (Figure 3, Additional file 6). Most of the clades grouped as previously observed. All the TNL reference genes grouped into the TNL cluster (red), including Gro1.4 (*Solanum tuberosum*), N (*Nicotiana glutinosa*), and KR1 (*Glycine max*); however, these proteins tended to cluster separately from other TNL cassava genes. RPS4 (*Arabidopsis thaliana*), for example, clustered separately from all

the other TNL members, as was previously reported in potato [30].
 The CC-1 clade (blue) harbored more than half of the total NBS-LRR genes, and most of the reference R genes clustered inside this group as well. The introduction of the reference genes, however, influenced the topology of the tree, and resulted in a division within this group into two separate clades, CC-1a and CC-1b (Figure 3,



Additional file 6). CC-1a grouped 58 cassava CNL genes. As was observed inside the TNL group, most of the reference R genes tended to cluster apart from the cassava resistance candidates, although two functionally validated genes, FOM-2, (*Cucumis melo*) and PI8 (*Helianthus annuus*), showed sequence similarity to some cassava genes. FOM-2 clustered together with NB000657 and NB034199 with high bootstrap support, while PI8 was part of a sub-branch that

contained several cassava genes. The CC-1b clade had 27 cassava genes; when adding the reference R genes, the topology of this subgroup broke apart (Additional file 6). Most of the reference genes in this clade belonged to grass species (*Hordeum vulgare*, *Oriza sativa*, or *Triticum aestivum*) and, thus, it was not a surprise that none of these clustered together with any cassava genes. The NBS-LRR family in grasses has a markedly different evolution that is

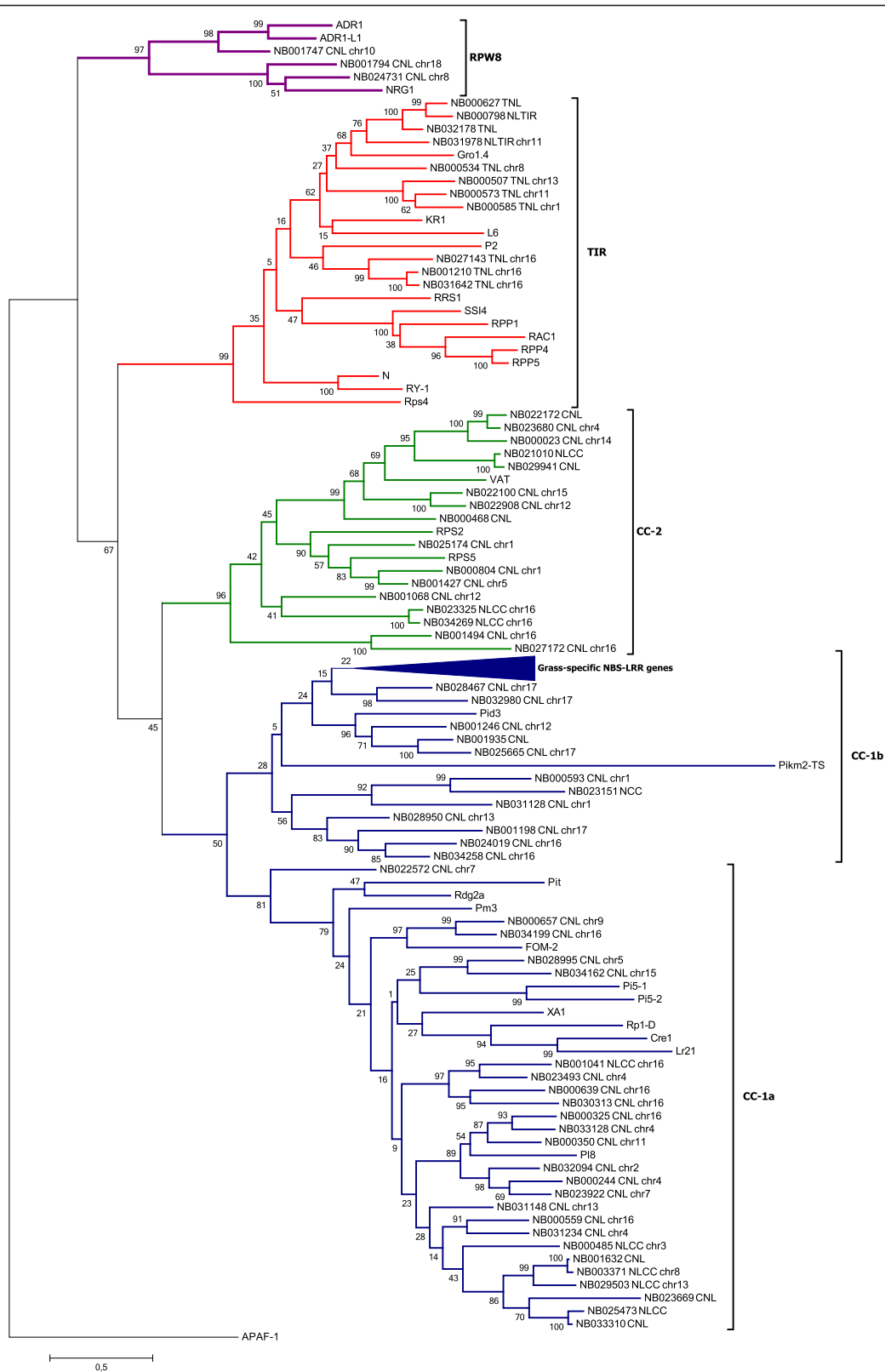


Figure 3 (See legend on next page.)

