Current Plant Biology 17 (2019) 2–16

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

Transcriptome profiling reveals genetic basis of disease resistance against *Corynespora cassiicola* in rubber tree (*Hevea brasiliensis*)

C. Bindu Roy[a,*], Haipei Liu[b], Anantharamanan Rajamani[a], Thakurdas Saha[a]

[a] Rubber Research Institute of India, Kottayam, 686 009, Kerala, India
[b] School of Agriculture, Food and Wine, The University of Adelaide, Adelaide, South Australia, Australia

ARTICLE INFO

Keywords:
Corynespora leaf disease
Rubber tree
RNA-Seq analysis
Digital gene expression
Disease resistance proteins

ABSTRACT

Corynespora leaf disease caused by *Corynespora cassiicola* (Berk. & Curt.) is one of the major diseases responsible for significant yield loss in rubber trees (*Hevea brasiliensis*). Next-generation sequencing based transcriptomic study of two rubber clones: RRII 105 (susceptible) and GT 1 (moderately resistant) were performed to understand the molecular basis of host tolerance to fungal diseases. Genes encoding disease resistance proteins, leucine-rich repeat proteins and genes involved in carbohydrate metabolic processes were significantly up-regulated in GT 1 upon infection, but were either completely suppressed or down-regulated in RRII 105. Transcription factor activity was a major molecular function triggered in both inoculated clones. Gene Ontology analysis revealed that majority of the transcripts was enriched for defense response, response to stimulus and stress. Higher expression of 118 transcripts with complete ORFs was identified in inoculated GT 1, indicating their possible role in disease resistance. In addition, both unique and common simple sequence repeats (SSRs) were identified. *In silico* analysis revealed 191 informative SSRs differentiating the two clones. Variant calling in control and disease GT 1 transcriptomes with reference to RRII 105 revealed over one lakh putative base substitutions. Microarray was used to validate the results obtained on transcriptional responses. Biotic stress overview from MapMan analysis revealed stronger activation of defense-related genes, receptor-like kinases and transcription factors. This study presents the first comprehensive transcriptome of resistant and susceptible rubber clones in response to *C. cassiicola*. The newly identified differentially regulated genes and sequence variation provide critical knowledge for understanding the genetic basis of disease resistance and marker development.

1. Introduction

Rubber tree (*Hevea brasiliensis*) is one of the world's most important strategic crops, producing latex of commercial utility. India ranks second in the productivity of natural rubber (NR) and sixth in production of NR in the world [1]. However, one of the major constraints in higher productivity of NR is the susceptibility of high yielding clones to fungal diseases. Corynespora leaf disease caused by *Corynespora cassiicola* is a major threat to NR production in South East Asian rubber growing countries, which contribute to about 98 per cent of total rubber production. The pathogen infects young refoliating leaves, leading to defoliation thereby extending the immaturity period of rubber trees. The widely cultivated clone RRII 105 developed at the Rubber Research Institute of India, although a high yielder is highly susceptible to *C. cassiicola*, whereas the Indonesian clone GT 1 is resistant to the pathogen in the existing Indian condition.

Plant resistance to diseases is associated with a number of defense responses, activated by the host upon infection with pathogens [2]. Identification of host genes involved in defense responses, is important to understand plant resistance mechanisms against phytopathogens [3,4]. Although we have effective fungicides and efficient spraying technology to combat the disease incidence, modern breeding methods employing biotechnological tools are also essential for minimizing crop loss due to disease, thereby increasing the productivity of rubber trees. Despite the fact that several expressed sequence tags (ESTs) have been generated and gene-based markers have been developed, functional genomics studies in rubber in response to fungal pathogens is still in its infancy. Differential display and subtractive hybridization were employed earlier to identify the disease responsive genes in rubber and characterize them *via* candidate gene approach [5]. This was laborious

* Corresponding author at: Plant Pathology Division, Rubber Research Institute of India, Kottayam, 686 009, Kerala, India.
E-mail addresses: binduroy@rubberboard.org.in (C.B. Roy), haipei.liu@adelaide.edu.au (H. Liu), ananth@rubberboard.org.in (A. Rajamani), saha@rubberboard.org.in (T. Saha).

https://doi.org/10.1016/j.cpb.2019.02.002
Received 10 January 2019; Received in revised form 6 February 2019; Accepted 24 February 2019

and time consuming, which resulted in identifying only a few transcripts that were differentially regulated during disease development.

Next-generation sequencing (NGS) technologies have opened new avenues into the field of genomic research for its unprecedented level of sensitivity and high-throughput nature [6]. Whole-transcriptome analysis with total RNA sequencing (RNA-Seq), combined with well-developed bioinformatics tools, provides a highly-reliable approach to study gene expression profiles [7]. It is often used to construct the complete transcriptome of an organism either by reference-based or *de novo* assembly [8,9]. Recently, *de novo* assembly of transcriptomes using short reads has been successfully employed to provide data of high accuracy [10–12]. RNA-Seq captures all the transcriptional responses in the cell and reflects the functional elements that are being actively expressed at a given time point. It provides a better approach to gene expression profiling that allows us to reveal transcriptional complexity at a faster pace [13]. Therefore, a comprehensive profile of host transcriptome during disease development is essential to decipher the genetic basis of disease resistance. However, such data has not been available for *Hevea - C. cassiicola* interaction.

In this study, we have undertaken the first global analysis of rubber transcriptome, using a resistant clone GT1 and a susceptible clone RRII 105, in both control and pathogen-challenged conditions. The digital expression of disease responsive genes were profiled and the molecular pathways involved were analysed. To get an in-depth understanding about the differentially regulated genes identified through the NGS study at each time point of pathogen infection and disease development, microarray experiment was performed. MapMan analysis for each time point further illustrated the critical pathways and genes involved at different stages of disease development. The data obtained will serve as a valuable public resource for disease resistance in rubber tree. A genomic resource is also generated for simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) markers, which is highly important for marker-assisted breeding for Corynespora leaf disease resistance in rubber.

2. Materials and methods

2.1. Fungal pathogen isolation and challenge inoculation

The pathogen *C. cassiicola* was isolated from infected leaves of rubber showing typical railway track symptom (Fig. 1). Isolation was made by plating surface-sterilized pieces of diseased tissue on potato dextrose agar medium and purified by single-spore culture [14]. A virulent isolate Cc 102 was chosen for the study. Virulence of the isolate was tested by performing *in vivo* challenge inoculation on young leaves of rubber plants and observing for disease symptom within 48 to 72 h. Conidial suspension was prepared from ten-day-old cultures. At three-whorl stage, *in vivo* inoculation with the spore suspension (4×10^4 spores/ml) was performed on two healthy rubber clones: RRII 105 and GT 1, a *Corynespora* susceptible and resistant clone respectively. Young refoliating leaves (45 leaves) from one plant that was representative of normal growth of each clone was used for inoculation. After inoculation, plants were covered with transparent polythene bags to maintain humidity. Nine leaves were collected after 6, 12 and 24 h following inoculation, cleaned, snap-frozen in liquid nitrogen and stored at −80 °C. Nine young refoliating leaves from unchallenged healthy plants (0 h) of both clones were also collected, cleaned, snap-frozen in liquid nitrogen and stored at −80 °C as control samples.

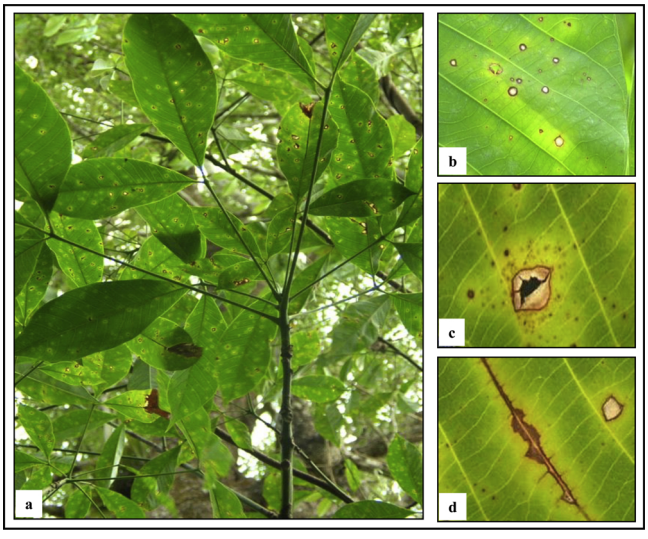

Fig. 1. An infected rubber plantation showing immature leaves with typical disease symptoms on rubber (a); leaf spots surrounded by yellow halo (b); circular concentric spots (c); infection on vein producing typical railway track symptom (d).

2.2. RNA isolation from leaf samples

Total RNA was extracted from the leaf samples using Ambion RNAqueous kit (Thermo Fischer Scientific). RNA quantity was measured using the NanoDrop ND-2000C UV–vis spectrophotometer (NanoDrop Technologies). Quality check was performed using Agilent 2100 Bioanalyzer (Agilent Technology, Inc.). Samples with RNA Integrity Number (RIN) greater than 8.5 were used for library construction. RNA isolated from control plant of RRII 105 and GT 1 were designated as C1 and C2 respectively. Equimolar amount of RNA from the three treatments (6, 12 and 24 h following challenge inoculation with *C. cassiicola*) for each clone were pooled together for cDNA synthesis to ensure that all the genes expressed at different time points in response to pathogen infection would be equally represented in a single library. The treated pooled RNA samples were named as T1 and T2 for RRII 105 and GT 1 respectively.

2.3. Transcriptome library construction, sequencing and de novo transcriptome assembly

Transcriptome library for sequencing was constructed according to the Illumina TruSeq RNA library protocol. Messenger RNA (mRNA) was purified from one microgram of total RNA using oligo-dT beads (TruSeq RNA sample preparation kit, Illumina). Sequencing was performed on Illumina HiSeq 2000 sequencing platform. After the sequencing was completed for both Read 1 (forward strand) and Read 2 (reverse strand), raw data was extracted using the proprietary Illumina pipeline software to generate short read information in FASTQ format (http://www.illumina.com/support/sequencing/sequencing_software.ilmn). Additional quality control was performed using SeqQC V2.1 (http://genotypic.co.in/SeqQC.html). The adapter sequences and low quality bases were trimmed and filtered. After trimming, reads with a length less than 50 bp were removed. Filtered high quality reads were used for further analysis. *De novo* assembly of reads into contigs was performed using Velvet 1.2.07 (https://www.ebi.ac.uk/~zerbino/velvet/) [15]. Parameters such as observed insert length and expected coverage were estimated using an initial draft assembly. The hash length was set as 49 and 53 for the controls and treated samples respectively. Minimum contig length was considered to be 100 bases. Contig assembly was followed by a transcriptome assembly with default parameters using Oases transcriptome assembler (version 0.1.21) (http://www.ebi.ac.uk/~zerbino/oases) [16]. The insert length obtained from Velvet was used for Oases with a minimum transcript length of 200 bases. As *de novo* assemblies are highly dependent on k-mer lengths, multiple assemblies were run to arrive at an optimal k-mer length for a better assembly.

2.4. Functional annotation of the rubber transcripts

Various databases were chosen to extract maximum information based on sequence and functional similarity. The data collected include Plant Metabolic Pathway (PlantCyc Enzymes database v2.0 - www.plantcyc.org), protein level sequence similarity information (UniProt: Swiss-Prot and TrEMBL databases) [17], nucleotide level sequence information (Viridiplantae mRNA database from NCBI GenBank), Clusters of Orthologous Groups (COG), functional classifications (KOG proteins from COG database) [18] and information on protein domains for distantly related proteins which do not have similarity at sequence level (Pfam database v26.0) [19]. Similarity search was performed using BLAST + v2.2.25 software [20]. Transcripts were then subjected to BLASTx for nucleotide query and protein database analysis by querying against the proteome of *Arabidopsis thaliana*, *Populus* spp. and *Ricinus communis* and the predicted proteome of the fungus *Corynespora cassiicola* (E-value cut off = 0.001) to eliminate fungus sequences. The expression levels of the rubber transcripts were normalised as Reads Per Kb per Million (RPKM) [21].

