Physiol Mol Biol Plants (October–December 2017) 23(4):767–777
DOI 10.1007/s12298-017-0470-7

RESEARCH ARTICLE

Development of transcriptome based web genomic resources of yellow mosaic disease in *Vigna mungo*

Rahul Singh Jasrotia[1,2] · Mir Asif Iquebal[1] · Pramod Kumar Yadav[2] · Neeraj Kumar[1] · Sarika Jaiswal[1] · U. B. Angadi[1] · Anil Rai[1] · Dinesh Kumar[1]

Received: 24 March 2017 / Revised: 6 September 2017 / Accepted: 11 September 2017 / Published online: 18 September 2017

Abstract *Vigna mungo* (Urdbean) is cultivated in the tropical and sub-tropical continental region of Asia. It is not only important source of dietary protein and nutritional elements, but also of immense value to human health due to medicinal properties. Yellow mosaic disease caused by Mungbean Yellow Mosaic India Virus is known to incur huge loss to crop, adversely affecting crop yield. Contrasting genotypes are ideal source for knowledge discovery of plant defence mechanism and associated candidate genes for varietal improvement. Whole genome sequence of this crop is yet to be completed. Moreover, genomic resources are also not freely accessible, thus available transcriptome data can be of immense use. *V. mungo* Transcriptome database, accessible at http://webtom.cabgrid.res.in/vmtdb/ has been developed using available data of two contrasting varieties viz., cv. VM84 (resistant) and cv. T9 (susceptible). De novo assembly was carried out using Trinity and CAP3. Out of total 240,945 unigenes, 165,894 (68.8%) showed similarity with known genes against NR database, and remaining 31.2% were found to be novel. We found 22,101 differentially expressed genes in all datasets, 44,335 putative genic SSR markers, 4105 SNPs and Indels, 64,964 transcriptional factor, 546 mature miRNA target prediction in 703 differentially expressed unigenes and 137 pathways. MAPK, salicylic acid-binding protein 2-like, pathogenesis-related protein and NBS-LRR domain were found which may play an important role in defence against pathogens. This is the first web genomic resource of *V. mungo* for future genome annotation as well as ready to use markers for future variety improvement program.

Keywords *Vigna mungo* · Differential expressed genes · Gene ontology · Markers · MYMIV · Transcriptome

Introduction

Vigna mungo (Urdbean), the orphan legume crop under '*Fabaceae*' family, is an inexpensive key source of dietary protein and also considered as "the poor man's protein" among different pulse crops (Kakati et al. 2010). This legume is widely cultivated in the tropical and sub-tropical regions of Asian continent for its edible seeds. Seeds of *V. mungo* are important source of dietary protein, starch, mineral elements, vitamins, higher level of folate and iron (Sharma et al. 2011). *V. mungo* has medicinal properties like, antihyperlipidemic and antihyperglycemic relevant in treatment of type 2 diabetes (Kaur et al. 2015). This pulse is also important due to anti-cardiovascular disease and anti-cancerous properties (Campos-Vega et al. 2010).

There is steep decline in the productivity in *V. mungo* due to several biotic stresses among which stress due to viruses are predominating (Naimuddin et al. 2011). Among the viral diseases, Yellow mosaic disease is the most destructive and widely distributed, caused by Mungbean Yellow Mosaic India Virus (MYMIV). This disease is transmitted by the vector *Bemisia tabaci* (whitefly) (Kundu et al. 2015). MYMIV belongs to family *Geminiviridae* and genus *Begomovirus*, having bipartite genomes (Reddy et al. 2015).

✉ Dinesh Kumar
Dinesh.Kumar@icar.gov.in; dineshkumarbhu@gmail.com

[1] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India

[2] Department of Computational Biology & Bioinformatics, Sam Higginbottom University of Agriculture, Technology & Sciences (SHUATS), Allahabad 211007, India

Management of MYMIV is often done through foliar spray of dimethoate which has concerns for health hazards beyond prescribed concentration (Davies et al. 2008). In order to develop viral resistant varieties, pathways and genes associated with these traits must be deciphered.

Since whole genome sequencing of this crop is yet to be done, the available transcriptome data can be an economical and valuable source of candidate gene discovery. Though very limited studies, where just two candidate genes (Qazi et al. 2007) and in another study, having virus challenged transcriptomic response confined to a single genotype (Ganguli et al. 2016) are reported, there is a need to carry out such investigations on contrasting genotypes which would be a better source of candidate gene discovery. Moreover, there is no report on molecular markers from differentially expressed transcripts having potential to be used as functional domain markers.

The present work aims at profiling differentially expressed genes in leaf tissue of two contrasting genotypes, i.e., VM84 (resistant) and T9 (susceptible) of *V. mungo*, discovery of genic region SSRs, SNPs and Indels, transcription factors and prediction of microRNAs targets along with the development of transcriptome based web genomic resources.

Materials and method

Pre-processing and de novo assembly

Four sets of RNA-seq data of *V. mungo* was retrieved from NCBI SRA database. First two datasets pertained to contrasting genotypes viz., VM84 (infected resistant, SRX1082731) and VM84 (mock inoculated control resistant, SRX1032950) and second two datasets were T9 (infected susceptible, SRX1058327) and (mock inoculated control, SRX1058325). All these Illumina datasets were pre-processed using Trimmomatic tool version 0.33 for removal of low quality reads (phred-score $\leq$20 and reads with ambiguous bases 'N'), trimming of bases from 5′ and 3′ end and adaptor sequences (Bolger et al. 2014). Quality assessment of raw and trimmed reads was carried out using FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Further, these clean and filtered reads were used for de novo assembly using short read assembly program Trinity Release v2.0.6 (Haas et al. 2013). CAP3 assembler was used to reduce redundancy (Huang and Madan 1999).

Identification of differential expressed genes

Mapping and alignment of high quality reads onto reference de novo assembly was performed by using bowtie tool (Langmead et al. 2009), to obtain the read density which provides the gap free alignment. Reads with maximum two mismatches were allowed to map the de novo assembled transcriptome. Abundant estimation and calculation of expression values were performed using 'RNA-Seq by Expectation–Maximization (RSEM)' tool (Li and Dewey 2011). Calculation of expression values of each transcript were based on fragments per kilobase of exon per million mapped reads (FRKM). EdgeR package (Empirical analysis of Digital Gene Expression in R) (Robinson et al. 2010) of Bioconductor was used to calculate the differential expressed genes applying filtering criteria of p value 0.01 and log fold change$_2$10 to get the significant genes.

