Genomics 112 (2020) 2041–2051

Contents lists available at ScienceDirect

Genomics

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

Original Article

Plant virus interaction mechanism and associated pathways in mosaic disease of small cardamom (*Elettaria cardamomum* Maton) by RNA-Seq approach

Aamir Khan[a,1], K. Johnson George[b,1], Rahul Singh Jasrotia[a], Sharon Aravind[b], U.B. Angadi[a], Mir Asif Iquebal[a], K.P. Manju[b], Sarika Jaiswal[a], P. Umadevi[b], Anil Rai[a], Dinesh Kumar[a,*]

[a] ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India
[b] ICAR- Indian Institute of Spices Research, Kozhikode, Kerala 673012, India

ARTICLE INFO

Keywords:
Small cardamom
Transcriptome
Molecular markers
Genomic resources
GRN

ABSTRACT

Small cardamom (*Elettaria cardamomum*), grown in limited coastal tropical countries is one of the costliest and widely exported agri-produce having global turnover of > 10 billion USD. Mosaic/marble disease is one of the major impediments that requires understanding of disease at molecular level. Neither whole genome sequence nor any genomic resources are available, thus RNA seq approach can be a rapid and economical alternative. *De novo* transcriptome assembly was done with Illumina Hiseq data. A total of 5317 DEGs, 2267 TFs, 114 pathways and 175,952 genic region putative markers were obtained. Gene regulatory network analysis deciphered molecular events involved in marble disease. This is the first transcriptomic report revealing disease mechanism mediated by perturbation in auxin homeostasis and ethylene signalling leading to senescence. The web-genomic resource (SCMVTDb) catalogues putative molecular markers, candidate genes and transcript information. SCMVTDb can be used in germplasm improvement against mosaic disease in endeavour of small cardamom productivity.

Availability of genomic resource, *SCMVTDb*: http://webtom.cabgrid.res.in/scmvtdb/

1. Introduction

Right from 14th century, spices are well known for its role in global business having socio-political-economic dimensions [1]. Even today, spices are one of the costliest (per unit weight) agri-germplasm having global market value > 30 times than that of rice and wheat. Among the two cardamoms, namely, small and large, former is much widely used for culinary and medicinal purpose [2–7]. Small or green cardamom (*Elettaria cardamomum*) is a perennial, herbaceous, rhizomatous, monocot plant belonging to *Zingiberaceae* family which is mostly cultivated in the tropical regions of the world. It is also known as "*queen of spices*", which is the world's third most expensive spice after vanilla and saffron. Though this spice is grown in limited coastal tropical countries like Guatemala, India, Sri Lanka, Tanzania, El Salvador, Vietnam, Laos, Cambodia and Papua New Guinea but it is one of the widely exported commodity and its production and price affects most part of the globe [8,9]. It has been projected that current global turnover (10 billion USD) is going to rise further in tune of 30 and 40 billion USD by 2023 and 2040, respectively. If productivity of this valuable crop can be enhanced by supplementation of genomics approach, it will have better price benefit for global consumers. The dynamics of productivity difference (2.5 folds) between top two producers namely, Guatemala and India reflects there is enough scope of genetic improvement for productivity enhancement in many other countries also [8,10].

Various abiotic and biotic stresses are impediment in productivity. Among the abiotic stresses, rain and soil fertility are relevant [11]. Among biotic stresses, viruses, bacteria and fungi are common [12]. Mosaic or marble disease is one of the major biotic stresses adversely affecting the productivity [13]. It is caused by *Cardamom mosaic virus (CdMV)* which is the species under the genus *Macluravirus* of the family *Potyviridae*. It is prevalent in India, Sri Lanka, Guatemala and many other countries [9,14]. Pale green stripes of leaf and its margin gives a distinct mosaic pattern and often disease is caused by group of viruses which is spread by its vector among which banana aphid (*Pentalonia nigronervosa*) is most common [15]. There is a gradual loss in productivity (70%) and in three years, it becomes totally unproductive

* Corresponding author at: ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, PUSA, New Delhi 110012, India.
E-mail address: Dinesh.Kumar@icar.gov.in (D. Kumar).
[1] Authors contributed equally.

https://doi.org/10.1016/j.ygeno.2019.11.017
Received 11 February 2019; Received in revised form 29 May 2019; Accepted 22 November 2019
Available online 23 November 2019
0888-7543/

[16]. Disease management of *CdMV* is a major challenge due to limited success in disease resistant variety development [13], thus alternative genomic approach is needed to supplement such efforts. However, neither whole genome sequence of this crop is available nor any web-genomic resource for initiation of genetic application of improvement. Though limited tissue specific transcriptomic studies [17] have been carried out, there is only few report on abiotic stress (drought) response [18,19] but there is no report of any biotic stress response in this crop. Since small cardamom having bigger genome (2.8 Gbp) [20], is yet to be sequenced, thus transcriptome based approach can be a cost effective and immediate alternate to cater the need of molecular markers for association studies required for varietal improvement [21].