2.5. Differential gene expression (DGE) and gene ontology (GO) annotation

The DGE study for four sets of comparisons (C1 vs. C2, C1 vs. T1, C2 vs. T2 and T1 vs. T2) was carried out using DESeq software [22]. The parameters used for DGE included: 1) transcripts with corrected P-value $\geq$ 0.05 were considered to exhibit significant differential expression; 2) transcripts with a log2 (fold-change) value $\geq$ 1 were considered to be up-regulated; 3) transcripts with a log2 (fold-change) value $\leq$ -1 were considered to be down-regulated and 4) transcripts with a log2 (fold-change) value between 1 and -1 was considered as not regulated. Significantly expressed transcripts were annotated with reference of *Arabidopsis thaliana*, *Populus* spp. and *Ricinus communis*. The corrected P-value was calculated by Benjamini Hochberg procedure [23]. To obtain a final annotation for each transcript, the annotations from each database were analysed using the BLAST scoring system [20] to obtain the best annotation for each transcript. The order of preference for obtaining the best annotation was protein database Swiss-Prot, PlantCyc, KOG and TrEMBL database. Unannotated transcripts were assigned with GenBank Nucleotide database and further they were scanned for the PFAM database for functional assignment.

Gene Ontology (GO) [24] was used to summarize possible functional classification of the unigenes *via* assignment of gene identifiers with the BLASTx alignments to the corresponding *Hevea* reads. Distribution of biological processes and molecular functions obtained using GO annotation was done with GO stat program [25]. For further comparative analysis, four sets of clustered references were generated. Transcripts of both the samples used in comparison (*i.e.* between the two controls C1 and C2; two treatments T1 and T2; control C1 and treatment T1; control C2 and treatment T2) were merged using CD-Hit at 95% identity and coverage. Merged transcripts were used for further analysis.

2.6. GO term enrichment

GO enrichment is a systematic method of categorizing genes or proteins into groups of gene ontology terms. It was used to select specific group of proteins from large pool of differentially expressed genes. The enrichment analysis was performed using AgriGO – Singular Enrichment Analysis Tool [26], which focuses mainly on agricultural species. GO enrichment was performed for differentially expressed proteins from samples C1 vs. T1, C2 vs. T2 and T1 vs. T2. Separate enrichment analysis was carried out for all samples using hypergeometric test with Benjamini-Hochberg FDR correction and a p-value < 0.05. Significantly enriched GO terms were cross-compared between samples to identify common and unique terms for further analysis.

2.7. Clustering of stress associated transcripts and heat map generation

Following GO enrichment analysis, the protein IDs of differentially expressed transcripts were extracted for cluster analysis. Expression value of these protein IDs were extracted from samples C1T1, C2T2 and T1T2. These unique proteins were subjected to hierarchical clustering method in order to group them based on their expression pattern [27]. The colour bars revealed the intensity scales with red to green ranging from the highest (5.0) to the lowest level of expression (-5.0).

2.8. Detection of complete open reading frames (ORFs) in resistant clone transcripts

Transcript sequences derived from the resistant clone GT 1 inoculated with *C. cassiicola* (T2) were assembled and clustered using CD-HIT tool (http://weizhongli-lab.org/cd-hit) with 90% identity to eliminate duplicate transcripts. These clustered transcripts were subjected to ORF (Open Reading Frame) detection using TransDecoder (http://transdecoder.sourceforge.net). Ambiguous sequences with "N", transcripts with multiple ORFs, and ORFs with less than 1 Kb in length were eliminated. The gene description of high quality ORFs was obtained through BLAST homology search (www.ncbi.nlm.nih.gov/blast) against *Ricinus communis* protein data set from UniProt database (www.uniprot.org).

2.9. Simple Sequence Repeats (SSR) identification and in silico SSR polymorphism detection

The perl script program MISA [MIcroSAtellite identification tool (http://pgrc.ipk-gatersleben.de/misa/)] was used to identify common SSRs between C1 vs. C2 and T1 vs. T2 samples as well as unique SSRs for each treatment. SSRs were detected by considering 100 bp flanking sequences both upstream and downstream of the SSRs. The SSRs were categorized based on the repeat types as mono-, di-, tri-, tetra-, penta- and hexa-nucleotide with a minimum threshold of 10, 6, 5, 3, 5 and 5 repeat units respectively. Simple sequence repeat regions with 100 bp on either side were extracted from C1, C2, T1 and T2 transcript sequences. BLAST alignment was carried out between C1 and C2, T1 and T2. From the sequence alignment, sequences which were not having 100% identity as well as 100% length coverage were extracted. These sequences were considered as non-identical sequences. Further, only upstream and downstream regions of these non-identical sequences (excluding SSR region) between C1 and C2, T1 and T2 were aligned. From this new alignment, transcript sequences which were identical in both upstream and downstream except the SSR region were identified for extracting SSR polymorphism. In order to generate informative SSR markers, minimum threshold of the repeat motifs for three different SSR types (di-, tri-, and tetra-nucleotide) was set as 7, 5 and 4 respectively. The gene descriptions of these transcripts containing polymorphic SSRs were identified using BLAST homology search against *Ricinus communis* protein database, as the transcripts from other studies on rubber showed maximum sequence homology with the database of *R. communis* [11,12].

2.10. Mapping reads and variant calling

In the absence of a complete genome reference for rubber, the assembled transcripts were considered as reference sequence for variant calling. Variant calling was performed between C1 and C2 as well as T1 and T2, *i.e.* between the susceptible (RRII 105) and resistant (GT 1) clones without and following challenge inoculation. Reads were mapped back to the assembled transcripts using BWA mapping tool [28]. Variants were called using SAMtools v0.1.7a (http://samtools.sourceforge.net) [29]. The variants were filtered with a minimal mapping quality (MAPQ) $\geq$25, read depth (D) $\geq$5 and strand level evidence. It was essential that the locus having variants should come from

at least two reads mapping to both the strands. In order to avoid sequencing errors, SNPs with more than two alleles were removed from the study. Transcripts with filtered SNPs were subjected to ORF prediction to obtain non-synonymous SNPs.

2.11. Microarray gene signatures for disease tolerance

RNA samples from RRII 105 and GT 1 under both control and pathogen-challenged conditions isolated at different time points were the experimental material for microarray analysis. Library was constructed according to NEB Next Ultra Directional Library Prep (NEB). 600 ng of total RNA was taken for mRNA isolation, fragmentation and priming. The fragmented and primed mRNA was further subjected to first and second strand synthesis. The double stranded cDNA was cleaned up using HighPrep PCR (Magbio) beads. Subsequently, cDNA was subjected to a series of enzymatic reactions that repair frayed ends, phosphorylate the fragments, add a single nucleotide 'A' overhang and ligate adaptors. The adaptor ligated cDNA was cleaned up using HighPrep PCR beads. After ligation, PCR amplification (12 cycles) was performed to enrich the adaptor-ligated fragments, which were also cleaned using HighPrep PCR beads. The prepared library was quantified and validated for quality by running an aliquot on D1000 Tape Station Kit (Agilent). The adapter positive fragments were quantified using KAPA qPCR quantification kit (KAPA Biosystems).

In addition to the 10,063 transcripts identified from the transcriptome sequencing (6000 transcripts common to all four samples and 4063 unique NGS transcripts), 459 genes previously identified from our study as disease responsive transcripts from *Hevea brasiliensis* were also used to develop probes on an array in 8 x 60 K format (Agilent). A total of 24,860 probes (60 bp length) in both sense and antisense orientations were designed. Microarray analysis was performed for 62,976 features along with 536 Agilent control features for optimal background subtraction to avoid errors. Based on the intensity of probe-target hybridization, the relative gene abundance was assessed. Expression fold-change values were provided in log base 2. To identify upregulated and downregulated genes, the log2 (fold-change) thresholds were set as ≥ 1 or $\leq$-1 respectively. GO terms based on molecular functions, cellular components and biological processes were analysed. K-means clustering using GeneSpring was performed and 20 clusters were generated based on grouping of similar data points into different expression patterns. MapMan was used to perform microarray data visualization and present biological pathway diagrams.

3. Results

3.1. Rubber clones in response to inoculation

Resistance of GT 1 to *C. cassiicola* was tested through *in vitro* challenge inoculation (through detached leaf technique) on immature leaves with Cc 102, a virulent strain of the pathogen. The same isolate was also used for challenge inoculation of RRII 105, which is considered to be a susceptible clone [30–32]. Based on the size of the lesion produced, it was evident that GT 1 was resistant (Fig. 2) when compared to RRII 105. In RRII 105, severe disease symptoms developed within 48 to 72 h of inoculation. For GT 1, although hypersensitive response was visible on the leaves, disease symptoms in the form of raised spots, lesions and the characteristic railway track symptom on veins was not observed even after 72 h. After 96 h of infection, typical leaf spot symptom started to develop on the leaves of both these clones. *C. cassiicola* was re-isolated from the lesions to confirm pathogenicity of the isolate.

3.2. Sequencing outputs and De novo assembly

To obtain an overview of the rubber transcriptome and for an initial comparison between disease-resistant and disease-susceptible rubber clones in response to *C. cassiicola* infection, four libraries (C1, C2, T1 and T2) were constructed for paired-end sequencing. Illumina sequencing was performed and four datasets with 100-bp raw reads were generated. Raw reads were subjected to quality control using SeqQC. High quality (> Q20) bases were more than 97% in both forward and reverse (paired-end) reads. Percentage of unresolved bases (Ns) was observed to be very minimal (0.002%). High quality processed paired-end reads were used to assemble into contigs and further into transcripts. A total of 134.2 million high quality reads (> 70% of bases with > 20 phred score) were generated (Supplementary Table S1). After filtration of low-quality and adapter sequences, a total of 26.59, 26.07, 41.89 and 39.6 million raw reads were generated for C1, C2, T1 and T2 respectively. These processed paired-end reads with high quality were used for further analysis (Supplementary Tables S1 and S2).

Fig. 2. *In vitro* challenge inoculation on young coppery brown staged detached leaves of clones RRII 105 (a) and GT 1 (b) showing variation in size of lesion produced. Based on symptom development it is evident that RRII 105 is highly susceptible and GT 1 is tolerant to *Corynespora cassiicola.*

De novo assembly was performed using Velvet and Oases program for generating contigs (minimum length of 100 nucleotides) and transcripts (minimum length of 200 nucleotides) (Supplementary Table S2). An average of 124,137 contigs and 105,405 transcripts with an average length of 440 and 1373 were generated. Over 321 million clean reads from C1, C2, T1 and T2 transcriptomes were assembled and 124,138 contigs were obtained. An average of 105,405 transcripts was generated per library with a mean length of 1373 bases (Fig. 3). The N50 value on an average was 2097 for the four libraries and the assembled reads were 88% on an average. The raw paired-end sequence data is deposited in the NCBI Short Read Archive (SRA) database under the accession number SRA061336.

3.3. DGE analysis between different rubber clones and treatments

A comparison of transcript abundance between RRII 105 control (C1) and treated (T1) samples revealed a set of differentially expressed genes in response to stress. Fifty-eight transcripts were significantly down-regulated (log2 fold-change of -5.052 to -2.133) and 574 were significantly up-regulated (log2 fold-change of 2.282–9.358) in RRII 105 upon pathogen infection (Supplementary Table S5).

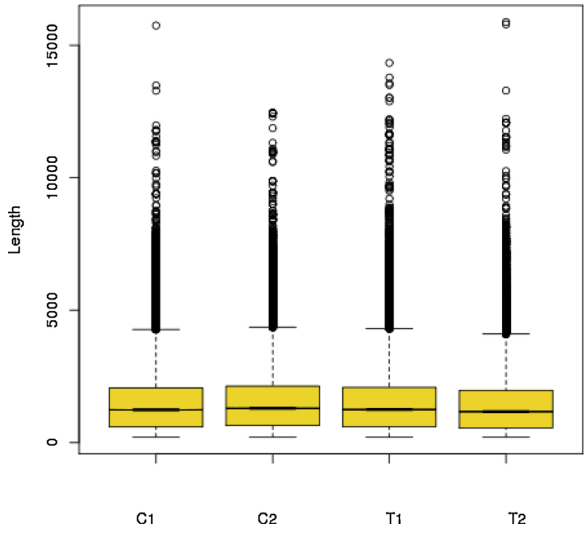

Fig. 3. Box plot showing length distribution of assembled transcripts. Box and whiskers show that transcript length are uniformly distributed in control (C1, C2) and treated (T1, T2) assemblies. Data points in C1 and T2 indicate that the maximum transcript size is over 15 Kb.