Annotations and functional categorization of differential expressed genes

Homology search of differential expressed genes against NCBI non-redundant database (nr.32) (ftp://ftp.ncbi.nlm.nih.gov/blast/db/) with Blastx algorithm having threshold expected value 1e-3 was done using Standalone local ncbi-blast-2.2.31+ (Camacho et al. 2009). Blast2Go Pro version 3.1 software (Conesa et al. 2005) was used for functional categorization and gene ontology of genes in three sub categories i.e. biological process, molecular function and cellular component. Kyoto encyclopaedia of genes and genomes (KEGG) analysis of unigenes was performed by Blast2Go software which provides the pathways and enzyme classification number (EC). Transcriptional factors (TF) were identified from de novo assembly of *V. mungo* using PlantTFDB (Jin et al. 2016) (http://planttfdb.cbi.pku.edu.cn/).

SSR and SNP mining

Identification of SSRs was performed by using perl script of MISA-MIcroSAtellite identification tool (Thiel et al. 2003). Ten repeating units for mono-, 6 repeating units for di- and 5 repeating units for tri-, tetra-, penta- and hexa nucleotide criteria were taken for mining the SSR markers and PRIMER3 tool (Untergasser et al. 2012) was used to design primer sets for database entry.

Also, we identified the SNPs and Indels from *V. mungo* resistant VM84 and susceptible T9 cultivars separately. All reads of different samples were aligned and mapped to de novo transcriptome reference assembly using Burrows–Wheeler Aligner (BWA) tool (Li and Durbin 2009). For calling SNPs and Indels in the mapped reads, SAMtools package was used (Li et al. 2009). The filtering criteria used for mining SNPs/Indels were 15× depth (Uitdewilligen et al. 2013), quality score of 30 (Liu et al. 2014; Yu and Sun 2013) and 2 SNPs were allowed within the flanking region of 50 bp on each side.

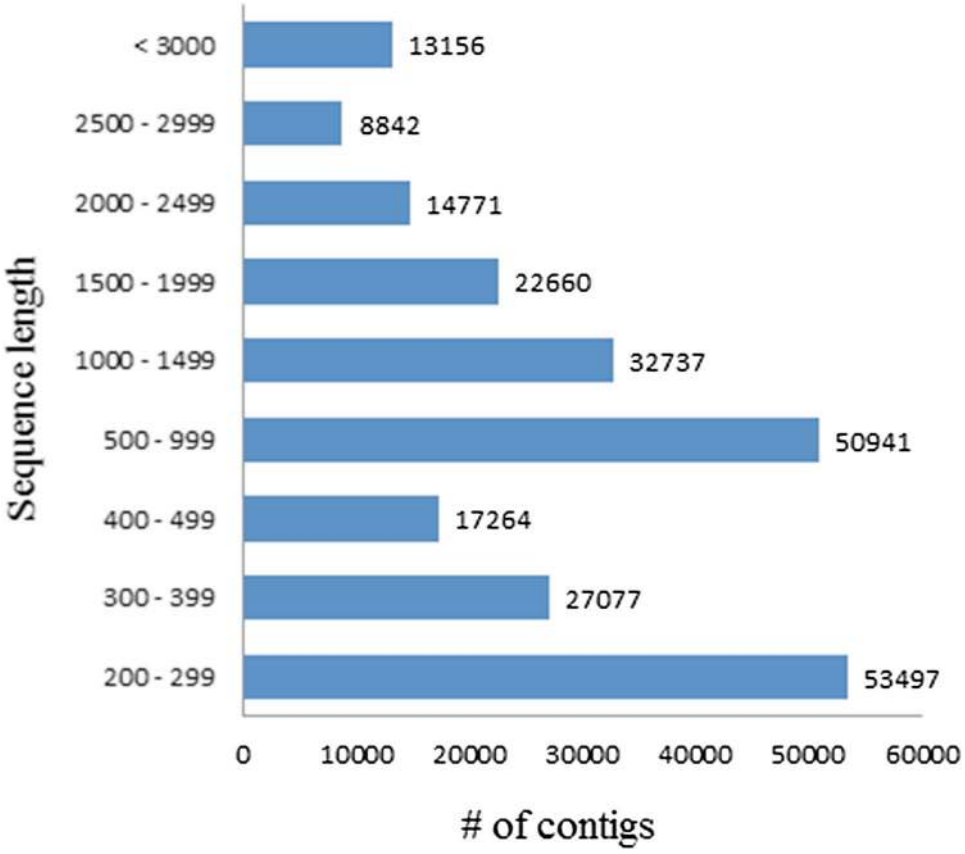

Fig. 1 Sequence length distribution of de novo transcriptome assembly of *V. mungo*

MiRNA target prediction of DEGs

The miRNA target predictions of *V. mungo* from differential expressed genes were done against the *fabaceae* family crops. 1545 mature miRNA sequences were retrieved from miRBase database release 21 (Kozomara and Griffiths-Jones 2014) (http://www.mirbase.org/). MiRNA Target prediction of *V. mungo* DEG's was carried out by using MiRanda-3.3a tool (http://www.microrna.org/microrna/getDownloads.do).

Development of web genomic resources

All the analysed results were catalogued in the online relational database of *V. mungo*, available at http://webtom.cabgrid.res.in/vmtdb/. It is based on "three-tier architecture" having a client tier, middle tier and database tier. To browse and query the database user can go through the developed web pages in client tier. Different tables having transcripts/contigs, SSRs, variants, miRNA targets corresponds to MySQL database in the database tier. For database connectivity, query execution and fetching, server side scripting has been done using PHP in the middle tier. Primer3 has also been integrated at the backend for primer generation of the selected SSR markers.

Results and discussion

Quality control and de novo assembly

After removing 4,332,250 poor qualities reads from the four sets of data, 154,742,250 (97.2%) cleaned and good quality reads with phred score ≤Q20 were retained for de novo assembly. A total of 308,523 contigs using short read algorithm were generated by Trinity assembler. After applying CAP3 assembler on Trinity assembly results for the removal of redundant sequences, 240,945 contigs were finally taken for further analysis. Total GC content found in assembly was 39.74% with N50 of 1759 bp. Minimum and maximum contigs lengths were 201 and 22,103 bp respectively. Figure 1 represents the sequence length distribution of the contigs produced by CAP3 assembly.