Present work aims to investigate the molecular mechanism of plant virus interaction to elucidate the infection process along with deeper understanding of plant immune response mechanism mediating pathogenesis. It further aims at identification of key candidate genes by construction of gene regulatory network involved in control of associated pathways in mosaic disease of small cardamom. It also aims at mining of putative molecular markers (SSR, SNP and InDels) from key candidate genes to develop web genomic resource for future association studies.

2. Materials and methods

2.1. Collection, maintenance of aphid colony and inoculation

Initially apterous form of the aphids (*Pentalonia caladii*) colonizing cardamom plants under field conditions were collected. In order to rule out the viruliferous nature and also to obtain a pure colony, individual females were separately transferred, on to young healthy leaves (4–5) of cardamom (variety: Appangala 1) made in the form of funnels. Subsequently, the funnels were maintained under dark with the bottom end immersed in water to prevent desiccation. The female aphids parthenogenitically produced younger ones. The colonies were periodically transferred to fresh cardamom leaf funnels until build-up of sufficient population for transmission studies. The Appangala strain of *CdMV* was used which is maintained on Appangala 1 variety of small cardamom at ICAR-IISR, Calicut, Kerala, India (11.2588° N, 75.7804° E). Viral transmission was done by fourth or fifth instar apterate aphids. Since the virus is reported to be stylet-borne, thus non-persistent mode of transmission was adopted. Prior to transmission, the aphids were subjected to a fasting period for 1 h and later allowed to feed (acquisition access period) on detached young symptomatic leaf exhibiting characteristic katte symptoms in a test tube for 30 min for acquiring the virus. After acquisition period, the aphids were used for inoculating the test plants. The actively growing leaves including the newly emerged leaf and surrounding 2–3 leaves were rolled to form a funnel and 10 viruliferous aphids were transferred to each funnel using camel's hair brush taking care that the stylet is intact and unbroken. Further, the aphids were allowed to feed on the test plants overnight (14–18 h) after which the plants were sprayed with an insecticide. The inoculated plants were observed for development of symptoms periodically (Fig. 1) [22].

2.2. RNA isolation, cDNA synthesis, library construction and sequencing

Leaf tissue samples of both control and virus infected plants were collected. Ten biological replicates samples were pooled to minimize variability across sample for RNA isolation [23]. Due to polyphenol and polysaccharide contents in cardamom tissues which seriously affect the extraction of RNA, thus we have used published small cardamom specific RNA extraction protocol having combination of RNeasy Plant Mini Kit (Qiagen) and CTAB (cetyl trimethylammonium bromide) along with quality and quantity check as described by Nadiya *et al* in 2015 [24]. Then gel electrophoresis was used to check the integrity of total RNA and concentration of RNA was measured by using Nanodrop-2000 spectrophotometer (Thermo, Inc.). All RNA samples were evaluated for integrity on Bioanalyzer 2100 (Agilent) and the RIN (RNA Integrity Number) values were 6.5 and 6.8 for the control and treated samples, respectively. Synthesis of cDNA was done by kit (Roche, Inc.). RNA-Seq libraries were prepared and sequencing was done by using Illumina HiSeq™ 2000 platform with 101 nucleotide paired end reads using manufacturer's protocol. The libraries were labelled as control (normal) and infected (Cardamom mosaic virus infected) and submitted to the SRA of NCBI having BioProject: PRJNA474822 (BioSamples: SAMN09374398, SAMN09374399).

2.3. Pre-processing and de novo assembly

Paired-end sequencing data of control and virus infected plants of *E. cardamomum M.* was generated by using Illumina Hiseq technology with 2*101 bp. Assessment and visualization of raw reads was performed using FASTQC tool [25]. Removal of adaptors sequences, overrepresented sequences, low quality reads, trimming of bases from 5′ and 3′ and phred-score ≤ 20 were done using Trimmomatic tool [26]. Further, high quality reads of both samples were used for *de novo* transcriptome assembly using Trinity assembler v2.0.6 [27].

2.4. Identification of differential expressed genes

Bowtie tool was used for mapping of control and virus infected reads onto *de novo* assembly of small cardamom [28]. RNA-Seq by Expectation-Maximization (RSEM) tool was used for calculation of expression [29]. EdgeR (Empirical analysis of Digital Gene Expression in R) tool of R bioconductor was used for identification of differential expressed genes (DEGs) in virus infected samples in comparison to control samples with stringent parameters such as FDR 0.05 and $\log_2$fold change value as 2 [30].