3.4. GO analysis of DGEs

The differentially expressed genes were functionally classified according to their gene ontology, which is an international standardized gene functional classification system with three main categories: biological process, molecular function and cellular component. Among the differentially expressed genes between the control samples (C1 vs. C2), a significant number of transcripts were grouped under defense response followed by proteolysis, signal transduction and carbohydrate metabolic process, which were annotated under biological process. Most represented transcripts involved in ATP and ADP binding activities were assigned under molecular functions. Similarly, significant number of transcripts related to membrane proteins was grouped under cellular components in C2. (Fig. 4a). Among the differentially expressed genes between GT 1 and RRII 105 upon pathogen infection (T1 vs. T2), the genes involved in defense response and regulation of transcription were most represented under biological process. ATP binding and DNA binding activities were most represented under molecular functions. Most of the transcripts encoding membrane protein were categorized under cellular components (Fig. 4b). In the genes that showed differential expression following challenge inoculation in the susceptible clone RRII 105 (C1 vs. T1), the most represented transcripts under biological process were related with regulation of transcription. Under molecular function, DNA binding and ATP binding were most represented. Significant number of transcripts related to nucleus and membrane were grouped under cellular component (Fig. 4c). For the genes differentially expressed between GT1 and RRII 105 under challenged condition, transcripts were mostly involved in regulation of transcription and proteolysis under biological process. Under molecular function, similar to other comparisons, transcripts were mostly involved in ATP binding and DNA binding. Transcripts falling under cellular component showed highest percentages in membrane and nucleus (Fig. 4d).

3.5. GO enrichment analysis

GO terms from differentially expressed proteins were extracted from different comparison between the four libraries. For C1T1 comparison, since there was no sufficient number of down-regulated genes in T1, no enrichment was performed. Only up-regulated genes of C1T1 were subjected to GO term enrichment. Most of the up-regulated genes in T1 were classified under transcription factors (TF) (e.g. WRKY, NAC, GATA and other putative TF proteins), zinc finger protein and putative uncharacterized proteins (Table 1). For the susceptible clone RRII 105 comparison C2T2, more major transcription factor genes (WRKY, NAC, GATA) were triggered and up-regulated than the ones in GT 1. Other than the upregulation of transcription factors, ATP binding proteins along with some putative uncharacterized proteins were also found (Table 2). When considering T1T2 comparison, the highly enriched GO terms largely fell into the following major categories: (i) defense response (GO:0,005,952), (ii) response to stress (GO:0,050,896) and (iii) response to stimulus (GO:0,006,950) (Fig. 5). However, almost equal representation of transcripts was observed for defense response category both in T1 and T2. The transcripts grouped under response to stimulus and response to stress GO terms were more enriched in T1 (susceptible clone RRII 105) compared to T2 (resistant clone GT 1). Besides these, other GO categories containing cellular catabolic process, RNA biosynthetic process, carbohydrate metabolism etc. were enriched only in T2. Disease resistant proteins and leucine rich repeat containing protein are classified under defense response in plants. These proteins were down-regulated in T1 in comparison to T2. In total seven disease resistant protein (RGA2, RPS2, RPM1), five leucine-rich repeat containing protein and seven TMV resistance protein N were up regulated in T2, which is a significant observation with regard to clonal response to pathogen (Table 3). Significantly enriched GO terms specifically in T2 were grouped into following main categories (i) heterocycle catabolic process (ii) carbohydrate metabolic process and (iii) RNA biosynthetic process. Multidrug resistance protein, ATP-binding cassette transporter (ABC) and dynamin domain containing protein were classified under heterocycle catabolic process. These proteins were also classified under nucleobase, nucleoside and nucleotide metabolic process. Some of the differentially induced genes in T2 encode pathogenesis-related (PR) proteins including chitinases and glucanases that are capable of degrading cell wall components of microbial pathogens (Supplementary Table S6).

3.6. Clustering of stress associated transcripts and heatmap generation

Based on conceptual translation of the transcript data, a total of 156 differentially stress associated protein IDs were identified from both susceptible and tolerant clone upon challenge inoculation. Expression value of these protein IDs was extracted from samples C1T1, C2T2 and T1T2. Hierarchical cluster was generated based on their relative expression pattern (Fig. 6). In C1T1 comparison, up-regulation of polyneuridine aldehyde esterase, homeobox protein, NAC-domain containing protein, WRKY transcription protein, GATA transcription factor, Zinc finger protein and aquaporin were observed. Down-regulation of cytochrome P450, glyceraldehyde 3-phosphate dehydrogenase, ketol-acid reductoisomerase, della protein, salt tolerance protein and chlorophyll A/B binding protein was also noticed (Fig. 6). Hierarchical clustering of stress associated transcripts in relation to C2T2 revealed upregulation of genes including WRKY transcription factor, GATA transcription factor, guanylate kinase, NAC domain containing protein, and chitin inducible gibberellin responsive protein. The down-regulated proteins in C2T2 were mainly those involved in cytochrome, serine pyruvate amino transferase and polyneuridine aldehyde esterase (Fig. 6). For T1T2, it was interesting to note up-regulation of several transcripts such as betaglucosidase, NAC domain containing protein, WRKY transcription factor, GATA transcription factor, cellulose synthase A catalytic subunit, and disease resistance protein RGA2. Down-regulation of only a few stress-associated transcripts, namely DNA mismatch repair protein, phosphoprotein phosphatase and major latex protein were noticed in T1T2 comparison (Fig. 6). It could also be observed that a large number of encoded proteins were uniquely up-regulated in the treated samples of the resistant clone GT 1 (T2). The transcripts were namely cellulose synthase A catalytic subunit 6 [UDP-forming], phosphomannomutase, starch branching enzyme II, sucrose

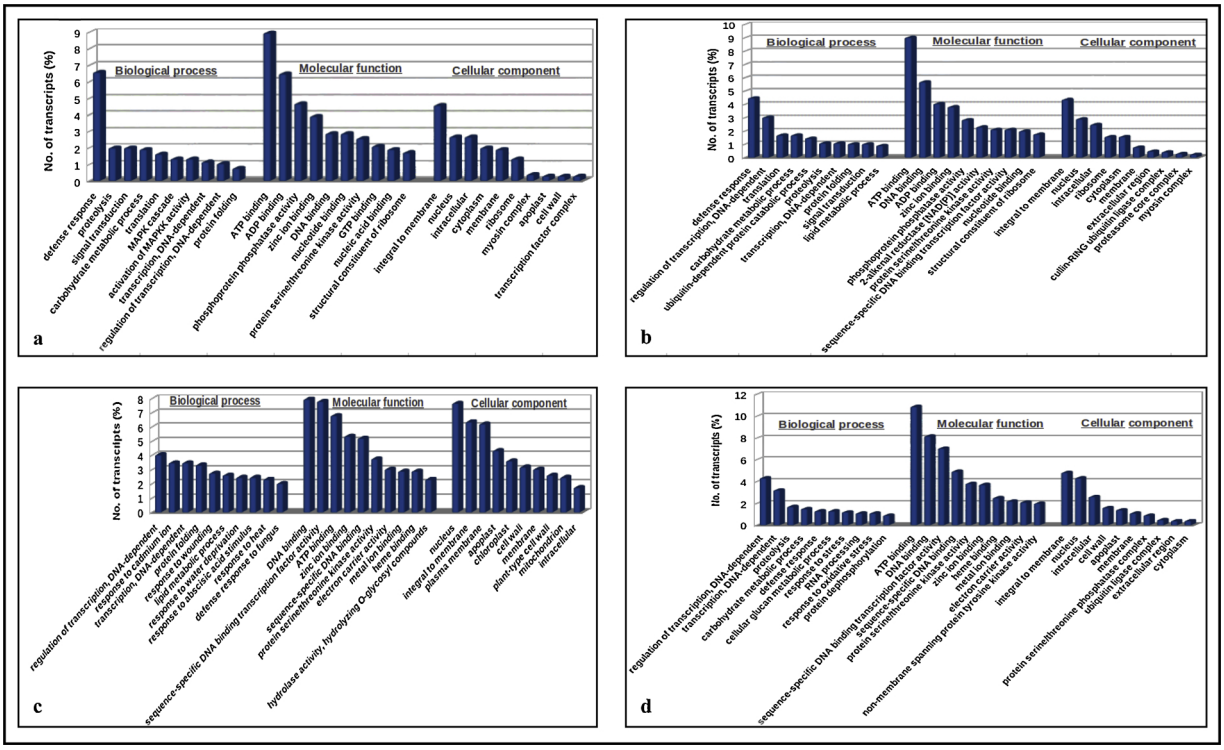

Fig. 4. Comparison of functionally classified (biological process, molecular function and cellular component) transcripts belonging to RRII 105 (susceptible) and GT 1 (tolerant) clones in control (C1 and C2) and following challenge inoculation (T1 and T2). Bar diagram demonstrate comparisons between (a) two clones in control condition (C1 vs. C2); (b) two clones in pathogen challenged condition (T1 vs. T2); (c) control and challenged samples of RRII 105 (C1 vs. T1); (d) control and challenged samples of GT 1 (C2 vs. T2).

Table 1
Gene Ontology (GO) enrichment analysis of up- and down-regulated proteins between C1T1. Since only a small number of transcripts were detected as down-regulated in this group, GO enrichment analysis was not performed for this data. The number at the beginning of the protein name indicates the occurrence of that particular protein in this group.

C1T1 Up (with GO enrichment)	C1T1 Down (without GO enrichment)
17 WRKY transcription factor	8 Putative uncharacterized protein
13 Zinc finger protein	8 Cytochrome P450
12 Putative uncharacterized protein	4 Root phototropism protein
10 NAC domain-containing protein	3 R2R3-MY transcription factor
7 Xyloglucan endotransglucosylase/ hydrolase protein A	3 Photosystem II 22 kDa protein
7 Transcription factor	3 Ketol-acid reductoisomerase
6 UDP-galactose transporter	2 UDP-glucosyltransferase
4 Light-inducible protein CPRF-2	2 Sigma factor sigb regulation protein rsbq
3 GATA transcription factor	2 Flavonol synthase
2 Xyloglucan endotransglucosylase/ hydrolase protein 9	2 Chlorophyll A/B binding protein
2 Homeobox protein	1 UDP-glucuronosyltransferase
2 Aquaporin PIP2.2	1 Transcription factor
1 Tonoplast intrinsic protein	1 Sialin
1 Patatin T5	1 Salt-tolerance protein
1 Auxin:hydrogen symporter	1 Pentatricopeptide repeat-containing protein
1 ATP synthase 9 mitochondrial	1 Metal ion binding protein
1 1-deoxyxylulose-5-phosphate synthase	1 Magnesium-chelatase subunit H
	1 Isoflavone reductase
	1 Glycine dehydrogenase
	1 Glyceraldehyde 3-phosphate dehydrogenase
	1 Glutamyl-tRNA reductase
	1 Fructose-bisphosphate aldolase
	1 DELLA protein SLR1

Table 2
Up- and down-regulated proteins between C2T2 after Gene Ontology (GO) enrichment analysis. Since only a small number of transcripts was detected as down regulated in this group, GO enrichment analysis was not performed for this data. The number in the beginning of the protein name indicates the occurrence of that particular protein in this group.