(See figure on previous page.)

Figure 3 Phylogeny of a subset of cassava NBS-LRR proteins with functional resistance genes from other species. Phylogenetic analysis of the NBS domain was carried out by the Maximum likelihood method using cassava NBS-LRR proteins as well as cloned functional R genes from different species including *Arabidopsis*, rice, and more (Additional file 3). For ease of visualization, we only used a subset of cassava genes that represents each clade (see Additional file 6 for the full tree). Red, green, purple, and blue represent TIR, RPW8, CNL-2, and CNL1 clades, respectively. The bottom clade in the CNL1-1b group is compressed, because it groups resistance proteins that are specific to grasses and does not cluster together with any cassava NBS-LRR protein. Members of this clade include *Pi* genes from *Oryza sativa*, *Lr* genes from *Triticum aestivum*, and *Mla* genes from *Hordeum vulgare*, for example.

represented by a significant underrepresentation of TNL genes [59].

The RPW8 clade (purple), containing three cassava genes, clustered with two ADR1 genes from *Arabidopsis* and the N-required gene 1 (NRG1) from *Nicotiana benthamiana*. Two sub-clades were evident; one contained ADR1 genes and NB001747, and the other contained NRG1, NB001794, and NB024731.

The last clade, CC-2 (green, previously reported as CNL-B, [22]), contained 36 cassava members. Only three reference genes fell into this group: Virus aphid Transmission (VAT, from *Cucumis melo*) and Resistance to *Pseudomonas syringae* protein 5 and 2 (RPS5 & RPS2). This group was previously reported as part of the CCR clade [29], although we did not obtain the same result.

Gene mapping

Physical chromosomal positions were established for 205 (~63%) of the NBS-LRR genes (the rest are on unanchored scaffolds) using their nucleotide sequences and the v5 assembly, and visualized using Circos and Map Chart (Figure 2, Additional file 7, Additional file 8). CNL genes were present on all the cassava chromosomes with at least one representative, while the distribution of TNL genes was more limited, having genes on only 9

chromosomes (Figure 2). We must consider, however, that 37% of the genes remain unmapped, so these estimates may be inaccurate.

It is clear (Figure 2, Figure 4) that the distribution of NBS-LRR genes is not even among the chromosomes and that they tend to form clusters. This clustered arrangement has been thought to facilitate sequence exchange through recombinational mispairing [34]. To identify NBS-LRR clusters, we used a previous definition [29] that a NBS-LRR cluster has two or more genes that are closer than 200 kb and separated by no more than eight non-NBS-LRR genes. Using this approach, we identified 39 clusters containing 143 NBS-LRR genes. Thus, 62 (30%) are singleton genes that do not map near other resistance genes. The size of the clusters varied across the genome from 2 to 10 members; the clusters can be classified further as homogeneous or heterogeneous based on how related the members of each cluster are (Additional file 9).

Chromosome 16 has the highest number of R genes (40, ~20% of mapped genes) distributed in 9 clusters plus 9 singletons. The number of members per cluster in this chromosome varies from 2 to 10. Cluster 35, for example, contains 10 genes belonging to the CC-1a clade (Figure 1, Additional file 9) with homology to RGA-2

