Development of genic-markers for yellow mosaic virus and bruchid resistance traits in black gram (*Vigna mungo* (L.) Hepper)

By
AVI RAIZADA

LIFE01201404006

Bhabha Atomic Research center, Homi Bhabha national institute, Mumbai

A thesis submitted to the
Board of Studies in Life Sciences

In partial fulfillment of requirements
for the Degree of
DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE

December, 2020

Homi Bhabha National Institute[1]

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Avi Raizada entitled <"Development of genic-markers for yellow mosaic virus and bruchid resistance traits in black gram (*Vigna mungo* (L.) Hepper)" > and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman - Name & Signature with date Dr. P. Suprasanna 16/02/2021

Guide / Convener - Name & Signature with date Dr. J. Souframanien 17/2/2021

Co-guide - Name & Signature with date (if any) -

Examiner - Name & Signature with date Dr. P. Jayamani 15/2/2021

Member 1- Name & Signature with date Dr. P.S. Variyar Prasad.S.V 17/2/2021

Member 2- Name & Signature with date Dr. Anand Ballal 17/2/2021

Member 3- Name & Signature with date

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 14-02-2021

Place: BARC, Mumbai

Signature
Co-guide (if any)

Signature 18/2/2021
Guide

[1] This page is to be included only for final submission after successful completion of viva voce.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Avi Raizada

Name & Signature
of the student

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Avi Raizada

Name & Signature
of the student

List of Publications arising from the thesis

Journal

1. Transcriptome sequencing, de novo assembly, characterization of wild accession of blackgram (*Vigna mungo* var. *silvestris*) as a rich resource for development of molecular markers and validation of SNPs by high resolution melting (HRM) analysis. Avi Raizada and Souframanien J. BMC Plant Biology, **2019**, 19., 358 DOI10.1186/s12870-019-1954-0.
2. An improved method for rapid isolation of DNA and RNA from polyphenolics rich leaves, flowers and roots of blackgram [*Vigna mungo* (L.) Hepper] for detection of begomovirus infection and RT-PCR. Avi Raizada and Souframanien J. Electronic Journal of Plant Breeding, **2019**, 10.,167-176. DOI: 10.5958/0975-928X.2019.00020.6.
3. Cross species amplification of mungbean derived resistance gene-SSR markers and analysis of genetic variations in blackgram [*Vigna mungo* (L.) Hepper]. Avi Raizada and Souframanien J. Genetic resources and crop evolution. 2020, https://doi.org/10.1007/s10722-020-01064-6.

Communicated:

- Comparative Transcriptomic Analysis Revealed Complex Molecular Mechanisms Underlying Pests, Pathogens Resistance and Seed Development in Wild and Cultivated Blackgram. Avi Raizada and Souframanien J. Scientific Reports (Under review)
- Genic-Single nucleotide polymorphism high resolution melting (SNP-HRM) analysis for genetic variation analysis in blackgram [*Vigna mungo* (L.) Hepper]

germplasm. Avi Raizada and Souframanien J. Crop genetic resources and evolution (Under review)

- Development and Mapping of Genic-markers Linked to Bruchid Resistance Quantitative Trait Loci (QTLs) in Blackgram (*Vigna mungo* (L.) Hepper). Avi Raizada and Souframanien J. Molecular Breeding (Under Review).
- Draft genome sequence of the pulse crop blackgram [*Vigna mungo* (L.) Hepper] reveals potential R-genes. Souframanien J and Avi Raizada. Scientific Reports (Under Review)
- Map Based Cloning of a Dominant Rust Resistance Gene and Mapping its Duplicated Paralogues in Cultivated Groundnut (*Arachis hypogaea* L.) Suvendu Mondal, Pallavi Bhat, Avi Raizada, K. Mohamed Shafi, Song Hui, Anand M. Badigannavar, Ramanatha Sowdhamini. (Under preparation)

Chapters in books and lectures notes

Conferences

1. "Development of SNP markers from transcriptome sequence of black gram (*Vigna mungo* L. Hepper)" presented at National symposium on nutritional security Conference in IIPR Kanpur, India (2017).
2. "Development of e-SSR markers from transcriptome sequence of black gram (*Vigna mungo* (L.) Hepper)" presented at International conference -Crops Genomics: Present and future, CEG in ICRISAT Hyderabad, India (2017).
3. "Exploiting geminivirus-plant host interacting gene(s) for developing transcriptomic markers associated with Yellow mosaic disease resistance in black

gram (*Vigna mungo*)" presented at First National Genetic Congress (NGC)- ICAR-IARI, Delhi, India (2018).

4. Comparative transcriptome analysis reveals genes associated with bruchid resistance in wild black gram" presented at First National Genetic Congress (NGC)- ICAR-IARI, Delhi, India (2018).
5. "Development of genic-SNP markers from transcriptome sequence of wild black gram (*Vigna mungo* var. *silvestris*)" presented at National symposium on Frontiers in Sustainable Agriculture by DAE-BRNS in BARC, Mumbai, India (2018).
6. "Development of Genic-SNP markers from transcriptome dataset and genotyping by High Resolution Melting (HRM) in black gram (*Vigna mungo* L. Hepper)" presented at National Conference on Recent Trends in Biotechnology and Bioprospecting in VISM, Gwalior, India (2018).

Others

Avi Raizada

Name & Signature of the student

DEDICATIONS

(Optional)

Dedicated to Param Purush Purandhani Hujur Data Dayal Maharaj Ji, My Parents, Shiny, Lavi, Leo, Dr. J. Souframanien and Dr. Shikha Sharma

ACKNOWLEDGEMENTS

(Optional)

I am using this opportunity to express my deepest gratitude to my guide Dr. Jegadeesan Souframanien for his valuable guidance, constructive criticism and generous advice throughout the project work. Without his supervision and constant support this project would not have been possible. I would like to express my gratitude to all the members of my doctoral committee namely Dr. P Suprasanna , Dr. P. Variyar and Dr. Anand Ballal for their suggestions and critical evaluation.

It is my pleasure to express sincere thanks to Dr. P Suprasanna for giving valuable help and suggestions at all stages of work. I am also very thankful to Dr. P. Variyar and Dr. Anand Ballal for their constant encouragement. I am genuinely grateful to all the members of NABTD; Dr. J. G. Manjaya, Dr. Sudhir Gupta, Dr. Sudhir Singh, Dr.Suvendu Mondal, Mr. Gautam Vishwakarma, Dr. Archana Joshi Saha, Dr. V. J. Dhole, Dr. P. Dhanskekar and Mr. Pandri for their invaluable help and sincere support. I am also sincerely thankful to Dr.Suvendu Mondal, Dr. Sudhir Singh and Mr. Gautam Vishwakarma for helping in carrying out molecular biology work done in this thesis. I would like to express my deep thanks to Dr. Suman Bakshi for her constant support and cooperation. Heartiest thanks are due for Prashanti Patel for her continuous support throughout the project. Special thanks to my friends Dr. Shikha Sharma, Anjali, Sayli and Shridhar for their love and invaluable help.

I am totally indebted to my family for everything. No words of acknowledgment will be sufficient for them.

Contents

List of Figures

List of Tables

Chapter 5

Summary and future perspective

Pulses are one of the most important dietary components of human diet. These are rich sources of proteins and thus, are important for vegetarian population in fulfilling their protein needs. Among the different pulses, blackgram (*Vigna mungo*) is one of the most popular pulse, (also known as king of pulses) and contains easily digestible non-flatulence proteins. Currently, blackgram is gaining international significance and is being grown in countries such as Thailand, Japan etc., for various purposes. Besides being nutritional, it is also recognized for its medicinal significance for being antihyperglycemic (treatment of type 2 diabetes) , anti-cardiovascular and anti-cancerous. Blackgram is extensively grown in Asian countries, of which India is the largest producer of blackgram. Blackgram production is very much affected by biotic and abiotic stresses. Among biotic constraints, yellow mosaic disease (YMD) and pests (bruchid) attack are the most common, causing a huge loss to standing crop and stored seeds, respectively. YMD is caused by a geminivirus and transmitted by whitefly. While bruchids (*Callosobruchus* species) are storage pests causing damage to stored seeds. Chemical control methods are not economical and have concerns associated with food and environment safety. Thus, host-plant resistance (HPR) can be a viable option for durable protection from pathogens and pests. Very limited molecular research has been conducted in blackgram which included development of 3 genetic linkage maps, few YMD linked markers, identification and mapping of bruchid resistance associated QTLs and few transcriptomic studies on YMV invasion and bruchid egg deposition. These preliminary studies laid a foundation for further research and future directions are needed in blackgram to accelerate blackgram varietal development programmes. Further, additional constraints in blackgram improvement programmes include lack of genome sequence, genotype specific YMD resistance linked markers, difficult to use bruchid QTLs markers, lack of bruchid resistance sources and numerous virus strains breaking resistance of elite cultivars. Hence complete and clear understanding of molecular mechanism for resistance

along with the development of genomic resources are needed for accelerating blackgram breeding programmes. Wild progenitor of blackgram (*Vigna mungo* var. *silvestris*), Trombay wild, TW (INGR10133) is a resistance source for bruchid infestation (*C. maculatus*) with antibiosis mechanism. Till date, no genic-markers were reported for blackgram and thus in the present work, TW is molecularly characterized through RNA sequencing and a preliminary idea on YMD resistance mechanism is also established.

Genomic resources for blackgram were developed in the form of transcriptome dataset and markers were derived from the same. The transcriptome sequencing of Trombay wild (TW) generated 2.97 GB data accounted by 2,970,339,385 bases corresponding to 19,690,124 HQ reads (SRR 5931432, SRX3091690). After assembly of HQ reads, 40,178 transcript contigs (TCs) were obtained which were further analyzed by bioinformatics tools. A total of 38,753 CDS were predicted from 40,178 TCs by ORF Finder that showed similarities with *Phaseolus vulgaris, Glycine max* and *Citrus clementina* when annotated by BLASTx algorithm (NCBI), GO and KEGG. A total of 1621 simple sequence repeats (SSRs) were identified in 1339 TCs with a frequency of one SSR per 11.1 kb. PCR primer pairs were designed for 1171 SSR loci. An amplification rate of 58% was achieved. HQ reads alignment of TW and TU94-2 blackgram resulted in identification of 1844 SNPs. Positional distribution showed the presence of 1291 SNPs in CDS and 518 in UTRs. The transition (Ts) type SNPs were 1129 and transversion (Tv) type SNPs were 716 with a Ts/Tv ratio of 1.57 for TW and the most abundant SNP being of A/G type. PCR primers were designed for 1749 SNP loci. HRM assay was standarized in blackgram for SNP genotyping with a validation rate of 78.8%. SNPs were also classified into 4 classes based on Tm difference in HRM assay and non-synonymous SNPs were analysed *insilico* for predicting SNP effect on protein kinetics by I-TASSER. To the best of our knowledge, this is the first report on transcriptome characterization, markers development and standardization of HRM assay for

SNP genotyping of wild accession of blackgram. Besides these markers, genomic-SSR primers flanking RGAs on mungbean scaffolds were designed from mungbean genome that showed 70% transferability in blackgram. In addition, SSR and SNP markers developed in this study (31 SSR, 19 SNP and 30 mungbean SSR) were deployed in genetic variation analysis of blackgram germplasm. Host cellular factors involved in geminivirus infection cycle were targeted for marker development. A total of 177 genic-primers based on cellular proteins were designed from TU94-2 blackgram contigs and tested for their amplification in 4 blackgram genotypes by PCR.

Previously reported YMD linked molecular markers in blackgram are not tightly linked and are genotype specific. No marker has been reported for the YMD resistant cultivar KU96-3 grown in northern region. For targeting specific genes related to YMD resistance trait, several strategies were deployed. Characterization of YMD resistant TU94-2 transcriptome (48,291 contigs) for resistance gene analogues using DRAGO 2 pipeline resulted in identification of 392 contigs containing R gene domains. A total of 341 host cellular factors with redundancy (including isoforms) involved in geminivirus infection cycle (based on literature survey) were used for designing 177 gene specific primers. This is the first report on targetting cellular factors for YMD linked marker development in blackgram. In initial screening (Parental survey), a total of 610 primers designed in this study were used, resulting in 50% PCR amplification. Out of 153 polymorphic markers, only 3 markers (TWSSR10, VGMSSR86 and VGMSSR547) showed desired amplification pattern in BSA followed by screening in 218 RILs of mapping population. Phenotyping for YMD reaction was conducted at Trombay in 2 different seasons using infector row method and RILs showed segregation in 1:1 ratio that was confirmed by chi-square test. Linkage analysis using JOINMAP software showed linkage of 1 genic-SSR marker (TWSSR10) with YMD resistance trait at a distance of 36.7 cM. Validation results in blackgram germplasm for these

genic-SSR markers were similar to that observed for previously reported YMD marker CEDG180. Gene function annotation for these marker sequences using NCBI nr database showed homology of TWSSR10 (TW contig 2139) and VMgSSR86 (TU94-2 Contig 1833) marker sequences, respectively with transcription factor TCP9 that positively regulates isochorismate synthase 1 of salicylic acid synthesis pathway and with formin-like protein 2 that has been hypothesized for role in pathogen perception. Formins binds plasma membrane-localized phosphoinisotides and Rho-guanosine triphophatases (GTPases) which are involved in plant-pathogen-associated processes, including pathogen entry and immune activation. Also, on mungbean genome, resistance genes TMV Resistance gene N and enhanced disease resistance 2 were found flanking to VMgSSR86 and VMgSSR547 marker sequences, respectively. These results suggest that, these 3 genic-SSR markers may not be directly linked to YMD resistance gene(s) but may be involved in defense response pathway and associated with QTLs for YMD resistance trait. To get an insight into the molecular mechanism operating in YMD resistant KU96-3 cultivar in response to YMV invasion, expression pattern of 16 genes with significant roles in geminivirus infection cycle were analysed by qRT-PCR. In resistant background under infected condition, 10 genes showed upregulation which included transcription factor MYB1R1 and WD repeat-containing protein DWA2, replication protein A (RPA), LRR-containing protein DDB, Tobamovirus multiplication protein 1, LRR-RLK, autophagy-related protein 8C, argonaute 1 (AGO1), eukaryotic translation initiation factor 4B3 and Importin subunit alpha-2. Among these, autophagy-related protein 8C, argonaute 1 (AGO1) and importin subunit alpha-2 are the most important ones which indicate that the resistance or defense response could be due to activation of autophagosome mediated degradation of viral proteins, RNA silencing of target viral transcripts or genome and importin mediating virus particles movement in host plant. While transcripts for 6 genes which were otherwise reported to be upregulated in other

studies, were found downregulated in this study in resistant background that included DNA repair protein RAD4, Allene oxide cyclase 4, SUPPRESSOR OF GENE SILENCING 3, RNA-directed DNA methylation 3, RNA-directed DNA methylation 4 and Eukaryotic translation initiation factor 4E (eIF4E). The low expression of these genes may be due to the fact that the study was conducted at early stage of infection while the genes could be activated at later stages of infection.