C2T2 Up (with GO enrichment)	C2T2 Down (without GO enrichment)
25 WRKY transcription factor	2 Polyneuridine-aldehyde esterase
22 ATP binding protein	1 Serine-pyruvate aminotransferase
20 Transcription factor	1 Cytochrome P450
20 Putative uncharacterized protein	
13 NAC domain-containing protein	
5 Multidrug resistance protein 1	
3 Poly [ADP-ribose] polymerase	
3 Pectinesterase (EC 3.1.1.11)	
3 Mitochondrial chaperone BCS1	
3 GATA transcription factor	
2 Inorganic pyrophosphatase	
2 Guanylate kinase	
2 Chitin-inducible gibberellin-responsive protein	
2 Chaperone ClpB	
2 ATP-dependent RNA helicase	
2 ATP-binding cassette transporter	
1 Xyloglucan endotransglucosylase/hydrolase protein 22	
1 TATA-box binding protein	
1 Cation-transporting ATPase plant	

phosphate synthase, disease resistance protein, leucine-rich repeat-containing protein, TMV resistance protein, major allergen Pru ar, chaperone ClpB, mitochondrial chaperone BCS1, dynamin, multidrug resistance protein, ferrochelatase, NAC domain-containing protein and

GO Information				CM		ID:667040821		ID:735237528	
No	GO Term	Onto	Description	1	2	Ajusted Pvalue	Num	Ajusted Pvalue	Num
1	GO:0006952	P	defense response			2.9e-16	18	1.4e-11	20
2	GO:0050896	P	response to stimulus			3.5e-08	22	4.1e-05	28
3	GO:0006950	P	response to stress			3.5e-08	19	3e-05	24
4	GO:0044248	P	cellular catabolic process			---	---	7e-06	20
5	GO:0009141	P	nucleoside triphosphate metabolic process			---	---	3e-05	13
6	GO:0009144	P	purine nucleoside triphosphate metabolic process			---	---	3e-05	13
7	GO:0009259	P	ribonucleotide metabolic process			---	---	3e-05	14
8	GO:0009199	P	ribonucleoside triphosphate metabolic process			---	---	3e-05	13
9	GO:0009150	P	purine ribonucleotide metabolic process			---	---	3e-05	14
10	GO:0009205	P	purine ribonucleoside triphosphate metabolic process			---	---	3e-05	13
11	GO:0006163	P	purine nucleotide metabolic process			---	---	3e-05	14
12	GO:0006351	P	transcription, DNA-dependent			---	---	4.7e-05	36
13	GO:0032774	P	RNA biosynthetic process			---	---	4.9e-05	36
14	GO:0051252	P	regulation of RNA metabolic process			---	---	9.3e-05	34
15	GO:0006355	P	regulation of transcription, DNA-dependent			---	---	9.3e-05	34
16	GO:0009117	P	nucleotide metabolic process			---	---	0.00044	14
17	GO:0006753	P	nucleoside phosphate metabolic process			---	---	0.00044	14
18	GO:0055086	P	nucleobase, nucleoside and nucleotide metabolic process			---	---	0.00087	15
19	GO:0046483	P	heterocycle metabolic process			---	---	0.0058	15
20	GO:0016070	P	RNA metabolic process			---	---	0.0062	36
21	GO:0005975	P	carbohydrate metabolic process			---	---	0.033	26
22	GO:0045449	P	regulation of transcription			---	---	0.041	34
23	GO:0019219	P	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process			---	---	0.041	34
24	GO:0031326	P	regulation of cellular biosynthetic process			---	---	0.042	34
25	GO:0009889	P	regulation of biosynthetic process			---	---	0.042	34

Fig. 5. Comparison of significantly enriched GO terms (top 25 entries) of down-and up-regulated transcripts from T1T2. Colour pattern: Red = High level enrichment, Orange = Medium level enrichment, Yellow = Low level enrichment and Grey = missing. CM = Cross comparison. 1 = T1T2 down-regulated GO terms. 2 = T1T2 up-regulated GO terms. Num: Number of GO entries. Adjusted p-value = P values for enriched GO terms.

vacuolar protein sorting-associated protein VPS4 (Table 4).

3.7. Complete open reading frames (ORFs) in T2 transcripts

In order to identify full-length genes possibly conferring resistance to *C. cassiicola* in the resistant clone GT 1, 113,863 transcript sequences from T2 were assembled and processed for complete ORF detection. A total of 8881 complete ORFs having length > 1 kb and with no ambiguities were identified. Among these, 118 transcripts were significantly more abundant compared with challenge inoculated RRII 105 (T1), which are considered to be potential genes actively involved in disease resistance in tolerant clone GT 1 (Supplementary Table S7). It was interesting to note that 17 of the complete ORFs with higher expression values were from the earlier selected 50 unique stress responsive transcripts from T2. These complete ORFs include cellulose synthase, sucrose phosphate synthase, TMV resistance protein, chaperone ClpB, mitochondrial chaperone, dynamin, multidrug resistance protein, NAC domain-containing protein and vacuolar protein sorting-associated protein VPS4 could be obtained (Table 4).

3.8. Simple sequence repeats (SSRs) identified in transcripts

Our transcriptome data was mined to identify the transcripts bearing different repeat motifs. A total of 41,456 sequences were examined covering a size of 75.76 Mb for identifying the SSRs. From within this region, 688,979 bases were covered by the SSR and the total number of SSRs identified was 32,323. It was observed that 3270 sequences contained more than one SSR. The relative abundance of SSR was found to be 503.59 SSRs/Mb. An average total sequence length of 6347.6 bp was contributed by each SSR per Mb of total sequence analyzed. A few of the longest SSR motifs observed in the transcriptome sequence data set were (A)87, (GA)38, (TTA)18, (TAA)18, (TTTA)8, (AATC)8, (TCAAC)7, (CTATT)7, (TCAAC)7 and (GCACCA)9. The number of common SSRs and unique SSRs across four libraries are summarized in Supplementary Table S11).

3.9. In silico SSR polymorphism detection

Using our in-house pipeline, SSR polymorphism between C1 and C2, T1 and T2 was identified (Supplementary Table S12). Three different SSR motifs (di-, tri- and tetranucleotide repeat motifs) were selected for *in silico* polymorphism analysis. In total, 82 polymorphic SSRs were identified between C1 and C2 (42 dimer, 37 trimer and 3 tetramer). Between the pathogen-challenged libraries T1 and T2, 109 polymorphic SSRs were identified, of which 67 were dimer, 38 trimer and 4 tetramer. Altogether 191 transcripts were found with polymorphic SSRs between the two rubber clones RRII 105 and GT 1 (Supplementary Table S13). Of these, sixteen transcripts were found common bearing 9 di-nucleotide, 6 tri-nucleotide and 1 tetra-nucleotide repeats. The number of repeat motifs ranged from 7 to 25 for di-nucleotide, 5 to 18 for tri-nucleotide and 4 to 8 for tetra-nucleotide repeats. The most prevalent repeat motif identified for di-nucleotide was AG/CT. For trinucleotide, the most frequent motif was AGG/CCT. Out of the 191 transcripts, gene descriptions were obtained for 127 transcripts through BLASTx homology search with the *R. communis* protein database. For the remaining 64 transcripts, 46 were putative uncharacterized protein while 18 had no match.

3.10. Single nucleotide polymorphisms (SNPs) discovery from transcripts

Nucleotide sequence variation were searched between the two control libraries (C1 and C2) and two treatment libraries (T1 and T2). Common and unique variations at single base level were observed in both these conditions. SNP calling predicted a total of 38,020 SNPs as non-synonymous in the T2 vs. T1 comparison. Transition and transversion ratio (Ts/Tv) of T2 vs. T1 was found to be 1.38 (Ts = 80,119; Tv = 58,035; total = 138,154) (Fig. 7). The total SNPs detected were 110,706 in C1 vs. C2, 116,985 in C2 vs. C1, 136,793 SNPs in T1 vs. T2 and 138,524 in T2 vs. T1 with stringent filtering criteria (Supplementary Table S14).

Table 3
Up- and down-regulated proteins between T1T2 after Gene Ontology (GO) enrichment analysis. The number in the beginning of the protein name indicates the occurrence of that particular protein in this group.

T1T2 Up (with GO enrichment)	T1T2 Down (with GO enrichment)
23 Transcription factor	8 Leucine-rich repeat-containing protein 2
18 ATP binding protein	4 Disease resistance protein RGA2
16 Beta-glucosidase	3 Leucine-rich repeat containing protein
13 Putative uncharacterized protein	2 Putative uncharacterized protein
7 TMV resistance protein N	2 Major allergen Pru av
5 NAC domain-containing protein 21/22	1 Phosphoprotein phosphatase
5 Multidrug resistance protein 1	1 Major latex protein
5 Disease resistance protein RGA2	1 DNA mismatch repair protein mlh1
4 Mitochondrial chaperone BCS1	
4 Dynamin	
3 Vacuolar protein sorting-associated protein VPS4	
3 Major allergen Pru ar	
3 ATP-binding cassette transporter	
2 Trehalose-6-phosphate synthase	
5 Leucine-rich repeat-containing protein	
2 Heat shock protein binding protein	
2 GATA transcription factor	
2 Cullin-1	
2 Class I chitinase	
2 Chaperone clpb	
2 Cellulose synthase A catalytic subunit 6	
1 Sucrose phosphate syntase	
1 Starch branching enzyme II	
1 Phosphomannomutase	
1 NAC domain-containing protein	
1 Multidrug resistance-associated protein 2	
1 MTA/SAH nucleosidase	
1 Leucine-rich repeat-containing protein 2	
1 Heat shock factor protein	
1 GMP synthase	
1 Ferrochelatase (EC 4.99.1.1)	
1 Disease resistance protein RPS2	
1 Disease resistance protein RPM1	
1 Cellulose synthase	

3.11. Microarray gene signatures for disease tolerance

Understanding the expression pattern of important stress-related genes at different time intervals following challenge inoculation (0 h, 6 h, 12 h and 24 h) is essential for pinpointing their involvement in disease tolerance. Therefore, microarray-based gene expression technique was adopted to get an overview on the expression profiles of selected genes based on functional classification and gene ontology, thereby giving a comprehensive understanding of the genes involved in different stages of infection leading to disease development in the susceptible clone (Supplementary Table S16). In the resistant clone GT 1, 368 and 524 disease resistance related genes were up- and down-regulated respectively; whereas in the susceptible clone RRII 105, 1104 and 660 disease resistance related genes were up- and down-regulated respectively (Table 5). It was observed in GT 1, majority of the transcripts up-regulated belonged to TFs and defence response but on the contrary majority of the down-regulated transcripts were related to response to stress. In RRII 105, although TFs, defence response-related and response to stimulus-related transcripts were the majority that were up-regulated, down regulation of transcripts related to response to stress and defence response were noticed (Table 5). This clearly established the importance of these disease resistance related genes conferring resistance to GT 1.

A comparison of the genes involved in disease resistance from the global data obtained through transcriptomic study as well as that of the microarray study was made. It was interesting to note that the results obtained through the NGS study was validated with the microarray analysis. In the resistant clone GT 1 as compared to the susceptible clone RRII 105 in control condition, it was observed that 246 and 234 common genes were up- and down-regulated respectively in the NGS dataset. When the up- and down regulated transcripts from the microarray dataset were mapped back to the NGS data, four genes each were found in common within the NGS and microarray dataset (Fig. 8) indicating these genes could be associated with innate resistance. However, following challenge inoculation with *C. cassiicola* in the resistant clone, 202 and 40 genes from the microarray dataset were up- and down-regulated respectively when mapped back to the NGS dataset (Fig. 9). Conversely, in the susceptible clone RRII 105, upon challenge inoculation with *C. cassiicola*, 365 and 92 genes from the microarray dataset were up- and down-regulated respectively when mapped back to the NGS dataset (Fig. 10; Supplementary Table S17).

MapMan was used to integrate and visualize the DEGs according to their functions in metabolic pathways. They were categorized into those involved in hormone signalling (auxin, brassinosteroid, ABA, ethylene, SA), cell wall, beta glucosidase, proteolysis, R genes, PR proteins and secondary metabolites under defense genes, transcription factor (ERF, bZIP, WRKY, MYB), heat shock protein and secondary metabolites (Supplementary Table S18; Fig. 11). A closer look at the MapMan profile clearly depicted low levels of R gene expression in the susceptible clone RRII 105 even with the progression of disease from 6 to 24 h. However, in the resistant clone GT 1, the level of expression of R gene was found to be increasing with the progression of disease indicating higher expression of R genes in GT 1 as disease progressed from 6 to 24 h. Similar was the case with up-regulation of PR proteins, cell wall transcripts, SA, TFs, HSPs and secondary metabolites in GT 1 as compared to RRII 105 (Fig. 11).

All the 38,281 differentially expressed genes were analyzed for time-scale variation in their expression at 6, 12 and 24 h following infection. They were classified into 20 clusters based on their pattern of expression (Supplementary Table S19.

4. Discussion

With advances in functional genomics, proteomics and host-pathogen interaction studies, molecular mechanisms of disease tolerance is being unravelled. Transcriptome sequencing in rubber using NGS platform was performed by several researchers for identification of tissue-specific genes involved in various metabolic pathways and development of SSR and SNP markers for linkage map construction ([33,34]; Li et al., 2012). Hurtado Páez et al. [12] reported the involvement of genes conferring tolerance to South American Leaf Blight (SALB) disease caused by *Microcyclus ulei*. However, more efforts are needed to elucidate the complexity of host tolerance of rubber tree in response to one of its major pathogens *C. cassiicola*. The present study of disease transcriptome of rubber in response to *C. cassiicola* infection reports molecular signatures providing vital clues related to disease resistance in rubber. Digital gene expression profiling during *C. cassiicola* infection in rubber through RNA-Seq has served as a platform to generate enormous information about the genes involved in host tolerance during disease development and identification of SSR motifs and SNPs in transcripts to be used in the development of molecular markers for disease tolerance in rubber trees.