Abundance estimation and identification of differential expressed genes

A total of 10,672, 10,209, 475 and 746 differentially expressed genes (DEGs) were obtained in analysis of transcriptomic sets, viz, infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9, respectively. Up and

Table 1 Upregulated and downregulated genes at each stage specific comparison study

Comparison stage	Upregulated	Downregulated	Total
Infected VM84 vs. infected T9	10,285	387	10,672
Infected VM84 vs. control VM84	9971	238	10,209
Infected T9 vs. control T9	260	215	475
Control VM84 vs. control T9	377	369	746

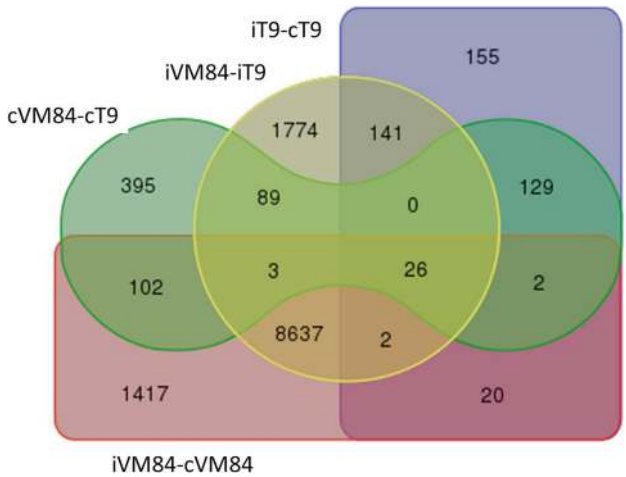

Fig. 2 Common and unique DEGs of all the four datasets

down regulated genes in each set of comparison is presented in Table 1. Common and unique DEGs of all the four sets are depicted by Venn diagram (Fig. 2). A total of 26 unigenes were common in all the four sets used in comparison, while 1774, 1417, 155, 395 unigenes were unique to infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9, respectively. Maximum of 8637 unigenes were common in sets of infected VM84 vs. control VM84 and infected VM84 vs. infected T9.

Venn diagram shows that no DEGs were common in infected VM84 vs. infected T9, infected T9 vs. control T9 and control VM84 vs. control T9 (Fig. 2) reflecting different response mechanism operating in these contrasting genotypes. Maximum numbers of DEGs were observed in VM84 infected vs. T9 infected and VM84 infected vs. VM84 control i.e., 10,672 and 10,209, respectively (Table 1).

We observed higher expression of Salicylic acid-binding signalling protein in resistant variety VM84 as compared to susceptible T9 in response to MYMIV infection. This signalling further triggers Mitogen-Activated Protein Kinase (MAPK) gene in cascade to trigger plant defence mechanism (Slaymaker et al. 2002). This has already been reported in other crops like tobacco, *Arabidopsis thaliana* and tomato (Pedley and Martin 2005).

We found an interesting observation that few pathways and associated genes of yellow mosaic virus are well conserved in other crop families like leguminosae and solanaceae. For example, Salicylic acid pathway is also reported in tomato (Miozzi et al. 2014), chilli (Kushwaha et al. 2015), capsicum (Gongora-Castillo et al. 2012) and tobacco (Czosnek et al. 2013). Similarly, NAC gene family is also reported in chilli pepper biotic stress resistance pathway (Oh et al. 2005). MAPK pathway is also reported in tomato against other geminivirus like crinivirus (Kaur et al. 2017).

Among the 26 differentially expressed genes in all the four datasets, upregulated terpene synthase was obviously expected due to its terpene molecule generalized role in defence against biotic stresses (Gershenzon and Dudareva 2007). Similarly, MYB TF is also highly expressed which is known for mediating and controlling both biotic as well as abiotic stress (Ambawat et al. 2013). Among other common set of genes, viz., U-box domain-containing protein 33-like isoform X2, cap-binding protein-like protein, mitochondrial import receptor were found highly expressed which is expected in metabolically active tissue against MYMIV challenge (Endo and Kohda 2002). Rest 21 are hypothetical proteins.

We found many pathogenesis-related proteins in our study, such as PR1 gene whose role is already implicated in plant immunity in very same crop of *V. mungo* (Kundu et al. 2015). Similarly, PR4 protein which was found in our study also and known for protecting the DNA and RNA from degradation during the host–pathogen interaction is also reported in *Capsicum chinense* (Guevara-Morato et al. 2010). These pathogenesis-related proteins were found highly expressed in infected VM84 vs. infected T9, infected VM84 vs. control VM84 and control VM84 vs. control T9.

In the three datasets, viz, infected VM84 vs. infected T9, infected VM84 vs. control VM84, and control VM84 vs. control T9, all the NBS-LRR proteins were found highly expressed. We found higher expression of resistant (R) genes which are well known in plant offering specific resistance. These proteins have nucleotide-binding site leucine-rich repeat (NBS-LRR) domains which provide resistance to plant from various pathogens like viruses, bacteria, fungi etc. (Głowacki et al. 2011). Conformational modifications in NBS domain stimulate the exchange of ATP with ADP which further triggers downstream signaling (DeYoung and Innes 2006). This clearly indicates that