Bruchid pests pose a huge troll on blackgram seeds during storage. No bruchid resistant variety has been reported in blackgram till date. Trombay wild urd (INGR10133) (TW) is a well known source of bruchid resistance in blackgram native to trombay hills and used in the present study. As of now, it is known that in TW, the resistance trait is under the control of two dominant duplicate genes and the QTLs associated with bruchid resistance have been identified and mapped. The resistance mechanism has been reported to be larval antibiosis which is constitutively expressed as dried seeds showed resistance to bruchids infestation. But biochemical basis of resistance and markers linked to QTLs which can be of practical use have not been reported. RNA sequencing results for TW developing seeds revealed upregulation of bruchid resistance and other defense related genes such as LRR-RLKs, acid phosphatase, 7S globulins, vicilins, thaumatin, miraculin, thioredoxin, hydroperoxide lyase (HPL, CYP74B). These upregulated defense related genes have been reported to confer resistance to bruchids in other studies and few of them have been validated successfully by qRT-PCR in this study. For bruchid resistance linked marker development, of the 434 SSR primers screened in parental survey, 235 and 106 primers showed amplification and polymorphism (48 VMgSSR, 34 TWSSR and 24 MRGSSR), respectively. A total of 33 polymorphic primers were used for screening the mapping population of 104 RILs. Reconstruction of genetic linkage map using genotypic data of these 33 new genic markers and 428 previously reported markers of earlier study (Gupta et al., 2008) [4] resulted in mapping of 11 markers in the reconstructed map. Comparative analysis of

reconstructed map and previously reported map showed high similarity between both with very few differences. The sequences of molecular markers in both the linkage maps were similar except at few locations. New genic-SSR markers that were found located near to the QTLs are VMgSSR61 and TWSSR13 for QTL Cmrdp1.1, TWSSR66 for QTL Cmrdp1.2, VMgSSR44 for QTL Cmrdp1.3, TWSSR62 for QTL Cmrdp1.4 and MRGSSR54 and VMgSSR9 markers for the QTL Cmrdp 1.6. Marker validation in germplasm showed 100% discrimination of resistant TW from all susceptible genotypes for markers TWSSR14, TWSSR68 and TWSSR87. SSR markers TWSSR13 and TWSSR16 showed >90% discrimination ability. To identify the candidate genes for resistance, genic-SSR marker sequences were searched for homology with adzuki bean genome. The marker TWSSR167 showed homology with vicilin-like seed storage protein which is reported for conferring resistance to bruchids. Although marker TWSSR14 sequence showed homology with homeobox-leucine zipper protein ATHB-13 which is not related to bruchid resistance, but defence related genes were found in upstream and downstream regions to this marker sequence in adzukibean genome such as beta-D-glucanase, endochitinase, TMV resistance protein N-like and LRR repeat protein and receptors. Similarly, TWSSR15 marker showed similarity with vacuolar-sorting receptor 1-like, but the flanking regions on adzukibean genome showed presence of defense related proteins such as wound-induced protein-like, thaumatin-like protein, galactosidase and LRR receptor-like protein kinase.

Based on these studies on genomic resources development, molecular markers and resistance mechanism for YMD and bruchid resistance, the following important conclusions were drawn:

From wild accession of blackgram, transcriptome dataset was generated and 3350 genic-markers were developed.

1. Protocol for SNP genotyping by HRM assay was standardized in blackgram.
2. A total of 2305 DEGs were identified by RNA sequencing of wild and cultivated blackgram immature seeds and RNAseq results were validated by qRT-PCR.
3. One genic-SSR marker (TWSSR10) derived from wild transcriptome was found linked to YMD trait at a distance of 36.7 cM.
4. In YMD resistant KU96-3 cultivar, RNA silencing and autophagy based molecular resistance mechanism has been speculated.
5. A total of 4 new genic-SSR markers were mapped in proximities to bruchid resistance QTLs in reconstructed genetic linkage map of blackgram.
6. RNA sequencing of wild blackgram developing seeds and gene expression analysis (qRT-PCR) showed anti-insect compounds such as acid phosphatase, vicilin, trypsin inhibitor and miraculin to be responsible for bruchid resistance in wild blackgram.

Future perspectives

1. Transcriptome dataset and genic-markers developed in this study will aid in accelerating molecular research in blackgram and is expected to enhance efficiency of marker assisted selection (MAS) in blackgram breeding programme.
2. RNAseq analysis resulted in identification of DEGs related to YMD and bruchid resistance whose function could be studied by transgenics.
3. The genic-markers developed for bruchid resistance could be useful for MAS.
4. Nucleotide variations present in DEGs related to YMD and bruchid resistance can be exploited for marker development.

Summary

Black gram (*Vigna mungo* L. *Hepper*) is an important pulse crop those production is majorly affected by biotic stresses particularly yellow mosaic disease (YMD) and bruchid pest infestation. Progress in black gram improvement programmes is getting hampered due to lack of genomic resources and hence, limited molecular research. To date research updates constitutes few genotype specific YMD linked markers, limited transcriptomic studies to study YMD resistance mechanism and biochemical basis of bruchid resistance is not yet worked out. Only one study reported transcript dynamics of black gram upon bruchid oviposition on developing seeds. In this study, transcriptome dataset of wild black gram (TW) was generated by NGS technology (RNA-Seq) and used for developing genic-markers. A total of 2.97 GB data was produced, 19,690,124 HQ reads were obtained, 38,753 (96.4%) CDS and 40,178 TCS were predicted. Transcriptome dataset was characterized for CDS, ORF, SSR, SNP through several bioinformatics tools and databases such as NCBI nr protein database, Blast2GO, KEGG database. A total of 1621 SSRs in 1339 TCS and 1844 SNPs were identified. PCR primers were designed for 1171 SSR and 1749 SNP loci and an amplification rate of 58% and 85% were achieved. SNP genotyping was performed with a validation rate of 78.8% by HRM assay. In total 31 SSR markers (transcriptome derived) and 19 SNP markers were used to study genetic variations in 27 different black gram genotypes. Besides genic-markers, 118 genomic SSR-markers were developed from mungbean genome scaffolds that harboured RGAs and a transferability rate of 70% was observed in 44 different black gram genotypes. The marker TWSSR10 was mapped at a distance of 36.7 cM from YMD trait and was found to have homology with the transcription factor TCP9 that is known to be involved in salicylic acid biosynthesis. Total 2306 DEGs (CDS) including 1116 upregulated and 1190 downregulated were identified from TW and TU94-2 transcriptomes. RNA-Seq of TW and TU94-2 developing seeds showed upregulation of insecticidal and autophagy related transcripts that were validated by qRT-PCR. Gene expression studies in KU96-3 (YMD resistant) and TAU-1 (YMD susceptible) genotypes under YMV infected conditions showed high expression of autophagy related genes and argonaute. New genic-SSR markers in close proximity to bruchid resistance QTLs were developed and mapped on reconstructed linkage map. Besides generating huge genomic resources, this study gives a preliminary idea on YMD resistance mechanism that might be activated in KU96-3 cultivar upon YMV invasion, hints on biochemical basis of bruchid resistance in wild black gram and identified genic-SSR markers that could be helpful in bruchid resistance breeding programmes.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Preamble

Food is one of the basic necessities of life and essential for sustenance of life. Food provides energy and nourishment. All kinds of foods come directly or indirectly from plants, of which pulses contribute considerable amount and are majorly taken along with cereals.

1.1. Overview

Blackgram (*Vigna mungo*) is an important pulse crop of Asian countries known for its protein rich seeds. In India, it is cultivated in an area of about 3.26 mha with a production of 1.74 mt (Gupta, 2012)[1]. Blackgram production is primarily hampered as a result of narrow genetic base and losses due to biotic and abiotic stresses. Yellow mosaic disease (YMD) and pests (bruchids) pose a heavy toll to standing crop and stored seeds, respectively. YMD caused by Geminivirus, is transmitted by whitefly and results in an annual loss of more than US$300 million (Varma et al., 1992)[2], while storage pest bruchids (*Callosobruchus* species) cause damage to entire seed lot within a period of 3-4 months. Host-plant resistance (HPR) can be a viable alternative to chemical control methods for both pests and pathogen attacks. Blackgram is lagging behind in molecular research compared to other legumes and very limited progress has been made which includes development of 3 genetic linkage maps based on Simple sequence repeats (SSR), Random amplified polymorphic DNA (RAPD), Amplified fragment length polymorphism (AFLP), Inter simple sequence repeats (ISSR) and Single nucleotide polymorphism (SNP) markers (Chaitieng et al., 2006; Gupta et al., 2008; Somta et al., 2019)[3,4,5], few YMD linked markers: VMYR1, YR4, CYR1 (Basak et al., 2005; Maiti et al., 2012)[6,7], CEDG180, CEDG185, VR9 (Gupta et al., 2013; Rambabu et al., 2018; Naik et al., 2017)[8,9,10], YMV1 (Souframanien and Gopalakrishna, 2006)[11], identification of bruchid resistance associated QTLs (Souframanien et al., 2010) [12] and few transcriptomic studies upon YMV invasion and bruchid egg deposition (Ganguli et al., 2016; Baruah et al., 2017; Kundu et al., 2019)[13,14,15]. The probable reason for slow

development in blackgram molecular breeding programmes could be attributed to the lack of whole genome sequence. Moreover, the reported YMD resistance linked markers are genotype specific with loose linkage to the resistance gene. The reported markers for bruchid quantitative trait loci (QTL) are based on AFLP, ISSR, RAPD which are impractical for routine large scale screening. This is compounded by the lack of bruchid resistance sources and rapid evolution of viruses causing breaking of resistance in current elite cultivars. Therefore, to develop varieties with long-term durable resistance to different strains of pathogens and pests, exploration of new YMD and bruchid resistance sources, complete and clear understanding of molecular mechanism of resistance, along with development of genomic resources are needed. Wild progenitor of blackgram (*Vigna mungo* var. *silvestris*), Trombay wild, TW (INGR10133) has been reported to harbour genetic resistance against bruchid infestation (*C. maculatus*) with antibiosis resistance mechanism (Souframanien and Gopalakrishna, 2007)[16]. Next generation sequencing technologies such as RNA-sequencing provide opportunities for exploring global gene networks and molecular mechanisms underlying different traits, plant responses to stimuli (Baruah et al., 2017; Kundu et al., 2019) [14,15] along with generation of genomic resources for EST/genic markers development which have the inherent advantage of direct gene tagging. Till date, no genic-markers have been reported in blackgram as compared to other *Vigna* species such as mungbean and cowpea. The present study was, therefore, proposed to enrich available genomic resources of blackgram for enhancing molecular research and was undertaken with the following objectives:

Objective 1. Development of genic-markers from transcriptome sequences

Objective 2. Identifying DNA marker(s) linked to yellow mosaic disease resistance gene

Objective 3. Studying differential gene expression between bruchid resistant and susceptible genotypes

1.2. Blackgram [*Vigna mungo* (L.) Hepper]

Blackgram (*Vigna mungo* (L.) Hepper) is a self-pollinating diploid (2n=2x=22) legume crop with genome size of 574 Mbp (0.56pg/1C). It belongs to the family *Leguminosae* or *Fabaceae* and placed under the genus *Vigna* and subgenus *Ceratotropis* (Gupta et al., 2008)[4]. It originated from *Vigna mungo* var. *silvestris* and is a native of India (Vavilov, 1926; Lukoki et al., 1980)[17,18]. It is extensively cultivated in different regions of Asia. Blackgram is known as the "king of the pulses" due to its delicious taste, low flatulence, readily digestible proteins and richness of many nutritional components. Blackgram is composed of proteins (25-26%), carbohydrates (60%), fat (1.5%), minerals, amino acids, vitamins, folate and iron (Sharma et al., 2011)[19]. The high lysine level in the protein makes it an ideal supplement to cereals and also a good source of proteins (Kakati et al., 2010)[20]. Interestingly, some medicinal properties were also reported for blackgram which are antihyperlipidemic, antihyperglycemic relevant in treatment of type 2 diabetes (Kaur et al., 2015) [21] and anti-cardiovascular, anti-cancerous properties (Campos-Vega et al., 2010)[22]. It also acts as a mini-fertilizer depository by maintaining and restoring soil fertility through atmospheric nitrogen fixation due to symbiotic association with *Rhizobium* bacteria. It is a short duration crop (Delic et al., 2009) [23] with flowering time of 30-50 days and maturing time of 60-90 days after sowing. Blackgram production is hindered by numerous factors such as lack of genetic variability, poor harvest index, absence of suitable ideotypes for different cropping systems (Souframanien et al., 2017) [24] and susceptibility to various biotic stresses such as YMD and bruchid pests (Naimuddin et al., 2011; Duraimurugan et al., 2011) [25,26] and abiotic stresses such as salinity, drought etc (Sahoo and Jaiwal, 2008)[27].

1.3. Yellow mosaic disease (YMD)

Yellow Mosaic Disease (YMD) is the most common and devastating viral disease of legumes caused by several strains of yellow mosaic virus and transmitted by whitefly (*Bemisia tabaci*). YMD occurs in several legumes such as blackgram (*Vigna mungo*), soybean (*Glycine max*), mungbean (*Vigna radiata*), frenchbean (*Phaseolus vulgaris*) and mothbean (*Vigna aconitifolia*) (Ramesh et al., 2017; Dikshit et al., 2020) [28,29]. Percent yield loss ranged from 5 to 100% depending upon time of infection, disease severity, susceptibility of cultivars, population of whitefly and developmental stage of crop at which infection occurred (Rathi, 2002) [30]. *Mungbean yellow mosaic virus* (MYMV) and *Mungbean yellow mosaic India virus* (MYMIV) has been reported to cause up to 85–100% yield loss (Varma and Malathi, 2003)[31].