The timing of recognition for *C. cassiicola* infection on rubber is not clearly known, which makes it challenging to standardize suitable time points for molecular analysis. Therefore, we started this investigation by examining the infection process from six to 144 h after challenge inoculation of rubber leaves with the pathogen in order to determine the optimal sampling time after inoculation. Quantification of hypersensitivity, fungal growth and disease severity evaluation reported expression of hypersensitive reaction symptoms in the susceptible clone

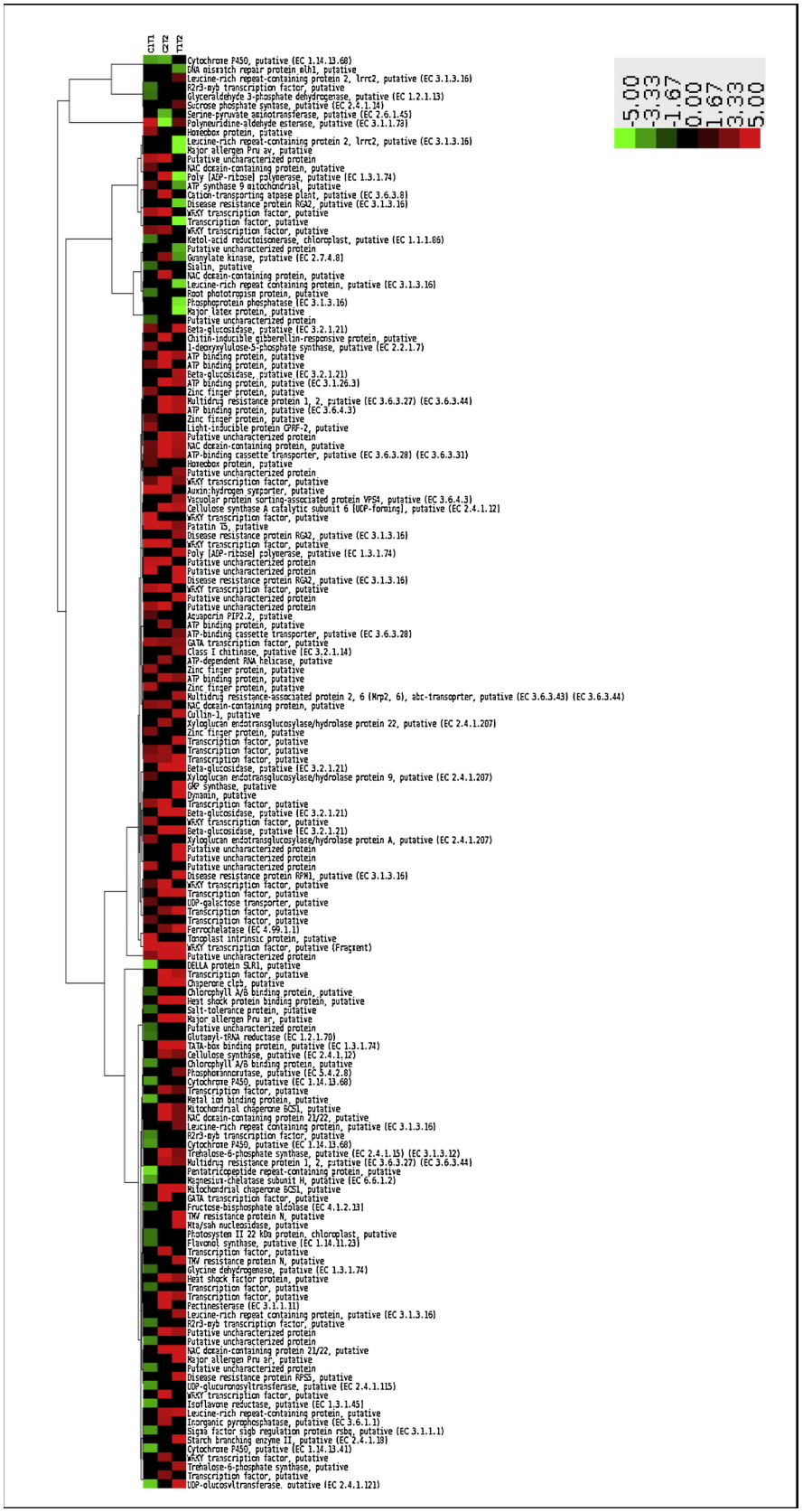

Fig. 6. Differentially expressed genes clustered using hierarchical clustering for visualizing expression patterns between control and treated samples. Gene expression is represented by different colour bars (red: up-regulation, green: down-regulation and black: missing values). C1 = RRII 105 control, C2 = GT 1 control, T1 = RRII 105 *Corynespora* challenged, T2 = GT 1 *Corynespora* challenged.

Table 4
Up-regulated unique transcripts identified in *Corynespora* challenged GT 1 (resistant clone) (T2) transcriptome. Majority of the transcripts fall under GO enriched terms: defense response, response to stimulus, carbohydrate metabolic process and regulation of gene expression. Twelve transcripts possessing complete ORF are marked with * (asterisk).

Transcript ID*	Transcript length (bp)	Log2 fold-change	Description
T2_Locus_13871_Transcript_2/5	3212	4.63	Cellulose synthase A catalytic subunit 6 [UDP-forming], putative (EC 2.4.1.12)
*T2_Locus_4098_Transcript_10/14	3208	3.44	Cellulose synthase, putative (EC 2.4.1.12)
T2_Locus_9326_Transcript_5/6	1292	3.46	Phosphomannomutase, putative (EC 5.4.2.8)
T2_Locus_20094_Transcript_5/6	758	5.08	Starch branching enzyme II, putative (EC 2.4.1.18)
*T2_Locus_9483_Transcript_7/11	3715	2.88	Sucrose phosphate syntase, putative (EC 2.4.1.14)
T2_Locus_3163_Transcript_22/31	3793	3.97	Disease resistance protein RGA2, putative (EC 3.1.3.16)
T2_Locus_5235_Transcript_44/49	2634	3.63	Disease resistance protein RGA2, putative (EC 3.1.3.16)
T2_Locus_895_Transcript_2/4	5173	5.04	Disease resistance protein RGA2, putative (EC 3.1.3.16)
T2_Locus_9502_Transcript_7/10	3688	5.84	Disease resistance protein RPM1, putative (EC 3.1.3.16)
T2_Locus_12428_Transcript_7/11	1356	4.86	Disease resistance protein RPS5, putative
T2_Locus_785_Transcript_44/49	2126	2.81	Leucine-rich repeat-containing protein 2, lrrc2, putative (EC 3.1.3.16)
*T2_Locus_16377_Transcript_8/11	2672	5.26	TMV resistance protein N, putative
T2_Locus_3268_Transcript_53/57	2227	4.72	TMV resistance protein N, putative
*T2_Locus_4372_Transcript_63/69	5794	2.94	TMV resistance protein N, putative
*T2_Locus_583_Transcript_86/86	4944	3.66	TMV resistance protein N, putative
*T2_Locus_7904_Transcript_2/2	4260	3.51	TMV resistance protein N, putative
T2_Locus_26457_Transcript_1/1	498	5.17	Major allergen Pru ar, putative
T2_Locus_29423_Transcript_1/2	867	6.73	Major allergen Pru ar, putative
T2_Locus_8748_Transcript_1/1	921	5.18	Major allergen Pru ar, putative
*T2_Locus_2277_Transcript_2/2	3156	5.98	Chaperone ClpB, putative
T2_Locus_22008_Transcript_1/2	1663	4.61	Mitochondrial chaperone BCS1, putative
*T2_Locus_3468_Transcript_1/2	1826	3.31	Mitochondrial chaperone BCS1, putative
T2_Locus_15936_Transcript_1/2	1375	4.92	MTA/SAH nucleosidase, putative
T2_Locus_4430_Transcript_3/6	817	6.58	GMP synthase, putative
*T2_Locus_1308_Transcript_19/22	2835	3.03	Dynamin, putative
*T2_Locus_4687_Transcript_12/16	4796	3.62	Multidrug resistance protein 1, 2, putative (EC 3.6.3.27) (EC 3.6.3.44)
T2_Locus_472_Transcript_6/12	5366	4.42	Multidrug resistance-associated protein 2, 6 (Mrp2, 6), ABC-transporter, putative (EC 3.6.3.43) (EC 3.6.3.44)
T2_Locus_11227_Transcript_1/1	1487	5.66	Ferrochelatase (EC 4.99.1.1)
T2_Locus_13330_Transcript_3/6	1997	4.36	NAC domain-containing protein 21/22, putative
*T2_Locus_13822_Transcript_2/2	1571	3.35	NAC domain-containing protein 21/22, putative
*T2_Locus_716_Transcript_3/5	2012	2.90	Vacuolar protein sorting-associated protein VPS4, putative (EC 3.6.4.3)

RRII 105 after 24 h of challenge inoculation and the disease severity increased from 48 h of infection. However, in the resistant clone GT 1, the hypersensitive reaction could be observed only after 48 h of inoculation. As a consequence, up to 24 h time-point was included in the study to capture the defense responses. Therefore, samples at 6, 12 and 24 h post infection were selected to investigate early-intermediate transcriptional changes, in order to elucidate host genetic responses. Our study provides the first large-scale investigation of gene expression changes that occur when rubber plant is inoculated with *C. cassiicola*, and is the first to compare and pathogenic interactions in susceptible and resistant genotypes of rubber.

4.1. DGE analysis between different rubber clones and treatments

In the absence of a complete reference genome and insufficient number of gene sequences in public database for *H. brasiliensis*, functional annotation of the transcripts is a challenging task. Identification of differentially expressed genes provides clues on the defense mechanisms and biochemical pathways involved in plants to combat the pathogen. Disease resistance is a complex trait that involves multiple complex molecular mechanisms. Analysis of gene expression through transcriptome profiling of RRII 105 and GT 1 in response to C. cassiicola infection revealed that they responded in a clone-dependent manner upon infection. A comparison of gene expression between C1 and T1 identified genes specifically triggered only in susceptible clone (RRII

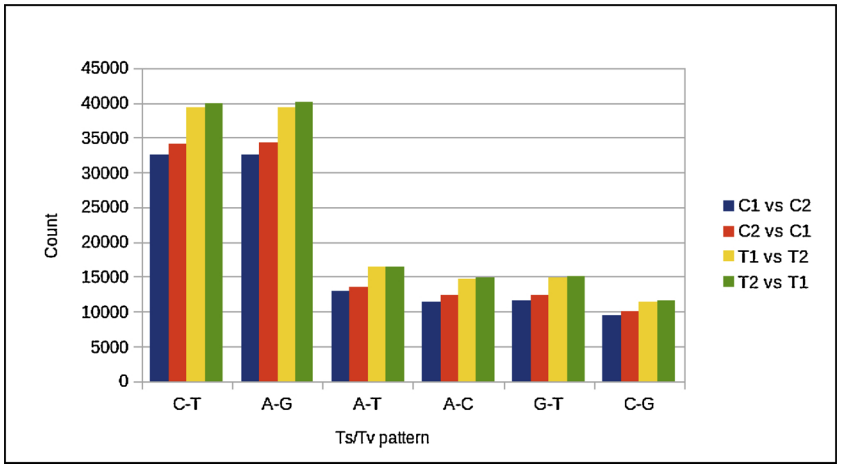

Fig. 7. Single nucleotide polymorphisms detected between RRII 105 (susceptible) and GT 1 (tolerant) in control and *Corynespora* challenged samples in the following comparisons: C1 vs. C2 (C1 as reference), C2 vs. C1 (C2 as reference), T1 vs. T2 (T1 as reference), T2 vs. T1 (T2 as reference). Number of base substitutions are represented in bar diagram.

Table 5
Details on the number of differentially regulated genes in RRII 105 and GT 1 in response to *C. cassiicola* infection affecting pathways related to disease resistance as identified through microarray analysis.

Function name	Total transcripts		Up-regulated		Down-regulated	
	RRII 105	GT 1	RRII 105	GT 1	RRII 105	GT 1
Transcription factor	463	224	329	108	134	116
Signalling pathway	107	77	69	42	38	35
Response to stimulus	364	123	275	54	89	69
Defense response	474	215	295	107	179	108
Response to stress	356	253	136	57	220	196
Total	1764	892	1104	368	660	524

105) in response to pathogen infection (Supplementary Table S3). It was observed that majority of the upregulated transcripts belonged to transcription factor group (WRKY, NAC and GATA). In addition zinc finger protein, putative uncharacterized protein, light inducible protein, homeobox protein and aquaporin were also found to be upregulated (Table 1). It is presumed that these TFs being regulatory proteins, triggered during disease development in T1, play a crucial role in resistance mechanism in the susceptible clone RRII 105. It has been reported that adaptability of plants to various stressful environmental conditions is achieved by reprogramming their transcriptome in a dynamic and temporal manner, through enforcement of a network of various transcription factors [35].

