Indian J. Genet. Plant Breed., (2023); 83(4): 573-586
https://doi.org/10.31742/ISGPB.83.4.14
ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Identification long non-coding RNAs, miRNAs and their targets in witches' broom disease tolerance in cocoa (*Theobroma cacao* L.)

Antara Das[1], Aparna Veluru[*], Alokesh Das[2] and R. Tava Pandian[3]

Abstract

A large part of transcripts is non-coding, which is transcribed from junk DNA; long non-coding RNAs and micro-RNAs regulate the expression levels of mRNAs. For the first time, we identified lncRNAs and miRNAs with their regulatory role in the disease tolerance of *Theobroma cacao*. In this study, about 2616 lncRNAs and 153 miRNAs were identified from 10 RNA-seq data representing healthy and witches' broom diseased tissues of cocoa. Around 604 lncRNAs are differentially expressed among healthy and diseased tissues. LncRNAs targeted 9692 mRNAs; 8827 are cis-acting, and 765 are trans-acting. Among targeted mRNAs, 281 are disease resistance-related transcripts, and 211 transcription factors (TFs) belong to more than 50 TF families, which were found to be involved in the regulation of the disease tolerance process. The identified 153 miRNAs belong to 27 miR families, and around 5337 mRNAs are targeted by the miRNAs, among them 114 codes for TFs and 170 codes for disease resistance protein. Ethylene responsive factor, bHLH, WRKY, MYB, bZIP, GTE, GATA, and heat stress transcription factors are the dominant TFs targeted by lncRNAs, and miRNAs play vital roles in disease progression and tolerance. A total of 55 lncRNAs-miRNAs interacting pair is identified, which were working on endogenous target mimics (e-TMs) mechanism and influenced the expression of 955 mRNAs. The ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) biological pathway analysis reveal that ncRNAs and their targets mRNAs code for transcription factors and genes that are involved in the disease tolerance processes, including synthesis of disease resistance proteins, amino acids, antibiotics, intracellular proteins that directly or indirectly recognize pathogen effectors are essential for plant biotic stress condition. The present study provides lncRNA and miRNA-based regulatory insight into the genes governing disease progression and tolerance in cocoa.

Keywords: Long non-coding RNA, micro-RNA, miRNAs-lncRNAs interaction, *Theobroma cacao* L.

Introduction

Non-coding RNAs, their importance, and their regulatory role in the gene expression of eukaryotic organisms were realized recently. The lncRNAs get involved in almost every plant biological process during its growth and development, including abiotic and biotic stresses (Shiv et al. 2023; Supriya et al. 2022). So far, several studies have been carried out for lncRNA identification and studying their role in several plants. The regulating role of lncRNAs during plant biotic stress was also well evidenced. Arabidopsis infected with *Fusarium oxysporum* triggered the expression of several lncRNAs, many of which were associated with antifungal immunity (Zhu et al. 2014). Joshi et al. (2016) identified around 931 differentially expressed lncRNAs upon infection with *Sclerotinia sclerotiorum* in canola. The expression of cotton lncRNAs against fungal infection was reported by Zhang et al. in 2018. In rice, around 567 lncRNAs were identified in response to bacterial pathogen *Xanthomonas oryzae* pv infection. o*ryzae (Xoo*) (Yu et al. 2020) and target analysis of some of these lncRNAs revealed their involvement in the jasmonic acid (JA) signaling cascade, which provide resistance against bacterial blight (Yu et al. 2020).

Division of Crop Improvement, ICAR-Central Plantation Crops Research Institute, Kasaragod 671 121, Kerala, India.

[1]Present Address: Division of Biotechnology, ICAR-Central Institute of Subtropical Horticulture, Regional Research Station, Malda 732 101, West Bengal, India.

[2]Department of Botany, Rampurhat College, Birbhum 731 224, West Bengal, India.

[3]Division of Plant Protection, ICAR-Central Plantation Crops Research Institute, Regional Research Station, Vittal 574 243, Karnataka, India.

***Corresponding Author:** Aparna Veluru, Division of Crop Improvement, ICAR-Central Plantation Crops Research Institute, Kasaragod, Kerala, India, E-Mail: aparna.cpcri@gmail.com

How to cite this article: Das A., Veluru A., Das A., Pandian R.T.P. 2023. Identification long non-coding RNAs, miRNAs and their targets in witches' broom disease tolerance in cocoa (*Theobroma cacao* L.). Indian J. Genet. Plant Breed., **83**(4): 573-586.

Source of support: Nil

Conflict of interest: None.

Received: April 2023 **Revised:** Aug. 2023 **Accepted:** Sept. 2023

MicroRNAs are a class of regulatory ncRNAs having a length of 20–24 nucleotides. They regulate gene expression, participate in various biological routes, conserve genome integrity, metabolism, growth, and development, and most importantly, adaptive responses towards abiotic and biotic stresses. Experiments carried out in many crops evidenced the conservation of miRNAs across the plant genera and their functional association with the plants. Even though the regulatory role of miRNAs was majorly proven during abiotic stress, its role in biotic stress was also inevitable. Reported studies have demonstrated the role of miRNAs during microbial infections in plants and mammals. In plants, Navarro et al. (2006) said the first miRNA, miR393, in *Arabidopsis*, plays against bacteria by regulating the auxin signaling pathway. Likewise, differential expression of miRNA patterns was reported in wheat upon infection with powdery mildew fungus (Xin et al. 2011). In apples, pathogen-induced host-specific miRNA (Md-miR395and Md-miR156ab) expression was seen after leaf spot fungus *Alternaria* infection, which negatively regulates the target genes *MdWRKYN1*, *MdWRKY26*, and induced pathogenicity (Zhang et al. 2017). Similarly, differential expressions of miRNAs were also noticed with healthy and susceptible sugarcane cultivars for smut disease (Su et al. 2017).

Besides sequestration/degradation of targeted mRNAs and miRNAs, lncRNAs also act as precursors for miRNAs. Co-expression of the lncRNA-miRNA complex also revealed the participation of lncRNAs in target miRNA degradation. Studies during recent years have demonstrated the emerging job of lncRNAs as miRNA regulators by functioning as endogenous target mimics (eTMs). Genome-wide analysis techniques identified some of the candidate eTMs from sequenced plants. Wu et al. (2013) identified eTMs for around 20 miRNAs in rice and *Arabidopsis*. In *Cajanus* the role of lncRNAs as e-TMs was discovered for flower, pod, and seed development (Das et al. 2019, 2020). Borah et al. (2018) discussed lncRNAs as e-TMs that help plants to withstand low nutrient conditions. The lncRNAs also participate in plant biotic stress conditions by acting as eTMs. For example, Tomato lncRNAs slylnc0195 and slylnc1077 expressed in response toTYLCV infection have eTMs that suppress miR166 and miR399 correspondingly to repress the plant defense response (Wang et al. 2015). Similarly, conserved eTM site for miR482b was identified in lncRNAs of tomato (lncRNA23468, lncRNA01308, and lncRNA13262); over-expression of any of these lncRNA results in decreased miRNA (miR482b) expression, and increased target gene NBS-LRRs, expression, effects in ensuring better resistance to *Phytophthora infestans* (Jiang et al. 2019). Likewise, lncRNA39026 contains an eTM site, which helps to decoy miR168a and influence the target gene expression (Pathogenesis-related genes) in tomatoes to combat *P. infestans* infection (Hou et al. 2020).

Cocoa (*Theobroma cacao* L.) is an important plantation grown widely in Central and South America, Africa, and Asia. Globally, five to six million farmers are directly involved in cocoa cultivation, and about 40 to 50 million people are getting livelihood security from the crop (End et al. 2014). In cocoa, witches' broom disease (WBD) is the most destructive disease that causes significant economic loss (50–90%) to the harvest, and *Moniliophthora perniciosa*, a basidiomycete fungus, causes the disease. This disease is more prevalent in the south and Central America and Caribbean countries where most cocoa germplasm existed (End et al. 2014). The disease causes direct production losses by infecting flower cushions and mature pods. The distal tissues of infected parts show the symptoms of hypertrophy and hyperplasia, the proliferation of axillary shoots due to loss of apical dominance, which subsequently develops into broom-like make-up called a green broom (Meinhardt et al. 2008). Successive pathogen attacks (WBD) kill the infected plants (Gramacho et al. 2016). WBD pathogen *M. perniciosa* was sequenced using Sanger and NGS methods (Meinhardt et al. 2014; Mondego et al. 2008). The genome of *M. perniciosa* contains more active transposable elements, which account for more variability within species, resulting in different biotypes with alerted virulence (Pereira et al. 2015; Meinhardt et al. 2014). The genome annotation of the WBD pathogen proved its ability to produce hormones, secondary metabolites, toxins, and reactive oxygen species (ROS) detoxification systems. More cytochrome P450 monooxygenases are essential for detoxification, adaptation, and disease development (Mondego et al. 2008).

The WBD pathogen has biotrophic and necrotrophic phases. The biotrophic phase grows in the apoplast region of the plant cell and absorbs nutrients; during this stage, it secretes several enzymes that encode ROS detoxification and other effectors to overpower plant defense (Mondego et al. 2008). Pathogen induces hormonal imbalances within the host to get the soluble sugars within the apoplastic region (Scarpari et al. 2005). Still, when supply disturbs or decreases, the pathogen promotes cell death and senescence, producing dry brooms, where the mycelia survive and defend themselves against host defensive compounds. Transcriptomic studies have been conducted to identify the differences between biotrophic and necrotrophic using microarrays, ESTs, and qPCR (Rincones et al. 2008). The pathogen *M. perniciosa* can produce hormones like abscisic acid (ABA), jasmonic acid (JA), indole acetic acid (IAA), and salicylic acid (SA) (Kilaru et al. 2007). Transcriptome analysis performed on infected tissue showed that the fungus triggers hormonal imbalances by disturbing host genes involved in the biosynthesis or signaling of hormones (auxin, gibberellin, cytokinin, and ethylene) (Teixeira et al. 2014). Ethylene stimulates tissue elongation, symptom development, and cell death in infected green brooms, and cytokinin's helps in loss of apical dominance (Teixeira et al. 2014; Scarpari et al. 2005). Pathogen *M. pernicious* interaction in resistant and susceptible cocoa material was studied

through microarrays and cDNA library sequencing (da Hora Junior et al. 2012; Gesteira et al. 2007). Host-pathogen interaction was also studied in green brooms using the RNA-seq technique (Teixeira et al. 2014).