Taxonomy, classification, host range and distribution of YMV

Viruses are major plant pathogens which constitute nearly 47% of plants infectious diseases (Anderson et al., 2004) [32] with more than 15 families reported, of which *Geminiviridae* is the second largest family. The *Geminiviridae* family includes a group of circular ssDNA viruses that affect monocots, dicots and weeds (Navas-Castillo et al., 2011)[33]. The International Committee on Taxonomy of Viruses has classified the *Geminiviridae* family into nine genera: *Becurtovirus, Begomovirus, Capulavirus, Curtovirus, Eragrovirus, Grablovirus, Mastrevirus, Topocuvirus*, and *Turncurtovirus* based on insect vector, genome organization, host range, and genome-wide pairwise sequence identities (Zerbini et al., 2017)[34]. These viruses can be monopartite or bipartite, present in the old or new world or can occur in both worlds with *Mastrevirus, Curtovirus* and *Begomovirus* being well characterized genera (Rojas et al., 2005) [35]. Bipartite genomes occurred as DNA-A and DNA-B components and mostly monopartite genomes were found associated with alphasatellites or betasatellites. YMV belongs to the genus *Begomovirus* which includes

members that are transmitted by whiteflies, infect dicotyledonous plants, and may be either bipartite or monopartite. The *Begomovirus* is the largest genus consisting of more than 180 species with several unassigned isolates (Fauquet et al., 2008)[36]. These caused most devastating plant diseases like leaf curls in cotton, pepper and tomatoes, mosaic and yellow mosaic of cassava, pulses and beans. YMD is predominant in the South-East Asian countries (Basak et al., 2005)[6]. Virus particles were first observed by Thongmeearkom et al., (1981)[37] and purified by Honda et al., (1983) [38]. In India two YMD causing viral species were reported, MYMIV found predominantly in northern, central, eastern part and MYMV confined to peninsular region of India (Haq et al., 2011) [39] which can easily be distinguished based on the nucleotide sequence identity (Fauquet et al., 2003)[40]. The viruses causing YMD of legumes across southern Asia are classified into four types: Mungbean yellow mosaic virus (MYMV), Mungbean yellow mosaic India virus (MYMIV), Horsegram yellow mosaic virus (HgMV) and Dolichos yellow mosaic virus (DYMV) and are collectively referred to as legume yellow mosaic viruses (LYMVs). International commodity trade, intercontinental transportation networks and a changing global climate are the major reasons for rapid spread of geminiviruses and its vector whitefly (Moffat, 1999)[41].

Geminiviruses genome, proteins and infection cycle

Geminiviral particles are 18–20 nm in diameter, 30 nm long, apparently consisting of two incomplete icosahedra joined together in a structure with 22 pentameric capsomeres and 110 identical protein subunits (Qazi et al., 2007)[42]. Geminivirus have circular single-stranded DNA (ssDNA) genomes enveloped by protein capsid to form geminate particles and double-stranded DNA (dsDNA) as replicative form which gets transcribed in the nucleus of infected plant cells (Rojas et al., 2005)[35]. Members of the genus *Begomovirus* have monopartite (one ~2.9 kb DNA) or bipartite genome (two ~2.6 kb DNAs referred to as "DNA-A" and "DNA-B"), are transmitted by whiteflies (e.g. *Bemicia tabaci* Gennadius), and infect

dicotyledonous plants. *Geminivirus* genomes are highly prone to mutations, recombination and reassortment thus resulting in wide diversity (Duffy and Holmes, 2008)[43]. Genomes are comprised of the origin of replication in a 5′ intergenic region, promoters for RNA polymerase II and several different transcription units. Generally, geminiviruses have bipartite genome organization with DNA-A and DNA-B components with a 180±200 nt common region (CR). This CR contains the viral DNA replication origin as well as the promoter for the leftward ORFs (ALs). All geminiviruse's replication origins contain a stretch of 31 bp with a characteristic invariant nanomer TAATATTAC and a characteristic secondary structure with GC-rich stem and an AT-rich loop. Geminivirus genomes encode 5-7 proteins with multi and diverse functions which are as follows: Coat protein (CP) for viral capsid, replication initiator protein (Rep) and replication enhancer protein (REn) or C3 for viral replication (Settlage et al., 2001)[44], V2, AV2, TrAP, C2, C4 and AC4 as TGS and PTGS suppressors (Trinks et al., 2005; Raja et al., 2008; Amin et al., 2011) [45,46,47] and movement proteins NSP and MP for virions movement and spread (Sanderfoot and Lazarowitz, 1996)[48]. In monopartite begomovirus, betasatellites encode βC1 and alphasatellites encode Rep, both being RNA silencing suppressors (Yang et al., 2011)[49]. After entry into the cell, viral genome gets replicated in nucleus which further can be (1) used as templates for another round of replication or transcription, (2) wrapped by viral movement proteins for transportation from the infected cell to adjacent cells through plasmodesmata (PD), or (3) encapsulated into infectious virions for long-distance virus transmission (Hanley-Bowdoin et al., 2013)[50]. Due to limited coding potential, they rely heavily on host proteins, thus manipulating host cellular machinery for their infection cycle (Hanley-Bowdoin et al., 2013)[50]. Few viral proteins also function in impairing RNA silencing mechanisms by preventing small RNA generation and suppress several components of TGS and PTGS

(Hanley-Bowdoin et al., 2013; Gnanasekaran et al., 2019) [50, 51] which are also associated with symptom development (Basu et al., 2018)[52].

Symptomology and epidemiology

Symptoms caused by YMVs are largely dependent on host species and susceptibility. The virus causes a range of symptoms, which include yellow mosaic spots, less flowering and pod development, necrosis, small and distorted pods with immature small sized seeds.In blackgram, there are two types of yellow mosaic symptoms depending upon the variety: 'yellow mottle' (generalized yellowing of the leaves) and 'necrotic mottle' (yellowing restricted to small spots which become necrotic) (Nair and Nene, 1974)[53]. The disease incidence depends on environment temperature and relative humidity which may directly influence vector population and its migration. The whitefly population increased with increase in temperature and relative humidity, whereas heavy rains and strong winds are detrimental to whiteflies.

Prevention and control of YMD

Control is based mainly on preventing the establishment of the whitefly vector population in the crop through application of insecticides but may have undesirable effects on health and environmental safety. Moreover, changes in agricultural practices, such as moving the cropping period out of times of high vector incidence (the wet period in late summer) to times of low vector incidence (dry season in early summer) can have short-term benefits. However, for long-term control of disease, natural host plant resistance is the best option although the available sources of resistance in most legume crops are limited.

1.4. YMV linked molecular markers reported in blackgram and marker validation

In 1980s, the advent of DNA-marker technology has dramatically accelerated the efficiency of plant breeding and its application in marker assisted genotype selection for several resistance genes (Kumar et al., 2011)[54]. Marker assisted selection (MAS) is a rapid,

economical and effective breeding approach for developing resistant cultivars of crops. Moreover, MAS is of paramount importance for pyramiding resistance genes where one gene may mask the action of another that makes the selection of plants with multiple resistance genes difficult. Thus, through molecular markers individual genes can be identified for gene pyramiding without pathogen inoculation and/or progeny testing. MAS has gained importance in blackgram due to lack of uniform field screening procedure, complex interaction among viruses, vectors and host, practical difficulties in creating artificial epiphytotic conditions and dependency on natural environmental conditions for YMD occurrence (Souframanien and Gopalakrishna, 2006; Sowmini and jayamani, 2014)[11,55]. Till date several molecular markers linked with YMD resistance trait have been developed in blackgram and few have been validated. Resistance gene analog (RGA) based tightly linked markers (Soybean, Kanazin et al., 1996)[56] 'VMYR1' (primer-RGA 1F-CG/RGA 1R), and CYR1 were developed for YMV tolerant genotype (natural mutant of YMV susceptible T9), its derived tolerant lines VM1 and VM4 (Basak et al., 2005) [6] and OBG-31 (YMV resistant) genotype (Panigrahi et al., 2016)[57], respectively. Marker VMYR1 DNA fragment showed homology with the NB-ARC domain which is a characteristic feature of disease resistance genes and other genes involved in nematode cell death and human apoptotic signaling (Basak et al., 2005)[6]. Remarkably, the difference in YMD reaction between blackgram cv.T9 and its natural mutant was presumed to be due to defect in the NB-ARC domain of putative disease resistance (R) gene (Kundagrami et al., 2009)[58]. Later on validation results showed complete linkage of CYR1 to the MYMIV resistance gene in resistant genotype AKU9904 (Maiti et al., 2012)[7]. But consistent susceptible reaction of AKU9904 to MYMIV at Kanpur over the years (Anjum Tuba et al., 2010) [59] restricted its use in breeding and ultimately its utility in mapping the MYMIV resistant gene. SSR based markers CEDG180, CEDG185 and VR9 were found tightly linked to YMD resistance in

blackgram genotypes DPU 88-31 (12.9 cM, MYMIV resistant, Gupta et al., 2013)[8], T9 (MYMV resistant, Rambabu et al., 2018) [9] and T9 (Naik et al., 2017)[10], respectively. ISSR ($ISSR811_{1357}$) sequence derived SCAR marker 'YMV1' was mapped at a distance of 6.8 cM in MYMV resistant genotype TU94-2 (Souframanien and Gopalakrishna, 2006)[11]. Different marker types (SSR, AFLP, RAPD and ISSR) from different crops were deployed for genetic variation analysis in blackgram accessions (Souframanien and Gopalakrishna, 2004; Gupta and Gopalakrishna, 2009) [60,61]. RAPD (OPN-OPF primers) and RAPD derived SCAR marker 20F/20R (developed from OPQ1 RAPD primer) were screened in germplasm lines and PU31 was identified as a useful variety for the development of markers linked to MYMV, UCLV (Urd bean leaf crinkle virus), wilt and powdery mildew resistance (Vishalakshi et al., 2017; Prasanthi et al., 2013; Chandrajini Devi et al., 2017)[62,63,64]. Three markers *viz.*, YR4, CYR1 and $SCAR_{ISSR\ 811}$ were found to be partially linked with YMD resistance trait in 14 blackgram genotypes (Sowmini and Jayamani, 2014)[55]. Although YMD linked markers were developed in blackgram but their numbers are very few and target resistance genes in specific mapping populations. These linked markers, thus may not be polymorphic in other genetic backgrounds and hence may not be suitable for MAS (Bernardo et al., 2013)[65]. Therefore, there is a need to explore diverse resistance sources for developing markers linked to different resistance genes followed by their validation in different genetic backgrounds. Till date three genetic linkage maps have been constructed for blackgram: first map with 148 marker loci (Chaitieng et al., 2006)[3], second map comprised of 428 markers (Gupta et al., 2008)[4] and third is a highly saturated map recently developed from 3675 SNPs identified by next generation sequencing (Somta et al., 2019)[5]. The comparison of blackgram map with adzuki bean map (Han et al., 2005) [66] revealed high level of conservation of the marker loci between the two genomes.

1.5. Genetics underlying YMD resistance in blackgram

Inheritance pattern of YMD trait is conflicting because of different results observed by researchers which could be due to different resistance sources and different viral species studied (Table 1.1).

Table 1.1. Details of blackgram genotype crosses and genetics of YMD resistance

Parent/cross	Genetics of resistance	Reference
-	Digenic recessive	Singh, 1980; Verma and Singh, 1986 [67,68]
-	Monogenic dominant	Kaushal and Singh, 1988; Gupta et al., 2005 [69,70]
Blackgram crosses	Monogenic recessive	Pal et al., 1991[71]
KMG189(R); VBN(Gg)2 (S)	Monogenic recessive	Basak et al. 2005; Sai et al., 2017 [6,72]
Co5 x VBN(Bg)4, Co5 x VBG66	Digenic and Trigenic dominant	Murugan and Nadarajan, 2012 [73]
-	Digenic dominant	Durga prasad et al., 2015[74]
MDU-1 x MASH-114, MDU-1 x VBN(Bg)6, MDU x PU31, MDU1 x Uttara, LPG752 x Mash-114, LPG752 x VBN(Bg)6, CO6 x VBN(Bg)6	Digenic dominant with epistasis	Thamodhran et al., 2016[75]
V. radiata x V. radiata var. Sublobata	Dominant and recessive epistasis	Singh and Sharma, 1983[76]
Wide cross of blackgram	Digenic recessive	Dwivedi and Singh, 1985 [77]
Mungbean x blackgram Mungbean x V. sublobata	Digenic recessive	Pal et al., 1991[71]
TNAU RED x VRM(Gg)1	Monogenic recessive	Sudha et al., 2013 [78]
V. radiata x V. umbellata	Major QTL	Mathivathana et al., 2019 [79]

R-Resistant; S-Susceptible

1.6. Plant host-virus pathogen interactions

Plants live in complex environments where they are exposed to numerous types of microbial pathogens, herbivorous insects and environmental changes. Plant resistance to viruses can be either non-host specific or host specific. Non-host specific resistance is common, robust and durable characterized by resistance to all strains of a viral species while host specific resistance is resistance of some genotypes of a plant species to particular viral pathogen which is otherwise susceptible to that viral pathogen. Host specific resistance is controlled by genes and can be either complete (inhibition of virus replication or propagation) or partial (less/delayed virus accumulation). To combat viruses, plants have evolved a tightly regulated multilayered defense system (Carr et al., 2019)[80].

1.6.1. Innate immunity: PAMP-triggered immunity (PTI) and Effector-triggered immunity (ETI)

Innate immunity of plants comprised of two different but interconnected branches, termed pathogen associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) (Jones and Dangl, 2006)[81]. PTI is the plant's primary response to microbes through non-specific recognition of conserved invariant pathogen- or microbe-associated molecular patterns (PAMPs/MAMPs) by plant cell-surface receptors. Series of cellular processes in PTI are MAP kinase signaling, transcriptional induction of pathogen-responsive genes, production of reactive oxygen species (ROS), synthesis of antimicrobials and deposition of callose to reinforce the cell wall at sites of infection, to prevent microbial growth (Jones and Dangl, 2006)[81]. ETI comes into play when some pathogens escape PTI through secretion of effector molecules in host apoplast or cytosol which are not recognizable by PTI but perceived by transmembrane or intracellular resistance (R) proteins, thus triggering ETI which is specific to pathogen effectors (Jones and Dangl, 2006)[81]. R genes/proteins are the ETI components whose activation suppresses microbial growth

through hypersensitive response (HR) a type of programmed cell death (PCD), which restricts the multiplication of the pathogen at the infection site (Muthamilarasan and Prasad, 2013; Ronde et al., 2014)[82,83]. ETI induced defense response comprised of initial ROS production, HR, followed by plant hormone production such as salicylic acid (SA), jasmonic acid (JA) or ethylene and further activation of pathogenesis related (PR) proteins. R genes can be expressed as single dominant genes having narrow recognition specificity. R proteins bear characteristic domains/motifs such as nucleotide-binding site leucine rich repeats (NBS-LRR), CC-NB-LRR (CC: coiled coil) and TIR-NB-LRR (TIR: Toll interleukin receptor) which are expressed dominantly such as N gene in tobacco against tobacco mosaic virus (TMV) (Whitham et al., 1996) [84] or as recessive alleles of genes that are critical for cellular processes such as Mo 1 from *Lactuca sativa* against lettuce mosaic virus (LMV) (Nicaise et al., 2003) [85]. Numerous dominant and recessive resistance genes have been identified for RNA viruses especially potyviruses and carmovirus but for geminivirus, resistance gene Ty-1 has been identified which code for RNA-dependent RNA polymerase (Voorburg et al., 2020)[86].

