Mining whole genomes and transcriptomes of Jatropha (*Jatropha curcas*) and Castor bean (*Ricinus communis*) for NBS-LRR genes and defense response associated transcription factors

Archit Sood · Varun Jaiswal · Sree Krishna Chanumolu · Nikhil Malhotra · Tarun Pal · Rajinder Singh Chauhan

Abstract Jatropha (*Jatropha curcas* L.) and Castor bean (*Ricinus communis*) are oilseed crops of family Euphorbiaceae with the potential of producing high quality biodiesel and having industrial value. Both the bioenergy plants are becoming susceptible to various biotic stresses directly affecting the oil quality and content. No report exists as of today on analysis of Nucleotide Binding Site-Leucine Rich Repeat (NBS-LRR) gene repertoire and defense response transcription factors in both the plant species. *In silico* analysis of whole genomes and transcriptomes identified 47 new NBS-LRR genes in both the species and 122 and 318 defense response related transcription factors in Jatropha and Castor bean, respectively. The identified NBS-LRR genes and defense response transcription factors were mapped onto the respective genomes. Common and unique NBS-LRR genes and defense related transcription factors were identified in both the plant species. All NBS-LRR genes in both the species were characterized into Toll/interleukin-1 receptor NBS-LRRs (TNLs) and coiled-coil NBS-LRRs (CNLs), position on contigs, gene clusters and motifs and domains distribution. Transcript abundance or expression values were measured for all NBS-LRR genes and defense response transcription factors, suggesting their functional role. The current study provides a repertoire of NBS-LRR genes and transcription factors which can be used in not only dissecting the molecular basis of disease resistance phenotype but also in developing disease resistant genotypes in Jatropha and Castor bean through transgenic or molecular breeding approaches.

Introduction

Jatropha (*Jatropha curcas* L.) and Castor bean (*Ricinus communis* L.) are oilseed plant species that belong to family Euphorbiaceae. Both are economically important as sources of oil, which is used for the production of various industrial products like lubricants, cosmetics, medicines, including high quality biodiesel due to the presence of large proportion of unsaturated fatty acids. Off late the large-scale cultivation of selected genotypes of both the species have made them vulnerable to biotic stresses including diseases and pests, thereby affecting their oil yield potential [1, 2]. In the recent past, *Jatropha curcas* mosaic disease (JcMD) has been found to be prevalent in the plantations and is continuously reducing fruit yield and quality of Jatropha plants in the field [3–5]. Similarly fungal strains of *Alternaria alternate, Neoscytalidium dimidiatum*, *Botryosphaeria dothidea* and *Colletotrichum gloeosporioides* were reported to be responsible for infectious spots, root rot, black rot and anthracnose disease, respectively causing reduction in overall yield of Jatropha [6–9]. The viruses are also prevalent in Castor bean and reduce overall yield and quality [10–12]. Other pathogens like fungi and bacteria are also responsible for causing diseases in Castor bean. In Tanzania, fungal complex

A. Sood · V. Jaiswal · S. K. Chanumolu · N. Malhotra · T. Pal · R. S. Chauhan (✉)
Department of Biotechnology & Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173 215, H.P., India
e-mail: rajinder.chauhan@juit.ac.in

including *Alternaria* and *Fusarium* species were reported to be responsible for infection of inflorescence and capsules of Castor bean [13]. Diseases like fungal disease (*Verticillium* spp.) and bacterial disease (Wilt) are prevalent in Castor bean [14–16]. Management of various abiotic stresses in Jatropha and Castor bean is not economically viable through pesticides. Therefore, selection and development of disease resistant genotypes would be a sustainable strategy. However, no systematic breeding programme for disease resistance has been initiated as of today, both in Jatropha and Castor bean.

Plants have acquired resistance to many pathogens and pests due to the presence of disease resistance (R) genes that encode proteins which protect them from pathogenic organisms [17]. The research in the recent past on R-genes and downstream signal transduction mechanism has provided a strong base, which pave the way for their use in disease control [18, 19]. The bulk of R-genes in plants are from nucleotide binding site-leucine rich repeat (NBS-LRR) class, providing resistance to a large number of pathogens including parasites, fungi, bacteria, oomycetes, insects, and viruses [20–24]. NBS proteins are further classified into two sub categories based primarily on domains and motifs. Those having N-terminal domain with resemblance to the Toll and interleukin-1 receptors are designated as TIR proteins, and those without a TIR domain are categorized as non-TIR proteins [25]. A few of non-TIR proteins encode an N-terminal coiled-coil (CC) domain that may be involved in signaling and interaction of proteins [24, 26]. The NBS domains linked with both TIR and non-TIR proteins consist of a P-loop (Kinase-1), Kinase-2, Kinase-3, and some additional short motifs of unknown role [27]. The NBS domain functions by binding ATP [28], and the C-terminal leucine rich repeat (LRR) is implicated in pathogen binding and regulation of signal transduction [24, 25]. TIR domains are also involved in resistance specificity determination and signaling [25, 29].

All angiosperms have NBS-LRR encoding genes with differences between monocots and dicots. Although many NBS-encoding genes have been identified in *A. thaliana* that code for TIR domains [30], this subclass remain missing in cereal species [31, 32]. This finding suggests that since divergence which occurred >200 million years ago [33], TIR domain association with NBS-encoding genes was conserved in dicots but lost in monocot species.

The regulation of immunity and response to other stresses of plants in their natural habitat is enforced by a netwok of regulatory proteins or transcription factors which are considered as potential targets for engineering plant defense [34]. Transcription factors normally bind to the promoters of resistance genes and thus regulate their expression [35]. Many of the defense or disease resistance related transcription factors have been studied recently

including the TGA family of basic domain-leucine zipper (bZIP) proteins [36, 37], the MYB proteins [38], the ethylene responsive element binding factors [ERFs, having a DNA binding domain also reside in the APETALA2 (AP2) protein family], the WRKYs [39], and the Whirly family [34].

Identification and characterization of disease resistance genes, including NBS-LRRs is anticipated to accelerate the process of genetic improvement programmes and breeding for development of disease resistant varieties [40]. In the recent past, many resistance genes, including NBS LRR genes have been employed to produce genetically modified and transgenic disease resistant varieties. In case of Tobacco, N gene encoding TIR-NBS-LRR was transferred to develop transgenic lines which showed resistance to the mosaic virus [41]. Similarly, transgenic tobacco lines were developed using common bean TIR-NBS-LRR gene, RT4-4 exhibiting resistance towards mosaic virus from tomato or pepper [42]. In tomato, Bs2 gene encoding NBS-LRR protein has been transferred to develop resistance against bacterial spot disease [43]. Another gene responsible for bacterial blight resistance, Xa21 was introduced into Chinese rice varieties and the transgenic plants exhibited resistance to bacterial blight [44]. In case of wheat, the Pm3b gene has been introgressed which showed resistance against powdery mildew [45]. In another important study, RPS4 and RRS1, two NBS-LRR type R genes exhibited resistance to members of Brassicaceae and Solanaceae by providing immunity against various bacterial and fungal pathogens [46]. All these studies suggest that NBS-LRR genes hold potential in the development of disease resistant transgenics. There is no information as of today on identification of disease resistance genes (NBS LRR) or transcription factors regulating defense response in Jatropha and Castor bean, however, recent sequencing of their genomes and availability of transcriptomes [47, 48] have opened up avenues for detailed analysis of disease resistance genes, especially NBS-LRR genes and the transcription factors regulating defense response.