Comparing cDNA libraries in susceptible and resistant cocoa lines infected with *M. perniciosa* showed qualitative differences (Gesteira et al. 2007). During infection, the plant expresses a series of genes essential in plant defense responses, like PR- and NB-LRR proteins (Teixeira et al. 2014). In the resistant lines, the defense genes responded early and at higher levels than susceptible ones (Gesteira et al. 2007). Though susceptible ones produce defense molecules, they cannot restrict the pathogen effectively (da Hora Junior et al. 2012). Biochemical studies conducted during host-pathogen interaction revealed increased compounds like procyanidins, theobromine, caffeine, and phenolics, mainly tannins after primary infection (Scarpari et al. 2005; Chaves and Gianfagna 2007).

The present experiment was planned to study the role of ncRNAs, especially lncRNAs and miRNAs, in witches' broom diseased plants compared to the healthy ones. However, numerous studies are available about the control mechanisms of several biotic and abiotic stress factors in plants; information is almost scanty about the job of non-coding RNAs such as lncRNAs and miRNAs and their working mechanism in cocoa disease tolerance. Studying the expression pattern of cacao lncRNAs, miRNAs in healthy vs. diseased samples may prove their potential role in disease tolerance or progression. The interaction between lncRNAs and miRNAs via endogenous target mimics was also studied, and the role of identified eTMs on potential gene expression and regulation during the WBD disease. This is the first report in cocoa about identifying and characterizing lncRNAs and miRNAs, therefore an investigation to determine their interacting mechanism for WBD disease resistance.

Materials and methods

Transcriptomic data and pipeline for lncRNA identification

The transcriptome data from the witches' broom diseased and healthy plant tissues of cocoa (*T. cacao* L.) were used to identify long non-coding RNAs and microRNAs. The sequenced data related to the stated cocoa tissues containing National Center for Biotechnology Information (NCBI) accession numbers (SRR747762, SRR747765, SSR747772, SRR747773, SRR747774, SRR747775, RR747776, SRR747777, SRR747778, and SRR747779) was downloaded from NCBI. Using the sequence read archive (SRA) toolkit, SRA files were converted to FASTQ files and subjected to downstream analysis.

Pipeline for identifying lncRNAs

The methodology followed for identifying lncRNAs has been illustrated in Fig. 1. Subsequently, adapter trimming was done, and reads with lower quality were removed with Trimomatic 0.36 (Bolger et al. 2014) with default parameters. Selected FASTQ files of diverse experiments were aligned and mapped to the reference genome of *T. cacao* (Argout et al. 2017) using a splice read aligner Tophat 2.0. Cufflinks were used for assembling aligned reads. The final transcriptome was made after merging all output files with the help of Cuffmerge. Transcript abundance was calculated by accepted Binary Alignment Map (BAM) files using Cuffdiff (Trapnell et al. 2013).

Long non-coding RNA prediction

All selected transcripts were filtered using known coding sequences of cocoa. The transcripts having non-overlapping known coding genes were chosen for subsequent analysis. The non-coding RNA types (lincRNAs-intergenic lncRNAs; lncNAT- Natural antisense transcripts; intragenic lncRNAs) are recognized based on their location and strand information. Transcripts with >200 bp length and fragments per kilo-base of transcript per million mapped reads (FPKM) value >1 were identified and screened further to find the length of open reading frame (ORF) with ORF finder (utilizing in-house perl script). Transcripts containing <100 amino acids were further sorted for their coding potential, and the remaining transcripts coding potential was calculated with coding potential calculator (CPC2) (Kang et al. 2017) and coding-non-coding index (CNCI) programs (http://www.bioinfo.org/software/cnci) (Sun et al. 2013). The transcripts identified with <0.5 CPC or CNCI scores were subjected to downstream analysis. The Basic Local Alignment Search Tool (BLASTX) was performed with databases (NCBI non-redundant protein database, Swiss-Prot database, Kyoto Encyclopedia of Genes and Genomes (KEGG) protein database, and clusters of orthologous genes (COG) database. Later the transcripts showing substantial similarity with already identified/

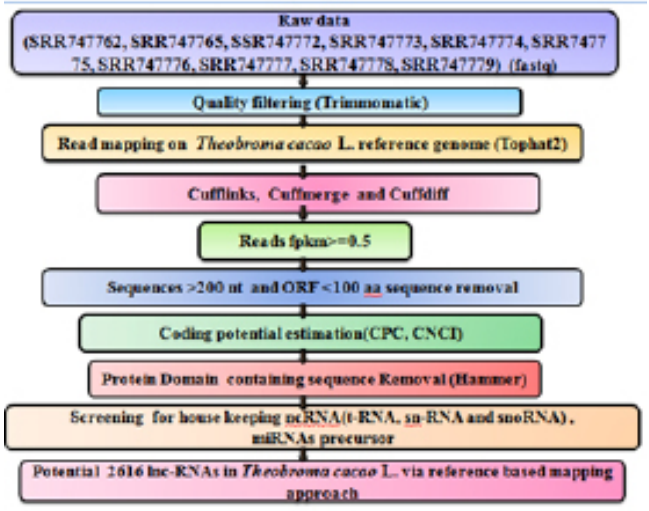

Fig. 1. Schematic representation of the methodology followed in the prediction of lncRNAs of *T. cacao*

known proteins were removed. Subsequently, the specified non-coding RNA types, such as miRNAs, rRNAs, snRNAs, tasiRNAs, tRNAs, and snoRNAs, were discarded to screen only lncRNAs. After studying the physical properties of identified lncRNAs, differential expression of lncRNAs with log2 fold was removed from the Cuffdiff results.

LncRNA targeted mRNA prediction

The mRNA targets (*Cis* and *trans*-acting mRNA) were identified for lncRNAs. Transcribed genes within a 10 kb window of lncRNAs were considered cis-target genes, whereas mRNAs having complementary sequences with lncRNAs and are coded by the genes that are away from the lncRNAs are considered trans-targets. BLASTn was used to identify the complementary sequences, keeping E-value ≤$1e^{-5}$ and identity ≥95%; RNAplex software(RNAplex dNG-50) was used to estimate complementary energy between two lines for subsequent screening and selection of potential *trans*-acting genes.

The mi-RNA prediction from the transcript data

From transcriptome data, non-coding transcripts were obtained for the prediction of miRNAs. From the database (miRBAse21), all the identified pre-miRNA and miRNA were downloaded and BLASTn against pooled non-coding transcripts with specification as less than three matches, 7-word size, and e-value cut-off of 1000.

To reform the non-coding property of transcripts, another time BLASTx was carried out against *the T. cacao* protein database (https://www.ncbi.nlm.nih.gov/assembly/GCF_000208745.1) by putting the sequence similarity cut-off value ≥80%, and the protein-coding sequences falling within the cut-off range were removed. The balance sequences were processed using CPC and CNCI concerning the NR database, and coding sequences were eliminated. The remaining non-coding transcripts were processed with the below parameters to find the miRNAs.

- The sequence should have the capacity to form a secondary stem-loop secondary structure with (Minimum Fold Energy Index) MFEI ≥0.41 and the least folding free energy.
- The sequence must have AU content ranging from 22–77%
- In one of the arms of the hairpin-loop structure, mature miRNA should present, and the miRNA sequence should be continuous without any loops or breaks
- The opposite miRNA sequence should not carry more than six mismatches
- The signature value of simple sequence repeat (SSR) should be R ≥2.5 with miRNA family
- Normalized values of Shannon entropy (NQ), base-pair distance (ND), and base pairing propensity (Npb) should be ≤ 0.45, ≤ 0.15, and ≤ 0.25, respectively.
- From several BLAST hits, we selected the sequences having the above characteristics along with the highest MFEI (Minimum Fold Energy Index) and R-values

The miRNA target prediction

For identification, miRNAs as subject and mRNA sequences of *T.cocoa* were employed as target query. The psRNATarget software (with default parameters) (Dai and Zhao 2011) was used to identify the miRNA targets. In the same way, miRNAs targeted by lncRNAs were also identified with the same program, but here, miRNAs and lncRNAs utilized as target queries and subjects correspondingly.

Detection of lncRNAs acting as candidate endogenous target mimics

The endogenous target mimics (e-TMs) are built up due to pairing among miRNAs and lncRNA. The eTMs between lncRNA and miRNA were discovered using different specifications followed by Wu et al. (2013). Using the enlisted specifications, the putative eTMs were identified using the software psRobot. By using the Vienna RNA package RNAfold web (http:/rna.tbi.univie.ac.at/) lncRNAs and miRNAs secondary structures were identified

Prediction and visualization of interaction among and between non-coding and coding RNAs was performed through the detection and imaging of interaction network build between lncRNAs, miRNAs, and their target mRNAs were done using Cytoscape (http://www.cytoscape.org/) (Shannon et al. 2003). Similarly, the visualization of the relationship between coding and non-coding RNAs was established with Cytoscape (Shannon et al. 2003). Gene Ontology (GO) by studying the interconnections among lncRNAs, miRNAs, and their mRNAs targets

Annotation of identified non-coding RNAs targeted mRNAs and lncRNA targeted mRNAs

Using the Blast2Go analysis pipeline, annotation was carried out for non-coding RNAs targeted and lncRNA-targeted mRNAs. Analysis was done for Gene Ontology (GO), and Kyoto Encyclopaedia of Genes and Genomes (KEGG) enrichment was carried out. Using Blast2GO v4.1 (Conesa and Götz, 2008), functional annotation was done for lncRNAs. Genes (Biological Process, Molecular Function, and Cellular Function) were classified based on GO definitions. To identify the role of lncRNAs in different biological pathways, KEGG Automatic Annotation Server (KAAS) analysis was done with the single-directional best-hit information procedure and default bit score threshold (Kanehisa 2002).