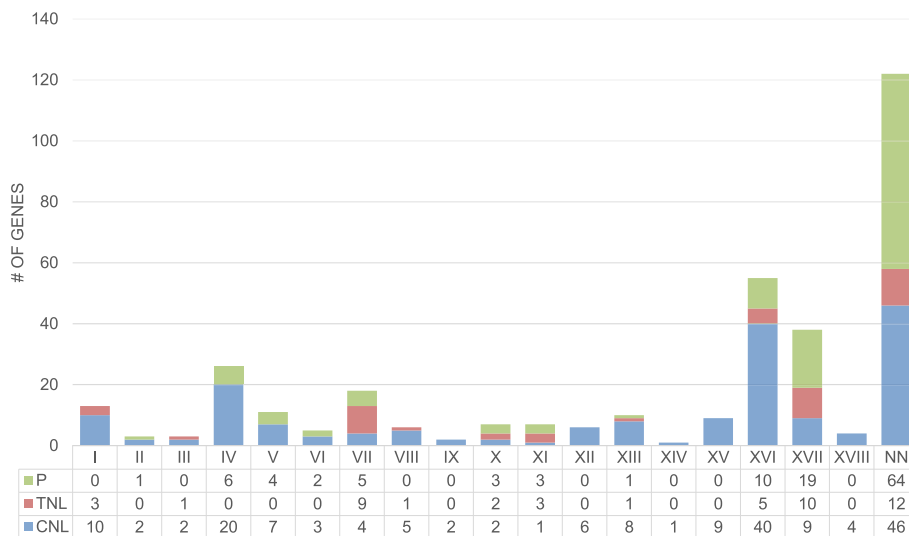


Figure 4 TNL, CNL, and partial genes distribution. Gene distribution for each class is shown across the cassava chromosomes. Bars are divided into CNL genes (blue), TNL genes (red), and partial genes (green).

(Resistance protein to *P. infestans* in tomato and potato). Multiple sequence alignment, followed by phylogenetic tree reconstruction of the proteins that belong to that group, shows that there are two subgroups within the cluster that represent two different origins (Additional file 9). Cluster 31 also carries 10 NBS-LRR genes, and belongs to the CC-2 clade with homology to putative resistance genes. There are only five TNL proteins in this chromosome, and only two of them are close enough to be considered a cluster by our criterion (cassava4.1_031642m and cassava4.1_001210m); these proteins encode close homologs to the TMV resistant protein N.

We only observed TNL clusters in chromosomes 7, 16, and 17. Most of the clusters comprise paralogs derived from the same recent common ancestor. It is less common to find TNL proteins clustered together with CNL proteins. In chromosome 17, for example, we found two neighbouring clusters (37 and 38) that encoded TNL genes; the first cluster carried 6 members of the TNL group with homology to TMV resistant protein N, and the second cluster contained 3 TNL and 6 CNL proteins. Two of the TNL genes lacked some domain (N_{TIR}, TN) and were very short, 633 and 480 bp, respectively, and the remaining gene, cassava4.1_027701m (975 bp), appeared to be a pseudogene caused by a frame-shift mutation. While these TNL proteins might be the remnants of previously functional genes that were defeated by pathogens, we cannot exclude the possibility of a sequencing/annotation error. The CNL proteins in the cluster belong to the CC-1b clade and are closely related.

We also checked the genome distribution of the RPW8-NBS-LRR proteins (purple clade in Figure 2). Three proteins belong to this clade, NB001747 (homologous to ADR1 *Arabidopsis* gene), NB001794, and NB024731 (homologous to NRG1 from *Nicotiana*). Genes that encode these proteins are distributed on chromosomes 10, 18, and 8, respectively, with sizes that range from 794 to 828. None of these genes are located close to another resistance protein. All the members of this clade have strong homologs in *Populus trichocarpa*, *Ricinus communis*, and *Jatropha curcas*.

Expression of NBS-LRR genes under biotic stresses

Recently a study on changes in the cassava transcriptome under Cassava Brown Streak Virus (CBSV) infection [53] found no significant differential expression of NBS-LRR genes in the Cassava genome one year after infection, either in a resistant or susceptible genotype. Only 235 NBS-LRR genes were identified in this study (based on conserved domains), contrasting with our finding of 327. We also found that some of the genes called as NBS-LRR in that study were miss-annotated and belonged to members of other families. FPKM (Fragments Per Kilobase of exon per Million fragments

mapped) values were obtained for the 327 NBS-LRR genes that we found and, confirming the observation of Maruthi et al., we saw no significant changes in expression between the infected and the control plant (Additional file 10).

Similar results were obtained in the plants infected with *Xanthomonas axonopodis* pv *manihotis* (causal agent of Cassava Bacterial Blight). In this experiment only one partial NBS-LRR cassava gene (cassava4.1_006209m) was differentially expressed during infection with the pathogen. There are however high expression of several of these genes across all conditions suggesting that they may still have a role in cassava's response to these pathogens.

Discussion

Cassava is a staple crop for millions of people in Africa, being their primary source of calories (FAO, 2003). This crop has a high yield potential under good conditions [60], yet it faces many biotic stresses. Given the importance of cassava, breeding for disease resistance is essential; the availability of the recently published cassava genome sequence allowed us to identify, classify, and map the NBS-LRR members, the biggest disease resistance gene family in plants.

According to our bioinformatics analysis the cassava genome carries a total of 327 NBS-LRR genes. From these, we annotated 99 as partial NBS-LRR genes that encode none or only a small part of the NBS domain, but they have high similarity with full-sized NBS-LRR genes. Partial genes may be the result of pseudogenization, given the rapid evolution of this gene family, but we cannot eliminate the possibility that these genes were incorrectly annotated because we did not perform a manual re-annotation to examine for sequencing errors. The 327 NBS genes found in the cassava reference genome represent 0.9% of the total number of coding sequences. The frequency of NBS sequences in the cassava genome fall within the range previously observed for other species (0.6% - 1.76%) [61].