Whole genome-wide investigation of NBS-LRR resistance genes and transcription factors in a plant genome can therefore, provide novel insights about the overall resistance architecture. NBS-LRR genes number vary in different plant species, irrespective to the genome size. For example Arabidopsis genome (125 Mb) contains 149 genes [49], Rice genome (420 Mb) contains 535 genes [32], Potato genome (840 Mb) has 438 NBS-LRR genes [50], Soybean genome (1,115 Mb) comprise 319 NBS-LRR genes [51], Populus genome (500 Mb) having about 400 NBS-LRR genes [52] and Cucumber genome (350 Mb) contains 57 NBS-LRR genes [53].

Whole genome analysis of defense-related transcription factors have been done in many plant species. In Chinese

cabbage genome, 291 putative AP2/ERF transcription factor regulating resistance against disease and biotic stresses were identified [54]. In the model plant Arabidopsis, 118 transcription factors of families APETALA2/ethylene responsive element binding proteins, MYB domain-containing proteins, C2H2 zinc finger proteins and WRKY domain showing response to defense elicitor, Chitin were identified using Affymetrix *Arabidopsis* whole-genome array [55]. In another example, Soybean, an important crop species, biotic stress response related trihelix-GT and bHLH transcription factors were identified and characterised using in silico approach [56]. No transcription factors specific to defense response have been identified in Jatropha and Castor bean till date. Comparative studies have suggested that plant genomes encode hundred of NBS-LRR genes, but a vast diversity in the total number and distribution of NBS-LRR genes and subclasses is there [23]. Guo et al. [23] compared the NBS-LRR gene complement of *Arabidopsis thaliana* and its relative *Arabidopsis lyrata* with both the species containing similar numbers of NBS-LRR genes. Plocik et al. [57] compared the NBS domain sequences of NBS-LRR resistance genes from *Helianthus annuus* (sunflower), *Cichorium intybus* (chicory) and *Lactuca sativa* (lettuce), suggesting that Asteraceae species having different R-genes families, comprised of genes related to both toll-interleukin-receptor homology (TIR) and coiled coil (CC) domain containing NBS-LRR resistance genes. Between two closely related species, chicory and lettuce, CC subfamily composition similarity was identified, while sunflower showed less resemblance in structure. A genome-wide comparative analysis of NBS-LRR genes in *Sorghum bicolor* and *Oryza sativa* revealed species-specific expansion of NBS-LRR genes that may directly explain variations in disease susceptibility of the corresponding species [58]. In another recent genome wide comparative study, analysis of NBS-encoding genes in *B. oleracea*, *B. rapa* and *A. thaliana* was performed which identified 157, 206 and 167 NBS-encoding genes, respectively and provided deep understanding to the evolutionary history of NBS-encoding genes after deviation of *A. thaliana* and the *Brassica* lineage [59].

Jatropha and Castor bean belong to the same family, Euphorbiaceae, however detailed analysis of NBS-LRR genes w.r.t. their domain architecture, expression analysis, comparative number, genome organization, phylogenetic relationship and status of transcription factors regulating resistant genes is completely lacking as of today. Therefore, the whole genomes and transcriptomes of both the species were analyzed and compared to identify whole complement of NBS-LRR genes, their genome location, characterization into Toll/interleukin-1 receptor NBS-LRRs (TNLs) or coiled-coil NBS-LRRs (CNLs),

transcription factors specific to defense response or disease resistance.

Materials and methods

Data collection

The transcriptome data of Jatropha and Castor bean were downloaded from Sequence Read Archive (SRA) module of NCBI with accession nos. SRR087417 and ERA047687 respectively. The whole genomes of Jatropha and Castor bean were downloaded (ftp://ftp.kazusa.or.jp/pub/jatropha/;http://castorbean.jcvi.org/downloads.php). Velvet software [60] was downloaded from (http://www.ebi.ac.uk/~zerbino/velvet/) for assembly of transcriptome SRA files (NGS data). For similarity search, all the available NBS-LRR mRNA sequences were downloaded from the GenBank module of NCBI. Perl program, pfam_scan.pl and Pfam library of hidden Markov models (HMMs) of protein families were retrieved from Pfam website (http://pfam.janelia.org/) for domains prediction in protein sequences translated from transcripts.

Identification of Pfam domains/families associated with NBS-LRR genes and transcription factors related to disease resistance

The domains/families associated with NBS region were considered in the study due to the conserved nature of NBS region. Pfam keyword search with 'NBS-LRR' and associated key words (Table 1) was performed in the Pfam database (Supplementary Fig. 1). All hits of domains/families from the keywords were manually checked for their role in plant defense response and included in 'Master list 1' (NBS-LRR). Transcription factors important in disease resistance were retrieved from the literature. It was found that 10 different transcription factor families were involved in disease resistance (Table 2). All 10 transcription factor (TF) families were searched in Pfam text search to find out domains associated with each family. All hits of domains/families from the keyword were checked for their role as transcription factors for plant defense response.

Table 1 NBS-LRR domains and their respective Pfam Ids

NBS-LRR domains (keywords)	Pfam Ids
NB-ARC	PF00931
TIR	PF01582
TIR2	PF13676
RPW8	PF05659
Dirigent	PF03018

Table 2 Transcription factors involved in plant disease resistance

Transcription factor family	Pfam Ids
MYB	PF00249, PF13921, PF14379, PF13837, PF12776, PF14215
WRKY	PF03106
ERF-type/AP2-EREBP	PF00847
CBF	PF02312, PF00808, PF03914
bZIP	PF00170, PF03131, PF07716, PF12498
SBP/SPL6	PF03110
NAC domain/NAM	PF02365, PF14303
TFIIA	PF03153, PF02268, PF02751
Homeo-domain	PF00046, PF05920, PF00157, PF13384, PF13565
Whirly	PF08536

Table 3 Comparative distribution of NBS-LRR disease resistance genes between Castor bean and Jatropha genomes

Characteristics	Castor bean	Jatropha
Genome size	$\sim$ 380 Mb	$\sim$ 410 Mb
Number of NBS LRR disease resistance genes present	121	91
Unique domain/Superfamily	Dirigent, protein kinases	RPW8
Occurrence of NBS domain	$\sim$ 0.4 %	$\sim$ 0.3 %

Only those domains were incorporated in the 'Master list 2 (transcription factors)' which had significant functional role as TF in plant defense response.

Identification of NBS-LRR genes and defense response associated transcription factors

The NBS-LRR mRNA sequences collected previously from NCBI were mapped on to the transcriptomes of Jatropha and Castor bean using BLAST in order to identify all NBS-LRR containing transcripts in the transcriptome other than predicted 91 NBS-LRR genes in Jatropha [48] and 121 in Castor bean [47]. The manually adopted Pfam IDs (domains/families) associated with NBS regions and transcription factors were also mapped using Pfam domain/family search against the transcriptomes of both Jatropha and Castor bean in order to identify all genes and transcription factors having the domain IDs from the NBS regions and families of transcription factors, respectively (Supplementary Fig. 1).