Results and discussion

Detection of lncRNAs in diseased and healthy tissues of T. cacao L. using transcriptome data

To find out the lncRNAs and miRNAs from transcriptome data of witch's broom diseased and healthy plants of *T. cacao,* we retrieved the existing transcriptome data from NCBI. After

trimming, 1952 million clean reads were mapped against the reference genome *T. cacao* with the TopHat2 and Bowtie program, and 1872 million (95.92%) reads could be mapped against a *T. cacao* genome. The FPKM value of all the genes and their isoforms were computed via Cufflink›s program. A total of 34,374 transcripts were obtained after the Cuffdiff program. Of those, 374 transcripts had lengths less than 200 nt. After removing the shorter transcripts (≤ 200 bp), CNCI and CPC were performed on the remaining 34,000 transcripts to check the coding potential, of which 8482 transcripts were found to have non-coding in nature, and the rest of them were (25,518) having coding potential. After CPC and CNCI, we again screened the 8482 transcripts via Transdecoder, Hammer, and Blastx with a complete protein dataset from Swissport in Pfamand. Finally, 2644 transcripts remained. After removing all other non-coding RNAs (snRNA, snoRNAs, tRNA, TasiRNA, microRNA, etc.) through the BLASTn, 2616 potential lncRNAs were finally identified. The methodology followed for finding the miRNAs and lncRNAs in *T. cacao* is shown in Fig. 1.

The lncRNAs characteristics

The average lengths of lncRNAs observed in cocoa were 1917 base pair, and most of them were in the range of 1000 to 2000 nt (1128) followed by 200 to 1000 nt (696), 2000 to 3000 nt (411) and 3000 to 4000 nt (175). The lncRNAs with exon numbers vary from 138, of which 14.94% were mono-exonic, 20% were di-exonic, 13.84% were tri-exonic, and 10.68% were tetra-exonic. Compared to mRNAs, the lncRNAs were AU-rich. Distribution of lncRNAs among the chromosomes was also observed; the results showed maximum presence in chromosome 2 and minimum in chromosome 8. Characteristic features of lncRNA were mentioned in Fig. 2 and supplementary files S1.

Identification of expressed lncRNAs in diseased vs. healthy plant tissues of cocoa

A total of 604 differentially expressed (DE) lncRNAs were identified among all samples (FC value >1 and < -1; *p-value* <0.005 and q-value <0.01). In plant tissue, the lncRNA expression differed from 11.28 to -15.16 fold. Among the 604 lncRNAs, 334 were upregulated, and 270 were downregulated in diseased tissues compared to healthy tissues. When we screened the differential expression based on FC values >2 and < -2 (*p-value* < 0.005 and q-value < 0.01), 223 differentially expressed among 91 were upregulated and 132 down-regulated in diseased tissues. The expression pattern between the tissues is revealed in Fig. 3 and Supplementary S1.

Variable expression observed for some of the lncRNAs in diseased and healthy plants were discussed here; for example, diseased plants showed almost three times higher expression of lncRNAs targeted mRNAs such as cytochrome b561 domain-containing protein (T.co-lnc_135),

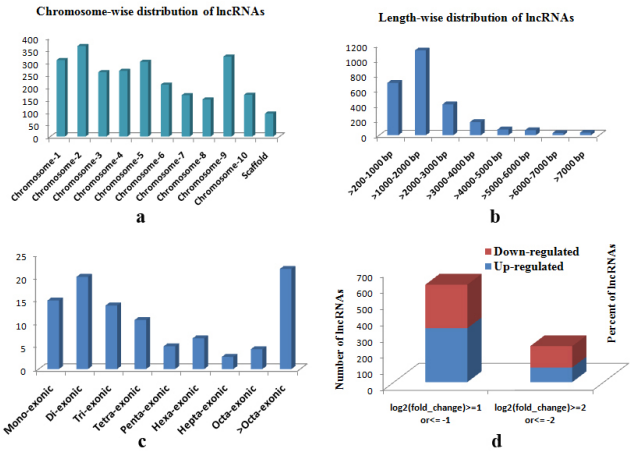

Fig. 2. Characteristics of *T. cacao* lncRNAs (a). Chromosome-wise distribution of lncRNAs; (b). Length-wise distribution of lncRNAs (c). Distribution of exon number of lncRNAs; (d). Differential expression of lncRNAs in healthy and diseased plants of cocoa under *Moniliophthora perniciosa* infection

and Cysteine-rich and transmembrane domain-containing protein-A (T.co-lnc_136), which are actively involved in detoxification of damaging reactive oxygen species (ROS) and neutralization of harmful compounds at the plasma membrane via chelation or redox-based mechanisms during stress conditions in plants (Venancio and Aravind 2010). Similarly, almost four times higher expression of lncRNA (T.co-lnc_160) with the target of HIPP-16 (heavy metal-associated isoprenylated plant proteins) was observed in diseased plants than in healthy ones. These HIPPs involved in heavy metal homeostasis and detoxification mechanisms, including abiotic and biotic resistance, were also noticed in plants. Zhang 1 et al. (2015) confirm the significant role of wheat heavy metal-associated isoprenylated plant protein (TaHIPP) in defense signaling pathways. On the contrary, another lncRNA (T.co-lnc_646) targeted HIPP-20 was found to have three times lower expression in *M. perniciosa* infected tissues than the healthy plants. CLAVATA3/ESR (CLE)-related protein 44 (CLE44) is an extracellular signal peptide that regulates the fate of cells by involving in functions like axillary shoot meristem initiation, maintenance of root meristem identity, phloem development, phloem, xylem, procambium histogenesis, regulation of cell differentiation. Enhanced levels of cocoa lncRNA T.co-lnc_176 with a target of CLE44 protein were observed in witches› broom diseased tissues over healthy ones. Overexpression, CLE44 resulted in unprecedented shrub-like dwarf growth in plants lacking epical dominance (Strabala et al. 2006). Similar symptoms were encountered in witches' broom-infected cocoa plants with overexpression CLE44. The HORMONOMETER profile findings of infected brooms showed altered auxin levels in cocoa (Teixeira et al. 2014). Under the study, enhanced levels of lncRNAs with auxin-responsive protein targets such as auxin-induced protein 22D, auxin-responsive protein

IAA20, auxin-responsive protein SAUR21, auxin-induced protein 15A, auxin-induced protein 15A-like, and auxin-responsive protein SAUR21 also noticed in diseased cocoa tissues. Increased expression levels of lncRNAs related to gibberellin-regulated-1 (T.co-lnc_1931) and 14 (T.co-lnc_970) and responsive proteins xyloglucan endotransglycosylase/hydrolase protein 8 (T.co-lnc_313) noticed in infected tissues of cocoa shows the involvement of hormone with the morphological alterations observed in infected cacao plants also proved by Teixeira et al. (2014). Phytosulfokine (PSK) plant peptide growth factors with multiple functions alter immune responses depending on the pathogen. It has been suggested that PSK integrates growth and defense signals to balance the competing metabolic costs of these plant responses. PSK integrates growth and defense signals to offset the competing metabolic costs of these plant responses (Sauter 2015). Altered expression levels of cocoa lncRNAs related to Phytosulfokine were noticed in healthy and diseased cocoas.T.co-lnc_1283 codes for putative phytosulfokine-6, involved in plant cell proliferation, differentiation, and organogenesis, was showed higher expression levels in diseased plants. At the same time, the other two lncRNAs, T.co-lnc_433 and T.co-lnc_81, codes for phytosulfokine and phytosulfokines3 were noticed with comparatively lower levels in diseased plants than in healthy ones. Diseased plants were also encountered with more menthol glutaredoxin-S6, an antioxidant compound involved in the cellular response to oxidative stress in bacterial and eukaryotic cells (Meyer et al. 2009).

Pathogenesis-related (PR) proteins are produced in the plant system as a mark of defense against pathogens. Increased expression levels of Pathogenesis-related proteins STH-2 (T.co-lnc_286), STH-21(T.co-lnc_1217), and PR-4 (T.co-lnc_1506) (as lncRNA targets) were seen in infected tissues of cocoa. Contrary to this, lower expression of other defense-related proteins, such as thaumatin-like protein (PR-5; T.co-lnc_753) and increased expression levels of another negative defense regulator DMR6-LIKE OXYGENASE 2 (T.co-lnc_439), which converts salicylic acid (SA) to 2, 3-dihydroxybenzoic acid was also noticed in the same tissue. The lncRNA targeted, NAC domain-containing protein 35, transcription factor (T.co-lnc_41), which acts as a floral repressor by negatively regulating CONSTANS (CO) expression in a GIGANTEA (GI)-independent manner, was found to show higher expression in diseased tissues (Yoo et al. 2007). Another T.co-lnc_922 with target NAC domain-containing protein 86, having a role in sieve element differentiation, enucleation, and cytosol degradation, showed more expression in infected tissues. On that, one more lncRNA targeted NAC domain-containing protein 104 (T.co-lnc_109) showed less expression in diseased tissues, which has a potential role in xylem development by negatively regulating secondary cell wall fiber synthesis and programmed cell death (Grant et al. 2010). The lncRNAs specific defense proteins such as, MLP-like protein 28 (T.co-lnc_652; T.co-lnc_655) and MLP-like protein 423 (T.co-lnc_477; T.co-lnc_625) exhibited the lower expression levels in diseased cocoa plants than the healthy ones. In Arabidopsis, cotton AtMLP-28 and GhMLP-28 were induced in response to *Plasmodiophora brassicae* and *Verticillium dahliae* infections (Yang et al. 2015).