No functional resistance genes have been cloned in cassava; genes found in this study, however, have strong homology with previously reported cassava Resistance Gene Candidates (RGCs) and NBS-LRR genes from other species. [35,58]. Lopez et al. [58] reported 12 Resistance Gene Candidates (RGC) in the cassava genome. The sequences for nine of these RGC regions were made available publicly. Eight of the nine RGCs aligned with >90% identity to NBS-LRR genes found in this study (Additional file 11). Additionally, the same study reported an RGC cluster at the end of linkage group J using a BAC library and RGC6 sequence as a probe. This region corresponded to the top of chromosome 4, which we found carries an NBS-gene cluster that contains the closest RGC6 homolog: cassava4.1_023508m (Additional file 12). More recently, Gedil et al. [35] reported

the sequence of several Resistance Gene Analogs (RGA) from different cassava varieties. All of the sequences reported as NBS-LRR-like were associated with 13 NBS-LRR sequences found in this study (Additional file 11).

Association studies and QTL identification for disease resistance are scarce in cassava. Most of these are related to Cassava Bacterial Blight (CBB), which is caused by different strains of the pathogen *Xanthomonas axonopodis* (*Xam*) [62-65]. One of these QTLs, which is associated with resistance to *Xam* strain CIO151 exclusively, explained 61% of the phenotypic variance and was located in linkage group U [65]. While the molecular marker associated with this QTL was not available, a nearby CAPS marker, DR11, is located at position ~16 Mb of chromosome 16, in the center of the largest NBS-LRR supercluster found in this study. Another major effect QTL was reported that confers resistance to Cassava Mosaic Disease (CMD) on cassava chromosome 8 [66]. This region, however, lacks any mapped NBS-LRR gene. More association studies on different diseases using different cassava genotypes may reveal a role for genes and clusters that we detected in this study.

Of the 228 full-length NBS-LRR genes, 181 belong to the CNL class, and 47 to the TNL class. This means that there are 3.8× more CNL than TNL genes. This ratio is indeed variable, and Leister (2004) [67] suggested that the over-representation of one of these groups could reflect the adaptation of the R genes to the predominant pathogens. For example, in *Oriza sativa* and *Sorghum bicolor*, members of the TNL family are present in a low frequency of approximately 1% [24,68]. In general, most grasses analysed contain only a few or no TNLs [59,69,70], which suggests that this class is specific for dicotyledons [28]. It is also interesting that most CNL genes from grasses presented in this study have no homologs among dicots (Figure 3, Additional file 6), which demonstrates that the evolution of NBS-LRR genes diverged significantly between monocots and dicots. Species of Brassicaceae, however, have a high percentage of TNLs: *Arabidopsis thaliana* (64%) and *Brassica rapa* (64%) [18,22]. Finally, there are some examples of ratios similar to what we found: in grapevine, for example, the proportion of CNL over TNL proteins is 3.8× [26] and in potato that ratio increases to 4.7× [29]. The over-representation of CNL in potato may be because CNL genes are typically responsible for resistance to *Pythophthora infestans* [29]. It was expected that the evolution of this family would be tightly linked with the pathogens affecting each species. Moreover, the rapid evolution of these genes may be visible among different cultivars from the same species in environments with different biotic stresses.

Previous studies showed that the CNL group forms two phylogenetic clades, the canonical one and the

CNL-R group, including members that encode an RPW8 domain in their N-terminal region. It is interesting, however, that the CNL branch in cassava does not include the RPW8 clade. We found that RPW8 genes were strongly separated from all other CNL genes, which was supported by strong bootstrap results (Figure 1, Figure 3). The RPW8 clade was described previously and referred to as CNL-A [22] or the CCR-NB-LRR encoding genes [71]. This family is thought to be one of the most ancestral of the major CC-NB-LRR clades [22], and it has been suggested to work differently than the more common CNL genes [71-73]. The ADR1 gene, present in this clade, is known to be an atypical CNL gene from *Arabidopsis*, which encodes abnormally conserved LRR domains and two conserved additional motifs in the NBS surroundings [74]. The homology and conservation of motifs is evident among proteins of this group, as shown by MEME (Additional file 13).

We tried to find close homologs from a set of known functional resistance genes (Figure 3, Additional file 6) for members of every clade, but there are a significant number of branches, especially in clade CC-2, that show no significant similarity to any of these well characterized genes. These genes might provide resistance to unknown cassava pathogens or may play a role in non-host resistance responses [75].