In order to reduce noise without compromising computational accuracy, two different tools, NOISeq and edgeR were used. EdgeR has advantage of TMM normalization to account for different sequencing depths between the samples and it uses Benjamini–Hochberg procedure to control the FDR. NOISeq has been found better due to its attribute of non-parametric as well as data adaptive approach based on count matrix [31]. It empirically models the noise distribution from the actual data by contrasting fold-change differences and absolute expression differences among samples within the same condition. In order to model the noise distribution it computes differences of both fold-change and absolute expression. It adapts to dataset size and efficiently controls the FDR. It has default normalization using RPKM. It has advantage of taking care of small numbers of replicates as well as genes having lower expression [32]. A *P*-value and threshold variance were set to 0.05 for identification of significant genes [33,34]. For comparative analysis, q value = 0.99 was used for identification of differential expressed genes by both tools. The values obtained by both the tools were compared to establish reliability of results obtained by edgeR.

2.5. Homology search, annotation and functional characterization

Homology search of differential expressed genes as well as *de novo* transcriptome assembly was performed against NCBI non-redundant database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/) using Blastx algorithm of standalone local ncbi-blast-2.2.31+ with threshold *E*-value 1e-3 [35]. This was followed by functional categorization, gene ontology, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways analysis, enzyme classes, domains and families search using Blast2GO Pro version 3.1 software [36]. PlantTFDB 4.0 was used for identification of transcriptional factors [37].

2.6. Mining of genic region putative molecular markers

Coding region putative DNA markers, namely, simple sequence

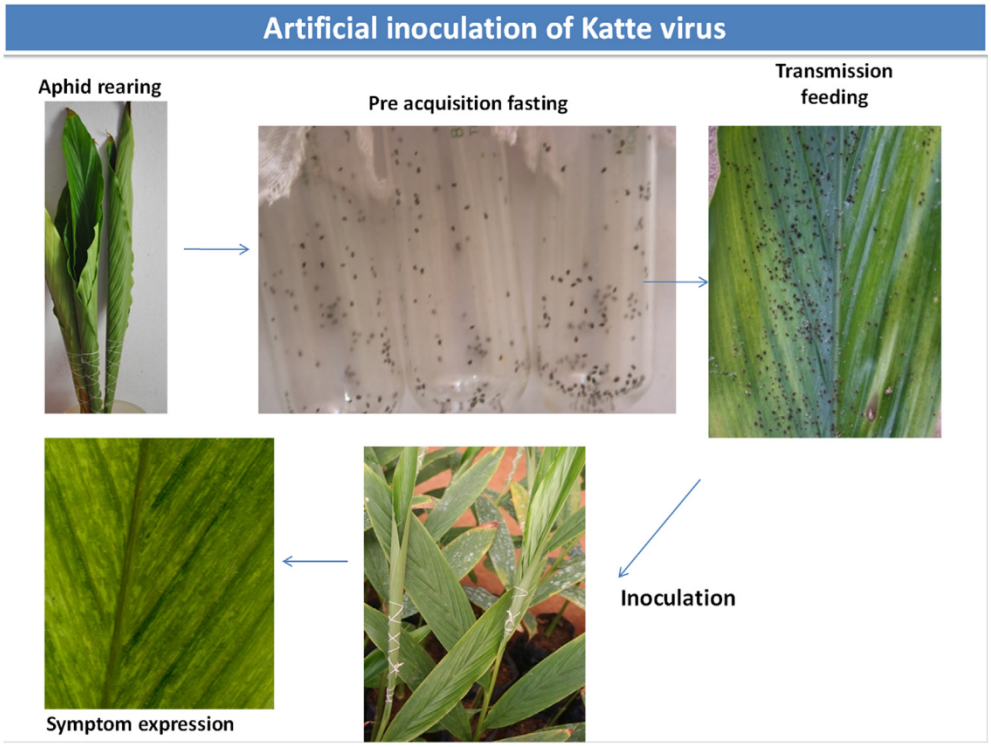

Fig. 1. Artificial inoculation of marble or katte virus.

repeats (SSR) markers were mined from *de novo* transcriptome assembly using perl script of MISA-MIcroSAtellite identification tool [38]. For SSR marker analysis, ten repeating units for mononucleotides, six for dinucleotides and five for trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides. Primers were designed in the flanking regions of the SSR loci using Primer3core executable [39].