Two lncRNAs, i.e., T.co-lnc_425 and T.co-lnc_1222, codes for allergen proteins peamaclein and major allergen Pru ar1 are seen in cocoa plants; the expression T.co-lnc_425 principles for of peamaclein is low in infected tissues as compared to healthy ones whereas the other lnc RNA, T.co-lnc_1222 responsible for major allergen Pru ar1 expression is more in diseased plants. The peptide snakin-2 (StSN2), isolated from potato tubers and active against fungal and bacterial plant pathogens (Berrocal-Lobo et al. 2002), is also noticed in cocoa plants. The lnc RNA, T.co-lnc_2025, responsible for snakin-2 peptide, showed 4 to five times higher expression levels in infected cocoa tissues over healthy plants.

Identification of lncRNAs, targeted mRNAs, including transcription factors

In cocoa, witches' broom diseased and healthy tissues, a total of 2616 lncRNA were identified, and they were found to target the 8827 cis mRNA target genes and 765 trans-target mRNAs. LncRNAs target a total of 9692 mRNAs. Among targeted mRNAs, 281 disease resistance-related transcripts and 211 TFs belonging to more than 50 TF families are found, which regulate the disease tolerance processes (Fig. 4). The identified transcription families were MYB, Ethylene responsive TFs, b-HLH, Nuclear TF-Y, TFIID, GATA, and WRKY, etc., which are reported to be actively involved in disease-related processes. TFs are crucial in responding to

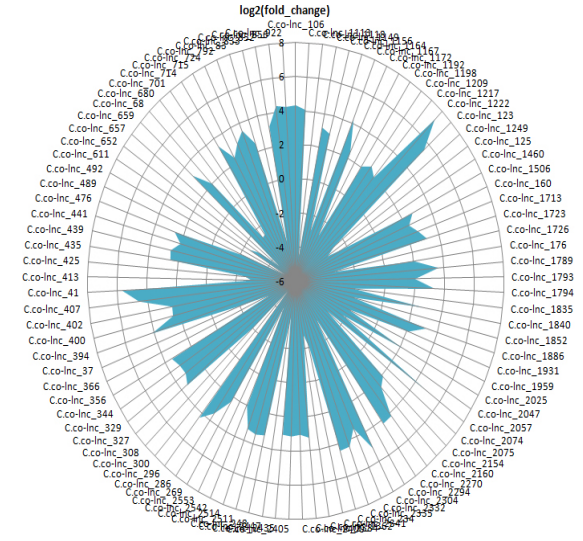

Fig. 3. Differentially expressed lncRNAs in diseased tissues compare to healthy tissues (Log2 fold=>2 or <=-2 with significant parameter 'yes'

biotic stresses. They act as positive and negative regulators in plants' complex defense response network processes. Transcriptomic studies conducted on WBD plants showed an up-regulation of WRKY and ERF transcription factors in infected tissues of cocoa (Teixeira et al. 2014). WRKY transcription factors (AtWRKY33) act as a positive regulator of resistance against *Alternaria brassicicola* and *Botrytis cinerea* in *Arabidopsis* (Zheng et al. 2006). AtWRKY53 and AtWRKY70 are involved in SAR's modulation (Wang et al. 2006). Likewise, MYB TF was found to have a role in the disease resistance process in sorghum against anthracnose and leaf blight (Baldoni et al. 2015; Ibraheem et al. 2015). Ethylene responsive factors are known to act either as repressors or activators of plant defense response against biotic stress. ERF proteins, such as ERF1 and ERF2, were known to be activators for the pathogen-induced plant defense process, while ERF3 and ERF4 are known to repress the gene expression and plant defense systems (Maruyama et al. 2013).

The lnc RNA targeted mRNAs were associated with different biological activities like replication of DNA (DNA ligase and DNA Helicase), chromatin alterations (topoisomerases), transcription (transcription initiation, elongation factor, etc.), cell division and expansion, vascular cells differentiation, protein trafficking, heat shock proteins, cellulose and callose synthase, etc. Also, several were associated with hormonal biosynthesis and signaling (inductive proteins of auxin, gibberellin-regulated proteins, auxin-responsive proteins -F- box proteins). The mRNAs, related protein kinases (serine/ threonine), B-box zinc fingers, F-box-LRR repeat proteins, and transporter proteins (sodium/potassium, sugar, and ABC transporters) were also targeted by lncRNAs. Studies on transcriptomic analysis of WBD-diseased tissue proved the enrichment of primary auxin-responsive genes. Likewise, genes related to biosynthesis (GA3ox), perception of gibberellic acid (GID1-like), and inactivation (GA2ox) were identified in infected tissue (Teixeira et al. 2014). The current evidence also supports the involvement/ role of lncRNA in hormonal synthesis and degradation in infected tissues. Around 281 disease-resistant proteins were identified as targeted by lncRNA in the current study. This included 84 genes with disease resistance RPP13-like protein orthologs, 45 disease resistance protein At4g27220 orthologs, 24, 10, and 13 genes with disease resistance protein RGA3, RGA2, and RGA1 orthologs, genes 20 and 14 each belong to At3g14460 and At1g61300 disease resistance protein and nine genes with disease resistance protein RPM1 orthologs, etc. (Fig. 4 and Supplementary Fig. 2).

GO analysis of lncRNAs

The gene ontology (GO) study (Supplementary Figs. 1 and 3) of predicted lncRNAs was performed. The enrichment study was categorized into three groups: GO annotation involved in biological, cellular, and cellular, molecular processes. Around 27.13 and 25.58% of lncRNAs were involved in biological, cellular, and metabolic processes. About 33% of lncRNAs were associated with biological regulation, response to stimulus, cellular component organization, and localization. Rest 14.21% of lncRNAs are involved in similar biological processes like development, signaling, and multicellular organismal function, positive and negative regulation of the biological process, reproductive process, reproduction, multi-organism process, growth, cell population proliferation, and immune system process. Under the molecular function category, many lncRNAs are involved in binding and catalytic activity, covering are involved in acute and catalytic activity, which cover 41.98 and 35.01%, respectively. Rest 23.01% of lncRNAs are involved in a similar type of activity like structural molecule activity, transporter activity, molecular function regulator, transcription regulator activity, translation regulator activity, nutrient reservoir activity, molecular transducer activity, antioxidant activity, molecular carrier activity, protein tag, and protein folding chaperone. In the cellular component category, 38.75% lncRNAs are under cell and cell parts. In membrane and membrane parts, 28.45% lncRNAs reside, and 21.64% lncRNAs reside in membrane parts. The rest of the lncRNAs have resided in protein-containing complex, membrane-enclosed lumen, extracellular region, cell junction, symplast, supramolecular complex, and nucleotide (Supplementary Fig. 3). involvement of lncRNAs in different pathways represented via Cytoscape (Fig. 5.) shows that many lncRNAs are involved in the same path, and some lncRNAs are involved in different pathways.

GO analysis of lncRNAs targeted mRNAs

In the GO study (Supplementary Figs 2 and 3), at the level of biological processes, it was noticed that 30.35 and 24.8% of total lncRNA targets correspondingly were engaged in the cellular and metabolic processes. It was observed that

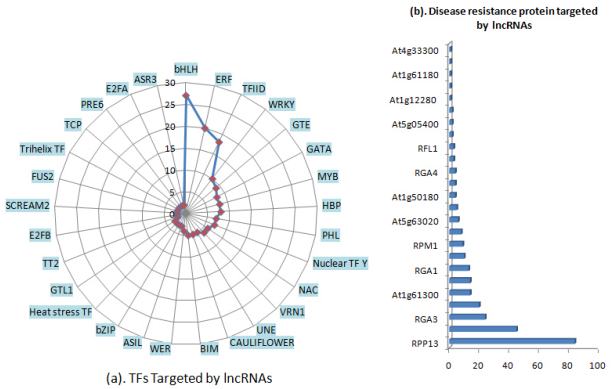

Fig. 4. (a). Statistics of lncRNAs targeted TFs families; (b). Disease resistance protein targeted by lncRNAs

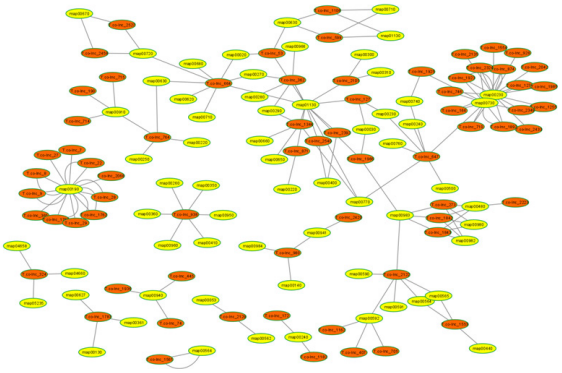

Fig. 5. lncRNAs with their interacting different pathways. Yellow colour represents the pathway and orangecolour represents the lncRNAs

31.56% of lncRNA targets were associated with regulating biological activities, response to stimuli, cellular localization, and compartmentalization. Rest 13.28% of lncRNA targets involve biological processes like developmental process, signaling, multicellular organismal process, positive and negative regulation of the biological processes, reproductive process, multi-organism process, growth, cell population proliferation, immune system process, detoxification, and rhythmic process. Under the molecular function of lncRNA, targeted mRNAs showed that most are involved in binding and catalytic activity, which contain 36.9 and 32%, respectively. Around 28.69% of lncRNA targets involve different activities such as transport, transcription, translation, and molecular functions. The remaining 2.5% of lncRNA targets are engaged in activities like molecular carriers, antioxidants, protein tag, molecular transducers, protein folding chaperones, and nutrient reservoir activities. The level of localization of lncRNA target mRNA is found that 38.75% reside in the cell and cell part. Around 22.66% of lncRNA target mRNA resides in the membrane and membrane parts. Around 20.56% of lncRNAs target mRNA reside in the organelle and organelle part. Rest 16.05% of the mRNAs targeted by lncRNAs reside in the protein complex, lumen, extracellular region, symplast, cell junction, nucleoid, and supramolecular complex (Supplementary Fig. 3).