Domain architecture of a protein can be explored through searching the sequence against the Pfam library of HMMs [61]. NBS-LRR genes and the transcription factors were identified according to domain architecture. All transcripts and transcription factors of Jatropha and Castor beans were translated into proteins (using canonical codon table) according to reading frames and then proteins were subjected to Pfam domain/family search to find out presence of domains. Finally, proteins matching with Pfam domains/families listed in the 'Master list 1' (refer section: Pfam NBS LRR domain identification) were selected as NBS proteins and corresponding transcript as NBS transcript. The proteins showing match to Pfam domains/families listed in the 'Master list 2' (refer section: Pfam NBS LRR domain identification) were selected as transcription factors associated with disease resistance. In-house PERL programs were used

to translate transcripts to proteins, Pfam domains prediction in translated proteins and comparison of predicted domains for their presence in Master list. Finally results were cross-checked manually (Table 3).

Location of NBS-LRR genes in sequence contigs

All predicted NBS-LRR genes of Jatropha and Castor bean were mapped to genome sequence contigs. BLAST search was used to map contigs on whole genomes with exact matching cut off. Position of contigs on genome was extracted from BLAST alignment output file. All analysis, BLAST search and extraction of contigs location were done through in-house developed PERL programs.

Identification of common and unique NBS-LRR genes and transcription factors in Jatropha and Castor bean genomes

To identify common and unique NBS-LRR genes and defense response related transcription factors between Jatropha and Castor bean, all predicted genes and transcription factors from both the species were used in similarity search. BLASTN was used for finding similarity among contigs of Jatropha and Castor bean with cut off values of equal to or more than 70 % within at least a length of 100 nucleotides (Supplementary Fig. 1). In-house PERL program was used to perform BLASTN and to extract results within mentioned cut off, further results were also cross checked manually.

Expression analysis of identified NBS-LRR genes and transcription factors

Transcript abundancy/quantification was carried out using RSEM [62]. RSEM is an user oriented software for quantification of transcript abundances from RNASeq Data. RSEM calculates abundance estimates and posterior mean estimates and 95 % credibility intervals for genes/isoforms. There are two measures which specify abundance estimates, one gives an estimate of the fragments number that can be derived from an isoform or gene [the expected counts (EC)], and the other

is the probable part of transcripts within the sample represented by the specified isoform or gene.

The expression profiles were obtained through pme_TPM (pme: Posterior mean estimates; TPM: transcripts per million) values. The TPM value is considered best over other metrics such as FPKM (Fragments Per Kilobase of transcript per million mapped reads) [63] and RPKM (reads per kilobase per million) [64] as it is not dependent on the mean expressed transcript length and so more comparable among diverse species and samples [62]. The transcript abundance of the contigs from the transcriptomes of Jatropha and Castor bean were calculated using pme_TPM parameter of RSEM package. All the parameters were kept default in the query option.

Identification of CNLs and TNLs in predicted NBS-LRR genes

Using PCOILS (http://toolkit.tuebingen.mpg.de/pcoils), the predicted NBS-LRR genes were further characterized into CNLs and TNLs with default parameters (Supplementary Fig. 1). PCOILS compares a sequence to previously identified parallel two-stranded coiled-coils and determines a similarity score.

Retrieval of disease resistance gene sequences of Castor bean and Jatropha

The NBS-LRR gene sequences of Castor bean were subjected to BLASTN against the Castor bean Database (http://blast.jcvi.org/erblast/index.cgi?project=rca1) and the database available at the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Similarly, the gene sequences of Jatropha were used as query against the Jatropha Genome Database where they were subjected to BLASTN (http://www.kazusa.or.jp/jatropha/cgi-bin/blast.cgi). This analysis was performed to predict position of these genes in respective sequence contigs. The protein sequences of both these plant genomes were also subjected to similar analysis using BLASTP.

Protein characterization, motif distribution and domain prediction

By using PCOILS, the predicted disease resistance proteins were characterized into CNLs and TNLs (http://toolkit.tuebingen.mpg.de/pcoils). The distributions of motifs in these proteins were predicted using MAST (http://meme.sdsc.edu/meme/cgi-bin/mast.cgi). Protein function domains of disease resistance genes were predicted using the NCBI Conserved Domain search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE) and the HMM search using Pfam (http://pfam.jouy.inra.fr/hmmsearch.shtml).

Results

Identification of NBS-LRR genes and defense response associated transcription factors in transcriptomes of Jatropha and Castor bean

In case of Jatropha, 45 potential NBS-LRR genes were identified by mapping Pfam domains and 7 by mapping NBS-LRR mRNA sequences with BLAST analysis out of which 5 showed common identity to both Pfam domain mapping and BLAST analysis (mRNA sequence mapping) (Supplementary file 1). Similarly in case of Castor bean, 47 potential NBS-LRR genes were identified where 44 genes were identified using mapping Pfam domains and 13 using mapping publicly available NBS-LRR mRNA sequences supported by BLAST analysis out of which 10 genes showed identity to both Pfam domain matching and BLAST hits (mRNA sequence mapping) (Supplementary file 2). A total number of 47 new NBS-LRR genes were identified in each Jatropha and Castor bean genomes in addition to previously identified NBS-LRR genes. All these newly identified NBS-LRR genes were confirmed through similarity search with previously reported NBS-LRR genes (Sato et al., 2011; Chan et al., 2010) for their uniqueness. Similarity search (70 % identity) provided that all identified NBS-LRR were new and not reported earlier. Similarly when Pfam domains specific to the Transcription factors involved in disease resistance or defense response were mapped on to the transcriptomes of both the species, 122 and 318 transcription factors were identified in Jatropha and Castor bean, respectively.

Location of NBS-LRR genes in genome sequence contigs

The identified NBS-LRR genes were mapped on to the respective genomes using BLAST search. In both Jatropha and Castor bean, the contigs showing multiple matches were manually curated to cover the entire query length. The query genes mapped on to the respective genome contigs showed identity in the range of 95–100 %. In case of Jatropha according to location, NBS-LRR genes were classified into three categories. First category had 28 genes, each of them were located in a single contig without any disruption in the coding sequence, thereby suggesting that these genes lacked introns. Second category had 7 genes and each were mapped on to single contigs with one or more gaps indicating insertion which may correspond to a intron. Third category comprised of 12 genes and each of these genes had match in more than one genomic contigs which implied that these genes were transcribed from different location and may have introns (Supplementary file 3). Out of 47 genes, 16 genes were found in cluster of two i.e. these 16 genes were present in 8 genomic contigs (Jc476461637, Jc476470256,

Table 4 Common NBS-LRR genes between Jatropha and Castor bean genomes

Transcript IDs of Castor bean	Transcript abundance (pme_TPM) of Castor bean	Transcript IDs of Jatropha	Transcript abundance (pme_TPM) of Jatropha	Percentage similarity
NODE_61167	2.27	contig_10049	9.16	78
NODE_57743	3.12	contig_10121	9.28	80
NODE_44733	2.65	contig_06992	3.91	86
NODE_50409	3.77	contig_00860	0.66	81
NODE_34103	10.46	contig_00810	0.64	80
NODE_44257	3.11	contig_00096	0.41	79
NODE_33311	5.99	contig_00073	0.40	82
NODE_25810	4.37	contig_00073	0.40	73