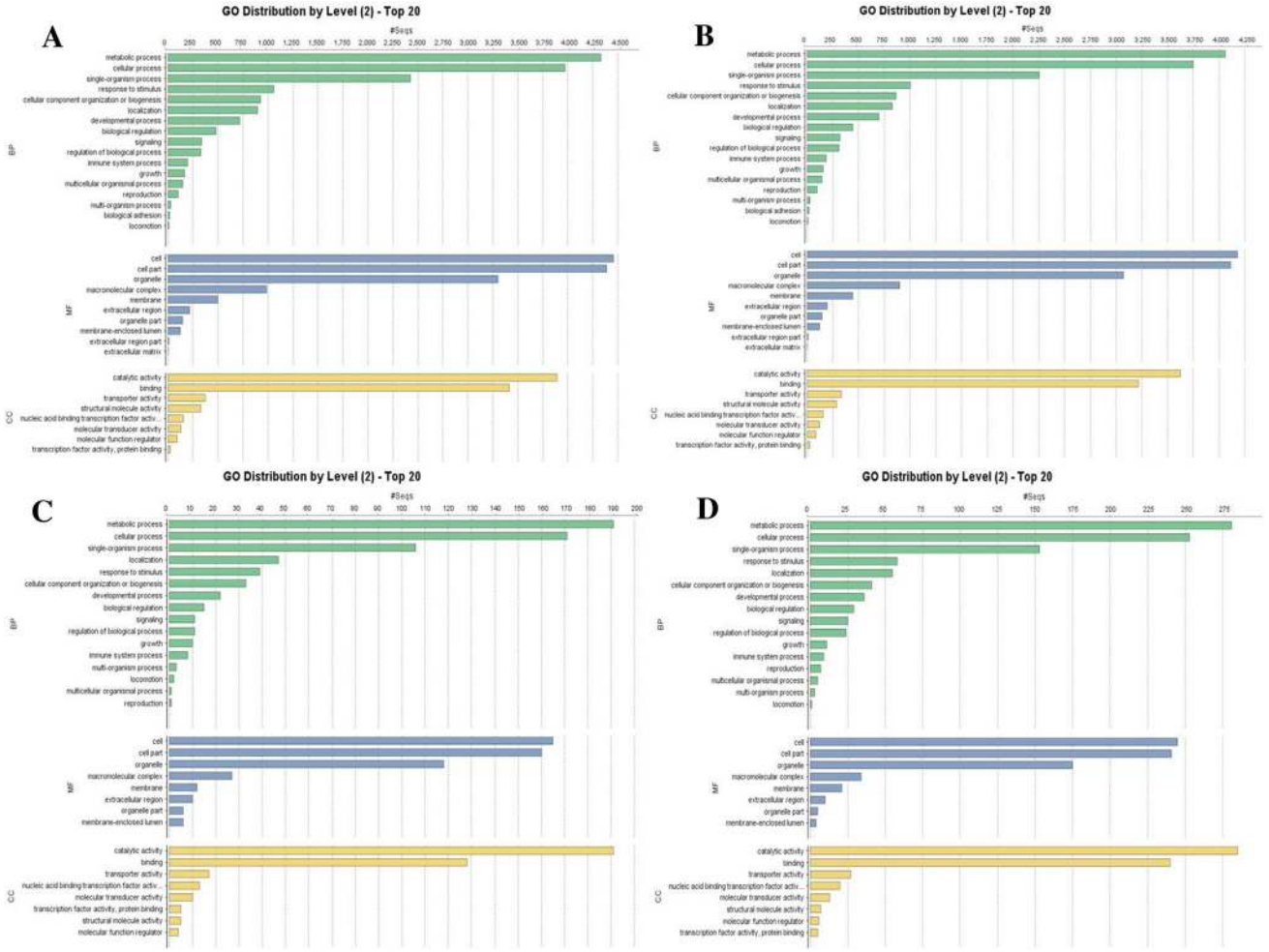

Fig. 3 Gene ontology of differential expressed genes of all datasets (comparative study of *V. mungo*). Green colour showed the biological process, blue colour represents the molecular functions and yellow colour represents cellular components. **a** Infected VM84 vs. infected T9; **b** infected VM84 vs. control VM84; **c** infected T9 vs. control T9; **d** control VM84 vs. control T9 (colour figure online)

MYMIV susceptibility in the genotype T9 is due to lack of R gene expression which is also reported in previous studies (Kundagrami et al. 2009).

Calreticulin is the special kind of protein of endoplasmic reticulum (ER) chaperone system, which was found up-regulated only in response to MYMIV challenge in resistant variety (VM84) indicating its critical role in plant immune system as reported in previous studies of Tobacco mosaic virus (TMV) (Caplan et al. 2009). Lack of N immune receptor may cause plant death in virus attack (Garg et al. 2015). Similarly, in chaperone system, Calmodulin (CaM) associate with the Ca^{2+} signalling has also been reported to be involved in plant defence against various pathogens (Cheval et al. 2013). Similar role has been reported for tobacco calmodulin-like protein having binding ability with dsRNA to act as viral related RNA silencing suppressors (RSSs) inhibiting RNAi which is the unique defence mechanism of host plants against the viral attacks (Nakahara et al. 2012).

Functional characterization and pathways analysis of DEG's

Blast homology search of all the four sets of DEGs revealed maximum hits with four species viz., *Phaseolus vulgaris, Glycine max, Solanum tuberosum* and *Solanum lycopersicum.* Blast results of all the four sets of differential expressed genes data and whole transcriptome assembly are given under the supplements heading of the database link (http://webtom.cabgrid.res.in/vmtdb/). We obtained four sets of unigenes 8107, 7679, 371 and 570 from infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9 stage, respectively, which were used for gene ontology and annotation. Gene ontology of DEG was further categorized into three i.e. biological process, molecular function and cellular component as depicted in Fig. 3.

Table 2 List of SSR markers obtained from de novo transcriptome assembly as well as differential expressed genes of stage study infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9

	De novo Transcriptome assembly	Inf VM84 vs. inf T9	Inf VM84 vs. con VM84	Inf T9 vs. con T9	Con VM84 vs. con T9
Sequences examined	240,946	10,673	10,210	476	747
Identified SSRs	46,499	2609	2425	177	262
SSR containing sequences	37,847	2125	1987	135	196
Sequences containing more than 1 SSR	6824	385	350	30	49
SSRs present in compound formation	2164	145	138	12	14
Mono	27,603	1452	1361	94	137
Di	9809	525	486	47	64
Tri	8530	604	551	33	59
Tetra	471	14	12	4	2
Penta	43	3	3	Nil	Nil
Hexa	43	11	12	Nil	Nil

Analysis of 240,945 contigs in PlantTF database revealed 87,195 (36.18%) transcription factors. Highest abundance was found for MYB family protein (9213 times) followed by bHLH (8704 times), NAC (5953 times), ERF (4520 times) and C2H2 (4293 times). Supplementary sheet of identified transcription factors are given in database link http://webtom.cabgrid.res.in/vmtdb/.

Transcriptional factors observed under study such as bZIP, WRKY, myelocytomatosis related proteins (MYC) and myeloblastosis related proteins (MYB) and NAC family, have been reported to play role in signal transduction during pathogen attack in other crops (Alves et al. 2014).

Using Blast2GO pro software, we identified the KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways. Differential expressed genes of all stages were mapped and found 134, 132, 72 and 86 pathways were involved in infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9 datasets, respectively. We found "Biosynthesis of antibiotics" pathway maximum number of time i.e. 149 and 147 in infected VM84 vs. infected T9 and infected VM84 vs. control VM84 respectively, where as "Starch and sucrose metabolism pathway" found the maximum number of times in infected T9 vs. control T9 and control VM84 vs. control T9, respectively.