1.6.2. Geminivirus interaction with plant host receptors and cellular proteins

Numerous studies showed interaction of geminivirus proteins with membrane localized and cytosolic receptors and other signaling molecules which are described in Table 1.2.

Table 1.2. Details of geminivirus-host interactions at cellular level

Geminiviral protein	Interacting host cellular factor	References
Tomato yellow leaf curl virus(TYLCV)-C4	BARELY ANY MERISTEM (BAM) 1 and 2	Rosas-Diaz et al., 2018 [87]
beet curly top virus (BCTV) C4	LRR receptor like kinase	Piroux et al., 2007[88]
cabbage leaf curl virus (CaLCuV) NSP	LRR receptor like kinase	Piroux et al., 2007[88]
Tomato yellow leaf curl Sardinia virus (TYLCSV)	CLAVATA1-related receptor-like kinase	Lozano et al., 2011[89]
CaLCuV Rep/TrAP/C2	GRIK, SnRK1 activating kinase	Shen and Hanley, 2006[90]
Tomato golden mosaic virus (TGMV)	SnRK1	Hao et al., 2003[91]; Shen et al., 2011[92]
BCTV, TGMV and TYLCSV C4	Shaggy-related kinase (AtSK2/SlSK)	Lozano et al., 2011; Piroux et al., 2007[88,89]
TYLCSV	SNF1-related kinase 2 (SnRK2.1)	Lozano et al., 2011[89]
CaLCuV	Plastid-nuclear signalling	Trejo-Saavedra et al., 2009[93]
CabLCV and TGMV TrAP	Transcription initiation factor Y4B (TIFY4B)	Chung and Sunter, 2014 [94]
TYLCV REn	SlNAC1/ATF1	Selth et al., 2005[95]
Wheat dwarf virus (WDV) and TYLCSV Rep A	GRAB2	Lozano et al., 2011[89]
TGMV and CaLCuVTrAP	PEAPOD2 (Transcriptional activator of CP promoter)	Lacatus and Sunter, 2009[96]
TYLCSV TrAP/C2	JDK	Lozano et al., 2011[89]
Tomato yellow leaf curl China virus (TYLCCNV) BetaC1	AS1	Yang et al., 2008[97]
TYLCV	heat shock protein 90 (HSP90) and suppressor of the G2 allele of Skp1 (SGT1)	Moshe et al., 2016[98]
TYLCV-V2	papain-like cysteine protease	Bar-Ziv et al., 2015[99]

1.6.3. Plant host defenses against geminivirus and geminivirus counter actions

With time plants and viruses both co-evolved parallelly in order to combat each other defense responses.

1.6.3.1. RNA Silencing

RNA silencing is an evolutionary conserved mechanism of gene expression regulation occurring at two levels: transcriptional gene silencing (TGS) through DNA methylation and

post-transcriptional gene silencing (PTGS) through RNA degradation or translational repression mediated by small RNAs (Vaucheret and Fagard, 2001) [100]. Plants deployed RNA silencing as a main defense strategy against viruses (Kong et al., 2018) [101]. RNA silencing RNA silencing involved production of viral double-stranded RNA (dsRNA) by the host RNA-dependent RNA polymerase (RDR) which can be considered as Virus-Associated Molecular Pattern (VAMPs) (Ruiz-Ferrer and Voinnet, 2009) [102]. These viral dsRNA are processed by Dicerlike (DCL) enzymes into virus-derived small interfering RNAs (vsiRNAs) and loaded onto the RNA silencing induced complex (RISC) by argonaute (AGOs). This act as RNA guide for suppressing viral transcription and replication (Ketting, 2011) [103]. In *Arabidopsis*, DCL3, DCL4, DCL2, AGO1, AGO2, AGO4, AGO5, AGO7, AGO10, RDR1, RDR2 and RDR6 are involved in virus-induced RNA silencing (Carbonell and Carrington, 2015; Wang et al., 2010) [104, 105]. Moreover, short distance (cell-to-cell) and long distance (e.g., phloem) spreading of silencing signals for establishing systemic antiviral immunity requires coordinated activity of AGO1/AGO2, DCL2, RDR6 and SGS3 (SUPPRESSOR OF GENE SILENCING 3) (Taochy et al., 2017) [106]. Interestingly, it was observed that PTGS and TGS mediated by endogenous sRNAs are key regulators of PTI and ETI (Sánchez et al., 2016) [107]. Geminivirus are inducers of RNA silencing owing to the bidirectional transcription of their genomes, aberrant read-through transcription and the overlapping ORFs which results in the production of double-stranded RNA (dsRNA)—a key instigator of the RNA silencing pathways (Shunmugiah et al., 2017) [108]. During geminivirus infection DCL2, DCL3 and DCL4 produced small interfering (siRNA) of 21, 24 and 22 nt-, respectively (Vanderschuren et al., 2007) [109]. The 21–22 nt siRNAs target viral coding regions and 24 nt siRNAs were found to be targeting IR of the viral genome (Akbergenov et al., 2006) [110]. Moreover, RNA silencing signals are amplified resulting in production of secondary siRNAs by host RNA polymerases IV and V and RNA dependent RNA

polymerase 2 (RDR2) leading to the DCL3-mediated production of 24 nt siRNAs (Wang et al., 2011) [111]. Although direct evidence for RDR2 in geminivirus infection is yet to be elucidated (Rodríguez-Negrete et al., 2009) [112], mutations involving RDR6 cause a little increase in viral DNA, suggesting that secondary siRNAs might play a vital role in the defense against viral infection (Wang et al., 2011) [111]. Transcriptional suppression and replication inhibition of geminivirus genome is mediated by DNA methylation such as cytosine methylation and histone modification (H3K9) (Pooggin and Hohn, 2003) [113]. Plant mutants of methylation pathway such as chromatin remodelers (*ddm1*), *AGO4* and RNA polymerase IV subunit (*nrpd2a*) displayed greater susceptibility to viral infection such as BCTV and CaLCuV infections. Moreover, mutants lacking non-CG methyl transferases (*drm1*/*drm2* and *cmt3*), H3K9 methyl transferases (*kyp2*/*suvh4*) and adenosine kinase (*adk1* and *adk2*) showed susceptibility to viruses.

1.6.3.2. Geminivirus suppress host RNA silencing

To counteract plant RNA silencing defense, some geminiviral proteins suppress RNA silencing and thus are named as viral suppressors of RNA silencing (VSRs). VSRs antagonize the RNA-silencing pathway at several steps such as (i) impairment of viral siRNA biogenesis by inhibiting DCL proteins and/or the activity of cofactors, (ii) sequestration of dsRNA/siRNA, (iii) promotion of AGO protein destabilization prior to RISC formation or (iv) transportation of the mobile silencing signal into the peroxisomes to disable plant defence (Incarbone et al., 2017) [114]. TYLCCNB-βC1 protein leads to sequesteration of calmodulin-like protein (rgs-CaM) to the nucleus, thus possibly preventing its degradation (Chung et al., 2014) [115]. This leads to rgs-CAM interference in RNA silencing mechanism at two stages: (1) by reducing the expression of RNA-dependent RNA polymerase 6 (NbRDR6) and (2) by interacting with SGS3 leading to its degradation through an autophagy pathway (Li et al., 2017) [116]. Hence, both induction of rgs-CaM and silencing of autophagy

related genes lead to enhanced susceptibility to geminivirus infection. Geminiviruses interfere with the proper functioning of the cellular methyl cycle at several steps (Yang et al., 2011; Zhang et al., 2011)[49,117]. Interference in cellular methyl cycle included transcriptional repression of DNA methyltransferases (Li et al., 2014) [118], inactivation of RDR2 (Hanley-Bowdoin et al., 2013)[50], up-regulation of host Werner exonuclease-like 1 (WEL1) (Trinks et al., 2005)[45], inactivation of SnRK1 and adenosine kinase (ADK) (Buchmann et al., 2009) [119]. Several studies worked out the mechanism of geminiviral VSRs for suppressing RNA silencing. TGMV-AL2 was found to inactivates serine/threonine kinase (SNF1) (Hao et al., 2003) [91] and ADK (Wang et al., 2005) [120], while BCTV-L2 inactivated ADK (Wang et al., 2005) [120]. TYLCCNV-C1/ TYLCCNB-bC1 inhibited *S*-adenosylhomocysteine hydrolase (SAHH) (Yang et al., 2011) [49] and BCTV C2/TrAP attenuated the degradation of SAM decarboxylase 1 (SAMDC1) (Zhang et al., 2011) [117]. The *Cotton leaf curl Multan virus* (CLCuMuV) C4 inhibited SAM synthetase (SAMS) (Ismayil et al., 2018) [121] while, C2/TrAP from *Beet severe curly top virus* (BSCTV), TGMV, and *Cabbage leaf curl virus* (CaLCuV) inhibited H3K9 histone methyltransferase SUVH4/KYP (Castillo-González et al., 2015) [122]. TYLCV V2 interacted with histone deacetylase HDA6 (Wang et al., 2018) [123] while, MYMV-AC4 suppressed the systemic phase of silencing through interaction with BAM1 which is a positive regulator of the cell-to-cell movement of RNAi (Carluccio et al., 2018) [124]. TYLCV-C4 protein interacted with BAM 1 and 2, which are Pm localized receptor like kinases (RLK) to hinder the cell-to-cell spread of RNA silencing (Rosas-Diaz et al., 2018)[87]. Thus, inhibition of the methyl cycle by geminiviruses poses a serious challenge to the engineering of RNA silencing mediated virus resistance. CabLCV-NSP redirected ASYMMETRIC LEAVES (AS) 2 from the nucleus to the cytoplasm to accelerate the decapping activity of DCP2 which indirectly inhibited accumulation of virus-derived small RNA (Ye et al., 2015) [125]. BCTV-C4/

EACMCV-AC4 protein suppressed the systemic phase of RNA silencing and restored replicational competency causing extensive hyperplasia (Mills-Lujan et al., 2015) [126]. These viral proteins act through plasmamembrane localization of C4-shaggy like kinases (SKs) (C4/AtSK complexes) which requires both N-myristoylation and palmitoylation for localization and stability of AC4-SK complexes at PM (Fondong et al., 2007; Piroux et al., 2007)[88, 127]. Some VSRs such as TYLCV V2 suppresses PTGS through interaction with SGS3 and/or competion for binding to dsRNA substrates (Fukunaga and Doudna, 2009) [128]. Recently, it was demonstrated that the TYLCV Ty-1 and Ty-3 resistance genes code for RDR which showed homology to *Arabidopsis* RDR3/4/5 (Verlaan et al., 2013) [129] and might be involved in vsiRNA biogenesis (Li et al., 2014)[118]. In turn plant hosts employed 'counter-counter defence' to down regulate suppression of RNA silencing by VSRs through their degradation. For instance, RING E3 ligase protein caused ubiquitination mediated degradation of the TYLCCV beta-C1(Li et al., 2014)[118].

1.6.3.3. Autophagy and programmed cell death (PCD)

Apart from major RNA silencing defense strategy, plants also deployed autophagy as antiviral defense to degrade viral virulence factors during geminivirus infection (Choi et al., 2018) [130]. In anti-viral autophagy, adaptor proteins ATG8, NBR1 (neighbor of BRCA1 gene 1), and Beclin1 recruits viral particles or viral proteins and guide the cargos to autophagosomes for degradation (Ding et al., 2018; Li et al., 2018) [131, 132]. Exportin 1 (XPO1) mediate degradation of TLCYnV-C1 by ATG8h (Li et al., 2020) [133]. The geminivirus C1-interacting kinases (GRIKs) are upstream activating kinases of SnRK1 (Shen et al., 2009) [134] which regulates autophagy (Soto-Burgos and Bassham, 2017) [135]. Therefore, it is possible that the C1 protein interacts with GRIKs, leading to the induction of autophagy by activating SnRK1. Activation of autophagy related genes upon geminivirus infection was observed for instance, induction of ATG8, ATG9, ATG12 against TYLCSV

and CabLCV (Reviewed by Bu et al., 2020) [136]. Autophagy can play both anti-viral and pro-viral roles in viral life cycle through targeting viral components for lysosomal degradation (i.e. xenophagy) or by negatively regulating PCD to prevent cell death for sustaining virus cycle (Kushwaha et al., 2019) [137]. Upon geminivirus attack, cell death response is hijackedto combat anti-viral autophagy by the expression of senescence- and cell death/HR-related transcripts without inducing a visible cell death phenotype (Ascencio-Ibáñez et al., 2008) [138]. For example in TYLCV infection, HSP90, suppressor of the G2 allele of Skp1 (SGT1) (Moshe et al., 2016) [98] gets affected, papain-like cysteine protease and two protein kinases (mitogen-activated protein kinase 4-MAPK4 and MAPK kinase 2-MKK2) gets inhibited (Bar-Ziv et al., 2015) [99]. While in CabLCV infection, up-regulation of WRKY40 and the Bax inhibitor (AtBI-1) leads to down-regulation of Bax-induced cell death (Watanabe and Lam, 2006) [139].

1.6.3.4. Ubiquitin proteasome system (UPS)

The ubiquitin proteasome system (UPS) comprising of ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin-ligase (E3) have been reported for interaction and degradation of geminiviral proteins through polyubiquitination followed by 26S proteasome degradation (Mandadi and Scholthof, 2013) [140]. For instance RING finger protein mediates the degradation of the CLCuMuB-bC1 protein through the ubiquitin proteasomal pathway (Shen et al., 2016) [141]. Several reports claimed interaction between host UPS machinery and viral proteins such as TGMV and TYLCSV Rep interaction with SUMO-conjugating enzyme 1 (SCE1) (Sánchez-Durán, 2011) [142] and C2 protein interaction with the SKP1 (Mandadi and Scholthof, 2013) [140]. Apart from its anti-viral feature, host UPS promote virus cycle through interaction of UPS components with viral proteins such as CLCuMuB-bC1 interaction with E2 ubiquitin-conjugating enzyme (UBC3) and S-phase kinase-associated protein 1 (SKP1) (Jia et al., 2016) [143]. Also,

phosphorylation influenced the stability of many viral proteins such as bC1 protein phosphorylation by tomato SlSnRK1 (Shen et al., 2011)[92]. Geminiviruses also exploit SUMO conjugating system to promote their multiplication by modulating host target proteins (Arroyo-Mateos et al., 2018) [144].