For variant calling, virus infected reads were mapped onto the *de novo* transcriptome assembly using Burrows-Wheeler Aligner (BWA) tool [40]. SAMtools package was used for calling SNPs (Single Nucleotide Polymorphism) and INDELs (insertion and deletion) [41]. Filtering criteria, namely, read depth ≥4, quality score ≥20 and flanking region 50 on both sides were used to identify significant SNP and Indels.

2.7. Construction of gene regulatory network

Gene regulatory network analysis was performed using Cytoscape 3.2.1 [42]. Expression Correlation plugin (http://apps.cytoscape.org/apps/expressioncorrelation) and network analysis program were used for construction of network. For analysis, highly upregulated and downregulated genes were consider for network analysis. Complex network was generated and hub genes were selected depending on the parameters like degree and between-ness centrality.

2.8. Validation of genes using qPCR

A total of seven differentially expressed genes were randomly selected for validation real-time PCR (qPCR). qPCR primers was designing using Primer3 [39] and Integrated DNA technologies (IDT) [43] online servers. For normalization, we have used elongation factor housekeeping gene. qPCR of selected genes was done using QuantiFast SYBR Green PCR master mix (Rotor gene Q apparatus, QIAGEN) [44] by 40 cycles with melt curve phase. After getting Ct values of each reaction, 2^-ΔΔCt method was used to calculate relative expression values [44].

2.9. Development of web genomic resource

SCMVTDb: Small Cardamom Mosaic Virus Transcriptome Database is based on "three-tier architecture" consisting of client tier, middle tier and database tier and available freely for non-commercial use at http://webtom.cabgrid.res.in/scmvtdb. In client tier, web pages have been developed using HTML and Javascript for user queries and browsing. In middle tier, scripting has been done using PHP for database connectivity, performing query and fetching data. Database tier has been developed using MySQL for storing information of DEGs, putative molecular markers (SSRs, SNPs and InDels) along with primers, blast results, transcription factors and KEGG pathways (Fig. 2).

3. Results and discussion

3.1. Pre-processing and assembly of sequence data

A total of 18,403,660 and 17,739,633 paired end read of control and infected samples, respectively were generated. After removal of 30,988 (control) and 30,945 (infected) low quality reads, remaining reads were further used for downstream analysis (Table 1). A total of 123,338 transcripts were generated by trinity with GC content 43.14% and N50 value of 1520 bp. The minimum and maximum lengths of assembled transcripts were 201 and 9327 bp, respectively. Maximum numbers of transcripts were found in the range of 200 to 299 bp length (34,020 transcripts), followed by 16,576 and 10,104 transcripts having the length between 300 to 399 and 400 to 10,104, respectively.

3.2. Abundance estimation and identification of differentially expressed genes

Paired end reads of both control and virus infected samples were mapped over *de novo* transcriptome assembly to calculate the expression values in terms of FPKM. These expression values can be used for identification of differentially expressed genes. In the present study, differential expressed genes were identified from NOISeq and edgeR