Jc476481852, Jc476483387, Jc476485273, Jc476487282, Jc476487650 and Jc476489371) (Supplementary file 3). Similarly in Castor bean, total 47 NBS-LRR genes were further classified into three categories according to location on the genome. In first category, there were 32 genes and each of them was located in a single contig without any insertion that means these genes lacked introns. Second category comprised 20 genes and each was mapped onto single contig with one or more gaps indicating insertion which may correspond to the intronic region. Third category had 5 genes with each having match in more than one contig indicating that these genes were transcribed from different locations and may have introns (Supplementary file 4). Clustered NBS-LRR genes were also observed in case of Castor bean, 4 genes were located on to a single contig (Rc124357167) i.e. were clustered together and 6 genes were found in cluster of two i.e. were present in 3 contigs (Rc124350718, Rc124354636 and Rc124357119) (Supplementary file 4).

Identification of common and unique NBS-LRR genes and transcription factors between Jatropha and Castor bean genomes

The identified 47 genes in Jatropha and Castor bean were aligned in order to analyze the common and unique NBS-LRR genes between Jatropha and Castor bean. This analysis was performed using BLASTN with cut off values i.e. identity >70 % and length of 100 bp. In order to identify common and unique genes, the genes from Castor bean were taken as database whereas the genes from Jatropha were taken as query. In case of Jatropha, 7 genes showed identity to Castor bean genes whereas in Castor bean 8 genes showed identity to Jatropha genes implying that 7 and 8 NBS LRR genes are common between Jatropha and Castor bean, respectively (Table 4). It was found that in case of Castor bean, out of 8 genes, 6 showed identity to each and specific gene from Jatropha whereas 2 genes from Castor bean showed similarity to 1 common (same) gene

from Jatropha. Further 40 and 39 genes were uniquely present in Jatropha and Castor bean genomes, respectively. In case of transcription factors (TF), castor bean TFs were taken as database whereas the TFs from Jatropha were taken as query. 70 transcription factors were found common in both Jatropha and Castor bean (Table 5). Further, 52 and 255 transcription factors were found to be uniquely present in Jatropha and Castor bean, respectively.

Transcript abundance of NBS-LRR genes and transcription factors associated with disease resistance

RSEM was used for the transcript abundancy measurements of identified set of NBS-LRR genes and transcription factors associated with the disease resistance mechanism. The expression profiles were obtained through pme_TPM (pme: Posterior mean estimates; TPM: transcripts per million) values using RSEM software package. In RSEM, posterior mean estimate (pme) is computed for each gene and isoform abundance, with a maximum likelihood (ML) estimate [62]. pme_TPM for both Jatropha and Castor bean transcriptome samples were generated. Expression profile of 47 NBS-LRR genes and disease resistance specific transcription factors was mined in both the species. The pme_TPM values of genes ranged between 0.4–133.54 and 2–62.84 in Jatropha and Castor bean, respectively (Supplementary files 5, 6). In case of Jatropha out of 47 genes, gene with id contig_14680 showed highest pme_TPM value i.e. 133.54 as compared to Castor bean where highest pme_TPM value was 62.84 from the gene NODE_36679. When common genes between Jatropha and Castor bean were analyzed, it was found that in Jatropha gene contig_10121 showed highest pme_TPM value of 9.28 while its corresponding gene in Castor bean i.e. NODE_57743 showed pme_TPM value of 3.12. In case of Castor bean, gene NODE_34103 showed the highest pme_TPM value i.e.10.46 as compared to Jatropha where the corresponding gene contig_00810 showed value of 0.64. The pme_TPM values for transcription factors ranged from

Table 5 Common disease resistance transcription factors between Jatropha and Castor bean

Transcription factor IDs of Castor bean	Transcript abundance (pme_TPM) of Castor bean	Transcription factor IDs of Jatropha	Transcript abundance (pme_TPM) of Jatropha	Percentage similarity
RcTF_54884	8.11	JcTF_00115	0.42	81
RcTF_19852	6.74	JcTF_00439	0.54	82
RcTF_53730	3.03	JcTF_00764	1.26	84
RcTF_44818	6.96	JcTF_00785	0.63	88
RcTF_20813	21.38	JcTF_00918	1.33	87
RcTF_20336	10.4	JcTF_00978	0.68	88
RcTF_58632	4.72	JcTF_01151	1.45	86
RcTF_16473	10.98	JcTF_01250	0.75	85
RcTF_41238	10.91	JcTF_01393	0.78	82
RcTF_24119	10.32	JcTF_01461	0.8	82
RcTF_69743	2.19	JcTF_01506	0.81	86
RcTF_49729	5.06	JcTF_01687	0.86	83
RcTF_24287	17.16	JcTF_01737	0.88	88
RcTF_28835	47.82	JcTF_01778	1.78	77
RcTF_42528	6.62	JcTF_01856	0.91	87
RcTF_26203	14.65	JcTF_01982	0.94	84
RcTF_32079	36.22	JcTF_02046	0.96	91
RcTF_39006	11.54	JcTF_02082	0.98	83
RcTF_2255	124.52	JcTF_02191	1	78
RcTF_29374	10.63	JcTF_02233	1.01	83
RcTF_37931	5.65	JcTF_02499	1.09	79
RcTF_22987	11.23	JcTF_02772	1.17	84
RcTF_21956	10.01	JcTF_02905	1.23	85
RcTF_14008	29.54	JcTF_03033	1.28	80
RcTF_16180	7.4	JcTF_03086	1.3	81
RcTF_6142	6.89	JcTF_03225	1.35	82
RcTF_36148	13.09	JcTF_03268	1.37	84
RcTF_35550	8.79	JcTF_03301	1.38	85
RcTF_49024	8.88	JcTF_03314	1.39	83
RcTf_1841	9.47	JcTF_03366	1.41	87
RcTF_19375	47.55	JcTF_03375	1.41	80
RcTF_92723	2.31	JcTF_03662	1.54	87
RcTF_49327	45.18	JcTF_03682	1.57	91
RcTF_47525	7.51	JcTF_04051	1.74	86
RcTF_45058	9.88	JcTF_04119	1.77	88
RcTF_30372	14.28	JcTF_04123	1.78	80
RcTF_20813	21.38	JcTF_04169	1.81	87
RcTF_59197	11.76	JcTF_04237	1.83	92
RcTF_30440	4.42	JcTF_04330	1.88	82
RcTF_58524	8.6	JcTF_04386	1.91	82
RcTF_32546	237.15	JcTF_04420	1.93	90
RcTF_43799	3.67	JcTF_04493	1.98	86
RcTF_53152	6.76	JcTF_04775	2.14	79
RcTF_2430	13.91	JcTF_04781	2.15	85
RcTF_70181	7.5	JcTF_04854	2.17	85
RcTF_68484	2.1	JcTF_05008	2.27	83
RcTF_41658	8.3	JcTF_05026	2.27	81