Identification and characterization of miRNAs in T. cacao

Around 153 miRNAs were identified from *T. cocoa* transcriptome data obtained from healthy and diseased tissues based on the pipeline mentioned in Fig. 6. All the identified miRNAs represent 27 families; most miRNAs belong to the miR168 family (21) followed by miR166 (12), miR171 (12), miR319 (12), miR396 (11), miR167 (9), miR393 (9) and miR390 (8). The complete information about identified miRNAs is shown in Fig. 7.

Identification of mRNA targeted by miRNA

The psRNATarget software was used to predict mRNA targets

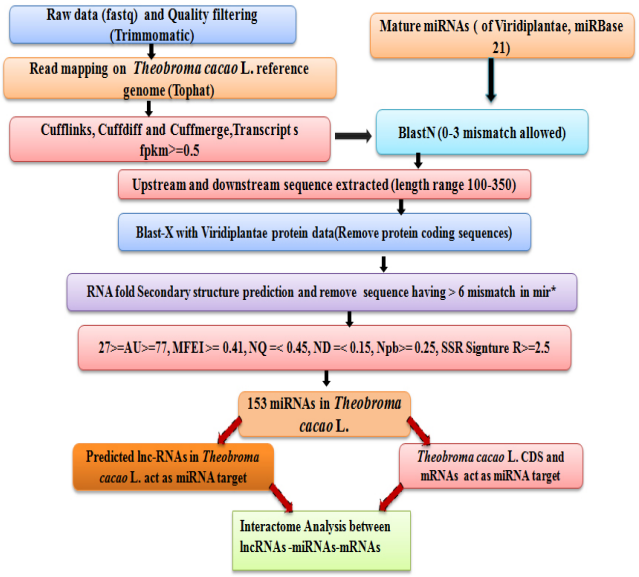

Fig. 6. Schematic representation of the methodology followed in the prediction of miRNAs, its targeted mRNAs and interaction between lncRNAs -miRNAs-mRNAs of T. cacao.

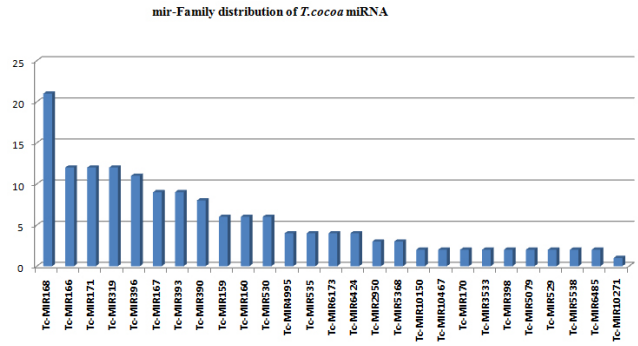

Fig. 7. miR Familywise distribution of predicted miRNAs

of identified 153 miRNAs. The mRNA sequences from the *T. cocoa* file were used as miRNA targets. The identified 153 miRNAs belonging to 27 miR families and 5337 mRNAs are targeted by the miRNAs, among them 114 codes for TFs and 170 codes for disease resistance protein. Ethylene-responsive factors, bHLH, WRKY, MYB, bZIP, GTE, GATA, and heat stress transcription factors are the dominant TFs, and they take a vital responsibility in disease tolerance processes (Figs 8 and Supplementary S5). Involvement of ERF TF in disease tolerance is evidenced by overexpression of ERF genes in JA, SA, or ET-mediated signal transduction pathways in plant species such as wheat (*TaPIE1*), rice (*OsEREBP1; OsERF83*), tomato (*Soly106),* and soybean (*GmERF113*) (Jisha et al. 2015; Huang et al. 2016; Zhao et al. 2017; Tezuka et al. 2019). Similarly, WRKYs were identified to be effective in microbe-associated molecular pattern-triggered immunity (PAMP-triggered immunity), effector-triggered immunity (ETI), or systemic acquired resistance (SAR). For example, the *CsWRKY50* gene in *Cucumis* is associated with the defense response against the *Pseudoperonospora cubensis* (Luan

et al. 2019). In grapes, the VvWRKY1 gene related to the JA pathway participates in biotic-stress tolerance against grape downy mildew fungi (Marchive et al. 2013). Another regulatory gene in the JA pathway, *CaWRKY27,* provides resistance against *Ralstonia solanacearum* in tobacco (Dang et al. 2014). *WsWRKY1* (Singh et al. 2017), *AcWRKYs* (Jing et al. 2018), *and GmWRKY31* (Dong et al. 2019) TFs imply versatile roles in response to pathogen stimuli in varied plant species. Another group of TFs, i.e., NAC, plays a dual role against different pathogen infections, such as hypersensitive mechanisms and ETI (Yuan et al. 2019). TF, *TaNAC8* in wheat positively protected plants from stripe rust infection (Xia et al. 2010). In maize, the tolerance against *Colletotrichum graminicola* infection

was induced by genes *ZmNAC41* and *ZmNAC100* from JA and SA pathways, respectively (Voitsik et al. 2013). *ONAC122* and *ONAC131* TFs in rice have essential roles in disease tolerance response via regulated expression of defense and signaling-related genes *OsLOX*, *OsPR1a*, *OsWRKY45*, and *OsNH1* (Sun et al. 2013). MYB TFs in disease resistance/tolerance against different pathogens was also studied in apples, pepper, grapevine, etc. (Zhang 2 et al. 2019). TFs of the bZIP family are also critical players in plant immunity due to their ability to control genes connected with PAMP-triggered immunity, ETI, and hormonal signaling networks (Norman et al. 2019).

The plant disease resistance (R) protein detects the existence of pathogens by identifying specific pathogen effector molecules produced during the infection (Martin et al. 2003). Most disease-resistance proteins are targeted by predicted miRNAs, which is a significant finding of our study. Most R proteins activated upon effector recognition fall into five different classes based on their combination of structural motifs. Though some R proteins may act as primary receptors of pathogen effector proteins, the majority of them play indirect roles in the process of pathogen resistance. The functions of diverse R proteins require protein degradation, phosphorylation, or specific localization within the host cell. Many R gene pathways share specific signaling components, whereas others seem pathway-specific (Martin et al. 2003).

The Cytoscape findings have revealed the relationship between miRNAs and their targets, and it was noticed that a single miRNA could interact with several mRNAs at a time. It was found that miR167b; miR167c interact with WRKY transcription factors and other mRNAs. miR168k, miR393b interacts with bHLH TFs. miR171a with MYB TF and miR396 interacts with Bzip, TCP TFs, and other mRNAs. Calmodulin-binding transcription activator and ethylene-responsive transcription factor RAP2-3 TF can be targeted by miR168c and miR168h, respectively.

GO annotation of miRNAs targeted mRNAs

Blast2Go analysis was performed to annotate the mRNAs targeted by miRNAs. The cellular and molecular components of the targets are represented in Supplementary Fig. 3. In the biological processes category, most marks, i.e., 56.81%, were associated with cellular cum metabolic processes. And the second majority 41.62% was implicated in regulating biological processes, stimulus-response, localization, signaling, biogenesis, negative and positive regulation of biological processes, multicellular organismal processes, the process of development, reproduction-related processes, as well as the organization of the different cellular components. While the rest, 1.56% are associated with growth, detoxification, rhythmic development, the proliferation of cells, and nitrogen consumption or utilization. In molecular category functions, catalytic and binding activities cover 45.48 and 43.20% of transcripts, respectively, and the remaining 11.30% are involved in different activities like transcription regulator, transporter, antioxidant, regulation of molecular function, molecular carrier, molecular transducer, nutrient reservoir, small molecule sensor, and structural molecule activity. The miRNA targets reside in various cellular components; most of them (proteins) localize in the cell, cell parts (36.97%), and the membrane and parts of the membrane (36.57%). Around 19.03% reside in the organelle and components of

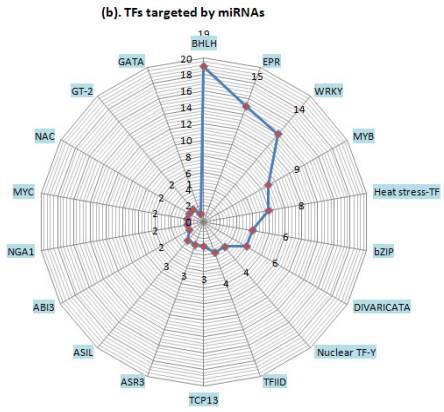

Fig. 8. Statistics of miRNAs targeted TFs families

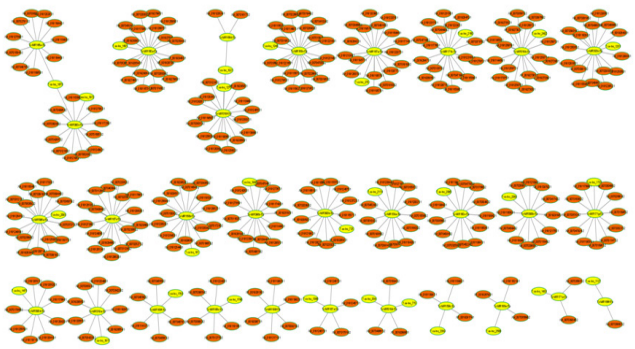

Fig. 9. eTMs (lncRNAs)-miRNAs-mRNAs Interaction, yellow color representing eTMs (lncRNAs) and miRNAs whereas orange color representing mRNAs