As mentioned previously, the cluster arrangement of NBS-LRR genes is considered to facilitate rapid gene evolution [76]. Several mechanisms have been proposed to contribute to the genomic diversity and distribution of this gene family: intragenic and unequal crossovers, gene conversion, positive and diversifying selection, and tandem duplications [67]. Most of the cassava NBS-LRR genes (70% of mapped genes) are located within a cluster; the biggest cluster is located in chromosome 16 with 10 CNL members. In homogeneous clusters like this, expansion is associated with tandem duplications. While most clusters are comprised of closely related genes, there are exceptions where members belong to different phylogenetic lineages; cluster 38 is an example, within which we found members of both TNL and CNL families. The formation of these heterogeneous clusters is thought to be the result of transposition, ectopic recombination, or chromosomal translocations [77]. As suggested before, this kind of genome evolution may be the result of positive selection for a higher complexity that can serve as the basis of new NBS-LRR – pathogen effector specificities [28,78].

In an effort to clarify the “cluster” definition, simulations were conducted to determine if the distribution observed in the NBS-LRR genes was caused by chance (see Methods). We observed that clusters of 2 and 3 genes occurred at the same frequency in a random sample of genes than when analysing NBS-LRR genes

(Additional file 14). For clusters containing more than 4 members, the difference in frequencies is clear, which suggests that, at least for cassava, only clusters with 4 or more members might be significant.

While the definitions used to detect clusters might be arbitrary, cassava NBS-LRR genes tend to lie in more evident superclusters, such as the 43 NBS genes on the end of chromosome 16 and the 19 genes in the middle of chromosome 17 (Figure 2). Collectively, these genes represent more than 30% of the total number of mapped NBS-LRR genes. Superclusters have been observed in other plants such as *Arabidopsis*, rice, and *Medicago*. In *Medicago*, an NBS-LRR supercluster represents more than 5% of all the genes present in the upper arm of the chromosome where it is located. In this scenario, the authors suggested that NBS superclusters may have played an important role in genomic remodelling during the evolution of those chromosome regions [16].

It is interesting that a high percentage of NBS-LRR genes are expressed constitutively in cassava leaves (72%). Moreover, 77% of the partial genes that might be considered as pseudogenes exhibit evidence of an RNA-seq expression. Whether these genes have an actual function or whether their expression is a temporary genome drag remains unclear. While not analysed in this study, the percentage of pseudogenes in NBS-LRR genes in plants can be very high. In rice, it was found to be as high as 55% [79]. Truncated NBS-LRR genes are often located close to intact NBS-LRR genes and are also clustered on specific chromosomes [16,30], a pattern that is followed commonly by the partial genes in cassava (black on Figure 2). The function of NBS-LRR pseudogenes are not well defined; they are usually only considered as genes that will be eliminated from the genome or sources of genetic diversity that may be used through recombination [19]. However, there may be a larger role for these genes. For example, in mice an expressed pseudogene played a role in maintaining the stability of its full-length homolog mRNA by interfering with the local silencing system [80]. In plants, truncated NBS-LRR peptides produced by alternative splicing (similar to what expressed pseudogenes look like) have a role in promoting disease resistance [81]. Uncovering the function of these expressed pseudogenes would be a major step to fully understanding plant-pathogen interactions.

Recently, studies of the cassava transcriptome under CBSV [53] and CBB [55] infection reported no significant differential expression of NBS-LRR genes among different time points and cassava genotypes. The lack of upregulation of these genes during infection is not surprising, and there are several explanations for this behaviour. Resistance to CBSV and CBB is considered to be quantitative and multigenic [1,53], so that NBS-LRR genes may not be involved in the resistance phenomena

at all. While this is a possibility, we should also consider that many NBS-LRR genes are expressed constitutively, meaning that NBS gene products will be present already in the plant cells to promote resistance even before the infection. Under this second scenario, NBS-LRR genes are not necessarily over-expressed to act in disease resistance. Additionally, when comparing susceptible and tolerant genotypes, gene expression might not be as relevant as the presence/absence of the specific resistance allele. We have to consider that the reference cassava genotype, AM560-2, is a partially inbred line derived from the Latin-American cassava cultivar MCOL-1505, that may lack NBS-LRR genes present in other genotypes. Moreover CMD and CBSV are recent diseases specific to Africa and are not present in the center of cassava domestication; comparing this analysis with some African genotypes would be valuable to see if evolution has caused divergence in the NBS-LRR gene family. Finally, high throughput methodologies, such as Resistance gene enrichment sequencing (RenSeq) [82] coupled with QTL or GWAS studies for other cassava diseases, would allow us to start mapping NBS-LRR clusters to specific pathogens.

Conclusions

We have identified 228 NBS-LRR type genes plus 99 partial genes related to the same family in the cassava genome. Information on the phylogeny of these genes and, most importantly, their physical positions on the chromosomes represent a valuable tool in future efforts to identify novel functional resistance genes in different cassava genotypes and other *Manihot* species. High throughput genotyping can also serve to explore the diversity of these regions across different genotypes. This kind of analysis would help decipher the recent evolution and dynamics of NBS-LRR genes in this clonally propagated crop.