Table 5 continued

Transcription factor IDs of Castor bean	Transcript abundance (pme_TPM) of Castor bean	Transcription factor IDs of Jatropha	Transcript abundance (pme_TPM) of Jatropha	Percentage similarity
RcTF_12179	15.78	JcTF_05255	2.43	86
RcTF_30440	4.42	JcTF_05708	2.75	89
RcTF_48430	12.96	JcTF_05776	2.79	89
RcTF_31599	19.53	JcTF_05940	2.93	86
RcTF_11012	19.15	JcTF_05953	2.93	89
RcTF_61146	3.66	JcTF_06110	3.08	90
RcTF_35772	5.8	JcTF_06232	3.16	85
RcTF_76068	10.84	JcTF_06247	3.18	82
RcTF_40708	11.05	JcTF_06299	3.21	79
RcTF_38145	11.11	JcTF_06326	3.23	93
RcTF_35340	5.07	JcTF_06768	3.58	86
RcTF_55941	2.16	JcTF_07041	3.88	83
RcTF_30401	2.24	JcTF_08297	5.55	83
RcTF_49389	16.86	JcTF_08810	6.37	92
RcTF_40676	11.13	JcTF_08950	6.79	84
RcTF_47609	4.89	JcTF_09429	7.7	75
RcTF_51270	5.93	JcTF_10466	10.77	89
RcTF_24005	18.95	JcTF_11496	14.23	92
RcTF_41259	10.75	JcTF_11676	15.4	85
RcTF_64484	5.14	JcTF_12223	17.84	84
RcTF_18048	14.93	JcTF_12486	19.08	86
RcTF_30480	13.19	JcTF_12728	20.85	89
RcTF_20625	9.05	JcTF_14789	142.48	87

0.42 to 289.67 and 1.74 to 237.15 for Jatropha and Castor bean, respectively (Supplementary file 7, Supplementary file 8). Similarly, on analyzing the transcript abundancy of transcription factors it was found that in case of Jatropha the transcription factor JcTF_15319 showed highest pme_TPM value i.e. 289.67 and in case of Castor bean the transcription factor RcTF_32546 showed highest value (237.15) of pme_TPM. On analyzing the common transcription factors between Jatropha and Castor bean, it was found that in Jatropha the transcription factor JcTF_14789 showed highest pme_TPM value of 142.48 whereas its corresponding transcription factor in Castor bean RcTF_20625 showed a value of 9.05. In case of Castor bean, the transcription factor RcTF_32546 gave highest pme_TPM value of 237.15 while its counterpart in Jatropha i.e. JcTF_04420 showed a value of 1.93. Further it was observed that 4 transcription factors showed higher pme_TPM values in Jatropha (range from 100 to 300) i.e. JcTF_14789, JcTF_14930, JcTF_15218 and JcTF_15319. Also in castor bean 5 transcription factors showed higher pme_TPM values (range from 100 to 300) i.e. RcTF_29450, RcTF_10530, RcTF_2255, RcTF_33998 and RcTF_32546.

Identification of CNLs and TNLs in identified NBS-LRR genes

For the prediction of CNLs and TNLS in the identified transcripts, PCOILS used sliding windows of 14 (green), 21 (blue), and 28 (red) and predictions were made based on coiled coil probability. In case of Jatropha, out of 47 identified NBS-LRR genes, 37 were predicted as TNLs and 10 were CNLs. Similarly in case of Castor bean, out of 47 identified NBS-LRR genes, 28 were predicted as TNLs and 19 were predicted as CNLs.

Domain architecture of NBS-LRR genes in Castor bean and Jatropha genomes

The disease resistance gene sequences of both the plant genomes were downloaded from the NCBI. The predicted disease resistance genes were reported as 121 and 91 in Castor bean and Jatropha, respectively. The genes were further subjected to query against their respective databases using BLASTN and BLASTP. Protein function domains and distribution of motifs in the disease

resistance genes were predicted by using the NCBI Conserved Domain search along with HMM search using Pfam. MAST was used for predicting the distribution of motifs. These programs were appropriate for only defining the occurrence or absence of TIR, NBS, and LRR domains, but they were unable to identify more dispersed patterns or smaller individual motifs, like as present in the CC domain. By using PCOILS, the disease resistance proteins were characterized into CNLs and TNLs. Out of 121 disease resistance genes of Castor bean, 80 were predicted as TNLs, and 41 as CNLs. Similarly, out of 91 disease resistance genes of Jatropha, 54 were predicted as TNLs, and 28 as CNLs. No significant result was obtained for 9 other genes.

Organization of disease resistance genes in Castor bean and Jatropha genomes

In Castor bean, 121 disease resistance genes were distributed among 121 contigs. Similarly, in Jatropha, out of 91 genes, 82 genes were distributed among 82 contigs and no significant result was found for 9 other genes. Further analysis revealed that 7 of the disease resistance genes present in Castor bean genome, viz. XM_002517562.1, XM_002517561.1, XM_002518665.1, XM_002517526.1, XM_002517548.1, XM_002521759.1, and XM_002529578.1 showed similarity to Jatropha genome, corresponding to genes XP_002517608.1, XP_002517607.1, XP_002518711.1, XP_002517572.1, XP_002517594.1, XP_002521805.1, and XP_002529624.1, respectively (Supplementary Tables S1, S2). Due to the presence of same domains and motifs, the genes were further clustered in varying sizes, comprising 2–4 genes in most clusters. Although both these plants showed almost similar type of motifs (Kinase 1, Kinase 2, Kinase 3, GLPL, MHDL, and AAA+) which were found to be conserved in their disease resistance genes, certain differences were also observed with respect to the presence of conserved domains, which included presence of dirigent domain/superfamily along with protein kinase domain in Castor bean genome, and RPW8 domain/superfamily which was found to be unique to Jatropha genome (Table 3).

Discussion

Castor bean and Jatropha are considered as promising biofuel crops. Commercial cultivation of selected genotypes of both these plant species has predisposed to a plethora of biotic stresses, including insect pests and fungal, viral and bacterial diseases. No systematic breeding efforts have been made towards the development of disease resistant genotypes in both the plant species. Since a large numbers of pest and disease resistance genes in various plant species belong to NBS-LRR family of proteins, which is highly conserved across kingdoms, there was a requirement to analyze and characterize disease resistance genes, particularly NBS-LRR genes and the defense related transcription factors in both these plants.