KEGG pathways analysis was performed to understand the biological function of differentially expressed genes. The major pathways found in our study were pertaining to phenylpropanoid biosynthesis, starch and sucrose metabolism. Starch and sucrose metabolism plays vital role in viral attack of crop as reported in studies of *Manihot esculenta* Crantz crop infected with cassava mosaic virus (Allie et al. 2014). Role of similar phenylpropanoid biosynthesis pathway in potato plant defence against *Potato virus Y* (PVY) infection has been reported (Kogovšek et al. 2016; Kundu et al. 2015).

Markers identification

We have identified SSR marker from de novo transcriptome assembly as well as all DEGs of the four datasets. We found 44,335 putative SSR markers from de novo transcriptome assembly. In DEG specific mining, we obtained 2464, 2287, 165 and 248 putative SSR markers in infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9 dataset, respectively (Table 2). Primers of these SSRs are present under the supplements heading of database link http://webtom.cabgrid.res.in/vmtdb/.

The genic region SSRs can be used as functional domain markers as well as in evolutionary, ecological, linkage mapping, quantitative trait loci (QTL) exploration, genetic diversity and comparative genomics studies. Similar use has been reported for linkage mapping in *V. mungo* (Gupta et al. 2008), varietal differentiation in black gram (*Vigna spp.*) (Kaewwongwal et al. 2015).

The bulk of the discovered genic region SSRs can be prioritized by narrowing down the number by mining dataset specific DEGs. Such SSRs (Table 2) can be used as functional domain markers for future association studies as reported in crop like *Pinus* (Alisoltania et al. 2016). Till date only a single genic region marker (RGA 1F-CG/RGA 1R) obtained in studies of contrasting genotypes *V. mungo* challenged with yellow mosaic virus (YMV) has been reported (Basak et al. 2005).

A total of 103,415 SNPs and 5890 Indels were mined from genic regions of all the four datasets (Table 3). Detailed information of SNPs and Indels with the flanking regions is present in the database under supplements heading (http://webtom.cabgrid.res.in/vmtdb/). It was observed that 4105 SNPs and Indels were common in all samples as shown in Fig. 4. Maximum number of variants in form of SNPs and Indels were found TR026348 (gene name: clustered mitochondria homolog).

These genic region SNP markers can be used in future association studies. Similar approach to obtain genic region SNP markers and trait improvement has been reported in other crops like *Ocimum basilicum* (Gupta et al. 2010), *Seasme indicum* (Bhattacharyya et al. 2014), *Elaeis guineensis* (Tranbarger et al. 2012) and *Camellia sinensis* (Sahu et al. 2012).

MicroRNA target prediction

MicroRNA controlling DEGs in each of the four datasets was obtained successfully. We found 178, 130, 113 and 125 miRNA of fabaceae family regulating 346, 303, 24 and 30 differential expressed genes of infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9 datasets, respectively.

Table 3 Number of SNPs and Indels shown in all samples of *V. mungo*

Samples	SNP	Indels	Total
VM84_infected	78,720	3267	81,987
T9_infected	9251	1089	10,340
VM84_control	7670	689	8359
T9_control	7774	845	8619

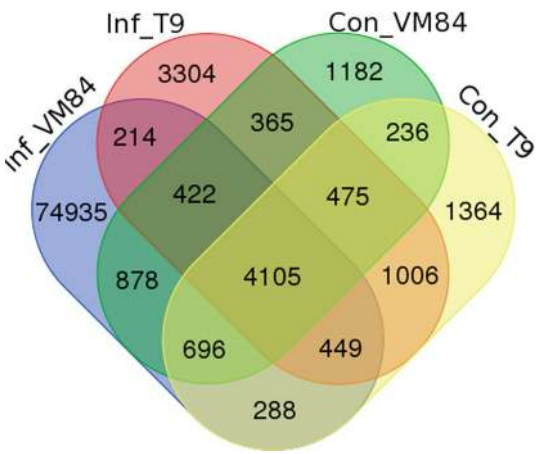

Fig. 4 Venn diagram of common and unique SNP and Indels in infected VM84, control VM84, infected T9 and control T9

List of Identified miRNA targets are present in the supplements section of database (http://webtom.cabgrid.res.in/vmtdb/).

Among these four datasets, only 12.6% miRNAs were found to have common targets as shown in Venn diagram (Fig. 5). Previous studies revealed that, both virus and host encodes miRNA which play critical role in host–pathogen interaction network and regulate gene at post transcriptional level. These miRNAs control and manipulate the biological process (Scaria et al. 2006). These miRNA based targeted seed region of UTR can be further investigated for polymorphism and association studies. Similar miRNA polymorphism based studies have been reported in many crops with trait improvement (Liu et al. 2013; Wang et al. 2013; Mishra and Bertino 2009).

Database development

The developed web transcriptomic resources of *V. mungo* catalogues detailed description of the transcripts/contigs, markers and miRNA targets, which is of immense use to *Vigna* breeders across the globe. The *V. mungo* Transcriptome database (VmTDB) provides transcriptomic information of two genotypes of *V. mungo* viz. VM84 (resistant) and T9 (susceptible) cultivars. It includes 240,945 assembled transcripts, differential expressed genes and expression values of four datasets, *i.e.,* infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9. Beside this, it provides the information related to pathways, simple sequence repeats, single nucleotide polymorphism,

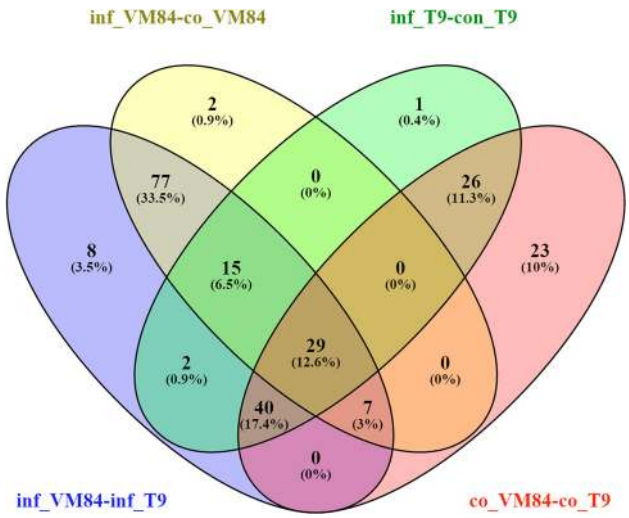

Fig. 5 miRNA target prediction of all datasets (comparative study of *V. mungo*). Stage infected VM84 vs. infected T9, infected VM84 vs. control VM84, infected T9 vs. control T9 and control VM84 vs. control T9 are shown in blue, yellow, green and pink colors, respectively (colour figure online)

transcriptional factors, domain and families and miRNA targets.