Additional files

Additional file 1: NBS domain identification pipeline. The process used for the identification of proteins encoding and NBS domain using hmm-pfam is presented.

Additional file 2: NBS-associated conserved domains identification pipeline. The process used for the identification of NBS-associated conserved domains using hmm-pfam is presented.

Additional file 3: List of reference R genes from different species. This file contains the list of the functional R genes proteins used as reference in the phylogenetic trees. Plant Resistance Gene database (prgdb.crg.eu/wiki) IDs, names, donor species, and protein type are shown.

Additional file 4: List of identified NBS-LRR genes. Includes the ID for all the NBS-LRR genes annotated in this study with additional information. Size of the protein in amino acids, best hit against the *Arabidopsis* genes, top hit after blast against the UNIREF 100 database, family code, domains present in

the gene, chromosome assignment, and the code used for the phylogeny analysis.

Additional file 5: NBS multiple alignment and subdomain conservation. A subset of CNL and TNL NBS domains was aligned to show the conserved subdomains; p-loop, kinase-2, kinase-3, and GLPLA.

Additional file 6: Phylogenetic tree plus reference R genes. A tree was calculated using the same parameters as in Figure 2, but using all the cassava NBS-LRR genes that carry a full NBS domain.

Additional file 7: Position of anchored NBS-LRR genes and cluster assignment. This file shows the genome position and cluster assignment of all the anchored genes across the 18 cassava chromosomes.

Additional file 8: Detailed position of each NBS-LRR gene on the chromosomes. TNL genes are shown in red, CNL on blue, and partial genes on black.

Additional file 9: CNL cluster with 10 members. The 10 genes are clustered together in a ~500 kb region. While the genes are very similar overall, there are two different sources of evolution (Red and Blue) as shown by DNA alignments b) and average distance tree. c) It is counterintuitive that members of red and blue groups are physically mixed. Moreover, the different "strand orientation" of the genes represents the complexity of evolution within NBS-LRR genes.

Additional file 10: FPKM values for NBS-LRR genes during CBSV infection. Expression values of NBS-LRR genes for control and CBSV infected plants are shown for two different cassava genotypes.

Additional file 11: Homology between NBS-LRR genes previously reported. Two different sets of NBS-LRR genes previously reported for cassava (Lopez et al. [58]; Gedil et al. [35]) were compared to the ones found in the present study using Blast.

Additional file 12: NBS-LRR cluster co-localizing with previously reported cluster. Eleven NBS-LRR homologs found in the tip of chromosome 4 share the same position as the previously proposed NBS-LRR cluster in linkage group J (Lopez et al. [58]).

Additional file 13: RPW8 motif conservation as revealed by MEME. Conserved motifs are shown as inferred by MEME. RPW8 genes show a remarkable conservation of the motifs that encode LRR domains (motifs 5, 6, 8, and 4).

Additional file 14: Stochastic cluster assignment simulation. Number of clusters per size were calculated for both the NBS-LRR genes and a simulated data set consisting of 1000 iterations of 205 random cassava genes.

Abbreviations

NBS-LRR: Nucleotide Binding-site and leucine-rich repeat; CC: Coiled-coil; TIR: Toll/interleukin 1 receptor; APAF-1: Human apoptotic protease-activating factor-1; CNL: CC-NBS-LRR; TNL: TIR-NBS-LRR; FPKM: Fragments per kilobase of exon per million fragments mapped; CBSV: Cassava Brown Streak Disease; CBSV: Cassava Brown Streak Virus; CBB: Cassava Bacterial Blight; HMM: Hidden Markov Model; QTL: Quantitative Trait Loci.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RL carried out the research, and drafted the manuscript. MTH and SP provided critical insights and revised the manuscript. JLJ supervised the study and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the project "Next Generation Cassava Breeding Project" through funds from the Bill and Melinda Gates foundation and the Department for International Development of the United Kingdom.

Author details

¹Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, USA. ²Institute for Genomic Diversity, Biotechnology Building, Cornell University, Ithaca, NY 14853, USA. ³US Department of Energy, Joint Genome

Institute, Walnut Creek, CA 94598, USA. ⁴United States Department of Agriculture, Agricultural Research Service (USDA-ARS) R.W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA.