1.6.4. Geminivirus movement in plant host

Plant host proteins involved in protein stabilization and transport are known to enhance or restrict cell to cell or long-distance movement of geminiviral particles. For monopartite geminiviruses, CP plays a crucial role in viral DNA transportation by interacting with cellular transporters (Sharma and Ikegami, 2009) [145]. While for bipartite begomoviruses, movement proteins (NSP and MP) interact with importin (nuclear transporter) for viral movement and spread (Guerra-Peraza et al., 2005) [146]. CabLCV-NSP and TYLCSV NSP recruits acetyltransferase (NSI) and Transport GTPase (NIG) for movement (Lozano et al., 2011; Carvalho et al., 2008) [89,147]. Bean dwarf mosaic virus (BDMV)-NSP-MP bind to histone H3 to form a viral DNA-H3-NSP-MP complex in the nucleus that facilitates the transport of the viral movement complex to the cell periphery and plasmodesmata (PD) (Zhou et al., 2011) [148]. Abutilon mosaic virus (AbMV)-MP exploits cellular membrane flow to transfer the complex from the endoplasmic reticulum (ER) to PD and into adjacent cells (Zhang et al., 2002) [149]. Importin alpha, Heat shock protein cognate 70 (HSC70), SET7/9, H3K4 methyltransferase, coatomer delta subunit (deltaCOP) are reported to be involved in geminiviral movement (Chandran et al., 2012; Krenz et al., 2012) [150, 151]. Phosphorylated and myristoylated C4 protein of TYLCCNV interact with a nuclear cargo protein (exportin a, XPO I) (Mei et al., 2018a) [152] and N-myristoylated EACMMV-AC4 protein functions in transporting viral DNAs (Fondong et al., 2007) [127].

1.6.5. Transcript dynamics of blackgram upon YMV infection

Very few transcriptomic studies were conducted in blackgram upon YMV invasion which showed intensive transcriptional reprogramming at several cellular levels resulting in either rapid switching to a defence mode in resistant genotype (incompatible interaction) or symptom development in susceptible host (compatible interaction). Extensive transcriptomic studies were conducted on MYMIV resistant line VM84 and susceptible genotype T9.

In resistant background, upregulated transcripts included pattern recognition receptors, enzymes scavenging ROS, PR genes, secondary metabolites pathway genes, NBS-LRR protein, HSP90, SGT1 (R-gene complex), MAPK6, SA induced protein kinases (SIPK) which stimulated N-gene mediated resistance. Other upregulated factors are transcription factors families such as ZF, WD40, bHLH and WRKY, UPS components, PR genes such as PR1, PR5 and PR17, enzymes for production of antimicrobials compounds, phenylalanine lyase (PAL), SA, phytoalexins, proline rich cell wall precursor, glycoprotein, serine/glycine hydroxymethyl transferase etc (Ganguli et al., 2016; Kundu et al., 2019)[13, 15].

Besides coding transcripts, dynamics of non-coding transcripts i.e. microRNAs (miRNAs) were also studied in resistant VMR84 genotype (Paul et al., 2014) [153]. In resistant background VMR84, miRNAs belonging to the families of miR156, miR159, miR160, miR166, miR398, miR1511, miR1514, miR2118 and novel vmu-miRn7, vmu-miRn8, vmu-miRn13 and vmu-miRn14 were found highly altered which could target NB-LRR, NAC, MYB, Zinc finger, CCAAT-box transcription factor, fructose 2-6 bisphosphate and HDZIP protein (Kundu et al., 2017) [154].

Proteomics and biochemical studies supports transcriptional reports that increased expression of ROS regulators (PRX, SOD, TRX, GST and APOX), Ca2+ concentration, Ca2+-responders (calreticulins (CRTs), calmodulins (CAM) and calcium homeostasis regulator (CHoR1)) and MAP kinases were observed in resistant genotypes (Chakraborty and

Basak et al., 2018a; Chakraborty and Basak et al., 2018b) [155,156]. Also, long non-coding RNAs (LncRNAs), putative genic SSR markers, SNPs and Indels, mature miRNA targets were predicted in differentially expressed unigenes and pathways of VMR84 genotype and submitted to *V. mungo* Transcriptome database, accessible at http://webtom.cabgrid.res.in/vmtdb/ (Jasrotia et al., 2017) [157].

1.6.6. Establishment and restriction of geminivirus infection cycle in legumes

Geminiviruses redirect the differentiated cell to ectopic division and expansion for its own multiplication. Numerous host cellular proteins are reported for their direct or indirect interaction with geminiviral proteins which include RBR (retinoblastoma-related protein) (Gutierrez et al., 2004) [158], proliferating cell nuclear antigen (PCNA), replication factor C, minichromosome maintenance protein 2 (Hanley-Bowdoin et al., 2013)[50] and cell differentiation regulators (ATHB7, ATHB12, CYCD1;1, and TRN1) (Mei et al., 2018b) [159]. For example, interactions leading to enhanced cell divisions are BCTV-C4 stimulated degradation of cyclin kinase inhibitors (ICK/KRP) through RING finger protein (RKP) (Lai et al., 2009) [160], TYLCCNV-C4 increased stability of CYCD1:1 by impeding its degradation by shaggy-like kinases (Mei et al., 2018a,b) [152,159] and membrane-bound CLAVATA1 (CLV1) mediated repression of WUSCHEL (WUS) (cell differentiation regulator) due to its interaction with TYLCV-C4 (Li et al., 2018) [132]. TYLCSV REn and CP proteins interact with GLO1 (Glyoxylase pathway), SKL2 (Shikimate kinase), PERM (Permease I-like protein), LeHT1 (Hexose transporter), AT4CL1 (4-coumarate:CoA ligase) and AOC1 (Allene oxide cyclase) (Lozano et al., 2011)[89]. NSP (CabLCV, TGMV and CdTV) gets phosphorylated by a proline-rich extension like receptor kinase to function as a pro-viral host factor for these bipartite begomoviruses (Florentino et al., 2006) [161]. Chilli leaf curl virus (ChiLCV) enhanced transcription of its genes through H3K4me3 of its genome mediated by Rep interaction with histone H2B monoubquitination machinery (UBC2 and

HUB1) (Kushwaha et al., 2017) [162]. Geminiviruses activate auxin inducible genes to promote cell proliferation and modulate differentiation in plants (Park et al., 2004) [163] such as for CLCuD (Wang et al., 2003) [164]. Also overexpression of SnRK1 was found associated with the expression of several autophagy marker genes (Baena-Gonzalez et al., 2007) [165]. In tomatoes, an ATPase (RPT4) and a DEAD-box RNA helicase (DEAD35) inhibit the transcription of ToLCNDV genes by RNA polymerase II (Sahu et al., 2016) [166]. Transient DEAD35-silencing in a tolerant cultivar resulted in increased susceptibility to ToLCNDV infection (Pandey et al., 2019) [167]. In *Arabidopsis* histone reader EML1 (EMSY-LIKE 1) suppresses CabLCV infection by inhibiting the association of RNA polymerase II with viral chromatin and represses viral gene expression (Coursey et al., 2018) [168]. WWdomain-containing proteins (WWPs) trapped NSP-interacting GTPase (NIG) in immune nuclear bodies (NBs) thereby affecting its nucleo-cytoplasmic trafficking capacity and nuclear export of viral DNA to the cytoplasm (Calil et al., 2018) [169]. Natural resistance factors (Ty-1 and Ty-3) that encode DFDGD-class RNA-dependent RNA polymerases restricted TYLCV infection in tomatoes (Verlaan et al., 2013) [129]. MYMV-AC4 VSR activity depended on its binding to plasmamembrane through palmitoylation at Cys residue (Carluccio et al., 2018) [124].

1.6.7. Symptoms other than yellow mosaic

TYLCCNB-bC1 protein modulates the expression of several leaf-specific and jasmonic acid responsive transcripts (Yang et al., 2008)[97]. Geminivirus AC4/C4 proteins downregulates several genes involved in anther and pollen development thus affecting male fertility (Mills-Lujan and Deom, 2010) [170]. In BSCTV infected *arabidopsis* plants, two transcription factors (ATHB12 and ATHB7) have been shown to be induced by abscisic acid, where ATHB12 expression was correlated with morphological abnormalities such as leaf curling, stunting, and callus-like structures (Park et al., 2011) [171].

1.6.8. Host microRNA in establishing and restricting geminivirus infection cycle

Upon virus attack, plant miRNA profile gets altered which either works against viral invasion, multiplication and propagation or sensitize the host for enhancing viral cycle (Yang and Li, 2018) [172]. PhasiRNAs with antiviral activity were generated by the action of DCL2-made 22-mer miRs on several disease related genes and factors like DCL4, AGO1, AGO4, AGO7, SGS3, RDR6, and DOUBLE-STRANDED RNA BINDING FACTOR 4 (DRB4) etc., (Deng et al., 2018) [173]. Alteration of host miRNAs profiles upon begomovirus infection was found to be related to viral symptoms development (Amin et al., 2011)[47]. A study hypothesized that upon geminivirus infection, induced miR159 cleaved the ACS8 transcript, thus reducing the ACC level with simultaneous upregulation of ACCO enzyme and the consequent decrease in ET level followed by the down-regulation of the ET responsive genes ERF2, ERF4, and ETR4 (Miozzi et al., 2014) [174].

1.7. Mixed infection of geminiviruses: Reversion of recovery and recombination mediated evolution of new strains

In some cases of viral infection (Pepper golden mosaic virus, PepGMV), plants showed recovery i.e. reduction of symptoms on the new leaves known as host recovery (Carrillo-Tripp et al., 2007) [175]. Transcripts related to the oxidative response (CAT, ANN4, ANN1 and GST1), PR-5, ethylene and jasmonic acid signaling (ACC, EIN3, OPR1 and LOX1) and the novel gene kiwellin ripening-related protein (PRR1) were speculated for pepper recovery from PepGMV infection (Góngora-Castillo et al., 2012) [176]. Later on reversion of recovery upon subsequent infection by another virus was observed owing to synergic interaction between both viruses as reported for PepGMV and Pepper huasteco yellow vein virus (PHYVV) (Rentería-Canett et al., 2011) [177]. DNA hypomethylation and hypermethylation were found to be associated with symptomatic infection and recovery, respectively (Coursey et al., 2018) [168]. The combined presence of Rep, Ren, and TrAP from mixed infection of

two geminiviruses might generate a strong silencing suppression from which the plant is unable to recover (Rodríguez-Gandarilla et al., 2020) [178]. Another possible outcome of mixed infections could be the emergence of newer strains from existing virus strains. CLCuD identified in the 1960s in the southern Asia post passing through four phases—pre-epidemic, epidemic, resistance breaking and post-resistance breaking, resulted in several recombinant versions (Zubair et al., 2017a, Zubair et al., 2017b) [179,180].

1.8. Transcripts dynamics studies in legumes upon geminivirus invasion

Numerous transcriptional and transgenics studies to get insight into the global gene networks and to study the function of individual host or viral proteins were conducted in tomato, tobacco, cotton etc. Transcriptomic studies conducted in crops for different geminivirus includes tomato upon ToLCNDV (Zhong et al., 2017) [181] and ToCV/TYLCV (Seo et al., 2018)[182], *Arabidopsis* against CaLCuV (Ascencio-Ibáñez et al., 2008) [138], pepper against PepGMV (Góngora-Castillo et al., 2012) [176] infections. Transcripts found to be upregulated upon geminivirus attack (TYLCSV, CaLCuV, ToLCNDV, PepGMV) were associated with nucleic acid and nitrogen compound metabolism, transcription factors (MYB, NAC), ubiquitin proteasome pathway, autophagy related genes (ATG8, ATG9 and ATG12), SA pathway, PCD, genotoxic stress, DNA repair, 1-aminocyclopropane-1-carboxylate oxidase (ACCO), jasmonic acid methyl transferase (JMT), PR proteins, *N*-acetyltransferase, small heat shock protein and dehydration-responsive protein (Chen et al., 2013) [183]. While downregulated genes were related to photosynthesis, respiration, RNA-polymerases II (synthesis of the precursors of mRNAs and most snRNAs and microRNAs), SAHH, ADKs, SAMs, 1-aminocyclopropane-1-carboxylate synthase 8 (ACS8),chloroplast Clp proteases (ClpP1), chloroplasts proteases such as FtsH and Deg, chloroplast MATK and PsbO protein (Miozzi et al., 2014) [174]. ABA has been found to limit the multiplication and movement of diverse viruses by increasing callose deposition (Adie et al., 2007) [184] or by

regulating miRNA and siRNA pathways (Zhang et al., 2008) [185]. These studies suggest that upon pathogen infection, energy requirements of the plants are increased which are fulfilled by primary and secondary metabolic pathways accompanied with production of secondary metabolites comprising of lignin and phenolics against microbial pathogens (Taguwa et al., 2015) [186]. Moreover, pathogens transform post translational modifications machinery of host to trigger susceptibility in host (Howden and Huitema, 2012) [187]. Downregulation of RNA polymerase II (Pol II) could be contemplated as plants attempt to counteract viral infections by limiting the production of viral transcripts or a decreased concentration of plant proteins involved in the antiviral response. Interestingly few reports confirmed that alphasatellites reduce accumulation of their helper virus and/or betasatellite DNAs as observed for TYLCCNV/TYLCCNB in presence of TYLCCNA (Luo et al., 2019) [188]. Therefore, exploring differentially expressed genes (DEGs) responding to TYLCCNA provides the opportunity to determine the biological functions of alphasatellites during the begomovirus/betasatellite infections. For instance a transcriptomic study suggests that TYLCCNA might modulate the TYLCCNV/TYLCCNB infection by regulating the 'starch and sucrose metabolism' and 'glycosaminoglycan degradation' KEGG pathways and via biosynthesis of secondary metabolites (Luo et al., 2019) [188].

1.9. BRUCHIDS

Bruchids (Coleopterous: *Bruchidae*) are small beetles having globular or triangular shaped body of dull colour with white, red or black markings. They belong to sub-family *Bruchidae* which includes several genera such as *Acanthoscelides, Bruchus, Bruchidius, Callosobruchus, Caryedon* and *Zabrotes*. Around 1300 species of bruchids are known, among which 20 species are known pests for legumes (Talekar, 1988) [189]. *Callosobruchus maculatus, C. chinensis, Acanthoscelidesobtectus* (Say) and *Zabrotes subfasciatus* (Boleman)

are the major seed bettles attacking both grain and tree legumes (Swella and Mushobozy, 2007) [190].