The *V. mungo* Transcriptome database (VmTDB) has six utility tabs such as *Home, Transcripts, Markers, Candidate genes, miRNA target and supplements.* In "*Transcript*" tab, user can search expression values, transcriptional factors, pathways, domains and families. In *"Expression profile",* search provides the information of expression values in the form of FPKM and their blast hits. In *"TF",* search provides the information of transcripts which showed similarities with transcriptional factors along with the hyperlinks attached to each transcript with PlantTFDB (Plant Transcription Factor Database). In *"Domain and families"* tab, information of identified domain and family of transcripts are hyperlinked with database of EMBL-EBI Interpro. In *"Pathways"* tab, it provides information of pathways with enzyme name and enzyme ID which were found in differential expressed gene of four datasets. Pathway maps are directly linked to KEGG database. *"Markers"* tab includes the option of SSR, SNP and Indels with details of these identified putative markers. Under the tab *"Candidate Genes",* information of differential expressed genes of all the four datasets are catalogued. Under the tab *"miRNA target"*, information of putative miRNA targets which regulates specific transcripts in all the four datasets are provided. Figure 6 shows the schematic diagram of VmTDB.

Conclusion

This is the first report of *V. mungo* transcriptome using two contrasting genotypes challenged against MYMIV depicting host-parasite interaction at molecular level. De novo transcriptome assembly exhibited 240,945 unigenes in our studies. A total of 22,102 DEGs were found having 20,893 up- and 1209 down- regulated genes. The study reveals systemic resistance pathways of *V. mungo* against viral infection. A total of 64,964 transcription factors and 546 miRNA targets were found in regulation of 703 DEGs in

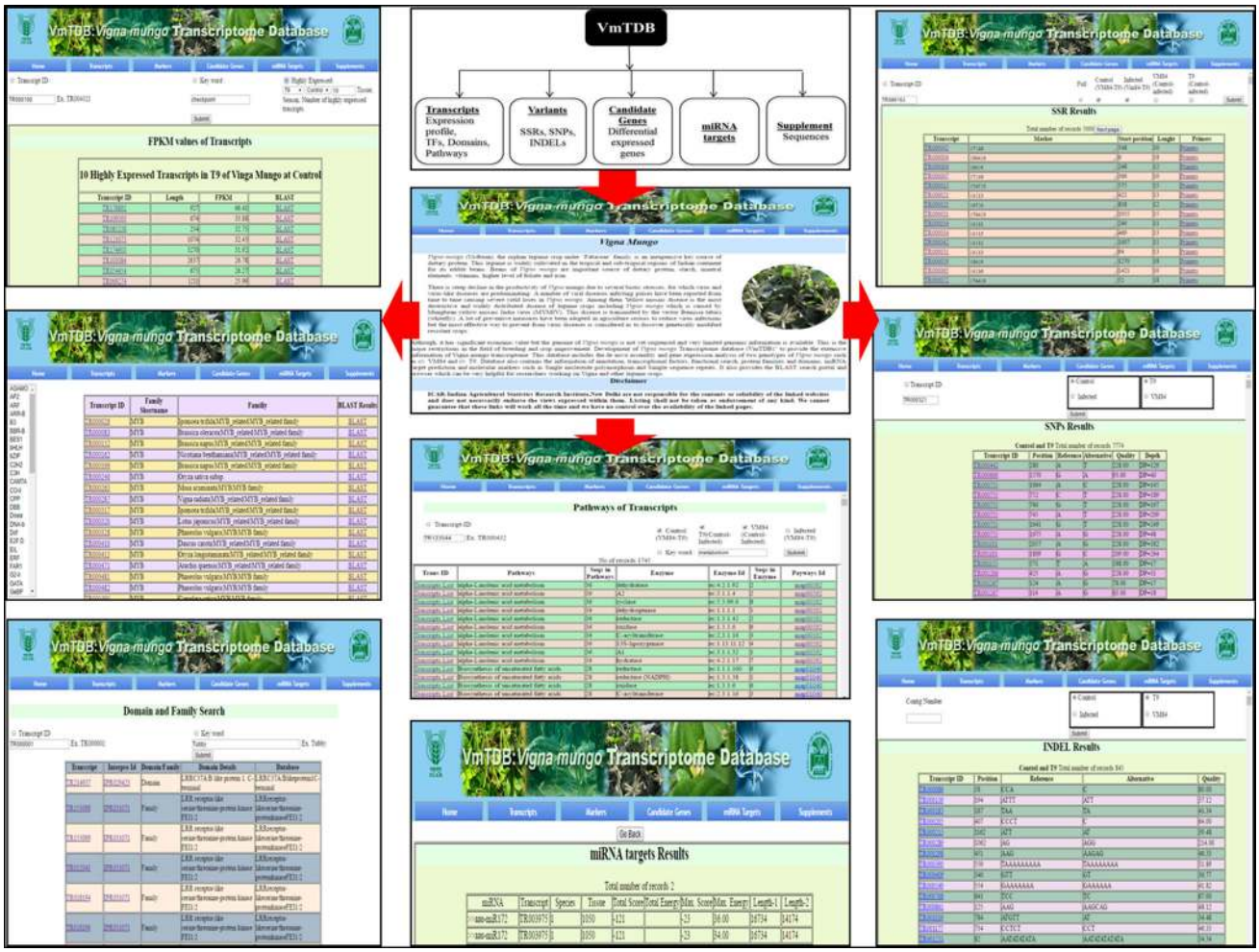

Fig. 6 Flow diagram of *V. mungo* Transcriptome database (VmTDB)

all datasets and 137 pathways. Genic region marker discovery revealed 44,335 SSRs and 109,305 SNPs and Indels. Obtained genomic resources have been populated in a database which is freely accessible for researchers and molecular breeders. The listed candidate and pathway genes can be targeted for future SNP discovery programs and this database would be a supplementing tool in varietal improvement program of this valuable nutritional health crop in the endeavour of accelerated genetic gain.

Acknowledgements We are thankful to Director, ICAR-IASRI, Indian Council of Agricultural Research, Ministry of Agriculture, Government of India, India for providing financial and infrastructural support to carry out this research.