Received: 18 November 2014 Accepted: 20 April 2015

Published online: 07 May 2015

References

- Hillocks R, Jennings D. Cassava brown streak disease: a review of present knowledge and research needs. *Int J Pest Manag.* 2003;49:225–34.
- Patil BL, Fauquet CM. Differential interaction between cassava mosaic geminiviruses and geminivirus satellites. *J Gen Virol.* 2010;91(Pt 7):1871–82.
- Alicai T, Omongo C, Maruthi M. Re-emergence of Cassava Brown Streak Disease in Uganda. *Plant Dis.* 2007;91:24–9.
- Legg JP, Jeremiah SC, Obiero HM, Maruthi MN, Ndyetabula I, Okao-Okuja G, et al. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res.* 2011;159:161–70.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell.* 2006;124:803–14.
- Jones JGD, Dangl JL. The plant immune system. *Nature.* 2006;444:323–9.
- McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 2006;7:212.
- Ting JP-Y, Davis BK. CATERPILLER: a novel gene family important in immunity, cell death, and diseases. *Annu Rev Immunol.* 2005;23:387–414.
- Tameling WIL, Elzinga SDJ, Darmin PS, Vossen JH, Takken FLW, Haring MA, et al. The Tomato R Gene Products I-2 and Mi-1 Are Functional ATP Binding Proteins with ATPase Activity. *Plant Cell.* 2002;14(November):2929–39.
- Kobe B. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol.* 2001;11:725–32.
- Qu S, Liu G, Zhou B, Bellizzi M, Zeng L, Dai L, et al. The broad-spectrum blast resistance gene Pi9 encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. *Genetics.* 2006;172:1901–14.
- Ellis JG, Lawrence GJ, Dodds PN. Further analysis of gene-for-gene disease resistance specificity in flax. *Mol Plant Pathol.* 2007;8:103–9.
- Collier SM, Moffett P. NB-LRRs work a "bait and switch" on pathogens. *Trends Plant Sci.* 2009;14:521–9.
- McDowell JM, Woffenden BJ. Plant disease resistance genes: recent insights and potential applications. *Trends Biotechnol.* 2003;21:178–83.
- Lam E, Kato N, Lawton M. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature.* 2001;411(6839):848–53.
- Ameline-Torregrosa C, Wang B-B, O'Bleness MS, Deshpande S, Zhu H, Roe B, et al. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* 2008;146:5–21.
- Meyers BC, Morgante M, Michelmore RW. TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. *Plant J.* 2002;32:77–92.
- 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;282:617–31.
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 1999;20:317–32.
- DeYoung BJ, Innes RM. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat Immunol.* 2006;7:1243–9.
- Dracatos PM, Cogan NOI, Sawbridge TI, Gendall AR, Smith KF, Spangenberg GC, et al. Molecular characterisation and genetic mapping of candidate genes for qualitative disease resistance in perennial ryegrass (*Lolium perenne* L.). *BMC Plant Biol.* 2009;9:62.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR – encoding genes in Arabidopsis. *Plant Cell.* 2003;15(April):809–34.
- 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;157:757–69.
- Monosi B, Wissner RJ, Pennill L, Hulbert SH. Full-genome analysis of resistance gene homologues in rice. *Theor Appl Genet.* 2004;109:1434–47.