Bruchid pests caused post harvest damage to nearly 5% of the world production of all cereal grains during storage on farms, elevators or warehouses, thus affecting seed quality, quantity and loss of viability (Duraimurugan et al., 2011)[26]. Traditional control methods include treatment of seeds with carbon disulfide, phosphine, methyl bromide, or by dusting with several other insecticides which are highly toxic and pose a threat to food safety. While plant-based extracts such as soy oil, maize oil, neem oil, hot pepper powder, custard apple extracts, and banana plant juice are slow in action, easily degradable, affect seed germination and non-target organisms (Sharma et al., 2012) [191]. The effectiveness of various oils such as eucalyptus, citronella, rosemary, cardamom and geranium against bruchids on pulses has been reported by several workers (Gorur et al., 2008) [192]. Disinfecting stored grains with 400 W microwaves for 28 seconds showed 100% mortality in all life stages of *Callosobruchus maculatus* , but the increase in temperature of seeds was associated with decreased germination (Purohit et al., 2013) [193].

Typically bruchid infestation begins in field when adult female beetle lays eggs on developing pods and larvae bore through the pod wall thus feeding on developing seeds (Mohan and Subbarao, 2000) [194]. After harvest, secondary infestation occurs during storage when newly emerged adults lay eggs on dried seeds thus resulting in total destruction of seeds within a period of 3 to 4 months. Both *C. maculatus* and *C. chinensis* exhibited similar life cycle and ecology which is about 28–30 days at 30˚C and 70% relative humidity (War et al., 2017) [195]. Adult bruchids mate within 24 hours of emergence, after which females can lay up to 100 eggs with 1–3 eggs/seed and large seed size accommodates more number of eggs. After oviposition, eggs get hatched after 6 days and larvae penetrate the seed testa and cotyledon followed by a developmental period involving five larval instars (Devi and Devi,

2014) [196]. Average time for adult emergence ranged from 20-30 days that depended on the host seed and storage conditions (temperature, humidity). Adults are either known to feed on pollen and flower nectar in the field for food rather than on seeds (Brier, 2007) [197] or do not feed during the brief adult stage (Guedes et al., 2007) [198]. Average life span of both adult *C. maculatus* males and females is 7 days with only a few able to survive for more than 2 weeks (Fatima et al., 2016) [199]. For a successful population build-up, a good quality substrate is needed which can be checked by adult bruchids (Guedes et al., 2007; Guedes and Yack, 2016) [198,200]. The adult females use various tactile, chemical, and physical cues such as multiple sensory modalities, egg-marking pheromones, and larval feeding vibrations from the seed to select suitable egg-laying substrate (Guedes and Yack, 2016) [200]. Yellow colored seeds are preferred over green or black seeds for oviposition and bruchid development (War et al., 2017) [195]. Some reports claimed no correlation between host preference and previous conditioning of the bruchids (*C. maculatus, Acanthoscelidesobtectus*and *C. chinensis)* on their hosts (Swella and Mushobozy, 2009) [190]. Also *C. maculatus* do not prefer seeds with a high protein-carbohydrate ratio and high fat content (Mphuru, 1981) [201].

1.9.1. Predominant species of *Callosobruchus*

In edible legumes, the species of *C. maculatus* and *C. chinensis*are found either singly or co-exist. But several studies reported *C. maculatus* to be dominant over *C. chinensis. C. maculatus* was observed to havemore egg laying and more adult producing ability than *C. chinensis* with shorter developmental period (Lale and Vidal, 2001) [202]. Lotka (1934) [203] studied competition between two species belonging to the same ecological niche and reached the following conclusions. i) each species inhibits its own potential increase more than that of the other and both continue to coexist, ii) the first species inhibits the potential increase of the second species and drives out the later from the given space, iii) the second

one drives out the first, iv) the inhibition of each species by the other is greater than its own inhibition.

1.9.2. Genetic basis of resistance and molecular markers developed in blackgram

Bruchid resistance trait has been identified in several closely related wild *Vigna* species including wild blackgram (*V. mungo* var. *silvestris*) (Sharma et al., 2013) [204]. *The* wild progenitor of blackgram *Vigna mungo var. silvestris has been well known for conferring resistance to C. maculatus through antibiosis mechanism which has been manifested as* reduced survival, longer developmental period and reduced body weight of emerged bruchid adults on dried seeds (Dongre et al. 1996; Souframanien et al. 2010)[12, 205]. The resistance trait was found to be under the control of 2 dominant duplicate genes designated as Cmr_1 and Cmr_2in*wild accession. Later on genetic linkage map of blackgram was constructed*using 428 markers and QTLs for bruchid resistance were mapped on the linkage map. Two QTLs, *Cmrae*1.1 and *Cmrae*1.2 for percentage adult emergence were mapped on linkage group (LG) 3 and 4, respectively and for developmental period, six QTLs were identified, with two QTLs (*Cmrdp*1.1 and *Cmrdp*1.2) on LG 1, three QTLs (*Cmrdp*1.3, *Cmrdp*1.4, and *Cmrdp*1.5) on LG 2, and one QTL (*Cmrdp*1.6) on LG 10 (Souframanien et al., 2010)[12]. Resistance to *C. chinensis* infestation in blackgram was reported to be under the control of a homozygous recessive gene (Fernandez and Talekar, 1990) [206].

1.9.3. Genetic basis of resistance and molecular markers developed in other *Vignas*

Wild species of mungbean have been well studied for resistance against various bruchid species (Kashiwaba et al., 2003) [207]. In mungbean wild accession *Vigna radiata* var. *sublobata*, TC1966 (Madagascar) resistance is under the control of a single dominant gene, *Br* (Fujii et al., 1989) [208] and has been utilized for developing bruchid-resistant mungbean cultivars (Somta et al., 2007) [209]. In mungbean, genetic linkage maps were constructed and bruchid resistance genes were mapped using RFLP, RAPD, SSR and SNP markers (Chen et

al., 2007; Chotechung et al., 2011) [210, 211]. In bruchid resistant TC1966 mungbean accession, four RAPD markers (OPW02, UBC223, OPU11, and OPV02) were found closely linked with the Br gene (Chen et al., 2007) [210]. The SSR marker DMB-SSR 158 was mapped at less than 0.1 cM distance to major QTL in TC1966 (Chotechung et al., 2011) [211]. In mungbean accessions V2709BG and V2802BG, resistance to *C. chinensis* (L.) and *C. maculatus* (F.) is maternally controlled by one major dominant gene with varying degrees of expressivity owing to the presence of modifiers (Somta et al., 2007) [209]. Two markers STSbr1 and STSbr2 were found closely linked to a major locus conditioning bruchid resistance in ACC41 resistant line (Miyagi et al., 2004; Sarkar et al., 2011) [212, 213]. These markers were found associated with polygalacturonase inhibitor genes (*Vradi05g03940-VrPGIP1* and *Vradi05g03950-VrPGIP2*) on chromosome 5 which could be the candidate gene for bruchid resistance in mungbean (Chotechung et al., 2016) [214]. Further *Br* locus was delimited through high-resolution mapping using SSR markers and *Br* locus linked EST-SSR marker DMB-SSR158 sequence showed homology with a polygalacturonase inhibitor (polygalacturonase-inhibiting protein PGIP), designated as *VrPGIP2*. *C*oding sequence comparison of *VrPGIP2* between four bruchid-resistant (V2802, V1128, V2817 and TC1966) and four bruchid-susceptible (KPS1, Sulu-1, CM and an unknown accession) mungbean lines revealed six SNPs with 3 SNPs corresponding to amino acid changes that may affect the interaction between PGIP and polygalacturonase. Bruchid resistance trait in a mungbean NIL VC6089A was associated with higher expressions of g39185 (resistant-specific protein), g34458 (gag/pol polyprotein), and g5551 (aspartic proteinase) (Lin et al., 2016) [215]. DEGs and sequence-changed protein genes (SCPs) for conferring resistance against bruchids were also identified (Liu et al., 2016) [216]. In ACC41, a QTL accounting for about 98.5% of bruchid resistance was identified (Mei et al., 2009) [217]. In mungbean accession Jangan, resistance was governed by a single dominant gene (Hong et al., 2015)

[218] and new molecular markers were identified. In cowpea, two recessive genes governing the resistance to *C. maculatus* were found (Adjadi et al., 1985) [219]. In mungbean accession V2709, resistance is controlled by single dominant locus with 2 markers, OPC-06 and STSbr2 mapped to the locus (named as Br2) at genetic distances of 11.0 cM and 5.8 cM, respectively (Sun et al., 2008) [220]. In mungbean accessions TC1966 and V2802, a major QTL associated with bruchid resistance was mapped on to chromosome 5 using a high resolution genetic map constructed with 6,000 SNP markers through genotyping by sequencing (GBS) method (Schafleitner et al., 2016) [221].

1.9.4. Methods of phenotyping

Various parameters for assessment of bruchid resistance are deployed which includes number of eggs laid/experimental unit, percentage of adult emergence, mean number of days for adult emergence, mean number of emergence holes per seed, percentage weight loss per seed, and percentage of undamaged seed (Redden and McGuire, 1983) [222]. Resistant and susceptible checks should be included in screening. Adult emergence is influenced by the atmospheric temperature and humidity which should be controlled to standardize the screening process. Also, mean adult insect emergence period (MDP) and percentage adult insect emergence (PAE) are the most sensitive parameters to determine *C. maculatus* resistance (Jackai and Asante, 2003) [223].

1.9.5. Sources of bruchid resistance in legumes

Several bruchid resistance sources were given in Table 1.3.

Table 1.3. Details of bruchid resistance sources reported in important legumes

Crop	Bruchid resistant accessions	References
Cowpea	TVu-2027, TVu-11952 and TVu-11953	Singh et al., 1985 [224].
Mungbean	TC1966, V2709 and V2802, Jangan, Zhonglv 3, Zhonglv 4, and Zhonglv 6, V1128 and V2817 , ACC 23 and ACC 41, PLM 262 and PLM 89, LM 131, V1123, LM 371, STY 2633, Km-2 and VBN2, VC1535-11-1-B-1-3-B, VC2764-B-7-2-B, VC2764-B-7-1-B, VC1209-3-B-1-2-B, and VC1482-C-12-2-B	Kitamura et al., 1988 [225]; Talekar and Lin, 1992 [226]; Hong et al., 2015 [218];Somta et al., 2008 [227]; Lambridges and Imrie, 2000 [228]; Srinivasan and Durairaj, 2007 [229]; Ponnusamy et al., 2014 [230]; Malaikozhundan and Thiravia, 2012 [231]; Asian Vegetable Research and Development Center AVRDC, 1988 [232].
Blackgram	*Vigna mungo* var. *silvestris* (INGR10133), VM2011, VM3529 and VM2164, Sindh Khed	Dongre et al., 1996 [205]; Sun et al., 2008 [220]; (AVRDC)
Adzuki bean	Transgenic expressing α-amylase inhibitor genes of tepary bean (*Phaseolus acutifolius* A. Gray)	Yamada et al., 2005 [233]

1.9.6. Host Plant Resistance

Association of pests and leguminous plants is an ongoing co-evolutionary process in which both have evolved parallely to counteract each others defensive systems. In this race, legumes have produced many toxic compounds to kill or deter bruchids and bruchids, in turn, have developed adaptive strategies to combat these toxic compounds. Bruchid pests species are very specific for their host as one insect species feeds on a very few seed species (Somta et al., 2007) [209]. Host plants confer resistance to pests in different ways such as antibiosis, antixenosis (non-preference) and/or tolerance through morphological, physiological and/or biochemical traits which can affect growth and development of insect pests (Edwards and Singh, 2006) [234]. The morphological traits in legumes include color, shape, size and texture of the pod and seed while the physiological and/or biochemical traits include

secondary metabolites and anti-nutritional compounds affecting the metabolic activity of bruchids (Somta et al., 2007) [209].

1.9.7. Physical basis of resistance

The first interaction between insect pests and host plants occurs at the time of oviposition by insect pests. Plants deploy several antixenosis traits which alter insects behaviour for oviposition that includes surface chemicals, plant volatiles, spines, hairs, etc. (War et al., 2013) [235]. Also the host plant/seeds avoid insect oviposition either directly or indirectly by killing the insect eggs to avoid hatching of the larvae, thus preventing future damage (War et al., 2017) [195]. Generally, female bruchids prefer smooth surfaced seeds rather than rough surfaced seeds for oviposition as observed in bruchid-resistant mungbean TC1966 seeds. These rough surfaced seeds possess a network of parallel and transverse ridges that restrict female bruchids from laying eggs on the seeds (Watt et al., 1977) [236]. Features such as dense hairy pods, hard seed coat and small seeds were found to be ovipositional deterrents for female bruchids (Somta et al., 2007; Mei et al., 2009) [209,217]. However, the role of seed coat and its size in conferring resistance has been ruled out by some researchers (Somta et al., 2008) [227]. Apart from these, a study claimed that egg laying trait alone cannot be considered for determining the criteria for selecting the genotypes for bruchid resistance (Ponnusamy et al., 2014) [230]. Indirect selection criteria for resistance to *Callosobruchus* in mungbean has been proposed based on the correlation between pest susceptibility and different seed parameters (seed weight and seed coat width) (Chakraborty et al., 2004) [237]. While green coloured seeds were prefered for oviposition against the black coloured seeds in blackgram (Sulehrie et al., 2003) [238], contrasting observations were found in cowpea lines (Chavan et al., 1997) [239].

