Genomics Data 7 (2016) 226–228

Contents lists available at ScienceDirect

Genomics Data

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

Lab Resource

Analyses of MYMIV-induced transcriptome in *Vigna mungo* as revealed by next generation sequencing

Sayak Ganguli [b,1], Avishek Dey [a,1], Rahul Banik [c], Anirban Kundu [a,2], Amita Pal [a,*]

[a] Division of Plant Biology, Bose Institute, Kolkata 700054, India
[b] Amplicon Biosciences, Kolkata, India
[c] The Biome, Kolkata, India

ARTICLE INFO

Article history:
Received 29 December 2015
Received in revised form 1 January 2016
Accepted 5 January 2016
Available online 7 January 2016

Keywords:
Vigna mungo
Transcriptome
Annotation
Recombinant inbred lines

ABSTRACT

Mungbean Yellow Mosaic Virus (MYMIV) is the viral pathogen that causes yellow mosaic disease to a number of legumes including *Vigna mungo*. VM84 is a recombinant inbred line resistant to MYMIV, developed in our laboratory through introgression of resistance trait from *V. mungo* line VM-1. Here we present the quality control passed transcriptome data of mock inoculated (control) and MYMIV-infected VM84, those have already been submitted in Sequence Read Archive (SRX1032950, SRX1082731) of NCBI. QC reports of FASTQ files generated by 'SeqQC V2.2' bioinformatics tool.

Resource table:	
Name of resource	VM84
Institution	Division of Plant Biology, Bose Institute, Kolkata 700054, India
Person who created resource	Sayak Ganguli, Avishek Dey, Rahul Banik, Anirban Kundu and Amita Pal
Contact person and email	Amita Pal, amita@jcbose.ac.in
Date archived/stock date	11.06.2015 and 07.07.2015
Type of resource	Recombinant inbred line of *Vigna mungo* resistant to MYMIV
Link to directly related literature that employed/validated this resource	Field Crop Res 135 (2012) 116–125
Information in public databases	• http://www.ncbi.nlm.nih.gov/sra/SRX1082731: Transcriptome Library of *Vigna mungo* RIL VM84 infected with MYMIV • http://www.ncbi.nlm.nih.gov/sra/SRX1032950: Transcriptome Library of mock inoculated *Vigna mungo* RIL VM84

1. Resource details

Yellow mosaic disease of blackgram (*Vigna mungo*) is caused by Mungbean yellow mosaic India virus (MYMIV). Irregular, chlorotic, yellow patches on the leaves indicate successful disease onset — the characteristic phenotype of MYMIV-infected susceptible plants. Cent percent yield loss occurs when MYMIV infects the host at the juvenile stage. MYMIV is transmitted through the whitefly, *Bemisia tabaci* Genn. [1]. It is one of the most devastating types of biotic stresses that causes up to 100% damage to a large number of leguminous crops. One candidate MYMIV resistance gene, CYR1, has been reported by Maiti et al. [2] and introgressed to develop several recombinant inbred lines (RILs) [3]. Here we report the transcriptome data of mock inoculated control and MYMIV infected resistant RIL, VM84.

1.1. Comparison of control and inoculated datasets based on reads and contigs

The total number of processed reads for the two samples was found to be 77.342016 for the control and 107.47377 million reads for the MYMIV inoculated cultivars, indicating a rise in approximately 30 million reads for the infected genotype; probably as a result of the expression of stress and defense pathway associated genes (Fig. 1). Following assembly of the reads into contigs this difference in expression between the control and the inoculated sets was found to be more evident as depicted in Fig. 2. However, average contig lengths

* Corresponding author.
E-mail address: amita@jcbose.ac.in (A. Pal).
[1] Both the authors contributed equally.
[2] Present address: Ramakrishna Mission Vivekananda Centenary College, Rahara, Kolkata 7000118, India.