Table 1. eTMs with their target's mRNA expressed during witches' broom disease development and progression of *T. cacao*

e-TMs		e-TMs target	Function of e-TM target
lncRNAs	miRNA		
T.co-lnc_2174	Tc-MIR159a	XM_007019358.2	*T. cacao* glycolipid transfer protein 3 (LOC18592554), mRNA
T.co-lnc_2339	Tc-MIR159b	XM_018113992.1, XM_018113997.1	*T. cacao* CSC1-like protein At3g21620 (LOC18612316), transcript variants X1 & X2 mRNA
T.co-lnc_1981	Tc-MIR160a	XM_018129457.1 XM_018129456.1	*T. cacao* probable disease resistance protein At1g58602 (LOC18585959), transcript variants X1 & X2 mRNA
T.co-lnc_1601	Tc-MIR160a	XM_018128572.1	*T. cacao* disease resistance RPP8-like protein 3 (LOC18586554), mRNA
T.co-lnc_2085	Tc-MIR160c	XM_007040918.2	*T. cacao* galactan beta-1,4-galactosyltransferase GALS3 (LOC18606982), mRNA
T.co-lnc_1873	Tc-MIR166a	XM_018113903.1	*T. cacao* sugar transporter ERD6-like 6 (LOC18610920), mRNA
T.co-lnc_1786	Tc-MIR166c	XM_018122433.1, XM_007030803.2	PREDICTED: *T. cacao* amidase (LOC18600400), transcript variants X1 & X2 mRNA
T.co-lnc_1021	Tc-MIR166d	XM_007016177.2	*T. cacao* tubulin beta-8 chain (LOC18590570), mRNA
T.co-lnc_2508	Tc-MIR166e	XR_001926790.1, XM_018116219.1	*T. cacao* phosphatidylinositol 3-kinase, root isoform (LOC18609484)
T.co-lnc_1127	Tc-MIR166f	XM_007038601.2	*T. cacao* membrane steroid-binding protein 2 (LOC18605546), mRNA
T.co-lnc_1386	Tc-MIR167a	XM_007017619.2	*T. cacao* floral homeotic protein DEFICIENS (LOC18591481), mRNA
T.co-lnc_278	Tc-MIR167a	XM_018124571.1, XM_018124572.1, XM_018124573.1	*T. cacao* heterogeneous nuclear ribonucleoprotein 1 (LOC18593376), transcript variants X1, X2 & X3, mRNA
T.co-lnc_1229	Tc-MIR167b	XM_018116941.1	*T. cacao* probable WRKY transcription factor 30 (LOC18605716), mRNA
T.co-lnc_1313	Tc-MIR167c	XR_001929488.1, XR_001929489.1	*T. cacao* probable WRKY transcription factor 23 (LOC18613724)
T.co-lnc_161	Tc-MIR168b	XM_018125450.1, XM_018125449.1	*T. cacao* proline-rich receptor-like protein kinase PERK13 (LOC18592467), transcript variant X1 & X2, mRNA
T.co-lnc_1248	Tc-MIR168c	XM_018114807.1	*T. cacao* cyclin-J18 (LOC18609991), mRNA
T.co-lnc_1783	Tc-MIR168f	XM_018119823.1, XM_018119826.1	PREDICTED: *T. cacao* F-box protein CPR30 (LOC18601593), transcript variants X& X4, mRNA
T.co-lnc_1837	Tc-MIR168h	XM_018114217.1, XM_007048653.2	*T. cacao* sec1 family domain-containing protein MIP3 (LOC18612069), transcript variants X1 & X2, mRNA
T.co-lnc_61	Tc-MIR168i	XM_018128890.1	*T. cacao* putative disease resistance protein RGA3 (LOC18587224), mRNA
T.co-lnc_2	Tc-MIR168k	XM_007017551.2	*T. cacao* transcription factor bHLH110 (LOC18591436), mRNA
T.co-lnc_2483	Tc-MIR170b	XM_007050375.2	*T. cacao* histone acetyltransferase GCN5 (LOC18613245), mRNA
T.co-lnc_1433	Tc-MIR171a	XM_007035831.2	*T. cacao* transcription factor MYB108 (LOC18603713), transcript variant X2, mRNA
T.co-lnc_977	Tc-MIR171c	XM_007032400.2	*T. cacao* probable E3 ubiquitin ligase complex SCF subunit sconB (LOC18601458), mRNA
T.co-lnc_1943	Tc-MIR171d	XM_007039336.2, XM_007039337.2	*T. cacao* GTPase-activating protein gyp7 (LOC18605981), transcript variants X1&X2, mRNA
T.co-lnc_2081	Tc-MIR171e	XM_018126511.1	PREDICTED: *T. cacao* cytokinin riboside 5'-monophosphate phosphoribohydrolase LOG8 (LOC18588088), transcript variant X1, mRNA
T.co-lnc_1720	Tc-MIR171g	XM_007030666.2	*T. cacao* adenylyl-sulfate kinase 3 (LOC18600301), mRNA
T.co-lnc_2437	Tc-MIR171h	XM_007038529.2	*T. cacao* BAG family molecular chaperone regulator 1 (LOC18605499), mRNA
T.co-lnc_1229	Tc-MIR2950	XM_007028427.2, XM_018122182.1, XM_007028429.2, XM_018122183.1	*T. cacao* kinesin-like protein KIN-7O (LOC18598763), transcript variants X1,X2, X3 & X4 mRNA
T.co-lnc_1969	Tc-MIR319a	XM_018120555.1	*T. cacao* F-box protein At5g07610 (LOC18599756), mRNA
T.co-lnc_2267	Tc-MIR319b	XM_007016035.2	*T. cacao* phosphoinositide phospholipase C 4 (LOC18590491), mRNA
T.co-lnc_304	Tc-MIR319d	XM_007018941.2	*T. cacao* probable N-succinyl diaminopimelate amino transferase DapC (LOC18592299), mRNA

T.co-lnc_770	Tc-MIR319d	XM_007048661.2	PREDICTED: *T. cacao* serine/threonine-protein kinase CDL1 (LOC18612071), mRNA
T.co-lnc_1276	Tc-MIR319f	XM_018113988.1, XM_018113989.1	PREDICTED: *T. cacao* xyloglucan galactosyl transferase MUR3 (LOC18613057), transcript variant X1 & X2, mRNA
T.co-lnc_1021	Tc-MIR319f	XM_018113988.1, XM_018113989.1	PREDICTED: *T. cacao* xyloglucan galactosyltransferase MUR3 (LOC18613057), transcript variants X1& X2, mRNA
T.co-lnc_1969	Tc-MIR319f	XM_018122498.1, XM_018122499.1, XM_018122500.1	*T. cacao* pentatricopeptide repeat-containing protein At1g80550, mitochondrial (LOC18595371), transcript variants X1, X2 & X3, mRNA
T.co-lnc_35	Tc-MIR390a	XM_018123071.1, XM_018123079.1, XR_001928461.1	*T. cacao* GTP-binding protein At3g49725, chloroplastic (LOC18611342), transcript variant X1& X3 mRNA, X2(misc-RNA)
T.co-lnc_725	Tc-MIR390b	XM_018124675.1, XM_007020393.2	PREDICTED: *T. cacao* WEB family protein At5g55860 (LOC18593256), transcript variants X1, X2, mRNA
T.co-lnc_2289	Tc-MIR390b	XM_007023323.2, XM_018123723.1	PREDICTED: *T. cacao*nuclear pore complex protein NUP96 (LOC18595402), transcript variants X1, X2, mRNA
T.co-lnc_1873	Tc-MIR390c	XM_007052033.2	*T. cacao* probable tRNA N6-adenosine threonylcarbamoyltransferase, mitochondrial (LOC18614323), mRNA
T.co-lnc_1872	Tc-MIR390c	XM_018121953.1	*T. cacao* zinc finger protein VAR3, chloroplastic (LOC108662303), mRNA
T.co-lnc_1475	Tc-MIR390d	XM_018117884.1	PREDICTED: *T. cacao* uncharacterized LOC18612971 (LOC18612971), mRNA
T.co-lnc_1384	Tc-MIR393a	XM_007032605.2 XM_007032606.2	PREDICTED: *T. cacao* uncharacterized LOC18601605 (LOC18601605), transcript variants X1, X3, mRNA
T.co-lnc_64	Tc-MIR393a	XM_007044288.2	PREDICTED: *T. cacao* homocysteine S-methyltransferase 1 (LOC18609262), mRNA
T.co-lnc_1201	Tc-MIR393c	XM_018115596.1	PREDICTED: *T. cacao* UDP-N-acetylglucosamine diphosphorylase 2 (LOC18607026), mRNA
T.co-lnc_1204	Tc-MIR393f	XR_001927219.1, XM_018118111.1	PREDICTED: *T. cacao* probable myosin-binding protein 5 (LOC18507050), transcript variants X1 (misc-RNA), X2, mRNA
T.co-lnc_1985	Tc-MIR396b	XM_018124263.1	PREDICTED: *T. cacao* bZIP transcription factor 44 (LOC18594145), mRNA
T.co-lnc_85	Tc-MIR396c	XM_007012171.2, XR_001929545.1	PREDICTED: *T. cacao* myb-like protein D (LOC18588030), transcript variants X1, mRNA

the organelle. In comparison, the remaining (7.39%) localize in the extracellular regions, protein-containing complex, membrane-enclosed lumen, supramolecular complex, symplast, cell junction, and nucleoid (Supplementary Fig. 3).