1.9.8. Biochemical basis of resistance

Secondary metabolites produced by plants are important defensive traits involved in plant defense due to their toxic effects on insect pests (War et al., 2013) [235]. These compounds can be either antibiotic or antixenotic for insects with either additive or synergistic mode of action (War et al., 2013) [235]. Some well known compounds include naringenins, vicilins, cysteine rich proteins (VrD1 or VrCRP), vignatic acids (A and B) and para-amino-phenylalanine (Somta et al., 2007) [209]. Extensive biochemical research in mungbean has resulted in identification of several biochemical compounds responsible for resistance to bruchids. These compounds are resistance-related proteins (chitinase, b-1,3-glucanase, and peroxidase), lignins, quinines, alkaloids, saponins, non-protein amino acids and polysaccharides, anti-nutritional seed proteins such as lectins, phytohemagglutinins (PHA), proteinase inhibitors, trypsin inhibitors and α-amylase inhibitors (Lattanzio et al., 2005) [240]. Toxic effects of several compounds present in plants on various insect pests belonging to Coleoptera, Homoptera, Diptera, and Lepidoptera were studied (Vandenborre et al., 2011) [241]. Insecticidal role of plant lectins such as canavalin (vicilin, 7S globulin), canatoxin (*Canavalia ensiformis*), zeatoxin (*Zea mays*), lectins (*Talisia esculenta* Radlk, *Gracilariaornata*Areschoug), galactose-specific lectin (African yam beans) have been studied (Vandenborre et al., 2011) [241]. These studies revealed pH stable nature of these lectins and their interference with digestion and absorption in the insect guts affecting pest insects growth and development (Vandenborre et al., 2011) [241]. Cyanogenic glycosides and phytic acids were also found to be associated with defense responses against bruchids (Lattanzio et al., 2005) [240]. Interestingly, relation of nutritive and digestive factors with bruchid pests development and survival were studied in soybean, garden pea etc with high protein-carbohydrate ratio, carbohydrate composition, saponin content and high fat and amylase contents (Mphuru, 1981) [201]. In mungbean accession TC1966, bruchid resistance

compound was found to be a cyclopeptide alkaloid composed of L-tyrosine, 3(*S*)-hydroxyl-L-leucine, L-phenylalanine, and 2-hydroxyisocaproic acid referred as Vignatic acid A with insecticidal property (Sugawara et al., 1996) [242]. Likewise, cysteine rich protein VrD1 with insecticidal property was characterized from mungbean accession VC6089 (Lin et al., 2005) [243]. The toxic substances beta-cyanoalanine and alphadiaminobutyric acid were found responsible for rendering pest resistance in haricot bean (Seifelnasr, 1991) [244]. The high phytic acid content in mungbean wild species *Vigna radiata* var. *Sublobata* was found to be associated with tolerance to YMD, powdery mildew disease (PMD) and storage pest bruchid (Dhole and Reddy, 2016) [245]. Apart from their anti-insect property, most of the chemical compounds are anti-nutritional which renders the seeds unfit for consumption (Modgil and Mehta, 1997) [246]. In blackgram accessions VM 2164 and Acc 2228, seed albumins and globulins adversely affected the egg deposition of the bruchids compared with the albumins of susceptible accessions Pag-asa 7. In mungbean accessions, VC6089A, TC1966, and the recombinant inbred line 59 (RIL59), the resistant-specific protein, gag/pol polyprotein, and aspartic proteinase were found associated with bruchid resistance (Lin et al., 2016) [215].

1.9.9. Counter adaptations in bruchid species

To combat/ escape plant defenses, bruchids have developed counter adaptations to most of the plant toxic chemicals (Chi et al., 2009) [247]. Bruchids have evolved metabolic pathways to bypass the enzyme block by producing high levels of mid-gut aspartic and cysteine proteinase to overcome protein anti-metabolites (Zhu-Salzman et al., 2003) [248]. High expression of major digestive cathepsin L-like cysteine proteases such as CmCPA and CmCPB in bruchids neutralize the effect of protease inhibitors (Koo et al., 2008) [249]. Similarly, over-production of glutathione S-transferases, cytochrome P450 monooxygenases

(P450s) and esterases in bruchids help in adaptation to plant/host toxic compounds (Chi et al., 2009) [247].

1.9.10. Trancriptomics of bruchid resistant genotypes

Very limited molecular research has been conducted in blackgram and other legumes with regard to bruchid infestation which serve as a preliminary data for future directions. Transcriptomic studies on bruchid-resistant and susceptible parents along with their offspring has resulted in identification of 91 DEGs, 408 nucleotide variations (NVs) in promoters of 68 DEGs and 282 NVs in exons of 148 sequence-changed-protein genes (SCPs) (Liu et al., 2016) [216]. Transcriptome sequencing of mungbean VC6089A (bruchid-resistant) led to the identification of 399 DEGs (Lin et al., 2016) [215]. Transcript dynamics of a mild tolerant blackgram variety (IC8219) in response to oviposition on developing seeds revealed differential expression of receptors, PR, lipoxygenase (LOX) etc genes which could be associated with bruchid resistance (Baruah et al., 2017)[14].

1.9.11. Breeding constraints for developing bruchid-resistant blackgram

Traditional and chemical control methods are expensive for resource poor farmers and biological control methods are not durable. This necessitates the search for sustainable solutions such as host plant resistance for developing resistant crops.This requires identifying resistance sources in germplasm, studying the underlying basis of resistance, mapping of resistance genes, development of molecular markers associated with resistance and their successful introgression in elite cultivars without compromising the yield and nutritional quality which is a major challenge faced by breeders all around the world (Keneni et al., 2011) [250]. Major constraints in breeding programme are lack of resistance sources, narrow gene pool due to extensive cultivation of few cultivars (Ladizinsky, 1985) [251], anti-nutrients present in resistant wild species (Acosta- Gallegos et al., 2008) [252], genetic drag of undesirable traits (pod shattering) (Watanasit and Pichitporn, 1996) [253] and

crossincompatiblity of wild species with cultivars (Keneni et al., 2011) [250]. Another major constraint for the breeders is the biotypic variation, i.e., genetic variability of the pest population which has led to the breakdown of resistance in legumes against bruchids (Fox et al., 2010) [254]. Therefore, gene pyramiding could help in developing cultivars with durable resistance to wide range of pest biotypes.

The identification of bruchid resistant genotypes by artificial bioassays and biochemical methodologies are expensive, time consuming and inconsistent (Somta et al., 2007; Srinivasan and Durairaj, 2007) [209,229]. The molecular markers for bruchid-resistance can increase the selection efficiency, reduce the number of selection tests, reduce the cost required for screening and is independent of phenotypes (Schafleitner et al., 2016) [221]. Thus, there is a need to explore and characterize different resistance sources for clear and complete understanding of resistance mechanism which will aid in introgression of target genes in cultivars.

CHAPTER 2

MATERIALS AND METHODS

2. Materials and Methods

2.1. Materials

2.1.1. Plant material for yellow mosaic virus disease resistance trait studies

Vigna mungo (cv. KU96-3) resistant to YMD was crossed with *Vigna mungo* (cv. TAU-1). A total of 218 lines of recombinant inbred population (F_8) from above parental cross were generated from a single F1 plant post-selfing and further advanced to F_8 generation by single seed descent method. Normal recommended agronomic practices were followed. For screening primers through end-point PCR, both parents, resistant and susceptible bulks and RILs were used. For gene expression experiments, three genotypes KU96-3,TU94-2 (both resistant to YMD) and TAU-1 (susceptible to YMD) grown in experimental field for natural YMV infection by white flies served as test/treated samples and that in glass house under protection for white-flies served as control.

2.1.2. Plant material for bruchid pest resistance trait studies

Wild blackgram accession (*Vigna mungo* var. *silvestris),* Trombay wild urd (TW) (Registration No. INGR10133) which is native of Trombay hills (Maharashtra) with bruchid resistance and *Vigna mungo* cv. TU94-2, a bruchid susceptible variety of the host institute were used for both screening primers through end-point PCR and differential expression of genes through quantitative reverse transcription polymerase chain reaction (qRT-PCR). For differential expression study, immature developing pods at 4 weeks post-flowering were taken and for screening markers through conventional PCR, genomic DNA extracted from 2 days old seedlings was used.

Note: Trombay wild is resistant to bruchid pests but susceptible to YMD and vice-versa is true for TU94-2 and KU96-3 genotypes.

2.1.3. Plant materials for genetic diversity study

A total of 44 blackgram genotypes collected from different parts of India and comprising of cultivars, landraces and one wild accession were analysed in the present study. The details of pedigree, reaction to YMD and bruchid infestation are presented in Table 2.1. The seeds were maintained as pure lines at Nuclear Agriculture and Biotechnology Division, BARC, Trombay, Mumbai, India.

Table 2.1. List of 44 blackgram genotypes used in the genetic diversity study

S. No.	Accessions	Pedigree	YMD reaction	PMD reaction
1	EC-168200	Exotic collection from AVRDC, Taiwan	R	-
2	PUSA-3	L-151 X T9	R	-
3	IPU02-43	DPU88-31 x DUR-1	R	HR (Akhtar et al., 2014) [255]
4	KU96-3	PU19 x NP21	R	R
5	KU96-7	-	R	-
6	IPU07-3	DPU88-31 x PDU-1	R	R
7	DPU88-31	PLU131 x T-9	R	S
8	TU94-2	TPU-4 x TAU-5 (Mutant of EC-1682000)	R	MR
9	Azad-1	-	R	-
10	LBG-752	-	S	MR
11	LBG-17	Netiminumu x Chikkuduminumu	S	R (Srivastava et al., 2011) [256]
12	LBG-693	-	S	-
13	LBG-623	-	S	R (Priyanka et al., 2017) [257]
14	TAU-1	T-9 x UM-196 (Mutant of No. 55)	S	HS
15	Trombay Wild	*Vigna mungo* var. *silvestris*	S	-
16	LBG-703	-	S	
17	LBG-20	-	S	R (Srivastava et al., 2011) [256]
18	T-9	Local selection from Bareilly U.P.	S	MR (Srivastava et al., 2011) [256]
19	Nayagarh	Local selection	R	-

20	Pant-U19	UPU-1 x UPU-2	R	R
21	PU31	UPU97-10 x DPU88-31	R	R (Akhtar et al., 2014) [255]
22	PLU-1		R	-
23	TU-43-1	Mutant of TU94-2	R	-
24	TU-55-1	Mutant of TU94-2	R	-
25	NDU-1	Sel. 1 x T9	R	-
26	TU-67	TAU-1 X KU96-3	S	-
27	WBG-17		S	-
28	WBG-57		S	S
29	WBG-13		S	-
30	COBG-653		S	S
31	PLU-710		S	-
32	Sharda mash		S	S
33	EC168058	Exotic collection from AVRDC, Taiwan	S	-
34	LBG-685		S	MR
35	LBG-709	-	S	R
36	Sheela		-	-
37	EC168234	Exotic collection from AVRDC, Taiwan	-	-
38	EC168242	Exotic collection from AVRDC, Taiwan	-	-
39	EC168243	Exotic collection from AVRDC, Taiwan	-	-
40	IPU-02-6		-	-
41	IPU-99-247		-	-
42	SPS-30		-	-
43	ANU-11		-	-
44	IPU-99-40		-	-

YMD-Yellow mosaic disease; PMD-Powdery mildew disease; R-Resistance; S-Susceptible

2.1.4. Chemicals, reagents, enzymes and kits

Lambda DNA EcoR I Hind III double digest DNA marker, RNase A, deoxynucleotide triphosphate (dNTPs) were obtained from Sigma-Aldrich (USA). Agarose, ethidium bromide, bromophenol blue and ficol were obtained from Sisco research laboratories Pvt. Ltd., Mumbai, India. Potassium acetate, sodium acetate, ammonium acetate, EDTA, calcium chloride, xylene cyanol were obtained from Sigma Chemical Co., USA. Other chemicals used

were either of molecular biology grade, AR (analytical reagent) or GR (guaranteed reagent) grade.

Reagents used for DNA extraction

Extraction buffer (pH8)and components for Dellaporta et al. method

Tris-Hcl	100 mM
EDTA	50 mM
NaCl	500 mM
SDS	20 % (w/v)
Potassium Acetate	5 M
Isopropanol	(75%)
Ethanol	70 %

Tris-EDTA buffer (TE) (pH8)

Tris-Hcl	10 mM
EDTA	0.1 mM

DNA extraction buffer for CTAB method

cetyl trimethylammonium bromide	2%
polyvinyl pyrrolidone	1%
Tris-HCl	100 mM
NaCl	1.4 M
EDTA	20 mM
TE Buffer	10 mM Tris, pH 8, 1 mM EDTA

Modified CTAB extraction buffer comprised of CTAB extraction buffer plus 0.1% beta mercaptoethanol. Gem-CTAB extraction buffer composed of CTAB extraction buffer with 1.4-2.0 M NaCl, 10 mM EDTA and 2-5% β-mercaptoethanol.

Reagent for gel electrophoresis

1 X TBE buffer (pH 8.3)

Tris	100 mM
Boric acid	83 mM
EDTA	1 mM

Gel loading dye

Bromophenol blue	0.1 %
Ficol	20 %
EDTA	10 mM

Reagents for RNA isolation

RNA extraction from immature seeds was done using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, USA) and for young leaves was done by RNA sure plant kit (Qiagen). Genomic-DNA was degraded after RNA extraction through DNase-I (Sigma-Aldrich, USA). For manual RNA extraction, 8 M Lithium chloride (LiCl) was prepared by dissolving 34g (MW: 42.4) of LiCl in 100 ml of milliQ water, autoclaved and storedat 4 ˚C.

The first-strant c-DNA synthesis was carried out with the help of PrimeScript™ RT-PCR Kit (Takara Clontech, USA). The quantitative real-time PCR was done using Sybr green master mix (10x) (Roche) according to manufacturer's manual. SNP analysis through HRM was performed using Type-it HRM kit (Qiagen) following manufacturer's manual.

DEPC treated distilled water (0.1%) was used for sterilizing pestle-mortar and spatula for RNA extraction experiment. For this, 1 ml of DEPC was mixed with 1000 ml of autoclaved distilled water and was shaken vigorously 2-3 times with intermittent settling of foam. In this DEPC treated water, pestle-mortar and spatula were kept dipped overnight. Next day, pestle-mortar and spatula were taken, dried in air and autoclaved before use.

2.1.5. Plasticware and instruments

Disposable polypropylene microfuge tubes (0.5 ml, 1.5 ml, 2 ml) and micropipette tips were obtained from Tarson, India and Axygen Scientific Pvt. Ltd., New Delhi, India. Thermal cycler used was Eppendorf Mastercycler gradient, Hamburg and Rotor Gene Q, Qiagen, Germany. Horizontal gel electrophoresis apparatus used was obtained from Hoefer Scientific Instruments, USA and Bangalore Genei Pvt. Ltd., Bangalore, India. PAGE was performed using vertical gel electrophoresis apparatus (BIORAD). Quantification and quality checking of extracted DNA and RNA were done with the help of Nanodrop ND 1000 spectrophotometer (Thermo Scientific, USA). Agarose gel pictures was documented through gel documentation system (Syngene, U.K.). For high resolution separation of PCR products, fine capillary electrophoresis was performed in Qiaxcel capillary electrophoresis machine (Qiagen). For centrifugation, centrifuge of up to 14,000 x g from Ependorff was used.