Similarly, expression study between C2 and T2 revealed the genes possibly conferring tolerance in resistant clone (GT 1) (Supplementary Table S4). As observed in the susceptible clone RRII 105, many of the major transcription factor genes (WRKY, NAC, GATA) were triggered and up-regulated also in the resistant clone GT 1 during disease development, which could be associated with enhanced immunity against the pathogen. Other than the upregulation of transcription factors, ATP binding proteins were also found to be significantly up-regulated (Table 2).

Genes showing differential regulation between the susceptible and resistant clone upon challenge inoculation was obtained by comparing T1 and T2 (Supplementary Table S5). This could be considered as a reflection of the clonal response to *C. cassiicola*. A few novel genes of interest were also observed, which could contribute to the response of the resistant clone to pathogen stress helping it to combat the pathogen.

The highly enriched GO terms which were common between T1 and T2 largely fell into the major categories defense response, response to stress and response to stimulus (Fig. 5). However, almost equal representation of transcripts was observed for defense response category both in T1 and T2. The transcripts grouped under response to stimulus and response to stress GO terms were more enriched in T1 (susceptible clone RRII 105) compared to T2 (resistant clone GT 1). This indicates that susceptible clone is much more sensitive to stress and stimulus than the resistant clone. Other GO categories containing cellular catabolic process, RNA biosynthetic process, carbohydrate metabolism etc. were enriched only in T2. Disease resistant proteins and leucine-rich repeat containing protein are generally classified under defense response in plants. These proteins were found to be significantly down-regulated in T1 in comparison to T2. Up-regulation of disease resistant protein, NAC, RGA2, leucine-rich repeat containing protein, GATA and chitinase were observed, which suggests about clonal response to pathogen (Table 3). GO terms enriched in T2 were mainly found to be TFs, ATP binding protein, Beta-glucosidase, TMV resistance protein, multidrug resistance protein, disease resistance protein, dynamin, ATP binding cassette (ABC) receptor, leucine-rich repeat containing protein, heat shock binding protein, chitinase etc. ABC transporters, driven by ATP hydrolysis belonging to a protein super family are membrane proteins involved in transportation of a variety of compounds across the plasma membrane. They are involved in extra-cellular excretion of toxic compounds in detoxification process, transport of steroids and their derivatives, lipid translocation, phytohormone transport and tolerance to heavy metal [36,37]. ABC transporters in plants are reported to be induced by external signals. They are essential for plant growth and development as well as tolerance to biotic and abiotic stresses indicating their role in different aspects of plant life cycle [38]. The role of tobacco ABC transporter NtPDR1 in plant defense was established by Sasabe et al. [39]. Glombitza et al. [40] also reported induction of ABC transporters by a variety of stress factors and pathogens. Similarly, ABC transporters in *Arabidopsis: AtPDR8, AtPDR12* and *Nicotiana NpPDR1* were found to be involved in plant defense response against various pathogens belonging to *Pseudomonas syringae*, *Peronospora parasitica, Phytophthora infestans, Golovinomyces orontii* and *Plectosphaerella cucumerina* [41]. In our study, higher expression of ABC transporter in tolerant plant (T2) induced by *C. cassiicola* could be considered as one of the factors conferring tolerance in GT 1.

An increase in carbohydrate is believed to be a metabolic signal that

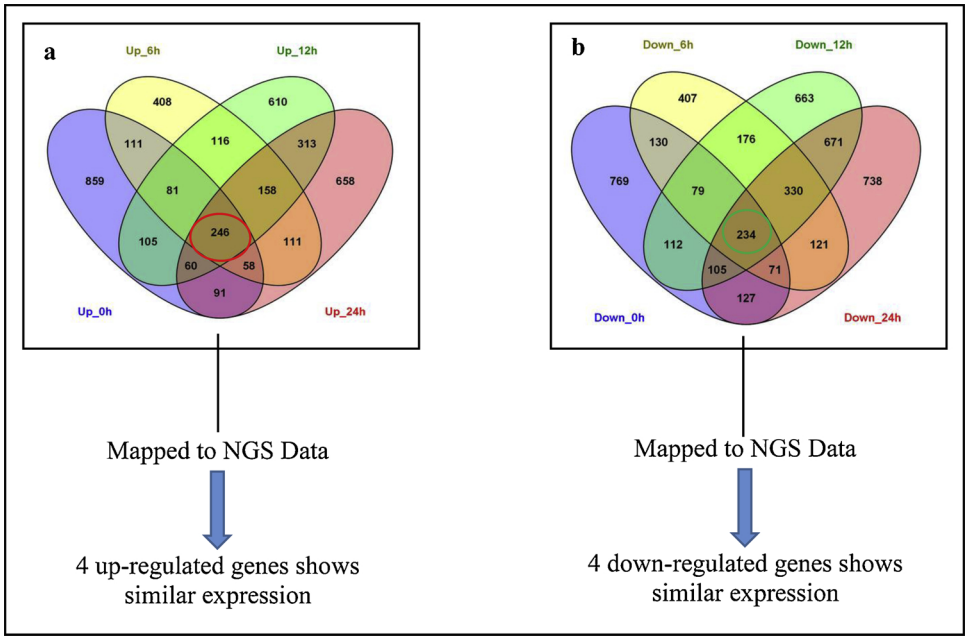

Fig. 8. Comparison of significant innate resistance genes/ transcripts identified at all-time points determined using Agilent microarrays. Venn diagram showing overlap of up regulated (a) and down regulated (b) transcripts in the resistant clone GT 1 as compared to the susceptible clone RRII 105 in control condition. Four genes/ transcripts each showing similar expression (up- and down-regulation) were identified when mapped back to the NGS data.

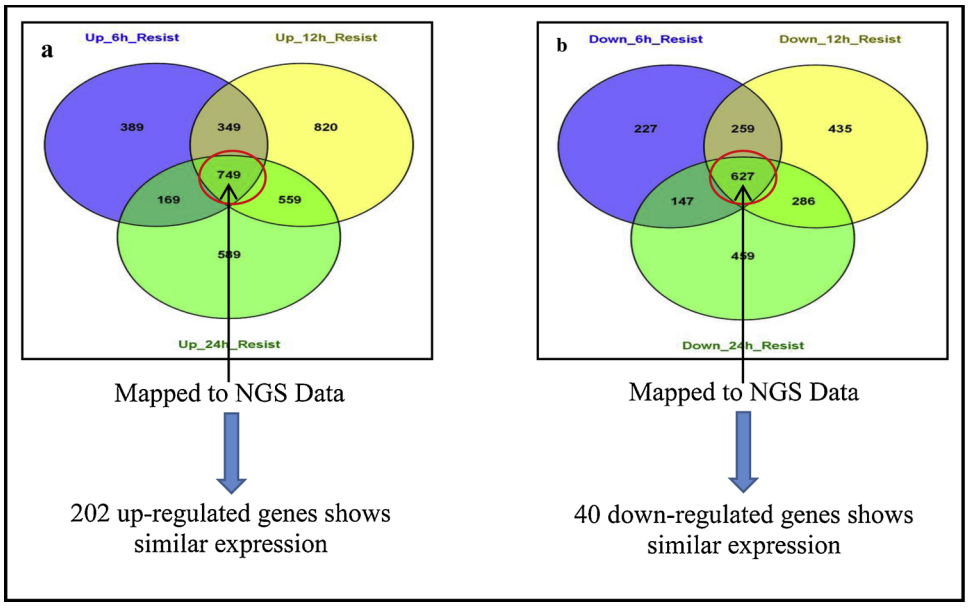

Fig. 9. Venn diagram showing sharing of up-regulated (a) and down-regulated (b) genes/ transcripts associated with progression of disease in resistant condition at different time points in the resistant clone GT 1 following challenge inoculation with *C. cassiicola*. 202 and 40 genes/ transcripts were identified as up- and down-regulated respectively when mapped back to the NGS data.

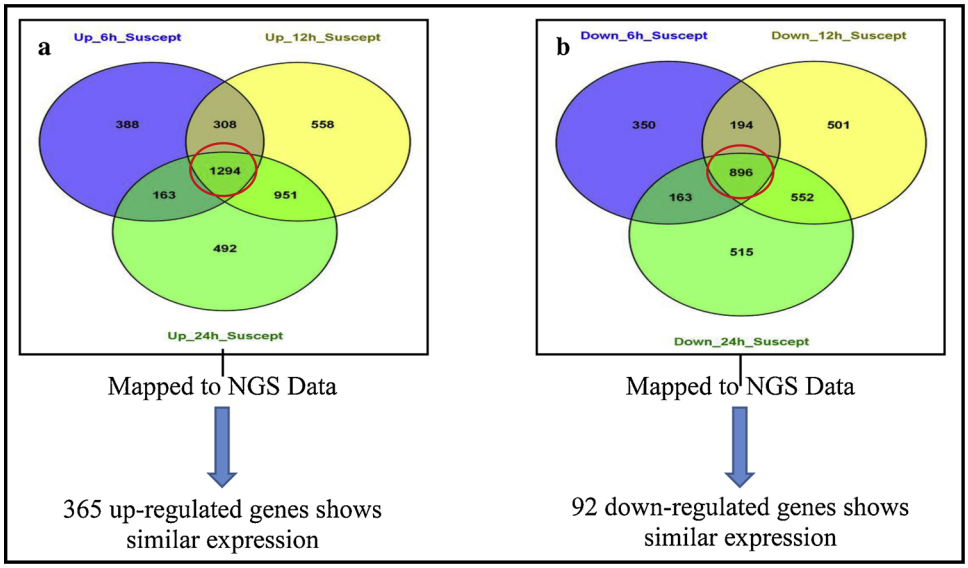

Fig. 10. Venn diagram showing sharing of up-regulated (a) and down-regulated (b) genes/ transcripts associated with progression of disease in susceptible condition at different time points in the susceptible clone RRII 105 following challenge inoculation with *C. cassiicola*. 365 and 92 genes/ transcripts were identified as up- and down-regulated respectively when mapped back to the NGS data.

induces the expression of defense-related genes and repression of photosynthesis. Cellulose synthase, chitinase, trehalose-6-phosphate synthase, starch branching enzyme II, beta-glucosidase, phosphomannomutase, sucrose phosphate synthase transcripts were annotated under carbohydrate metabolic process. Plants have the ability to recognize and respond to various pathogens, which lead to reprogramming the function at the cellular level for activating and deploying defense responses essential to arrest pathogen growth. Such responses are associated with increased demand for energy that is provided by primary metabolic processes. Expression of beta-glucosidase in resistant clone GT 1 is believed to be induced following *C. cassiicola* infection as a similar observation was reported in tomato in response to fungal infections [42].

Some of the significantly up-regulated genes in T2 encode pathogenesis-related (PR) proteins including chitinases that are capable of degrading cell wall components of microbial pathogens. The major role of plant chitinase is to hydrolyze the N-acetylglucosamine polymer chitin. We have also observed the up-regulation of chitinase gene in the susceptible clone RRII 105 following challenge inoculation. The differential expression of this enzyme is highly enhanced by fungi, bacteria, viruses and other biotic and abiotic stresses [43]. Plant susceptibility towards pathogen is highly reduced by chitinase in combination with several other pathogen resistance proteins, especially when the pathogen has chitinaceous cell wall, as in the case of *C. cassiicola*. Regulation of transcription under RNA metabolic process/RNA biosynthesis process was one of the key biological processes observed in T2. Like many other complex biological processes, plant defense responses upon pathogen infection involve transcriptional regulation of a large number of host genes [44]. It is also believed that many of these differentially regulated genes encode enzymes in a variety of primary and secondary metabolic pathways and the change of their synthesis may result in reprogramming of cellular metabolism helping in degradation of pathogen cell wall. Other differentially regulated plant host genes encode regulatory factors that are involved in the activation, suppression and modulation of various signalling pathways in plant cells upon pathogen infection. Thus, transcriptional regulation of plant host genes is an integral part of plant defense responses with a critical role in induced plant disease resistance [45].