Identification of lncRNAs as candidate endogenous target mimics

The psRNA Target analysis revealed that 55 differently expressed miRNA targets 955 mRNAs. A total of 55 lncRNAs-miRNAs interacting pair was identified were working on endogenous target mimics (e-TMs) mechanism and influencing the expression of 955 mRNAs. The results found in the current study support the chance that endogenous target mimics (eTMs) formation might occur between Tc-lncRNAs and miRNAs to regulate the expression of different genes during pathogen infection followed by the tolerance process within the plant. In Tomato, Jiang et al. (2019) identified the lnc RNAs containing eTMs for regulating miR482b responsible for decreasing disease resistance against *P. infestans*. Similarly, Gao et al. (2020) identified 13 lncRNAs as endogenous target mimics (eTMs) in melon. Das 1 et al. (2019, 2020) studied the role of e-TMs during flower, seed, and pod development in *Cajanus*. The part of each eTMs target mRNA's role is mentioned in Table 1. The lncRNAs act as eTMs, and their interaction results are shown in Fig. 9 and other information is given in Supplementary S6. Thus, it can be hypothesized that lncRNA controls the role of these mRNAs, including transcription factors, ultimately by the mechanism of endogenous target mimics (eTMs) via miRNAs.

Authors' contribution

Conceptualization of research (AV, AD); Designing of the experiments (AV, AD); Contribution of experimental materials (AD, AV, RTPP, AD); Execution of field/lab experiments and data collection (AD, AV, RTPP, AD); Analysis of data and interpretation (AD, AV); Preparation of the manuscript (AV, AD, RTPP, AD).

References

Argout X., Martin G., Droc G., Fouet O., Labadie K., Rivals E., Aury J.M. and Lanaud C. 2017. The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. BMC Genom., **18**(1): 730.

Baldoni E., Genga A. and Cominelli E. 2015. Plant MYB transcription factors: Their role in drought response mechanisms. Int. J. Mol. Sci., **16**: 15811-15851.

Berrocal-Lobo M., Segura A., Moreno M., López G., García-Olmedo F. and Molina A. 2002. Snakin-2, an antimicrobial peptide

from potato whose gene is locally induced by wounding and responds to pathogen infection. Plant Physiol.,**128**(3): 951-61.

Bolger A.M., Lohse M. and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform., **30**(15): 2114-20.

Borah P., Das A., Matthew J. M., Ali A., Alison R., Bentley and Pandey R. (2018). Long non-coding RNAs as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants. Genes, **9**(9): 459.

Chaves F.C. and Gianfagna T. J. 2007. Cacao leaf procyanidins increase locally and systemically in response to infection by *Moniliophthora perniciosa* basidiospores. Physiol. Mol. Plant Pathol., **70**(4-6): 174-179.

Conesa A. and Götz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics., 619832.

da Hora Junior B.T., de Faria Poloni J., Lopes M.A., Dias C.V., Gramacho K.P., Schuster I., Sabau X., Cascardo J. C., Di Mauro S. M., da Silva Gesteira A. and Bonatto D 2012. Transcriptomics and systems biology analysis in identification of specific pathways involved in cacao resistance and susceptibility to witches' broom disease. Mol. Biosyst., **8**: 1507-1519.

Dang F., Wang Y., She J., Lei Y., Liu Z., Eulgem T., Lai Y., Lin J., Yu L. and Lei D. 2014. Over expression of CaWRKY27, a subgroup IIe WRKY transcription factor of *Capsicum annuum*, positively regulates tobacco resistance to *Ralstonia solanacearum* infection. Physiologia Plant., **150**: 397-411.

Das A., Deepti N., Alim J.,Tribhuvan K.U.,Kumar K., Kumar D., Singh N. K. and Gaikwad K. 2019. Expressivity of the key genes associated with seed and pod development is highly regulated via lncRNAs and miRNAs in pigeonpea. Sci Rep., **9**(1): 1-14.

Das A., Saxena S., Kumar K.,Tribhuvan K. U., Singh N.K. and Gaikwad K. 2020. Non-coding RNAs having a strong positive interaction with mRNAs reveal their regulatory nature during flowering in a wild relative of pigeonpea (*Cajanus scarabaeoides*). Mol Biol. Rep., **47**(5): 3305-3317.

Dai X. and Zhao P. X. 2011. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res., **39**: 155-159.

Dong H., Tan J., Li M., Yu Y., Jia S., Zhang C., Wu Y. and Liu Y. 2019. Transcriptome analysis of soybean WRKY TFs in response to *Peronospora manshurica* infection. Genomics, **111**: 1412-1422.

End M. J., Daymond A. J. and Hadley P. 2014. Technical Guidelines for the Safe Movement of Cacao Germplasm. Revised from the FAO/ IPGRI Technical Guidelines No. 20 Montpellier: Global Cacao Genetic Resources Network (CacaoNet), Bioversity International.

Gao C., Sun J., Dong Y., Wang C., Xiao S., Mo L. and Jiao Z. 2020. Comparative transcriptome analysis uncovers regulatory roles of long non-coding RNAs involved in resistance to powdery mildew in melon. BMC Genom., **21**: 125.

Gesteira A. S., Micheli, F., Carels N., Da Silva A. C., Gramacho K. P., Schuster I., Macêdo J. N., Pereira G. A. and Cascardo J. C. 2007. Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora perniciosa*. Ann. Bot., **100**(1): 129-140.

Gramacho K. P., Luz E. D., Silva F. S., Lopes U. V., Pires J. L. and Pereira L. 2016. Pathogenic variability of *Moniliophthora perniciosa* in three agroecological zones of the cacao region of Bahia, Brazil. Crop. Breed. Appl. Biotechnol., **16**: 7-13.

Grant E. H., Fujino T., Beers E. P. and Brunner A. M. 2010. Characterization of NAC domain transcription factors implicated in the control of vascular cell differentiation in *Arabidopsis* and *Populus*. Planta, **232**(2): 337-52.

Hou X., Cui J., Liu W., Jiang N., Zhou X., Qi H., Meng J. and Luan Y. 2020. LncRNA39026 Enhances Tomato Resistance to *Phytophthora infestans* by Decoying miR168a and Inducing PR Gene Expression. Phytopathol., **110**(4): 873-880.

Huang Y., Zhang B. L., Sun S., Xing G. M., Wang F., Li M. Y., Tian Y. S. and Xiong A.S. 2016. AP2/ERF transcription factors involved in response to tomato yellow leaf curly virus in tomato. Plant Genom., **9**: 1-15.

Ibraheem F., Gaffoor I., Tan Q., Shyu C. R. and Chopra S. A. 2015. Sorghum MYB transcription factor induces 3-deoxyanthocyanidins and enhances resistance against leaf blights in maize. Molecules, **20**: 2388-2404.

Jiang N., Cui J., Shi Y., Yang G., Zhou X., Hou X., Meng J. and Luan Y. 2019. Tomato lncRNA23468 functions as a competing endogenous RNA to modulate NBS-LRR genes by decoying miR482b in the tomato-*Phytophthorainfestans* interaction. Hortic Res., **6**: 28.

Jing Z. and Liu Z. 2018. Genome-wide identification of WRKY transcription factors in kiwifruit (*Actinidia* spp.) and analysis of WRKY expression in responses to biotic and abiotic stresses. Genes Genom., **40**: 429-446.

Jisha V., Dampanaboina L., Vadassery J., Mithöfer A., Kappara S. and Ramanan R. 2015. Over expression of an AP2/ERF type transcription factor OsEREBP1 confers biotic and abiotic stress tolerance in rice. PLoS One, **10**: 1-24.

Joshi R. K., Megha S., Basu U., Rahman M. H. and Kav N. N. 2016. Genome Wide Identification and Functional Prediction of Long Non-Coding RNAs Responsive to *Sclerotinia sclerotiorum* Infection in *Brassica napus*. PLoS One, **11**(7): e0158784.

Kanehisa M.,Goto S., Kawashima S. and Nakaya A. 2002. The KEGG databases at GenomeNet. Nucleic Acids Res., **30**(1): 42-46.

Kang Y.J., Yang D. C., Kong L., Hou M., Meng Y. Q., Wei L. and Gao G. 2017. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res., **45**: 12-16.

Kilaru A., Bailey B. A. and Hasenstein K. H. 2007. *Moniliophthora perniciosa* produces hormones and alters endogenous auxin and salicylic acid in infected cocoa leaves. FEMS Microbiol. Lett., **274**: 238-244.

Luan Q., Chen C., Liu M., Li Q., Wang L. and Ren Z. 2019. CsWRKY50 mediates defence responses to *Pseudoperonospora cubensis* infection in *Cucumis sativus*. Plant Sci., **279**: 59-69.

Marchive C., Léon C., Kappel C., Coutos-Thévenot P., Corio-Costet M. F., Delrot S. and Lauvergeat V. 2013. Over-expression of VvWRKY1 in grapevines induces expression of jasmonic acid pathway-related genes and confers higher tolerance to the downy mildew. PLoS One, **8**: 1-8.

Martin G. B., Bogdanove A. J. and Sessa G. 2003. Understanding the functions of plant disease resistance proteins. Annu. Rev. Plant. Biol., **54**: 23-61.

Maruyama Y., Yamoto N., Suzuki Y., Chiba Y., Yamazaki K., Sato T. and Yamaguchi J. 2013. The Arabidopsis transcriptional repressor ERF9 participates in resistance against necrotrophic fungi. Plant Sci., **213**: 79-87.

Meinhardt L.W., Costa G. G., Thomazella D., Teixeira P. J., Carazzolle M. F., Schuster S. C., Carlson J. E., Guiltinan M. J., Mieczkowski P., Farmer A. and Ramaraj T. 2014. Genome and secretome analysis of the hemibiotrophic fungal pathogen, *Moniliophthora roreri*, which causes frosty pod rot disease of cacao, mechanisms of the biotrophic and necrotrophic phases. BMC Genom., **15**(1): 1-25.

Meinhardt L.W., Rincones J., Bailey B. A., Aime M. C., Griffith G.W., Zhang D. and Pereira G. A. 2008. *Moniliophthora perniciosa*, the causal

agent of witches' broom disease of cacao: what's new from this old foe? Mol. Plant Pathol., **9**(5): 577-588.