http://dx.doi.org/10.1016/j.gdata.2016.01.005

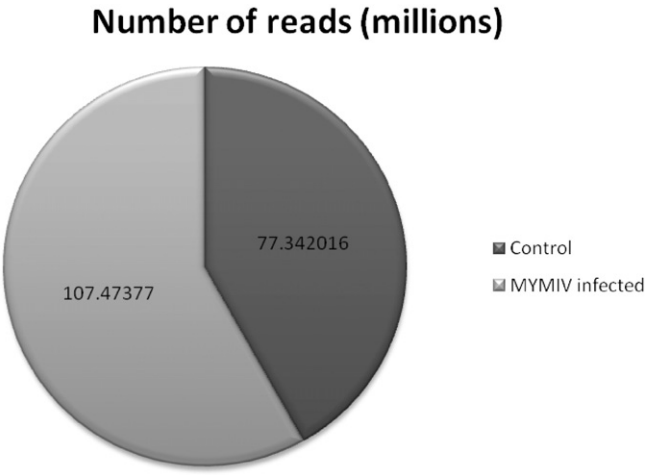

Fig. 1. Pie chart showing number of reads in control and MYMIV infected *Vigna mungo* RIL 84.

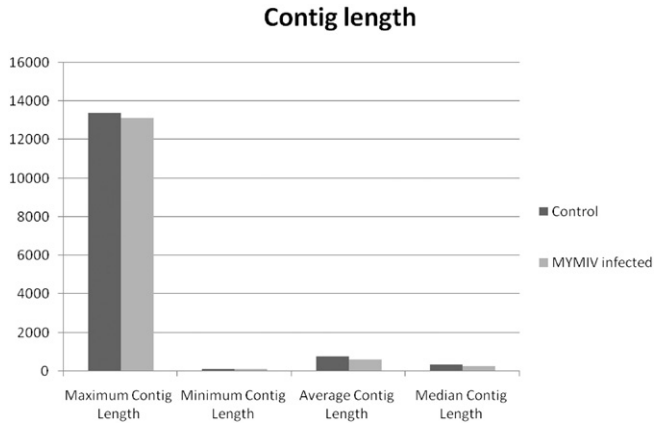

Fig. 3. Bar graph showing different length of contigs generated in control and MYMIV infected *Vigna mungo* RIL 84.

were found to be more or less proportional (Fig. 3), indicating that the difference in the number of contigs can be attributed to the differential expression of a few genes as well as expression of new genes as a result of infection and associated stress.

1.2. De-novo assembly and transcript generation

De-novo assembly of Illumina HiSeq2000 data was performed using velvet-1.2.10[2] and Oases_0.2.083 was used for transcript generation for various k-mers and concluded that hash lengths (k-mer) 55 (for control sample) and 57 (for MYMIV-infected sample) were better than others considering various parameters like the total number of transcripts generated, maximum transcript length, total transcript length and less number of N's. De-novo transcript statistics is presented in Table 1.

1.3. Transcripts annotation

In the absence of genomic information, *V. mungo* transcripts were annotated using the following databases:

i. *Medicago* Protein (Uniprot)
ii. *Soybean* Protein (Uniprot)
iii. *Cowpea* EST (NCBI).

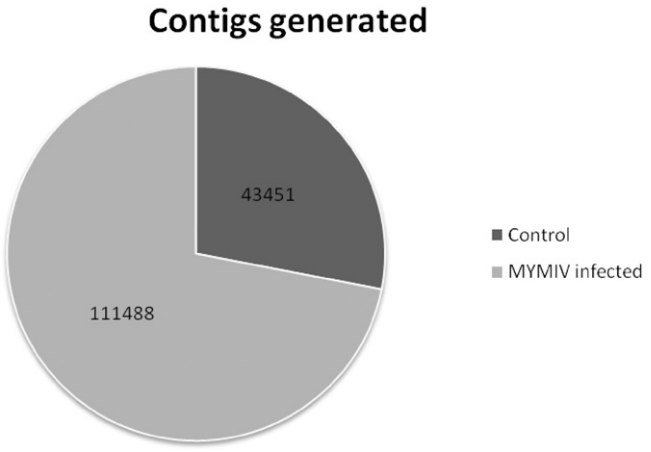

Fig. 2. Pie chart showing number of contigs generated in control and MYMIV infected *Vigna mungo* RIL 84.

The annotation statistics are shown in Table 2. Maximum transcript annotation was possible using soybean database.

2. Materials and methods

Plants samples (mock inoculated control and MYMIV infected *V. mungo*, line VM84) were collected and prepared following the method described by Kundu et al. [4].