Differential gene expression observed between resistant and susceptible clone in the absence of pathogen indicated that resistance

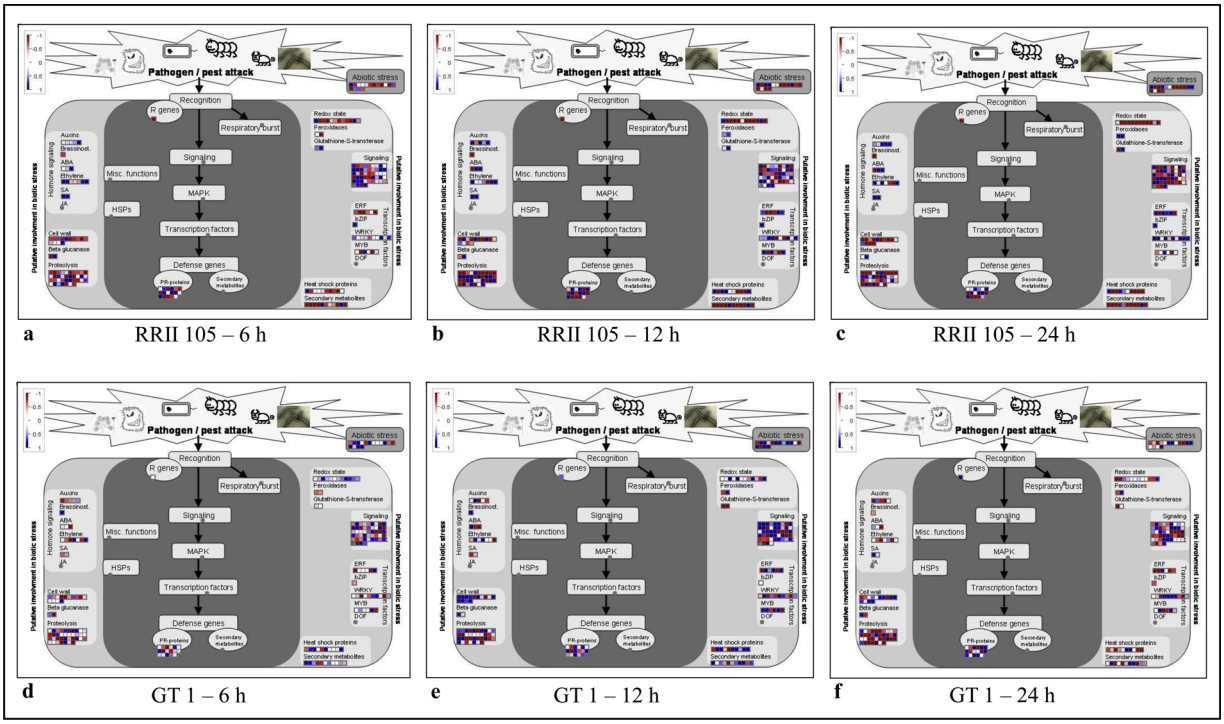

Fig. 11. An overview of activation of disease response as annotated in MapMan functional groups. Comparison of RRII 105 (a, b, c) with GT 1 (d, e, f) at different time points (6, 12 and 24 h) following infection with *C. cassiicola* is shown. Triggering of defence response subsequent to pathogen attack was through the involvement of transcription factors and other signaling mechanisms. Each square represents the normalized count value for single gene (heatmap on the left and right within each category).

mechanism had already been built-up in the resistant clone GT 1 as a priming mechanism. Therefore, GT 1 appears inherently tolerant to *C. cassiicola* compared to RRII 105. In GT 1, 19 unique NBS-LRR transcripts were up-regulated following challenge inoculation, which could be grouped under TIR-NBS-LRR and CC-NBS-LRR subfamilies of disease resistance genes prevailing in plant species having active role in pathogen detection (Supplementary Table S4). Conformational changes in the amino-terminal and LRR domains of plant NBS-LRR proteins due to interaction with either a modified host protein or a pathogen protein, might promote the exchange of ADP for ATP by the NBS domain, which could activate downstream signalling, leading to pathogen resistance [46]. This could be a reason that large number of ATP binding proteins was up-regulated in resistant clone GT1 upon infection (T2). Cullin ring ubiquitin ligase complex was uniquely observed in resistant clone GT 1 upon infection. This complex has been reported as key mediators of post-translational protein regulation [47] involved in disease resistance caused by microorganisms [48].

4.2. GO analysis of DGEs and GO enrichment analysis

The differentially expressed genes classified according to their gene ontology revealed their roles in regulation of biological processes, molecular functions and cellular component to minimize pathogen establishment in the host. The up-regulated DGEs were mainly enriched in the functional pathway that were mostly associated with disease resistance, defence response to fungus, proteolysis, signal transduction and protein serine-threonine kinase activity. During the process of infection, *C. cassiicola* secrete various virulence factors into the host cells to enable successful invasion. As far as the host is concerned, R genes are primarily responsible for recognising pathogenic secretions, which trigger plant immune system. In this study also disease resistance proteins (RGA2, RPM1, RPS2, RPS5), leucine-rich repeat containing protein, chitinase, multidrug resistance protein and TMV resistance protein were significantly induced in the resistant clone upon challenge inoculation with *C. cassiicola* (Tables 3 and 4). These induced genes have also been reported to sequester effectors and reactivate plant immune system [49]. These resistant genes were found to be more expressed in the resistant clone GT 1 compared to the susceptible clone RRII 105. Gene ontology category analysis revealed that defence-related proteins were highly enriched. Numerous DGEs related to various transcription factors (WRKY, NAC, GATA, putative TF proteins), zinc finger protein, disease resistance proteins, leucine rich repeat containing protein, chitinase, TMV resistance protein and ATP binding proteins showed clear differences between the two cultivars and most were up-regulated in the resistant clone following infection. Our analysis also showed that although the original expression status of some of the transcripts in both susceptible and resistant clone was very similar before infection, certain defence-related genes with pathogen-defence functions demonstrated constitutively higher expression in resistant clone GT 1 than in the susceptible clone RRII 105. This may play a critical role in defending the infection from progressing in the resistant clone. Induction of these pathogen-related proteins in plants when exposed to various pathogens has been well documented [50]. These were induced to a higher level in the resistant clone to trigger rapid activation of defence-responsive mechanisms in plants challenged with the pathogen.

The highly variable expression patterns of transcription factors revealed substantial changes of expression that may play crucial role in plant resistance to pathogens. Transcription factors, interacting with the transcriptional regulatory elements present adjacent to the genes that they regulate, possess the ability to govern the expression of many downstream genes to control diverse biological processes. An alteration in transcript abundance of TFs can trigger a cascade of reactions implicated in many physiological processes resulting in a substantial change in downstream gene expression [51]. In our study, numerous TF families (WRKY, NAC, GATA, putative TF proteins) were identified to be induced. These are known key regulators in the defence response [52–54]. Thus it is evident that different types of transcription factors in

different ways significantly regulate expression of a large set of disease-related genes.

4.3. Microarray gene signatures for disease tolerance

Microarray analysis is a comprehensive and high-throughput approach used to screen candidate genes and predict gene function. We performed microarray-based gene expression analysis (Agilent) to gain an overall picture of transcriptome-wide changes in rubber trees under *Corynespora* infection at different time intervals. The expression profiles of many genes obtained by RNA-seq and microarray analysis were in good accordance (Supplementary Table S17). Furthermore, the microarray data showed transcriptional changes of various well-known functional and regulatory genes; including transcription factors, kinases, heat shock proteins, late embryogenesis-abundant (LEA) proteins, osmoprotectant biosynthesis-related proteins, hormone-related proteins, transporters and detoxification enzymes. The study identified genes that are highly differentially expressed and involved in defense-related networks. These genes were mostly related to defense response, transcriptional regulation, signalling and metabolism.

Plants recognize pathogens through pathogen-associated molecular patterns (PAMPs) by host sensors, referred to as pattern-recognition receptors (PRRs) which initiate a series of defense responses called PAMP-triggered immunity (PTI). Most of these receptors belong to receptor like kinases (RLK) [55,56]. In our study LRR type RLK and serine threonine protein kinase were up-regulated in both clones. During defense, RLKs recognize pathogen-associated signals and trigger a broad range of downstream defense responses. Hormone signalling related genes (auxin, brassinosteroid, salicylic acid, ethylene, ABA) are major defensive players in the regulation of signalling networks underlying resistance to pathogens [57]. SA plays an extensive signalling role in plants, in defense against pathogens by activating expression of defense-related genes, inducing cell death and provoking systemic acquired resistance [58]. One of the features of plant defense response is production of PR proteins. In our study PR proteins (chitinases, thaumatin-like proteins, peroxidase, endoglucanase) were induced after inoculation with the pathogen, which is closely related with plant defense and fungal pathogenicity. PR proteins also play important roles in hypersensitive responses and contribute to SAR. Induction and greater accumulation of PR cell wall-degrading enzymes suggests that the degradation of cell wall components of pathogens is important defense reaction in *Hevea* against *C. cassiicola* at the early infection stage. It is speculated that the defense responses initiated at the early stage of fungal infection continue to play a role at the later stage of infection in the resistant clone. Genes related with cell wall, beta-glucanase, and proteolysis were induced in response to infection.

Our research identified up-regulation of a large number of genes encoding TFs (NAC, ERF, MYB, GATA, WRKY, LEA, bZIP), that are important upstream regulatory protein which play a crucial role in regulating plant responses to stresses and enhanced disease resistance against pathogens. NAC TF is reported to play critical roles in plant immune responses, basal defense and SAR [54]. ERFs are transcriptional activators and are known to be responsible for generating tolerance to stress in plants. MYB TF is large and functionally diverse and plays crucial roles in the interaction of regulatory networks that control development, metabolism and response to biotic and abiotic stresses [59]. A single TF can regulate the expression of multiple genes. Transcriptional control of stress-response gene expression is a crucial component of plant response to a range of stresses [60]. In our study, many DGEs encoding TF families were identified in the tolerant genotype. These can modulate the interaction of defense pathways activated by pathogen infection. Thus, it can be concluded that the presence of these TFs indicates that various signal molecules act to improve disease tolerance in the resistant clone GT 1. Most of these identified DGEs from microarray analysis were consistent with our transcriptome results.

4.4. SSRs and SNPs discovery from transcripts

SSR and SNP markers have become the preferred markers for many applications in genetic and genomic studies. As effective and stable markers, they play an important role in molecular aided selection and breeding. EST-derived SSRs are easily transferable across species, are distributed in coding sequences, can be related to functional genes and are widely used for comparative mapping of related crops or genetic diversity of wild and cultivated accessions [61]. Further, EST-SSRs also represent transcripts that contribute to important agronomic traits [62] and are useful for molecular marker assisted selection breeding, with molecular markers either originating from a gene of interest or co-segregating a gene with a desirable agronomic trait. Our transcriptome data identified a total number of 32323 SSRs with a relative abundance of 503.59 SSRs/Mb. The mean transcript length of these SSRs was 1838.1 bases and all types of repeat units (mono, di, tri, tetra, penta and hexa nucleotide) were observed. SSR polymorphism between the two rubber clones RRII 105 and GT 1 in control and challenged conditions was also identified. Similarly, nucleotide sequence variation was explored between the two control libraries (C1 and C2) and two treatment libraries (T1 and T2). Common and unique variations at single base level were observed in both these conditions. SNP calling identified both synonymous and non-synonymous SNPs. We believe that these SSRs and SNPs will be valuable genetic resources for constructing linkage maps, quantitative trait locus (QTL) mapping, genetic diversity and MAS breeding.

5. Concluding remarks

Our study presented a genome-wide gene expression profiling of *H. brasiliensis* during *C. cassicola* infection in two clones. A considerable number of differentially expressed genes that play major roles in the response to pathogen were identified using transcriptome sequencing. A huge amount of transcriptomic data obtained from microarray analysis has revealed disease-responsive and clone-dependent expression patterns at different time points. The biotic stress map view of MapMan analysis showed that genes related with signalling, transcription factors and secondary metabolism were rapidly induced, which was important for the defense response. Traditional rubber tree improvement is a lengthy and challenging process. Availability of the transcriptome will hopefully enable development of a resource of molecular markers that can be used to identify genes involved in disease tolerance thereby providing scope for improvement of rubber tree with disease resistance. This study is a significant step towards development of genomic resources for rubber and will accelerate functional genomic studies thereby facilitating marker-assisted selection breeding in rubber tree. However, the transcript responses to the progressive disease stress should be further studied in more genotypes of rubber showing extreme levels of resistance/ susceptibility to further detect key metabolic pathways, metabolites and genes for improving rubber disease resistance.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to thank Dr. James Jacob, Director (Research), Rubber Research Institute of India for his support and encouragement throughout the work. They acknowledge Genotypic Technology, Bangalore for generating RNA-seq and microarray data. Funding provided by Australian Government to CBR as Australian Endeavour Fellowship is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.cpb.2019.02.002.