25. Zhou T, Wang Y, Chen J-Q, Araki H, Jing Z, Jiang K, et al. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics*. 2004;271:402–15.
26. 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;280:187–98.
27. Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, et al. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol*. 2012;12:139.
28. Perazzolli M, Malacarne G, Baldo A, Righetti L, Bailey A, Fontana P, et al. Characterization of Resistance Gene Analogues (RGAs) in apple (*Malus x domestica* Borkh.) and their evolutionary history of the rosaceae family. *PLoS One*. 2014;9, e83844.
29. Jupe F, Pritchard L, Etherington GJ, Mackenzie K, Cock PJA, Wright F, et al. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics*. 2012;13:75.
30. 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.
31. Andolfo G, Sanseverino W, Aversano R, Frusciantè L, Ercolano MR. Genome-wide identification and analysis of candidate genes for disease resistance in tomato. *Mol Breed*. 2013;33:227–33.
32. Marone D, Russo MA, Laidò G, De Leonardi AM, Mastrangelo AM. Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) genes: active guardians in host defense responses. *Int J Mol Sci*. 2013;14:7302–26.
33. Joshi RK, Nayak S. Perspectives of genomic diversification and molecular recombination towards R-gene evolution in plants. *Physiol Mol Biol Plants*. 2013;19:1–9.
34. Friedman AR, Baker BJ. The evolution of resistance genes in multi-protein plant resistance systems. *Curr Opin Genet Dev*. 2007;17:493–9.
35. Gedil M, Kumar M, Igwe D. Isolation and characterization of resistant gene analogs in cassava, wild Manihot species, and castor bean (*Ricinus communis*). *African J Biotechnol*. 2012;11:15111–23.
36. Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, et al. The cassava genome: current progress, future directions. *Trop Plant Biol*. 2012;5:88–94.
37. Cassava I, Map G. High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (*Manihot esculenta* Crantz) from 10 Populations. G3 (Bethesda). 2014;5:133–44.
38. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(Database issue):D1178–86.
39. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39(Web Server issue):W29–37.
40. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42(Database issue):D222–30.
41. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*. 2011;39(Database issue):D225–9.
42. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res*. 2006;34(Web Server issue):W369–73.
43. McDonnell AV, Jiang T, Keating AE, Berger B. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*. 2006;22:356–8.
44. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids Res*. 1997;25:3389–402.
45. Larkin M, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947–8.
46. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–91.
47. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9.
48. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 2001;18:691–9.
49. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. 1992;8:275–82.
50. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
51. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.
52. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered*. 2002;93:77–8.
53. Maruthi MN, Bouvaine S, Tufan HA, Mohammed IU, Hillocks RJ. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. *PLoS One*. 2014;9:e96642.
54. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*. 2010;26:136–8.
55. Muñoz-Bodnar A, Perez-Quintero AL, Gomez-Cano F, Gil J, Michelmore R, Bernal A, et al. RNAseq analysis of cassava reveals similar plant responses upon infection with pathogenic and non-pathogenic strains of *Xanthomonas axonopodis* pv. *manihotis*. *Plant Cell Rep*. 2014;33:1901–12.
56. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562–78.
57. Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. *Genome Res*. 2011;21:2213–23.
58. López CE, Zuluaga AP, Cooke R, Delseny M, Tohme J, Verdier V. Isolation of Resistance Gene Candidates (RGCs) and characterization of an RGC cluster in cassava. *Mol Genet Genomics*. 2003;269:658–71.
59. Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen J-Q, et al. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol Genet Genomics*. 2010;283:427–38.
60. Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, et al. Cassava genome from a wild ancestor to cultivated varieties. *Nat Commun*. 2014;5:5110.
61. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ. Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Mol Genet Genomics*. 2009;281:609–26.
62. Jorge V, Fregene MA, Duque MC, Bonierbale MW, Tohme J, Verdier V. Genetic mapping of resistance to bacterial blight disease in cassava (*Manihot esculenta* Crantz). *TAG Theor Appl Genet*. 2000;101:865–72.
63. Jorge V, Fregene M, Vélez CM, Duque MC, Tohme J, Verdier V. QTL analysis of field resistance to *Xanthomonas axonopodis* pv. *manihotis* in cassava. *TAG Theor Appl Genet*. 2001;102:564–71.
64. Wydra K, Zinsou V, Jorge V, Verdier V. Identification of pathotypes of *xanthomonas axonopodis* pv. *manihotis* in Africa and detection of quantitative trait loci and markers for resistance to bacterial blight of cassava. *Phytopathology*. 2004;94:1084–93.
65. López CE, Quesada-Ocampo LM, Bohórquez A, Duque MC, Vargas J, Tohme J, et al. Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. *Genome*. 2007;50:1078–88.
66. Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, Jannink J-L, et al. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res*. 2014;186:87–96.
67. Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet*. 2004;20:116–22.
68. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The Sorghum bicolor genome and the diversification of grasses. *Nature*. 2009;457:551–6.
69. Tarr DEK, Alexander HM. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res Notes*. 2009;2:197.
70. Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, et al. Genome-wide investigation on the genetic variations of rice disease resistance genes. *Plant Mol Biol*. 2006;62:181–93.
71. Collier SM, Hamel L-P, Moffett P. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol Plant Microbe Interact*. 2011;24:918–31.
72. Peart JR, Mestre P, Lu R, Malcuit I, Baulcombe DC. NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr Biol*. 2005;15:968–73.

73. Chini A, Loake GJ. Motifs specific for the ADR1 NBS-LRR protein family in *Arabidopsis* are conserved among NBS-LRR sequences from both dicotyledonous and monocotyledonous plants. *Planta*. 2005;221:597–601.
74. Xiao S, Calis O, Patrick E, Zhang G, Charoenwattana P, Muskett P, et al. The atypical resistance gene, RPW8, recruits components of basal defence for powdery mildew resistance in *Arabidopsis*. *Plant J*. 2005;42:95–110.
75. Schulze-Lefert P, Panstruga R. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends Plant Sci*. 2011;16:117–25.
76. Hulbert SH, Webb CA, Smith SM, Sun Q. Resistance gene complexes: evolution and utilization. *Annu Rev Phytopathol*. 2001;39:285–312.
77. Malacarne G, Perazzolli M, Cestaro A, Sterck L, Fontana P, Van de Peer Y, et al. Deconstruction of the (paleo)polyploid grapevine genome based on the analysis of transposition events involving NBS resistance genes. *PLoS One*. 2012;7:e29762.
78. Chen Q, Han Z, Jiang H, Tian D, Yang S. Strong positive selection drives rapid diversification of R-genes in *Arabidopsis* relatives. *J Mol Evol*. 2010;70:137–48.
79. Luo S, Zhang Y, Hu Q, Chen J, Li K, Lu C, et al. Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol*. 2012;159:197–210.
80. Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, et al. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature*. 2003;423:91–6.
81. Mastrangelo AM, Marone D, Laidò G, De Leonardis AM, De Vita P. Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci*. 2012;185–186:40–9.
82. Jupe F, Witek K, Verweij W, Sliwka J, Pritchard L, Etherington GJ, et al. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J*. 2013;76:530–44.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