Meyer Y., Buchanan B. B., Vignols F. and Reichheld J. P. 2009. Thioredoxins and glutaredoxins: unifying elements in redox biology. Annu. Rev. Genet., **43**: 335-367.

Mondego J. M., Carazzolle M. F., Costa G. G., Formighieri E. F., Parizzi L.P., Rincones J., Cotomacci C., Carraro D. M., Cunha A. F., Carrer H. and Vidal R. O. 2008. A genome survey of *Moniliophthora perniciosa* gives new insights into witches' broom disease of cacao. BMC Genom., **9**: 1-25.

Navarro L., Dunoyer P., Jay F., Arnold B., Dharmasiri N., Estelle M., Voinnet O. and Jones J. D. G. 2006. A plant miRNA contributes to antibacterial resistance by repressing auxin signalling. Science, **312**: 436-439.

Noman A., Hussain A., Adnan M., Khan M. I., Ashraf M. F., Zainab M., Khan K. A., Ghramh H. A. and He S. 2019. A novel MYB transcription factor CaPHL8 provide clues about the evolution of pepper immunity against soil borne pathogen. Microb. Pathogen, **137**: 103758.

Pereira J.F., Araújo E.F., Brommonschenkel S. H., Queiroz C. B., Costa G. G., Carazzolle M. F., Pereira G. A. and Queiroz M.V. 2015. MpSaci is a widespread gypsy-Ty3 retrotransposon highly represented by non-autonomous copies in the *Moniliophthora perniciosa* genome. Curr. Genet., **61**(2): 185-202.

Rincones J., Scarpari L. M., Carazzolle M. F., Mondego J. M., Formighieri E. F., Barau J. G., Costa G. G., Carraro D. M., Brentani H. P., Vilas-Boas L. A., de Oliveira B. V., Sabha M., Dias R., Cascardo J. M., Azevedo R. A., Meinhardt L.W. and Pereira G. A. 2008. Differential gene expression between the biotrophic-like and saprotrophic mycelia of the witches' broom pathogen *Moniliophthora perniciosa*. Mol. Plant Microbe Interact., **21**(7): 891-908.

Sauter M. 2015. Phytosulfokine peptide signalling. J. Exp. Bot., **66**(17): 5161-5169.

Scarpari L.M., Meinhardt L.W., Mazzafera P., Pomella A.W., Schiavinato M.A., Cascardo J. C. and Pereira G. A. 2005. Biochemical changes during the development of witches' broom: the most important disease of cocoa in Brazil caused by *Crinipellis perniciosa*. J. Exp. Bot., **56**: 865-877.

Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J.T., Ramage D., Amin N., Schwikowski B. and Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res., **13**(11): 2498-504.

Shiv A., Krishna H., Sinha N., Priyadarshini P., Sahu S., Jain N., Singh P. K. and Prabhu K.V. 2023. Leaf rust responsive miRNA mediated regulation of *Puccinia triticina* genes during host pathogen interaction in wheat. Indian J. Genet. Plant Breed., **83**(2): 185-194.

Singh A. K., Kumar S.R., Dwivedi V., Rai A., Pal S., Shasany A.K. and Nagegowda D.A. 2017. A WRKY transcription factor from *Withaniasomnifera* regulates triterpenoid with anolide accumulation and biotic stress tolerance through modulation of phytosterol and defence pathways. New Phytol., **215**: 1115-1131.

Strabala T.J., O'donnell P. J., Smit A.M., Ampomah-Dwamena C., Martin E. J., Netzler N., Nieuwenhuizen N. J., Quinn B. D., Foote H.C. and Hudson K.R. 2006. Gain-of-Function Phenotypes of Many CLAVATA3/ESR Genes, Including Four New Family Members, Correlate with Tandem Variations in the Conserved CLAVATA3/ESR Domain. Plant Physiol., **140**(4): 1331-1344.

Su Y., Zhang Y., Huang N., Liu F., Su W., Xu L., Ahmad W., Wu Q., Guo J. and Que Y. 2017. Small RNA sequencing reveals a role for sugarcane miRNAs and their targets in response to *Sporisorium scitamineum* infection. BMC Genom., **18**(1): 325.

Sun L., Zhang H., Li D., Huang L., Hong Y., Ding X. S., Nelson R. S., Zhou X. and Song F. 2013. Functions of rice NAC transcriptional factors, ONAC122 and ONAC131, in defence responses against *Magnaporthe grisea*. Plant Mol. Biol., **81**: 41-56.

Sun L., Luo H., Bu D., Zhao G., Yu K., Zhang C., Liu Y., Chen R. and Zhao Y.2013.Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res.,**41**(17): e166-e166.

Supriya P., Kumar A., Archak S. and Bhat K. 2022. Computational identification of microRNAs and their target genes in sesame (*Sesamum indicum* L.). Indian J.Genet. Plant Breed., **82**(04): 469-473.

Teixeira P. J., Thomazella D.P., Reis O., do Prado P. F., do Rio M. C., Fiorin G. L., José J., Costa G.G., Negri V.A., Mondego J.M., Mieczkowski P. and Pereira G.A. 2014. High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. Plant Cell, **26**(11): 4245-69.

Tezuka D., Kawamata A., Kato H., Saburi W., Mori H., Imai R. 2019. The rice ethylene response factor OsERF83 positively regulates disease resistance to *Magnaporthe oryzae*. Plant Physiol. Biochem., **135**: 263-271.

Trapnell C., Hendrickson D.G., Sauvageau M., Goff L., Rinn J. L. and Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol., **31**: 46-53.

Venancio T.M. and Aravind L. 2010. CYSTM, a novel cysteine-rich transmembrane module with a role in stress tolerance across eukaryotes. Bioinfom., **26**(2): 149-152.

Voitsik A.M., Muench S., Deising H.B. and Voll L.M. 2013. Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection. BMC Plant Biol., **13**: 85.

Wang D., Amornsiripanitch N. and Dong X. 2006. A genomic approach to identify regulatory nodes in the transcriptional network of systemic acquired resistance in plants. PLoSPathog., **2**: e12.

Wang J., Yu W., Yang Y., Li X., Chen T., Liu T., Ma N., Yang X., Liu R. and Zhang B. 2015. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. Sci Rep., **18**5(1): 1-6.

Wu H.J., Wang Z.M., Wang, M. and Wang X. J. 2013. widespread long non-coding RNAs as endogenous target mimics for micro RNAs in plants. Plant Physiol., **161**(4): 1875-84.

Xia N., Zhang G., Sun Y.F., Zhu L., Xu L. S., Chen X. M., Liu B., Yu Y. T., Wang X. J. and Huang L.L. 2010. TaNAC8, a novel NAC transcription factor gene in wheat, responds to stripe rust pathogen infection and abiotic stresses. Physiol. Mol. Plant Pathol., **74**: 394-402.

Xin M., Wang Y., Yao Y., Song N., Hu Z., Qin D., Xie C., Peng H., Ni Z. and Sun Q. 2011. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. BMC Plant Biol., **11**: 1-3.

Yang C.L., Liang S., Wang H.Y., Han L. B., Wang F. X., Cheng H. Q., Wu X. M., Qu Z. L., Wu J. H. and Xia G. X. 2015. Cotton major latex protein 28 functions as a positive regulator of the ethylene responsive factor 6 in defense against *Verticillium dahliae*. Mol. Plant., **8**(3): 399-411.

Yoo S.Y., Kim Y., Kim S.Y., Lee J. S. and Ahn J. H. 2007. Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis*. PLoS One, **2**: E642-E642.

Yu Y., Zhou Y.F., Feng Y.Z., He H., Lian J.P., Yang Y.W., Lei M.Q., Zhang Y.C. and Chen Y.Q. 2020. Transcriptional landscape of pathogen-responsive lncRNAs in rice unveils the role of ALEX1 in jasmonate pathway and disease resistance. Plant Biotechnol J., **18**(3): 679-690.

Yuan X., Wang H., Cai J., Li D. and Song F. 2019. NAC transcription factors in plant immunity. Phytopath Res., **1**: 3.

Zhang Q., Li Y., Zhang Y., Wu C., Wang S., Hao L., Wang S. and Li T. 2017. Md-MiR156ab and Md-mir395 target WRKY transcription factors to influence apple resistance to leaf spot disease. Front Plant Sci., **8**: 1-14.

Zhang X., Feng H., Feng C., Xu H., Huang X., Wang Q., Duan X., Wang X., Wei G., Huang L. and Kang Z. 2015. Isolation and characterisation of cDNA encoding a wheat heavy metal-associated isoprenylated protein involved in stress responses. Plant Biol., **17**(6): 1176-86.

Zhang Y.L., Zhang C.L., Wang G.L., Wang Y.X., Qi C.H., Zhao Q., You C.X., Li Y.Y. and Hao Y.J. 2019. The R2R3 MYB transcription factor MdMYB30 modulates plant resistance against pathogens by regulating cuticular wax biosynthesis. BMC Plant Biol., **19**: 362.

Zhang, Han Z., Guo Q., Liu Y., Zheng Y., Wu F. and Jin W. 2014. Identification of Maize Long Non-Coding RNAs Responsive to Drought Stress. PLoS One, **9**(6): e98958.

Zhao Y., Chang X., Qi D., Dong L., Wang G., Fan S., Jiang L., Cheng Q., Chen X. and Han D.A. 2017. Novel soybean ERF transcription factor, GmERF113, increases resistance to *Phytophthora sojae* infection in soybean. Front. Plant Sci., **8**: 299.

Zheng Z., Qamar S.A., Chen Z. and Mengiste T. 2006. Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. Plant J., **48**(4): 592-605.

Zhu Q.H., Stephen S., Taylor J., Helliwell C.A. and Wang M. B. 2014. Long non-coding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. New Phytol., **201**(2): 574-584.