References

[1] ANRPC (Association of Natural Rubber Producing Countries), Natural Rubber Trends and Statistics, Kuala Lumpur, March (2018).
[2] D.O.J. Amaral, M.M.A. Lima, L.V. Resende, M.V. Silva, Differential gene expression, induced by salicylic acid and *Fusarium oxysporum* f. Sp. *Lycopersici* infection in tomato, Pesq Agropec Bras 43 (8) (2008) 1017–1023.
[3] L.F. Thatcher, J.P. Anderson, K.B. Singh, Plant defence responses: what have we learnt from *Arabidopsis*? Funct. Plant Biol. 32 (2005) 1–19.
[4] S. Desender, D. Andrivon, F. Val, Activation of defence reactions in *Solanaceae*: where is the specificity? Cell. Microbiol. 9 (2007) 21–30.
[5] C.B. Roy, C.K. Jacob, T. Saha, Towards identification of genes involved in host tolerance to corynespora leaf disease in rubber (*Hevea brasiliensis*), 5th International Conference on Plant Pathology in the Globalized Era (2009).
[6] O. Morozova, M. Hirst, M.A. Marra, Application of new sequencing technologies for transcriptome analysis, Annu. Rev. Genomics Hum. Genet. 10 (2009) 135–151.
[7] M. Jain, Next-generation sequencing technologies for gene expression profiling in plants, Brief Func. Genomics 11 (2011) 63–70.
[8] J.A. Martin, Z. Wang, Next-generation transcriptome assembly, Nat. Rev. Genet. 12 (2011) 671–682.
[9] M. Garber, M.G. Grabherr, M. Guttman, C. Trapnell, Computational methods for transcriptome annotation and quantification using RNA-seq, Nat. Methods 8 (2011) 469–477.
[10] K. Nakasugi, R. Crowhurst, J. Bally, P. Waterhouse, Combining transcriptome assemblies from multiple *de novo* assemblers in the allo-tetraploid plant *Nicotiana benthamiana*, PLoS One 9 (3) (2014) e91776, , https://doi.org/10.1371/ journal.pone.0091776.
[11] L. Salgado, D. Koop, D. Pinheiro, R. Rivallan, V. Le Guen, M. Nicolas, *De novo* transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers, BMC Genomics 15 (1) (2014) 236–250.
[12] U.A. Hurtado Páez, I.A. García Romero, S. Restrepo Restrepo, F.A. Aristizábal Gutiérrez, D. Montoya Castaño, Assembly and analysis of differential transcriptome responses of *Hevea brasiliensis* on interaction with *Microcyclus ulei*, PLoS One 10 (8) (2015) e0134837.
[13] V. Costa, C. Angelini, I. De Feis, A. Ciccodicola, Uncovering the complexity of transcriptomes with RNA-Seq, J. Biomed. Biotechnol. (2010), https://doi.org/10.1155/2010/853916.
[14] W.C. Ho, W.H. Ko, A simple method for obtaining single spore isolates of fungi, Bot. Bull. Acad. Sin. 38 (1997) 41–44.
[15] D.R. Zerbino, E. Birney, Velvet: algorithms for *de novo* short read assembly using *de Brujin* graphs, Genome Res. 18 (2008) 821–829.
[16] M.H. Schultz, D.R. Zerbino, M. Vingron, E. Birney, Oases: Robust *de novo* RNA-Seq assembly across the dynamic range of expression levels, Bioinformatics 28 (8) (2012) 1086–1092.
[17] M. Magrane, UniProt Consortium: UniProt Knowledgebase: a Hub of Integrated Protein Data. *Database (Oxford)*, 2011 (2011), https://doi.org/10.1093/database/bar009.
[18] R.L. Tatusov, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, S. Smirnov, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, D.A. Natale, The COG database: an updated version includes eukaryotes, BMC Bioinformatics 4 (2003) 41.
[19] A. Bateman, L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L. Sonnhammer, D.J. Studholme, C. Yeats, S.R. Eddy, The Pfam protein families database, Nucleic Acids Res. 32 (2004) D138–D141 (Database Issue).
[20] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinformatics 10 (421) (2009).
[21] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat. Methods 5 (2008) 621–628.
[22] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010) R106.
[23] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. 57 (1995) 289–300.
[24] The Gene Ontology Consortium, The gene ontology project in 2008, Nucleic Acids Res. 36 (2008) 440–444.
[25] T. Beissbarth, T.P. Speed, GOstat: find statistically overrepresented Gene Ontologies within a group of genes, Bioinformatics 20 (2004) 1464–1465.
[26] Z. Du, X. Zhou, Y. Ling, Z. Zhang, Z. Su, AgriGO: a GO analysis toolkit for the agricultural community, Nucl Acids Res. 38 (2010) 64–70.
[27] M.J.L. de Hoon, S. Imoto, J. Nolan, S. Miyano, Open source clustering software, Bioinformatics 20 (9) (2004) 1453–1454.
[28] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760.
[29] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup: The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[30] C.K. Jacob, Diseases of potential threat to rubber in India, Planters Chronicle 92 (1997) 451–461.
[31] M.J. Manju, S.P. Idicula, C.K. Jacob, K.K. Vinod, E. Prem, M. Suryakumar, R. Kothandaraman, Incidence and severity of Corynespora leaf fall (CLF) disease of rubber in coastal Karnataka and North Malabar region of Kerala, Ind. J. Nat. Rub. Res. 14 (2001) 137–141.
[32] J. Mathew, Clonal resistance of *Hevea brasiliensis* to Corynespora leaf fall disease, in: C.K. Jacob, P. Srinivas, C.B. Roy (Eds.), Corynespora Leaf Disease of Hevea brasiliensis – Strategies for Management, Rubber Research Institute of India, India, 2006, pp. 83–96.
[33] K. Triwitayakorn, P. Chatkulkawin, S. Kanjanawattanawong, S. Sraphet, T. Yoocha, D. Sangsrakru, J. Chanprasert, C. Ngamphiw, N. Jomchai, K. Therawattanasuk, S. Tangphatsornruang, Transcriptome sequencing of Hevea brasiliensis for development of microsatellite markers and construction of a genetic linkage map, DNA Res. 18 (2011) 471–482.
[34] Z. Xia, H. Xu, J. Zhai, D. Li, H. Luo, C. He, X. Huang, RNA-Seq analysis and *de novo* transcriptome assembly of *Hevea brasiliensis*, Plant Mol. Biol. 77 (2011) 299–308.
[35] S.P. Pandey, I.E. Somssich, The Role of WRKY Transcription Factors in Plant Immunity, Plant Physiol. 150 (4) (2009) 1648–1655.
[36] K. Yazaki, ABC transporters involved in the transport of plant secondary metabolites, FEBS Lett. 580 (4) (2006) 1183–1191.
[37] P.A. Rea, Plant ATP-binding cassette transporters, Annu. Rev. Plant Biol. 58 (2007) 347–375.
[38] J. Kang, J. Park, H. Choi, B. Burla, T. Kretzschmar, Y. Lee, E. Martinoia, Plant ABC transporters, Arabidopsis Book 9 (2011) e0153.
[39] M. Sasabe, K. Toyoda, T. Shiraishi, Y. Inagaki, Y. Ichinose, cDNA cloning and characterization of tobacco ABC transporter: NtPDR1 is a novel elicitor-responsive gene, FEBS Lett. 518 (2002) 164–168.
[40] S. Glombitza, P.H. Dubuis, O. Thulke, G. Welzl, L. Bovet, M. Gotz, M. Affenzeller, B. Geist, A. Hehn, C. Asnaghi, Crosstalk and differential response to abiotic and biotic stressors reflected at the transcriptional level of effector genes from secondary metabolism, Plant Mol. Biol. 54 (2004) 817–835.
[41] Y. Kobae, T. Sekino, H. Yoshioka, T. Nakagawa, E. Martinoia, M. Maeshima, Loss of AtPDR8, a plasma membrane ABC transporter of *Arabidopsis thaliana*, causes hypersensitive cell death upon pathogen infection, Plant Cell Physiol. 47 (2006) 309–318.
[42] A. Edreva, V. Sotirova, I.D. Georgieva, E. Stoimenova, R. Rodeva, N. Bogatzevska, Differential expression of β-glucosidase in tomato - stress stimuli systems, Acta Physiol. Plant. 22 (2000) 274–277.
[43] Z.K. Punja, Y.Y. Zhang, Plant chitinases and their roles in resistance to fungal diseases, J. Nemat. 25 (4) (1993) 526–540.
[44] P.J. Rushton, I.E. Somssich, Transcriptional control of plant genes responsive to pathogens, Curr. Opin. Plant Biol. 1 (1998) 311–315.
[45] C. Chen, Z. Chen, Potentiation of developmentally regulated plant defense response by AtWRKY18, a Pathogen-Induced *Arabidopsis* transcription factor, Plant Physiol. 129 (2) (2002) 706–716.
[46] B.J. DeYoung, R.W. Innes, Plant NBS-LRR proteins in pathogen sensing and host defense, Nat. Immunol. 7 (2006) 1243–1249.
[47] M.D. Petroski, R.J. Deshaies, Function and regulation of cullin–RING ubiquitin ligases, Nat. Rev. Mol. Cell Biol. 6 (2005) 9–20.
[48] W.M. Gray, Plant Defence: A New Weapon in the arsenal, Curr. Biol. 12 (10) (2002) 352–354.
[49] K. Tsuda, Katagiri, Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity, Curr. Opin. Plant Biol. 13 (2010) 459–465.
[50] M. Xin, X. Wang, H. Peng, Y. Yao, C. Xie, Y. Han, Z. Ni, Q. Sun, Transcriptome comparison of susceptible and resistant wheat in response to powdery mildew infection, Genomics Proteomics Bioinf. 10 (2012) 94–106.
[51] Y. Wang, X. Tao, X.M. Tang, L. Xiao, J.L. Sun, X.F. Yan, Comparative transcriptome analysis of tomato (*Solanum lycopersicum*) in response to exogenous abscisic acid, BMC Genomics 14 (841) (2013).
[52] H.S. Ryu, M. Han, S.K. Lee, J.I. Cho, N. Ryoo, S. Heu, A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defence response, Plant Cell Rep. 25 (2006) 836–847.
[53] T. Eulgem, I.E. Somssich, Networks of WRKY transcription factors in defence signaling, Curr. Opin. Plant Biol. 10 (2007) 366–371.
[54] M. Nuruzzaman, A.M. Sharoni, S. Kikuchi, Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants, Front. Microbiol. 4 (4) (2013) 248.
[55] P. Liu, W. Wei, S. Ouyang, J.S. Zhang, S.Y. Chen, W.K. Zhang, Analysis of expressed receptor-like kinases (RLKs) in soybean, J. Genet. Genomics 36 (2009) 611–619.
[56] M. Beck, W. Heard, M. Mbengue, S. Robatzek, The Ins and OUTs of pattern recognition receptors at the cell surface, Curr. Opin. Plant Biol. 15 (2012) 1–8.
[57] Y.X. Yang, G.J. Ahammed, C. Wu, S.Y. Fan, Y.H. Zhou, Crosstalk among jasmonate, salicylate and ethylene signaling pathways in plant disease and immune responses, Curr Protein Peptide Sci. 16 (5) (2015) 450–461.
[58] S. Lemarie, A. Robert-Seilaniantz, C. Lariagon, J. Lemoine, N. Marnet, M. Jubault, M.J. Manzanares-Dauleux, A. Gravot, Both the jasmonic acid and salicylic acid pathways contribute to resistance to the biotrophic clubroot agent *Plasmodiophora brassicae* in Arabidopsis, Plant Cell Physiol. 56 (11) (2015) 2158–2168.
[59] C. Dubos, R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, L. Lepiniec, MYB transcription factors in *Arabidopsis*, Trends Plant Sci. 15 (10) (2010) 573–581.
[60] K.B. Singh, R.C. Foley, L.O. Sanchez, Transcription factors in plant defense and stress responses, Curr. Opin. Plant Biol. 5 (2002) 430–436.
[61] K. Kumari, M. Muthamilarasan, G. Misra, S. Gupta, A. Subramanian, S.L. Parida, Development of eSSR-markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species, PLoS One 8 (6) (2013) 1–15.
[62] S. Dutta, G. Kumawat, B.P. Singh, D.K. Gupta, S. Singh, V. Dogra, K. Gaikwad, T.R. Sharma, R.S. Raje, T.K. Bandhopadhya, S. Datta, M.N. Singh, F. Bashasab, P. Kulwal, K.B. Wanjari, R.K. Varshney, D.R. Cook, N.K. Singh, Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus cajan (L.) Millspaugh], BMC Plant Biol. 11 (2011) 17–29.