Since the previously predicted NBS-LRR genes in Jatropha and Castor bean [47, 48] are quite small in number in comparison to other sequenced plant genomes with same range of genome sizes (For example, the genomes of *A. thaliana* and *V. vinifera* contain relatively higher number of NBS-LRR genes (ranging from 174 to 535), even though their genome sizes are in the order of 125 and 487 Mb, respectively [65]. We identified 47 new NBS-LRR genes in the transcriptomes of Jatropha and Castor bean from the available transcriptomes, while earlier identification of NBS-LRR genes was done through genome mining which may contain pseudogenes [47, 48]. The identified NBS-LRR genes in Jatropha and Castor bean were also mapped on to the respective genomes to have a clue about their physical location [66]. Some of the NBS-LRR genes can be frequently clustered in the genome due to segmental and tandem duplication [49, 67]. Consistent with these findings, some researchers have identified presence of NBS-LRR genes in clusters in Jatropha as well as Castor bean. In case of Jatropha, 16 genes were found in clusters of two genes whereas in castor bean 6 genes were in binary cluster and 4 genes were clustered in single contig. Our results indicate that there is more clustering in case of Jatropha as compared to Castor bean which may support the concept of novel resistance specificities through recombination or gene conversion and also rapid R gene evolution in Jatropha [40, 49]. Moreover the NBS-LRR genes present in clusters can be primarily targeted for breeding to develop disease resistant varieties. In both cases, several NBS-LRR genes were mapped with gaps which represent the presence of intronic region in these genes and is in consonance with the fact that most of the eukaryotic genes comprised of introns [68]. Further these intronic regions can be explored in spliced site studies for disease resistance [69–71].

Further, a comparative analysis between Jatropha and Castor bean showed common and unique NBS-LRR genes. It was found that 7 and 8 NBS-LRR genes were common between Jatropha and Castor bean, respectively. In Castor bean, out of 8 genes, 6 showed identity to each specific gene from Jatropha whereas 2 genes from Castor bean showed similarity to 1 common gene from Jatropha. The results are in line with the previous analysis of NBS-LRR genes and resistance gene analogues (RGAs) in Sweet potato and Arabidopsis which support the concept of recent duplication or have been conserved devoid of significant divergence [72, 73]. Common transcripts/genes can be

targeted in a cross generic or cross specific manner for enhancing the disease resistance potential of Jatropha and Castor bean [40, 54, 74]. The common genes identified from both organisms implies that these are conserved in nature and may be responsible for providing resistance to general disease conditions not specific to any particular pathogen.

The transcript abundancy was measured for newly identified set of NBS-LRR genes with the help of in silico expression analysis in order to support the identification of transcripts and their expression levels. A high variation was found in the expression values of identified genes in both Jatropha and Castor bean. The genes showing higher values of expression with more transcript abundance can be used to design and conduct the experiments for providing enhanced resistance to disease and pest conditions in Jatropha, Castor bean and related economically important plants of same family such as Rubber tree, Cassava, etc. [40, 74]. The identified NBS-LRR genes were further characterized into TNLs and CNLs and in both cases the number of TNLs were more compared to CNLs, as the TNLs were confined only to dicots [75]. Further, these N terminal domains i.e. TIR (TNLs) and CC (CNLs) were responsible for pathogen recognition which supports the resistance potential of the associated genes [76].

These investigations are the first attempt to identify transcription factors related to disease resistance or defense response in whole transcriptomes of Jatropha and Castor bean where 122 and 318 transcription factors were identified, respectively. Many of the transcription factors are being implicated in maintaining transcriptional reprogramming linked with plant defense and resistance response. An association among activating and repressing transcription factors from many families control the defense response expression of the target genes [77]. Transcription factors such as WRKY, bZIP, ERF, MYB and Whirly families bind to the promoters of the resistance genes and regulate expression level [34, 35, 37–39]. In comparison to conventional screening of cDNA libraries or EST sequencing, the computational transcription factors discovery approach provides fast, simple, consistent and precise methods to reveal the transcription factor families specific to disease resistance and defense response at both the whole genome and transcriptome levels. In case of Castor bean, the number of identified transcription factors related to the defense response is about 3 times more as compared to Jatropha as evident from their transcriptomes size. The comparison between the identified transcription factors was made between Jatropha and Castor bean in order to elucidate the common and unique number of transcription factors which showed that a large number of transcription factors (70) are common between Jatropha and Castor bean which also support the fact of either recent

duplication or conserved defense response mechanism in Jatropha and Castor bean [78–80].

In the past, many transgenic crop and model plants with improved disease resistance have been developed [81] by over expressing the defense related transcription factors. Over expression of WRKY and ERF transcription factors have resulted in developing disease resistant varieties of many plants [82]. Over expression of the defense associated transcription factors can provide resistance to many dissimilar pathogens also. Arabidopsis transcription factor MYB30 over expression has resulted in enhanced resistance to pathogenic bacteria and fungus in transgenic Arabidopsis and Tobacco [83]. Identification of transcription factors related to defense response or disease resistance is also of great significance in predicting the pathogen responsive promoter elements. Only a few pathogen responsive elements in the promoter regions have been identified. One most cited example is the presence of W-box in the promoter region of various genes activated by WRKY transcription factors [84–86]. In both Jatropha and Castor bean only 4–5 transcription factors showed the higher transcript abundance which signifies their role as potential targets for achieving or providing disease resistance. Those transcription factors can be considered on primary basis for manipulation of the genes associated to them to develop the resistant lines of Jatropha and Castor bean. A comparative study of varying expression profiles or variations in transcript abundance measurements of NBS-LRR genes and transcription factors associated to disease resistance between both the transcriptomes revealed that some NBS-LRR genes and transcription factors can be good candidates for enhancing the resistance potential of Jatropha and Castor bean. By using comparative analysis, the exploration of evolutionary fate of the NBS-LRR genes and transcription factors in the Euphorbiaceae family and the understanding of disease resistance between the important family members is anticipated [59].

A total of 121 disease resistance genes were predicted in the current version of Castor bean genome [47] in which 80 genes have been classified into TNLs and 41 into CNLs, which are the two important subfamilies of NBS-LRR proteins in plants, having Toll/interleukin-1 receptor (TIR) or coiled-coil (CC) motifs in the amino-terminal position of domain and also the NBS-LRR genes represent ∼0.4 % of all identified ORFs. Similarly, 91 disease resistance genes were predicted in Jatropha genome [48] in which 54 have been predicted as TNLs and 28 into CNLs and also the NBS-LRR genes represent ∼0.3 % of all predicted ORFs. Since the CNLs and TNLs are both involved in pathogen recognition [76], the prediction and classification of NBS-LRR proteins into CNLs and TNLs further support the disease resistance potential. The presence of TNLs is known exclusively only for dicots not for monocots [75]

which further support the motifs prediction as Jatropha and Castor bean both are dicotyledonous species. These results are in accordance with the previous classification of TNLs and CNLs for the novel identified NBS-LRR genes.

The detailed analysis revealed that 7 of the disease resistance genes present in Castor bean genome showed similarity to Jatropha genome, signifying that these genes emerged from the recent duplication or have been conserved devoid of significant divergence, as was found for NBS-LRR genes and RGAs in Sweet potato and Arabidopsis earlier [72, 73]. Furthermore, 60 % gene clustering was observed in both these plant species and the genes which were present in clusters consisted of same domains and motifs. Similar kind of motif patterns were observed in both these plants which also corroborates the concept of synteny [87], but certain differences with respect to the presence of conserved domains were also observed between two plant species, which included presence of dirigent domain/superfamily along with protein kinase domain in Castor bean genome, and RPW8 domain/superfamily in Jatropha genome.