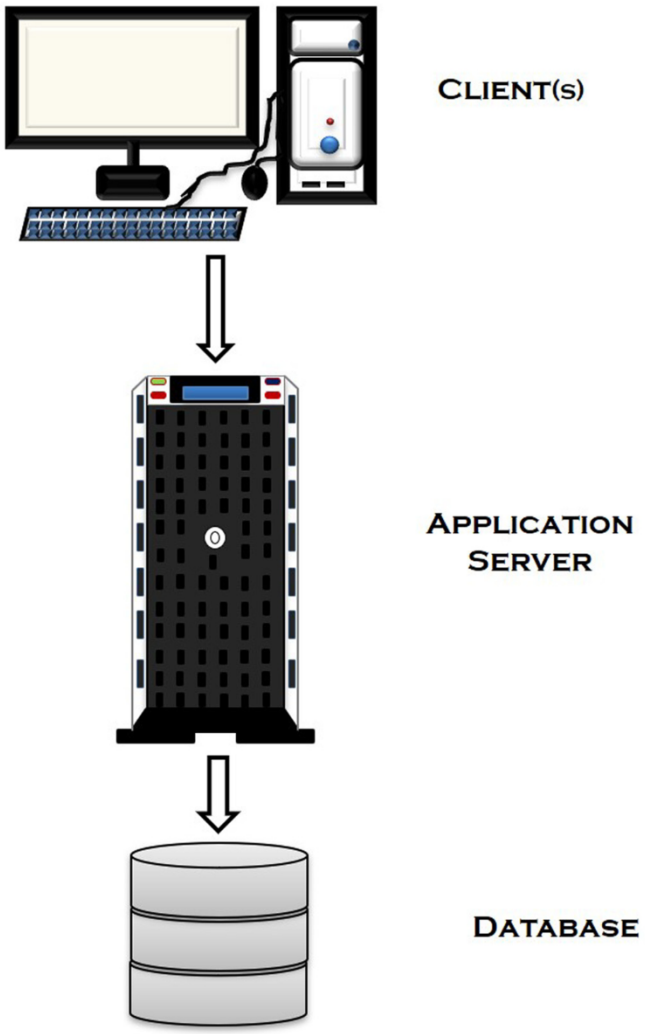

Fig. 2. Three-tier architecture of *SCMVTDb.*

methods. Using NOISeq, a total of 3519 DEGs were obtained at q value = 0.95. A total 5317 DEGs (2579 up and 2738 down regulated) were discovered from edgeR tool. The percentage similarity of common DEGs by both methods were 98.41% which reflects reliability of findings by edgeR. Findings by edgeR were used for further analysis.

3.3. Homology search, annotation and functional characterization

Out of 5317 transcripts, we found 4913 transcripts having similarity with other genes present in the non-redundant database, while 404 transcripts seemed to be novel, without having any similarity. A total of 1243 transcripts were involved in mapping and 2947 transcripts in GO annotation process. Top hit species distribution revealed maximum hits with *Musa acuminata* (3328), followed by *Elaeis guineensis* (2014) and *Phoenix dactylifera* (141) transcripts (Supplementary Table 1). Maximum similarity observed with these monocot plants due to conserved fatty acid biosynthetic pathways [45]. Blast2GO annotation results categorized these differentially expressed genes into three sub categories such as biological process, molecular functions and cellular components. Blast analysis of *de novo* transcriptome assembly of small cardamom was performed against non-redundant (NR) database which revealed that 81,966 (66.5%) out of 123,338 transcripts showed similarity with known genes in NR database. Maximum transcripts showed similarity with *Musa acuminata*, *Elaeis guineensis*, *Phoenix dactylifera* and *Vitis vinifera i.e.* 51,955, 3588, 2808 and 682, respectively (Supplementary Table 2).

3.4. Discovery of candidate genes involved in plant immune response

These DEGs represent key candidate genes and pathways which are associated in plant viral interaction. In this present study, important key differential expressed genes such as disease resistance RGA1 and RGA3, mitogen-activated kinase ANP1-like, pathogenesis related protein 1 (PR1), Thaumatin-Like Protein (PR5), NAC transcriptional factors, LRR receptor-like serine threonine- kinase, NBS-LRR disease resistance, glutathione S-transferase, heat shock cognate 70 kDa were found.

Previous study revealed that NIB genes encode RNA-dependent RNA polymerase for production and replication of viral RNA which has been used in development of transgenic small cardamom crop plant having *CdMV* resistance [46]. In our study, we found over expression of this gene in response to viral infection (> 9 fold). Leucine-rich repeat LRR domain present in the receptor-like kinase (RLK) interacts with other protein leading to signalling response. This RLK gene family protein plays an important role in pathogen recognition *via* signalling pathways which activates plant defense mechanism providing resistance against viral disease [47]. We also found several differential expressed genes which belonged to RLK gene-family. LRR receptor-like serine threonine-kinase FEI 1 gene of RLK family play role in regulation of cell wall [48]. RLK family genes, namely, receptor-like serine threonine- kinase SD1–8 and probable LRR receptor-like serine threonine- kinase At2g23950 which were found to be upregulated in our *in silico* analysis and were further confirmed in qPCR. Interestingly, Mitogen-activated protein kinase kinase kinase (MAPKKK) ANP1-like gene was found upregulated in *in silico* analysis as well as q-PCR results. ANP1 an "Arabidopsis homolog of NPK1" plays role in regulating innate immunity, cellular processes, cell-plate formation and progression of the cell cycle and development of crop [49,50]. Heat shock protein 81–1 (HSP81–1) belongs to heat shock protein 90 family and plays role in stabilization, newly synthesized protein folding and refolding of denatured proteins and also involved in signal transduction [51,52]. Resistance gene analog (RGA) plays role in disease resistance against root wilt disease of coconut [53]. We also found upregulation of RGA1 and RGA3 genes in *in silico* and qPCR.

Several other candidate genes identified from differential expression analysis which can be useful in development of mosaic disease resistant crops. For example, PR1 genes which are known to play an important role in immune system response in plant defense against various pathogens, namely, viral, fungal and bacterial [54]; PR5 [55]. We also found these genes to be upregulated in our study. PR1 gene reported in tomato when plants get infected with potato purple top phytoplasma showed upregulation [56]. NBS-LRR (nucleotide-binding site leucine-rich repeat) protein involved in the activation of kinases plays an important role in plant defense mechanism against pathogens invasion [53] but our study shows this protein to be downregulated.