Total RNA was extracted from control and infected leaves using Trizol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's protocol, followed by Dnase-I treatment (Sigma-Aldrich, USA) and purification in an RNeasy Plant Mini Kit (Qiagen, USA). Qualitative and quantitative assessments of the extracted RNA were done by an Agilent 2100 Bioanalyzer (RNA Nano Chip, Agilent). RNA samples were supplied to Genotypic Technologies Pvt. Ltd. (Bangalore, India) for preparation of transcript library and high throughput sequencing using Illumina HiSeq 2000 platform.

3. Verification and authentication

RNA sequencing has become a common method for analyses of functional plant genomics. Direct sequencing of mRNA provides a cost effective alternative to microarray technology for the analyses of gene expression for the entire transcriptome of a particular species [5]. Cell type specific transcript levels provide important research avenues for assessing the exact range of reads per sample for analyzing differential gene expression [6]. It was claimed that depth of coverage is directly

Table 1
De-novo *V. mungo* transcripts statistics.

Transcript statistics	Control sample	Infected sample
k-mer	55	57
Transcripts generated	49,720	103,842
Maximum transcript length	15,357	23,005
Minimum transcript length	200	200
Average transcript length	1688.2	1375
Median transcript length	939	3422.5
Total transcripts length	83,938,205	142,778,942
Total number of non-ATGC characters	536	897
Percentage of non-ATGC characters	0.001	0.001
Transcripts > = 200 bp	49,720	103,842
Transcripts > = 500 bp	42,048	76,066
Transcripts > = 1 kbp	33,281	54,945
Transcripts > = 10 kbp	42	77
N50 value	2254	2031
Percentage of reads used	96.48	93.32

Table 2
Annotation statistics of *V. mungo* transcripts.

Annotation	Control sample	Infected sample
Total transcripts	49,720	103,842
Transcripts annotated with *Medicago* database	30,497	53,091
Transcripts annotated with Soybean database	36,280	61,661
Transcripts annotated with Cowpea EST database	16,884	17,188
Total annotated transcripts	37,723	64,154
Percentage of annotated transcripts	75.87	61.78

proportional to the identification of new genes [5,7,8]. Li et al. [9] has established, using a negative binomial model of variations, that $\log_2$ fold change of two or more decreased the number of replicates to a maximum of six for effective identification of differentially expressed genes.

Acknowledgments

This work was supported by the Council of Scientific and Industrial Research, New-Delhi, India for the Emeritus Scientist's Scheme to AP (Sanction No. 21(0884)/12/EMR-II). AD is thankful to CSIR for a Research Associateship. We are thankful to the Director, Bose Institute for providing all infrastructural facilities.

References

[1] T.K. Nariani, Yellow mosaic of mung (*Phaseolus aureus* L.). Ind. Phytopathol. 13 (1960) 24–29.
[2] S. Maiti, S. Paul, A. Pal, Isolation, characterization, and structure analysis of a non-TIR-NBS-LRR encoding candidate gene from MYMIV-resistant *Vigna mungo*. Mol. Biotechnol. 52 (2012) 217–233.
[3] A. Kundu, A. Pal, Identification and characterization of elite inbred lines with MYMIV-resistance in *Vigna mungo*. Field Crop Res. 135 (2012) 116–125.
[4] A. Kundu, A. Patel, S. Paul, A. Pal, Transcript dynamics at early stages of molecular interactions of MYMIV with resistant and susceptible genotypes of the leguminous host, *Vigna mungo*. PLoS One (April 17, 2015)http://dx.doi.org/10.1371/journal.pone.0124687.
[5] J.C. Marioni, C.E. Mason, S.M. Mane, M. Stephens, Y. Gilad, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18 (9) (2008) 1509–1517.
[6] B.J. Blencowe, S. Ahmad, L.J. Lee, Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. Genes Dev. 23 (2009) 1379–1386.
[7] M. Griffith, L.O. Griffith, J. Mwenifumbo, R. Goya, S.A. Morrissy, et al., Alternative expression analysis by RNA sequencing. Nat. Methods 7 (2010) 843–847.
[8] J.M. Toung, M. Morley, M. Li, V.G. Cheung, RNA-sequence analysis of human B-cells. Genome Res. 21 (2011) 991–998.
[9] J. Li, R. Tibshirani, Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat. Methods Med. Res. 22 (2013) 519–536.