The NBS-LRR genes, the defense related transcription factors predicted in this study and domain architecture of previously identified NBS-LRR genes will supplement the disease resistance knowledge pool in both the bioenergy plant species so that better breeding and genomics-based interventions can be made for developing disease resistant varieties. Further, these in silico based analysis and comparison of NBS-LRR genes and transcription factors between Jatropha and Castor bean will reveal specific insights on the function, organization, conservation and evolution of the NBS–LRR resistance genes and defense response related transcription factors in related members of family Euphorbiaceae.

Conclusion

The study has led to the identification of 47 new NBS-LRR genes, in addition to 91 in Jatropha and 121 genes in Castor bean and 122 and 318 disease resistance specific transcription factors in Jatropha and Castor bean respectively, for the first time. The outcome of study is, therefore, of great practical importance in two major oilseed crops of industrial value. Since Jatropha and Castor bean are becoming susceptible to various diseases and biotic stresses, current findings can be used in the development of candidate gene markers intended for molecular breeding of disease resistance. The transcription factors specific to resistance or defense response can be targeted to engineer disease resistant varieties of Jatropha and Castor bean which share taxonomical and biochemical similarity.

Data archiving statement

References

1. Jayanna K (2006) Studies on Jatropha mosaic virus disease. M.Sc. Dissertation, University of Agricultural Sciences, Dharwad
2. Tewari JP, Dwivedi HD, Pathak M, Srivastava SK (2007) Incidence of a mosaic disease in *Jatropha curcas* L. from eastern Uttar Pradesh. Curr Sci 93(8):1048–1049
3. Aswatha Narayana DS, Rangaswamy KS, Shankarappa MN, Maruthi MN, Reddy CNL, Rekha AR, Murthy KVK (2007) Distinct Begmoviruses closely related to cassava mosaic viruses causes Indian Jatropha Mosaic Disease. Int J Virol 3:1–11
4. Raj SK, Kumar S, Snehi SK, Pathre U (2008) First report of Cucumber mosaic virus on *Jatropha curcas* L. in India. Plant Dis 92:171
5. Gao S, Qu J, Chua NH, Ye J (2010) A new strain of Indian cassava mosaic virus causes a mosaic disease in the biodiesel crop *Jatropha curcas* L. Arch Virol 155:607–612
6. Kumar R, Sinha A, Singh SR, Kamil D (2009) Incidence of a leaf spot disease in *Jatropha curcas* L. from Eastern Uttar Pradesh. J Mycol Plant Pathol 39(3):536–538
7. Machado AR, Pinho DB, Dutra DC, Pereira OL (2012) First report of collar and root rot of physic nut (*Jatropha curcas* L.) caused by *Neoscytalidium dimidiatum* in Brazil. Plant Dis 96(11):1697
8. Srinivasa Rao CH, Kumari MP, Wani SP, Marimuthu S (2011) Occurrence of black rot in *Jatropha curcas* L. plantations in India caused by *Botryosphaeria dothidea*. Curr Sci 100(10):1547–1549
9. Kwon JH, Choi O, Kim J, Kwak YS (2012) First report of anthracnose disease on *Jatropha curcas* L. caused by *Colletotrichum gloeosporioides* in Korea. J Phytopathol 160:255–257
10. Grieco F, Parrella G, Vovlas C (2002) An isolate of *Olive latent virus 2* infecting castor bean in Greece. J Plant Pathol 84:129–131
11. Parrella G, De Stradis A, Vovlas C (2008) First report of Olive latent virus 2 in wild castor bean (*Ricinus communis* L.) in Italy. Plant Pathol 57:392
12. Raj SK, Snehi SK, Gautam KK, Khan MS (2010) First report of association of Cucumber mosaic virus with blister and leaf distortion of Castor bean (*Ricinus communis* L.) in India. Phytoparasitica 38:283–289
13. Gahukar RT, Seguni ZSK (2009) First record of severe attack of the fungal complex on castor in Tanzania. Int Pest Cont 51(1):26
14. Esfahani MN, Monazzah M (2011) Identification and assessment of fungal diseases of major medicinal plants. J Ornam Hortic Plants 1(3):137–145
15. Smith EF, Godfrey GH (1921) Bacterial wilt of Castor bean (*Ricinus communis* L.). J Agric Res 21(4):255–261

16. Sabet KA (1959) Studies in the bacterial diseases of Sudan crops IV. Bacterial leaf-spot and canker disease of mahogany *(Khaya senegalensis* (Desr.) A. Juss. and *K. grandifoliola* C. DC). Ann Appl Biol 47(658–665):14

17. Chauhan RS, Sood A (2013) Comparative genomics in Euphorbiaceae. In: Bahadur B, Sujatha M, Carels N (eds) Jatropha: challenges for a new energy crop, vol 2. Springer, New York, pp 351–374

18. Klessig DF, Durner J, Shah J, Yang Y (1998) Salicylic acid-mediated signal transduction in plant disease resistance. In: Romeo JT, Kelsey R, Verpoorte DR (eds) Recent advances in phytochemistry. vol 32. pp 119–137

19. Zhang Y, Goritschnig S, Dong X, Li X (2003) A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1. Plant Cell 15(11):2636–2646

20. Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature 14;411(6839):826-833

21. Lorang JM, Sweat TA, Wolpert TJ (2007) Plant disease susceptibility conferred by a "resistance" gene. Proc Natl Acad Sci USA 104:14861–14866

22. Tarr DE, Alexander HM (2009) TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res Notes 2:197

23. Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D (2011) Genome-wide comparison of nucleotide-binding site-leucine-rich-repeat-encoding genes in *Arabidopsis*. Plant Physiol 157:757–769

24. Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54:23–61

25. DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 7(12):1243–1249

26. van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol 45:43–72

27. van der Biezen EA, Jones JDG (1998) The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biol 8:226–227

28. Tameling WI, Vossen JH, Albrecht M, Lengauer T, Berden JA, Haring MA, Cornelissen BJ, Takken FL (2006) Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. Plant Physiol 140:1233–1245

29. Luck JE, Lawrence GJ, Dodds PN, Shepherd KW, Ellis JG (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. Plant Cell 12:1367–1377

30. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317–332

31. Monosi B, Wisser RJ, Pennill L, Hulbert SH (2004) Full-genome analysis of resistance gene homologues in rice. Theor Appl Genet 109:1434–1447

32. Zhou T, Wang Y, Chen J, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271:402–415

33. Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH (1989) Data of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201–6205

34. Desveaux D, Subramaniam R, Després C, Mess JN, Lévesque C, Fobert PR, Dangl JL, Brisson N (2004) A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*. Dev Cell 6:229–240

35. Rushton PJ, Somssich IE (1998) Transcriptional control of plant genes responsive to pathogens. Curr Opin Plant Biol 1:311–315

36. Fan W, Dong X (2002) In vivo interaction between NPR1 and transcription factor TGA2 leads to salicylic acid–mediated gene activation in *Arabidopsis*. Plant Cell 14:1377–1389

37. Alves MS, Dadalto SP, Gonçalves AB, de Souza GB, Barros VA, Fietto LG (2013) Plant bZIP transcription factors responsive to pathogens: a review. Int J Mol Sci 14:7815–7828

38. Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V, Bansal KC (2012) Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. BMC Genomics 10:544

39. Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiol 150(1648–1655):15

40. Marone D, Russo MA, Laidò G, Leonardis AMD, Mastrangelo AM (2013) Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. Int J Mol Sci 14:7302–7326

41. Liu Y, Schiff M, Marathe R, Dinesh-Kumar SP (2002) Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. Plant J 30:415–429

42. Seo YS, Rojas MR, Lee JY, Lee SW, Jeon JS, Ronald P, Lucas WJ, Gilbertson RL (2006) A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus specific manner. Proc Natl Acad Sci USA 103:11856–11861

43. Tai TH, Dahlbeck D, Clark ET et al (1999) Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. Proc Natl Acad Sci USA 96:14153–14158

44. Gao L, Cao Y, Xia Z, Jiang G, Liu G, Zhang W, Zhai W (2013) Do transgenesis and marker-assisted backcross breeding produce substantially equivalent plants?—a comparative study of transgenic and backcross rice carrying bacterial blight resistant gene Xa21. BMC Genomics 14:738

45. Mascher F, Matasci C, Kneubuehler Y, Brunner S, Schori A, Keller B (2012) Evaluation of disease resistance in wheat supplemented with *Pm3b*. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 15–18

46. Narusaka M, Kubo Y, Hatakeyama K, Imamura J, Ezura H, Nanasato Y, Tabei Y, Takano Y, Shirasu K, Narusaka Y (2013) Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. PLoS One 8:e55954

47. Chan AP, Crabtree J, Zhao Q et al (2010) Draft genome sequence of the oilseed species *Ricinus communis* L. Nat Biotechnol 28(9):951–956

48. Sato S, Hirakawa H, Isobe S et al (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res 18:65–76

49. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809–834

50. Jupe F, Pritchard L, Etherington GJ et al (2012) Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics 13:75

51. Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, Van K, Lee SH (2012) Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol 12:139

52. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F (2008) Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. Plant Mol Biol 66:619–636

53. Wan H, Yuan W, Bo K, Shen J, Pang X, Chen J (2013) Genome-wide analysis of NBS-encoding disease resistance in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. BMC Genomics 14:109

54. Song X, Li Y, Hou X (2013) Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp *pekinensis*). BMC Genomics 14:573

55. Libault M, Wan J, Czechowski T, Udvardi M, Stacey G (2007) Identification of 118 Arabidopsis transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. Mol Plant Microbe Interact 20:900–911

56. Osorio MB, Buecker-Neto L, Castilhos G, Turchetto-Zolet AC, Wiebke-Strohm B, Bodanese-Zanettini MH, Margis-Pinheiro M (2012) Identification and *in silico* characterization of soybean trihelix-GT and bHLH transcription factors involved in stress responses. Genet Mol Biol 35:233–246

57. Plocik A, Layden J, Kesseli R (2004) Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce and chicory. Mol Phylogenet Evol 31(1):153–163

58. Li Y, Tessaro MJ, Li X, Zhang Y (2010) Regulation of the expression of plant resistance gene SNC1 by a protein with a conserved BAT2 domain. Plant Physiol 153(1425–1434):16

59. Yu J, Tehrim S, Zhang F et al (2014) Genome-wide comparative analysis of NBS-encoding genes between Brassica species and *Arabidopsis thaliana*. BMC Genomics 15:3

60. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829

61. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL (2000) The Pfam protein families database. Nucleic Acids Res 28:263–266

62. Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinform 12:323

63. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515

64. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 5:621–628

65. Yang S, ZhangX Yue JX, Tian D, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genomics 280:187–198

66. Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. Nat Rev Genet 13:85–96

67. Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39:285–312

68. Deutsch M, Long M (1999) Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res 27:3219–3228

69. Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. Proc Natl Acad Sci USA 97:1908–1913

70. Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol 58:267–294

71. Tang S, Riva A (2013) PASTA: splice junction identification from RNA-sequencing data. BMC Bioinform 14(1):116

72. Chen G, Pan D, Zhou Y, Lin S, Ke X (2007) Diversity and evolutionary relationship of nucleotide binding site-encoding disease-resistance gene analogues in sweet potato (*Ipomoea batatas* Lam.). J Biosci 32:713–721

73. Chen Q, Han Z, Jiang H, Tian D, Yang S (2010) Strong positive selection drives rapid diversification of *R*-genes in *Arabidopsis* relatives. J Mol Evol 70:137–148

74. Wang HZ, Zhao PJ, Xu JC et al (2003) Virus resistance in transgenic watermelon plants containing a WMV-2 coat protein gene. Acta Genet Sin 30:70–75

75. KimTH Kunz HH, Bhattacharjee S et al (2012) Natural variation in small molecule-induced TIR-NB-LRR signalling induces root growth arrest via EDS1-and PAD4- complexed R protein VICTR in *Arabidopsis*. Plant Cell 24:5177–5192

76. McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. Genome Biol 7:212

77. Eulgem T (2005) Regulation of the *Arabidopsis* defense transcriptome. Trends Plant Sci 10:71–78

78. Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol 130:639–648

79. Voitsik AM, Muench S, Deising HB, Voll LM (2013) Two recently duplicated maize NAC transcription factor paralogs are induced in response to Colletotrichum graminicola infection. BMC Plant Biol 13:85

80. Chen L, Zhang L, Li D, Wang F, Yu D (2013) WRKY8 transcription factor functions in the TMV-cg defense response by mediating both abscisic acid and ethylene signaling in *Arabidopsis*. Proc Natl Acad Sci USA 110(21):1963–1971

81. Shin R, Han JH, Lee GJ, Paek KH (2002) The potential use of a viral coat protein gene as a transgene screening marker and multiple virus resistance of pepper plants coexpressing coat proteins of cucumber mosaic virus and tomato mosaic virus. Transgenic Res 11(215–219):17

82. Salomon D, Sessa G (2012) Biotechnological strategies for engineering plants with durable resistance to fungal and bacterial pathogens, Tel Aviv University, Israel 329–339

83. Vailleau F, Daniel X, Tronchet M, Montillet JL, Triantaphylides C, Roby D (2002) A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. Proc Natl Acad Sci USA 99:10179–10184

84. Kirsch C, Logemann E, Lippok B, Schmelzer E, Hahlbrock K (2001) A highly specific pathogen-responsive promoter element from the immediate-early activated *CMPG1* gene in *Petroselinum crispum*. Plant J 26:217–227

85. Heise A, Lippok B, Kirsch C, Hahlbrock K (2002) Two immediate-early pathogen-responsive members of the *At*CMPG gene family in *Arabidopsis thaliana* and the W-box containing elicitor-response element of *At*CMPG1. Proc Natl Acad Sci USA 99:9049–9054

86. Choi PS, Kim YD, Choi KM, Chung HJ, Choi DW, Liu JR (2004) Plant regeneration from hairy-root cultures transformed by infection with Agrobacterium rhizogenes in *Catharanthus roseus*. Plant Cell Rep 22:828–831

87. Rawal HC, Singh NK, Sharma TR (2013) Conservation, divergence, and genome-wide distribution of PAL and POX a gene families in plants. Int J Genomics, vol. 2013, Article ID 678969, 10 pages, 2013. doi:10.1155/2013/678969