Table 1
Trimming report of control and virus infected samples of *E. cardamomum* M.

	Input read Pairs	Both surviving	Forward only surviving	Reverse only surviving	Dropped
Control	18,403,660	17,887,330 (97.19%)	445,919 (2.42%)	39,423 (0.21%)	30,988 (0.17%)
Virus infected	17,739,633	17,142,028 (96.63%)	530,469 (2.99%)	36,191 (0.20%)	30,945 (0.17%)

3.5. Discovery of pathways involved in viral pathogenesis

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis was performed and a total of 1638 transcripts were found to be involved in 114 pathways. Maximum transcripts (185) were found to be involved in Purine metabolism, followed by 138 and 95 transcripts in biosynthesis of antibiotics and thiamine metabolism, respectively (Fig. 3, Supplementary Table 3). Interproscan was used for identification of domains and families in differential expressed genes. A total of 1219 domains in 5069 transcripts were found. Maximum transcripts belonged to protein kinase-like domain (IPR011009), protein kinase domain (IPR000719), P-loop containing nucleoside triphosphate hydrolase (IPR027417), zinc finger, RING/FYVE/PHD-type (IPR013083) and concanavalin A-like lectin/glucanase domain (IPR013320) in 170, 150, 124, 81 and 69 transcripts, respectively. In InterProScan family search, a total of 807 families were found in 1735 transcripts. Maximum transcripts were present in protein families such as cytochrome P450 (IPR001128), cytochrome P450, *E*-class, group I (IPR002401), protein phosphatase 2C family (IPR015655), short-chain dehydrogenase/reductase SDR (IPR002347) and insect cuticle protein (IPR000618) in 30, 22, 20, 17 and 16 transcripts, respectively (Supplementary Table 4).

3.6. Prediction of transcription factors involved in plant immune response

Transcriptional factors were identified searching against PlantTFDB 4.0 using blastx tool. A total of 2267 transcriptional factors out of 5317 transcripts with e-value, 1e-05 were observed. In our analysis, bHLH, MYB and NAC transcriptional factors were found maximum number of times *i.e.*, 251, 233 and 154, respectively (Supplementary Table 5). Transcriptional factors play an important role in defense mechanism of plants against biotic, abiotic stresses and signal transduction during pathogen invasion [57]. In our blast results against plant transcriptional factor database, PlantTFDB 4.0, we found maximum transcripts hits with these bHLH, MYB, NAC, C2H2, WRKY transcriptional factors. MYB plays role in development and responses to biotic and abiotic stress, and has the capability to bind with DNA. We found several MYB isoforms in our study with high upregulation which can be further used for development of disease resistant crop [57,58]. C2H2 plays an important role in defense response and various physiological processes [59]. Basic helix-loop-helix (bHLHs) play role as a regulator of plant defense mechanism. It activates the different genes which are involved in plant responses to environmental signals like hormone signalling. We found several bHLHs TFs with upregulation in our study which can be used for developing resistant crops [60]. NAC transcriptional factors play an important role in abiotic as well as biotic stress tolerance which helps in regulating transcriptional reprogramming related with different type of plant stress responses. In biotic stress, it is one of the essential components of plant defense, plant immune system [61]. In our study, we found few isoforms of NAC transcriptional factors such as NAC29, NAC100, NAC72, NAC68, NAC76, NAC2, NAC74, NAC8, NAC78, NAC25, NAC domain containing 21/22, NAC transcription factor NAM-B2-like. For q-PCR, we selected two NAC29 and NAC100 [62] due to its high fold *in silico* analysis change values and results were in concordance. Several studies have been done on NAC100 and NAC29 transcriptions against fungal infection or other biotic stresses but no such study on virus attack is available, so these NAC transcriptional factors genes may be important for developing disease resistant varieties.

3.7. Mining of putative molecular markers

A total of 5317 and 21,735 putative SSR markers were mined respectively from differentially expressed genes and *de novo* transcriptome assembly, for which 1092 and 15,168 primers were computed using PRIMER3 tool (Table 2, Supplementary Table 6). We found 21,735 SSR markers which can be useful in variations and genetic diversity studies. There are < 450 SSR markers reported so far in this crop [63,64]. In absence of whole genome, these ready to use SSR markers can cater the need of genotyping [21]. Further these identified makers can be useful for polymorphism and cross taxon species studies. Though genic region SSRs are less polymorphic but have robust amplification efficiency and stable transferability. Being in coding region, they represent functional diversity of the gene [65]. They are also called functional domain marker (FDM). Such SSR-FDMs have been used in trait improvement in various crops like *Camellia sinensis* [66], papaya [67], *Ocimum basilicum* [68], chickpea [69], sugarcane [70], *Elaeis guineensis* [71] and *Seasme indicum* [72]. Besides trait improvement, such markers are also used in cultivar identification, population structure and QTL analysis [73].