Author's contribution DK, MAI, SJ, PKY and RSJ conceived theme of the study. RSJ, MAI and SJ did the computational analysis of generated data. RSJ, UBA and MAI contributed to database development. NK helped in SNP mining, RSJ, MAI, SJ drafted the manuscript. DK, MAI, SJ, PKY, AR edited the manuscript. All co-authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest Authors declare that there is no competing interest.

References

Alisoltania A, Ebrahimia S, Azariana S, Hematyara M, Shira B, Jahanbazic H, Fallahid H, Mousavi-Farde S, Rafieia F (2016) Parallel consideration of SSRs and differentially expressed genes under abiotic stress for targeted development of functional markers in almond and related *Prunus* species. Sci Hortic 198:462–472. doi:10.1016/j.scienta.2015.10.020

Allie F, Pierce EJ, Okoniewski MJ, Rey C (2014) Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. BMC Genom 15:1006. doi:10.1186/1471-2164-15-1006

Alves MS, Dadalto SP, Gonçalves AB, Souza GB, Barros VA, Fietto LG (2014) Transcription factor functional protein–protein interactions in plant defense responses. Proteomes 2:85–106. doi:10.3390/proteomes2010085

Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. Physiol Mol Biol Plants 19:307–321. doi:10.1007/s12298-013-0179-1

Basak J, Kundagrami S, Ghose TK, Pal A (2005) Development of Yellow Mosaic Virus (YMV) resistance linked DNA marker in *Vigna mungo* from populations segregating for YMV-reaction. Mol Breed 14:375–383. doi:10.1007/s11032-005-0238-6

Bhattacharyya U, Pandey SK, Dasgupta T (2014) Identification of EST-SSRs and FDM in sesame (*Sesamum indicum* L.) through data mining. Sch J Agric Sci 4:51–60

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. doi:10.1093/bioinformatics/btu170

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinform 10:421. doi:10.1186/1471-2105-10-421

Campos-Vega R, Loarca-Piña G, Oomah BD (2010) Minor components of pulses and their potential impact on human health. Food Res Int 43:461–482

Caplan JL, Zhu X, Mamillapalli P, Marathe R, Anandalakshmi R, Dinesh-Kumar SP (2009) Induced ER chaperones regulate a receptor-like kinase to mediate antiviral innate immune response in plants. Cell Host Microbe 6:457–469. doi:10.1016/j.chom.2009.10.005

Cheval C, Aldon D, Galaud JP, Ranty B (2013) Calcium/calmodulin-mediated regulation of plant immunity. Biochim Biophys Acta 1833:1766–1771. doi:10.1016/j.bbamcr.2013.01.031

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. doi:10.1093/bioinformatics/bti610

Czosnek H, Eybishtz A, Sade D, Gorovits R, Sobol I, Bejarano E, Rosas-Díaz T, Lozano-Durán R (2013) Discovering host genes involved in the infection by the tomato yellow leaf curl virus complex and in the establishment of resistance to the virus using tobacco rattle virus-based post transcriptional gene silencing. Viruses 5(3):998–1022. doi:10.3390/v5030998

Davies J, Roberts D, Eyer P, Buckley N, Eddleston M (2008) Hypotension in severe dimethoate self-poisoning. Clin Toxicol (Phila) 46:880–884. doi:10.1080/15563650802172063

DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 7:1243–1249. doi:10.1038/ni1410

Endo T, Kohda D (2002) Functions of outer membrane receptors in mitochondrial protein import. Biochim Biophys Acta 1592:3–14. doi:10.1016/S0167-4889(02)00259-8

Ganguli S, Dey A, Banik R, Kundu A, Pal A (2016) Analyses of MYMIV-induced transcriptome in *Vigna mungo* as revealed by next generation sequencing. Genom Data 7:226–228. doi:10.1016/j.gdata.2016.01.005

Garg G, Yadav S, Yadav G (2015) Key roles of calreticulin and calnexin proteins in plant perception under stress conditions: a review. Adv Life Sci 5:18–26. doi:10.5923/j.als.20150501.03

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3:408–414. doi:10.1038/nchembio.2007.5

Głowacki S, Macioszek VK, Kononowicz AK (2011) R proteins as fundamentals of plant innate immunity. Cell Mol Biol Lett 16:1–24. doi:10.2478/s11658-010-0024-2

Gongora-Castillo E, Ibarra-Laclette E, Trejo-Saavedra DL, Rivera-Bustamante RF (2012) Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (*Capsicum annuum*). Virol J 9:295. doi:10.1186/1743-422X-9-295

Guevara-Morato MA, de Lacoba MG, García-Luque I, Serra MT (2010) Characterization of a pathogenesis-related protein 4 (PR-4) induced in *Capsicum chinense* L3 plants with dual RNase and DNase activities. J Exp Bot 61:3259–3271. doi:10.1093/jxb/erq148

Gupta SK, Souframanien J, Gopalakrishna T (2008) Construction of a genetic linkage map of black gram, *Vigna mungo* (L.) *Hepper*, based on molecular markers and comparative studies. Genome 51:628–637. doi:10.1139/G08-050

Gupta S, Shukla R, Roy S, Sen N, Sharma A (2010) In silico SSR and FDM analysis through EST sequences in *Ocimum basilicum*. POJ 3:121–128

Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. doi:10.1038/nprot.2013.084

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G (2016) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. doi:10.1093/nar/gkw982

Kaewwongwal A, Kongjaimun A, Somtam P, Chankaew S, Yimram T, Srinives P (2015) Genetic diversity of the black gram [*Vigna mungo* (L.) *Hepper*] gene pool as revealed by SSR markers. Breed Sci 65:127–137. doi:10.1270/jsbbs.65.127

Kakati P, Deka SC, Kotoki D, Saikia S (2010) Effect of traditional methods of processing on the nutrient contents and some antinutritional factors in newly developed cultivars of green gram [*Vigna radiata* (L.) Wilezek] and black gram [*Vigna mungo* (L.) Hepper] of Assam, India. Int Food Res J 17:377–384

Kaur G, Somaiya R, Patel S (2015) Preventive and curative potential of *Vigna mungo* against metabolic syndrome in acute and chronic rat models. J Biol Sci 15:85–91. doi:10.3923/jbs.2015.85.91

Kaur N, Chen W, Zheng Y, Hasegawa DK, Ling K-S, Fei Z, Wintermantel WM (2017) Transcriptome analysis of the whitefly, *Bemisia tabaci* MEAM1 during feeding on tomato infected with the crinivirus, Tomato chlorosis virus, identifies a temporal shift in gene expression and differential regulation of novel orphan genes. BMC Genom 18:370. doi:10.1186/s12864-017-3751-1