Interestingly, even in a single genotype used in our experiment, variant mining results showed a total 154,217 variants (from virus infected samples with 149,097 SNPs and 5120 Indels) (Supplementary Table 7). In any plant genotype even if it is an inbreeding line, invariably there will be presence of heterozygosity which is called residual heterozygosity [74]. Moreover, in case of small cardamom, beside residual heterozygosity, potential source of high level variation in small cardamom is due to cross-pollination [75] also. Variation could be also due to presence of various isoforms in RNA Seq data [76]. This was further confirmed by higher abundance of SNPs in plant immune responsive genes. For example, maximum number of SNP and Indels in transcript ID c36155_g1_i1 - clathrin heavy chain 1-like gene (492) which is a well-known plant immunity related genes in crop maize reported to have > 10 variants [77]. Similarly, another transcript having higher number of variants represented by calpain-type cysteine protease DEK1 gene- transcript ID c36145_g1_i1 (399) is also reported to have large number of variants [77]. Similar higher abundance of isoform is also reported in maize immunity responsive genes ATP-dependent helicase BRM-like gene-transcript ID c36135_g3_i2 (341) (Brahma1 A0A1D6GYJ1_MAIZE) [78].

3.8. Mechanism of plant-virus interaction and gene regulatory network

Top 100 up and down regulated genes were selected to construct GRN describing plant-virus interaction mechanism. Such approach narrows down the number of DEGs to get key candidate genes using parameters of betweenness, centrality and degree (104 to 199). It further depicts hub genes playing major role in network mode. Among the top 20 hub genes (degree value > 195), 8 and 12 were found up and down regulated, respectively (Fig. 4, Supplementary Table 8).

Sensing of stress and subsequent immune response of host against *CdMV* was observed in our finding. Differential expressed genes receptor-like kinase (RLK) gene-family was found which are found to play role in regulation of cell wall [48]. Leucine-rich repeat LRR domain present in the RLK interacts with other protein leading to signalling response of pathogen recognition. It activates plant defense mechanism and provides resistance against disease [47]. Upregulated mitogen-activated protein kinase kinase kinase (MAPKKK) ANP1-like gene is known for various functions like innate immunity, cellular processes, cell-plate formation and progression of the cell cycle and growth [49,50].

Plant immune response was observed in our finding having hits with *Tobamovirus* protein families representing both up and down regulated transcripts. They are known to modulate host plant immunity in mosaic viral disease of Chinese yam [79]. It was interesting to observe EF-hand domain gene acting as hub gene in GRN constructed from our dataset. It is a part of calcium-calmodulin (CaM) system which is known sensing mechanism for various biotic and abiotic stress signalling by sequestering calcium ions. It is a major regulator of antiviral response [80]. TIFY 6B gene family mediates the Jasmonic acid pathway of stress response by JAZ (Jasmonate ZIM-domain) protein in *Brassica rapa* [81]. Heat shock protein 81–1 (HSP81–1) belongs to heat shock protein 90