Kogovšek P, Pompe-Novak M, Petek M, Fragner L, Weckwerth W, Gruden K (2016) Primary metabolism, phenylpropanoids and antioxidant pathways are regulated in potato as a response to potato virus Y infection. PLoS ONE 11:e0146135. doi:10.1371/journal.pone.0146135

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42(D1):D68–D73. doi:10.1093/nar/gkt1181

Kundagrami S, Basak J, Maiti S, Kundu A, Das B, Ghose TK, Pal A (2009) Agronomic, genetic and molecular characterization of MYMIV-tolerant mutant lines of *Vigna mungo*. Int J Plant Breed Genet 3:1–10. doi:10.3923/ijpbg.2009.1.10

Kundu A, Patel A, Paul S, Pal A (2015) Transcript dynamics at early stages of molecular interactions of MYMIV with resistant and susceptible genotypes of the leguminous host, *Vigna mungo*. PLoS ONE 10:e0124687. doi:10.1371/journal.pone.0124687

Kushwaha N, Sahu PP, Prasad M, Chakraborty S (2015) *Chilli leaf curl virus* infection highlights the differential expression of genes involved in protein homeostasis and defense in resistant chilli plants. Appl Microbiol Biotechnol 99(11):4757–4770

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi:10.1186/gb-2009-10-3-r25

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323. doi:10.1186/1471-2105-12-323

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Liu Q, Wang H, Zhu L, Hu H, Sun Y (2013) Genome-wide identification and analysis of miRNA-related single nucleotide polymorphisms (SNPs) in rice. Rice (N Y) 6:10. doi:10.1186/1939-8433-6-10

Liu J, McCleland M, Stawiski EW et al (2014) Integrated exome and transcriptome sequencing reveals ZAK isoform usage in gastric cancer. Nat Commun 5:3830. doi:10.1038/ncomms4830

Miozzi L, Napoli C, Sardo L, Accotto GP (2014) Transcriptomics of the interaction between the monopartite phloem-limited geminivirus tomato yellow leaf curl Sardinia virus and *Solanum lycopersicum* highlights a role for plant hormones, autophagy and plant immune system fine tuning during infection. PLoS ONE 9(2):e89951. doi:10.1371/journal.pone.0089951

Mishra PJ, Bertino JR (2009) MicroRNA polymorphisms: the future of pharmacogenomics, molecular epidemiology and individualized medicine. Pharmacogenomics 10:399–416. doi:10.2217/14622416.10.3.399

Naimuddin K, Akram M, Gupta S (2011) Identification of Mungbean yellow mosaic India virus infecting *Vigna mungo* var. *silvestris* L. Phytopathol Mediterr 50:94–100

Nakahara KS, Masuta C, Yamada S et al (2012) Tobacco calmodulin-like protein provides secondary defense by binding to and directing degradation of virus RNA silencing suppressors. Proc Natl Acad Sci USA 109:10113–10118. doi:10.1073/pnas.1201628109

Oh S-K, Lee S, Yu SH, Choi D (2005) Expression of a novel NAC domain-containing transcription factor (CaNAC1) is preferentially associated with incompatible interactions between chili pepper and pathogens. Planta 222(5):876–887

Pedley KF, Martin GB (2005) Role of mitogen-activated protein kinases in plant immunity. Curr Opin Plant Biol 8(5):541–547. doi:10.1016/j.pbi.2005.07.006

Qazi J, Ilyas M, Mansoor S, Briddon RW (2007) Legume yellow mosaic viruses: genetically isolated begomoviruses. Mol Plant Pathol 8:343–348. doi:10.1111/j.1364-3703.2007.00402.x

Reddy BVB, Obaiah S, Prasanthi L, Sivaprasad Y, Sujitha T, Krishna G (2015) Mungbean yellow mosaic India virus is associated with yellow mosaic disease of blackgram (*Vigna mungo* L.) in Andhra Pradesh, India. Arch Phytopathol Plant Prot 48:345–353. doi:10.1080/03235408.2014.888874

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. doi:10.1093/bioinformatics/btp616

Sahu J, Sarmah R, Dehury B, Sarma K, Sahoo S, Sahu M, Barooah M, Modi MK, Sen P (2012) Mining for SSRs and FDMs from expressed sequence tags of *Camellia sinensis*. Bioinformation 8:260–266. doi:10.6026/97320630008260

Scaria V, Hariharan M, Maiti S, Pillai B, Brahmachari SK (2006) Host–virus interaction: a new role for microRNAs. Retrovirology 3:68. doi:10.1186/1742-4690-3-68

Sharma OP, Bambawale OM, Gopali JB, Bhagat S, Yelshetty S, Singh SK, Anand R, Singh OM (2011) Field guide Mung bean and Urd bean. Department of agricultural and co-operation, NCIPM, ICAR, New Delhi, India http://www.ncipm.org.in/A3P/UI/HOME/PDF%20Files/Field%20Guide%20Mungbean%20&%20Urdbean.pdf. Accessed on 20 Sept 2016

Slaymaker DH, Navarre DA, Clark D, Pozo OD, Martin GB, Klessig DF (2002) The tobacco salicylic acid-binding protein 3 (SABP3) is the chloroplast carbonic anhydrase, which exhibits antioxidant activity and plays a role in the hypersensitive defense response. Proc Natl Acad Sci USA 99:11640–11645. doi:10.1073/pnas.182427699

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422. doi:10.1007/s00122-002-1031-0

Tranbarger TJ, Kluabmongkol W, Sangsrakru D, Morcillo F, Tregear JW, Tragoonrung S, Billotte N (2012) SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of *Elaeis guineensis*. BMC Plant Biol 12:1. doi:10.1186/1471-2229-12-1

Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, Eck HJV (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS ONE 8:e62355. doi:10.1371/journal.pone.0062355

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115. doi:10.1093/nar/gks596

Wang C, Ye J, Tang W, Liu Z, Zhu C, Wang M, Wan J (2013) Loop nucleotide polymorphism in a putative miRNA precursor associated with seed length in rice (*Oryza sativa* L.). Int J Biol Sci 9:578–586. doi:10.7150/ijbs.6357

Yu X, Sun S (2013) Comparing a few SNP calling algorithms using low-coverage sequencing data. BMC Bioinform 14:274. doi:10.1186/1471-2105-14-274