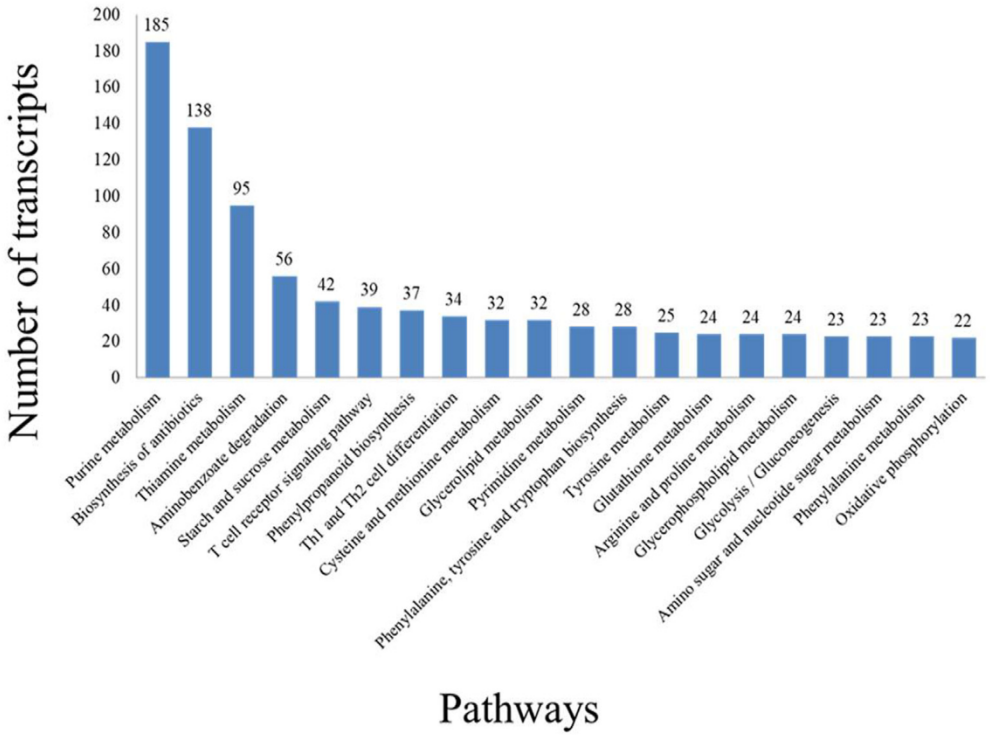

Fig. 3. Top 20 KEGG pathways of differentially expressed genes of *E. cardamomum* M.

family and plays role in stabilization, newly synthesized protein folding and refolding of denatured proteins and also involved in signal transduction [51,52]. Upregulated bZIP transcription factor TRAB1 is known to interact plant viviparous 1 (VP1) mediating abscisic acid induced pathways [82].

Secondary metabolites are produced in plants as a defense mechanism against various pathogens [83]. In our dataset, up regulation was found in cytochrome P450 gene which controls production of secondary metabolite including polyphenolic flavonoids having antiviral activities in plants [84]. We found differential expression of two pathogenesis-related (PR) proteins namely, PR1 and PR2. Such proteins are known for plant immunity against viral, fungal and bacterial pathogens [54]. PR genes has been used for crop improvement against viral disease like Iris yellow spot virus resistance in onion [55]. Protein phosphatase gene is upregulated and performs resistance against viral infection by inhibiting cell to cell movement of virus by inducing callose deposition through ABA signalling [85].

In *CdMV* disease, there is a gradual loss of host plant growth over a period of three years [16]. This growth retardation is mediated by perturbation in auxin homeostasis. This mechanism was also observed in our constructed GRN (Fig. 4) where ARF has been found as a hub gene. In case of Cauliflower mosaic virus infection, ARF (auxin response factor) has been reported to function as repressor and activator of auxin mediated pathway [86]. In case of tobacco mosaic virus, reprogramming of auxin-responsive pathway enhances the viral infection [87]. Hub gene, PIN in our GRN a known auxin transporter affecting growth of the plant. PIN gene family are distributed over different chromosomes having cis-element variation especially in polyploid species like cotton. Since they are known to play role in auxin homeostasis affecting plant growth and differentiation, thus they are acting as hub gene in GRN [88].

Viral infection in plant is witnessed with bi-directional traffic

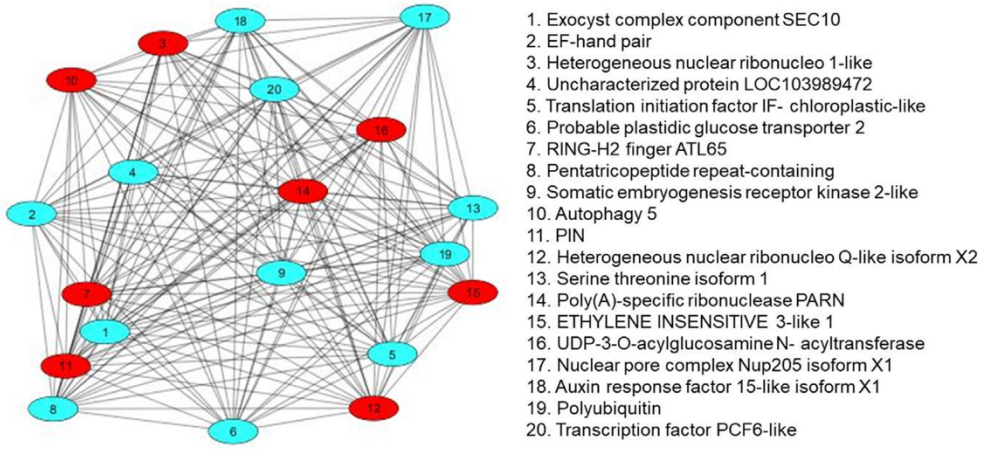

Fig. 4. Top 20 selected hub genes. Red color represents upregulated genes and cyan represents downregulated genes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)