2.2. Methods

2.2.1. Transcriptome sequencing of blackgram immature seeds

2.2.1.1. Sample preparation, transcriptome sequencing, *de novo* assembly and characterization

Transcriptome dataset of the wild accession of blackgram, TW (*Vigna mungo* var. *silvestris*) and cv. TU94-2 was generated in this study. Total RNA was isolated from immature seeds (4 weeks after flowering) of different TW and TU94-2 plants using RaFlextotal RNA isolation kit as per protocol. Extracted RNA was checked for quality and quantity using QubitFluorometer. The paired-end cDNA sequencing libraries were made by fragmenting mRNA, performing reverse transcription reaction and second-strand synthesis using IlluminaTruSeq RNA Sample Preparation V2 kit. Ligation reaction with paired-end adapters and index PCR amplification was performed. Quality assessment for library was done on Agilent Caliper LabChip GX Bioanalyser using DNA High Sensitivity Assay Kit.

Sequencing was done in a single lane using paired end sequencing chemistry on Illumina MiSeq platform (Xcelris Genomics Ltd. Ahmedabad). Infiltration, adaptors and low quality reads (QV < 20) were removed using Trimmomatic v0.30 (Bolger et al., 2014) [258]. High quality reads (HQ) were assembled into transcript contigs on CLC Genomics Workbench with default parameters (minimum contig length: 200, automatic word size: yes, perform scaffolding: yes, mismatch cost: 2, insertion cost: 3, deletion cost: 3, length fraction: 0.5, similarity fraction: 0.8). Coding sequences (CDS) were predicted with the selection of longest frame using ORF-predictor. The functional annotation of CDS were assigned by similarity search program, BLASTx through alignment of CDS to non-redundant protein sequence (nr) database of NCBI (http://www.ncbi.nlm.nih.gov) with reference to green plant database. The functions were assigned to transcripts with significant e value ≤1e-5. Mapping of CDS was performed using both Gene Ontology (GO) database using Blast2GO (Conesa et al., 2005) [259] and KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa and Goto, 2000) [260] (http://www.genome.jp/kegg).

2.2.1.2. Prediction of SSR and differentially expressed genes (DEGs)

SSRs with 2 to 6 nucleotides in size and a minimum of 6, 5, 4, 4, 3 contiguous repeat units for di-, tri-, tetra-, penta- and hexa-nucleotides respectively, excluding mononucleotides were predicted using WebSat (http://purl.oclc.org/NET/websat/) online software (Martins et al., 2009) [261]. Frequency and relative abundance of SSRs were calculated. Location of SSRs in the CDS or untranslated regions (5′ or 3′ UTR) of the gene was found out. The presence of ORF with the longest ORF starting with ATG codon was identified with the help of ORF Finder software (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). DEGs were identified by mapping HQ reads of TU94–2 (control) and Trombay wild (treated) to their respective set of CDS using CLC. FPKM was calculated using the formula: FPKM = 109 x C/ (N x L), where C is the numberof reads mapped onto the transcript contigs, N is the total number of

mappable reads in the experiment, L is the number of base pairs in the transcript contigs. Common hit accessions for DEGs were found using BLAST against Green Plant database. Log fold change (FC) values were calculated by the formula: FC = Log2 (Treated/Control). FC values greater than zero were considered up-regulated whereas less than zero were considered down-regulated.DEGs ranging from 3 to 12 and -3 to -12 were reannotated with the help of tBLASTn against nr database of NCBI as most of the DEGs remained uncharacterized with no function when Green Plant database was used as reference.

2.2.1.3. Identification of SNPs, characterization and functional significance

SNPs were identified by aligning HQ reads of TW (alternate) with the assembled transcript of TU94–2 using BWA version 0.7.5a with default parameters. After alignment, SAM (Sequence Alignment/Map) format files were converted to BAM (binary version of a SAM file) format using samtools and sort programs. The SNP calling was performed with these BAM files using mpileup program of samtools and varFilter program of bcf tools with default parameters except minimum read depth, maximum read depth and mapping quality were assigned to 20, 200 and 10, respectively. The SNPs were generated in vcf (Variant Call Format) file. SNPs bearing TU94–2 TCS were analysed for ORF and SNP position within transcript. SNPs were also characterized into transition and transversion types with further calculation of Transition (Ts) to transversion (Tv) ratio (Sablok et al., 2011) [262].

Sequence homology of SNPs bearing contigs was found out through BLASTn and BLASTX algorithms of NCBI. Functional and structural significance of non-synonymous SNPs were analyzed using the ITASSER automated web server (Zhang, 2008) [263] through prediction of *ab intio* three dimensional secondary protein structure and active catalytic domain binding sites with ligands. The protein model with correct topology and high confidence scores were choosen. Presence of ORF was found out using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and replacement of amino-acid were predicted

by aligning the amino-acid sequences of wild with cultivated form of proteins through Clustal Omega software (Sievers et al., 2011) [264].

2.2.2. DNA extraction

2.2.2.1. From seedlings

DNA was extracted from 1 day old seedlings following Dellaporta et al (1983) [265] method. Seedlings were crushed using mortors and pestles in the extraction buffer (500 µl) and transferred to 2 ml eppendorf tubes. To this, 200 µl of 20% SDS was added, vortexed and kept at 65˚C for 20 min in a heating block. Chilled 5 M potassium acetate (200 µl) was added, vortexed and kept on ice for 20 min. After incubation on ice, 500 µl of chloroform:iso-amyl alcohol (24:1) was added, mixed and centrifuged at 12000 rpm for 20 min at 4 ˚C. The supernatants were transferred to fresh 1.5 ml eppendorf tubes. Chilled iso-propanol (500 µl) was added and incubated on ice for 30-45 min. Tubes were centrifuged at 10000 rpm for 10 min at 4 ˚C. Supernatants were discarded and pellets were kept for complete drying in laminar air flow. After air drying of pellets, 500 µl of TE I buffer and 2 µl of RNase (100mg/ml) were added and incubated at 37 ˚C for 1 hour in heating block. Equal volumes (500µl) of chloroform:iso-amyl alcohol (24:1) were added, mixed and centrifuged at 12000 rpm for 20 min at 4 ˚C. Supernatants were transferred to fresh 1.5 ml eppendorf tubes and chilled absolute ethanol (1 ml) was added. Tubes were incubated on ice for 30-45 min and centrifuged at 10000 rpm for 10 min at 4 ˚C. Supernatants were discarded and pellets were kept for complete drying in LAF. The pellets were later dissolved in TE II buffer (50 µl).

2.2.2.2. From leaves

2.2.2.2.1. CTAB, modified CTAB and GEM-CTAB methods

DNA was extracted with these three methods by same experimental steps as described above with the exception of the extraction buffers. Fresh leaf tissue (100 mg) was crushed in 500 µl of CTAB Extraction Buffer. Slurry was transferred to 2 ml eppendorf tube and incubated at 65°C for 60 minutes in a heating block. After incubation, 500 µl of chloroform:isoamyl alcohol (24:1) was added and centrifuged for 20 minutes at 12000 rpm at room temperature. Supernatant was transfered to a new tube and treated with 5 µl of RNase at 32°C for 20 minutes. After RNase treatment, equal volume of chloroform:isoamyl alcohol (24:1) was added, vortexed for 5 seconds and then centrifuged for 1 min. at 14,000 x g to separate the phases. Transfered the upper aqueous phase to a new tube and precipitated the DNA by adding 0.7 volume cold isopropanol at -20°C for 15 minutes. After centrifugation (14,000 x g, 10 min), supernatant was decanted and the pellet was washed with 500 µl ice cold 70% ethanol. Pellet was air-dried and resuspended in 50 µl TE buffer (10 mM Tris, pH 8, 1 mM EDTA).

2.2.2.2.2. Modified GEM-CTAB method (DNA extraction from minimal tissue in short time with improved quality)

About 20 mg of fresh leaf tissue was ground to fine powder in liquid nitrogen using pre-chilled mortar and pestle. The samples were transferred to 0.5 ml eppendorf tubes and 200 µl of extraction buffer was added and incubated at 95 ˚C for 10 min in the thermal cycler (PCR machine for small tube and large scale extraction). To the homogenate, 0.8 volumes of chloroform:*iso*-amyl alcohol (24:1) was added and mixed gently by inversion. The samples were centrifuged at 10000 rpm for 7 min at room temperature and the top aqueous phase was transferred to a fresh tube. Cold isopropanol (0.7 volume) was added to the samples, mixed

and kept at -80 ˚C for 10 min. Samples were centrifuged for 10 min at 10,000 rpm. The nucleic acid pellet was washed twice with 80% and 70% cold ethanol. The pellet was air-dried in laminar air flow until all ethanol were removed and dissolved in 100 µl of nuclease free water. About 2 µl of RNase was added and the samples were incubated at 37 ˚C for 20 min. No further purification was done and eluted RNA was used for quantification and reverse transcription reaction.

2.2.3. Development of genic-SSR, genic-SNP, DEGs, gene-specific and genomic-SSR markers

2.2.3.1. Primer designing

After identification of SSRs in TW transcriptome dataset, primers were designed using WebSat (http://purl.oclc.org/NET/websat/) online software (Martins et al., 2009) [261] with default parameters. Primers were designed for all SNPs of TW using TU94–2 transcripts accessible from NCBI (submission ID. SRR 1616991, SRX710526 and study accession SRP 047502) with the help of Primer3 software (http://sourceforge.net/projects/primer3/) with default parameters and product size ranging from 100 to 250 bp. Expression study real-time primers were designed using Gene Runner v. 5.0.99 Beta (Hastings Software, Inc.) and primer 3 softwares with default parameters. Oligonucleotides were obtained from Eurofins Genomics India Pvt. Ltd., India and CUSABIO, China. Primers for host cellular factors involved in geminiviruses interaction from TU94-2 transcript contigs were also designed by Primer 3 software with default parameters.

Note: Default parameters - Optimum primer length was 20 mer (range: 18–27 mer), optimum annealing temperature was 60 °C (range: 57–63 °C) and GC content was from 20 to 80%.

2.2.3.2. Primer designing for genomic-SSR from mungbean based on RGA

A total of 23 putative resistance proteins identified in mungbean (*Vigna radiata* var. *radiata* cultivar:VC1973A) from whole genome shotgun (wgs) sequencing by Kang and co-workers

(Kang et al., 2014) [266] available in NCBI database (Accession: PRJNA243847, ID: 243847) was used in this study. Amino-acid sequence of proteins were downloaded from the UniProt (https://doi.org/10.1093/nar/gkw1099) and searched for sequence homology with mungbean wgs sequence (*Vigna radiata* var. *radiata*, taxid:3916) with the help of tBLASTn algorithm. Mungbean scaffolds were mined for SSRs and primer-pairs were designed with the help of websat (http://purl.oclc.org/NET/websat/) online software (Martins et al., 2009) [261].

2.2.3.3. Thermal cycling conditions

SSR Amplification

PCR reactions were carried out in a 25 µl reaction volume in an Eppendorf Master Cycler (Eppendorf, Hamburg, Germany) with following composition: 75 ng of genomic DNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 0.08% Nonidet P40, 0.2 mM dNTPs, 1.5 pmoles forward and reverse primers, and 0.5 unit *Taq* DNA polymerase (Fermentas Life Sciences). The amplification conditions were initial denaturation at 94°C for 3 min, 5 cycles at 94°C for 30 s, 56 to 46°C (-1°C each cycle) for 10 s, 72°C for 1 min followed by 35 cycles at 94°C for 30 s, 46°C for 1 min, 72°C for 1 min and ended up with a final extension at 72°C for 7 min. PCR products were resolved on 3% agarose gels in TBE buffer at 80 V and images were captured in a gel documentation system (Syngene, U.K).

SNP - Realtime HRM-PCR amplification and data analysis

HRM-PCR reactions were carried out according to manual of Type-it HRM PCR Kit (Qiagen) in Rotor-Gene Q cycler (Qiagen) in 20 µl volume, containing 2X HRM PCR master mix (HotStarTaqPlus DNA Polymerase, Eva green dye, optimized concentration of Q-solution, dNTPs, and $MgCl_2$), 1.5 pmoles of each forward and reverse primers and 70 ng of genomic DNA at thermal conditions given in Type-it HRM PCR Kit (Qiagen) with acquisition of fluorescence data at a temperature range of 65-95˚C with 0.1 °C increments in

temperature. The Type-it HRM cycling conditions were as follows: 2-step cycling program withinitialactivation of enzyme at 95˚C for 5min followed by 40 cycles of denaturation at 95˚C for 10 s, 55˚C for 30 s with no extension step and fluorescence data acquisition at a temperature range of 65-95˚C with temperature increment of 0.1 at every step of 2 sec. Shape of normalized HRM curves and melting curves were used to predict zygosity of genotypes. Primer-pairs specificity was checked by resolving HRM-PCR amplicons on 3% agarose gels to confirm the absence of primer-dimers and non-targeted amplicons. Raw HRM data was analysed with the help of inbuilt software in Rotor-Gene Q. For data quality control, PCR amplification was analysed through the assessment of the CT value and amplification efficiency (Wu et al., 2008) [267]. The amplification data which deviate from any of the two following criteria were not used for HRM analysis. HRM runs with CT value ≤30 with amplification efficiency > 1.4 only were considered for analysis. Raw HRM curves were analysed using the HRM analysis module as per Corbett Research manual. Realtime PCR HRM amplicons were resolved on 3% agarose gels to detect the presence of primer-dimers, and non-targeted amplicons. Melting temperature (Tm) difference was calculated manually with the help of negative derivative plots. Homozygotes melt in a single transition whereas, heterozygotes showed multiple melt phases and the order of melting was correctly predicted by nearest-neighbor calculations as A/A _T/T _ C/C _G/G (Liew et al., 2004) [268]. Class 1 SNPs are C/T and G/A transitions that produce C::G and A::T homoduplexes and C::A and T::G heteroduplexes, class 2 SNPs (C/A and G/T) are transversions that produce C::T and A::G heteroduplexes, Class 3 SNPs (C/G) produce C::G homoduplexes with C::C and G::G heteroduplexes and Class 4 SNPs (A/T) produce A::T homoduplexes with A::A and T::T heteroduplexes (Liew et al., 2004) [268]. SNPs were differentiated into 2 groups, homozygous and heterozygous based on normalised HRM curve. SNP amplicon sequences were checked for the presence of SSRs with the help of WebSat online software.

2.2.3.4. Statistical and Diversity analysis

Polymorphic markers were scored for each SSR allele as presence (1) or absence (0) of amplified products. Polymorphic information content (PIC) for each SSR marker was determined using the formula of Anderson et al.,(1993) [269]: PIC = 1 − S(Pij)2, where Pij is the frequency of the jth allele for the ith locus. Dendrogram based on Jaccard's similarity coefficients was generated after subjection to unweighted pair group method with arithmetic average (UPGMA) using software NTSYS-pc version 2.0 (Rohlf, 1998) [270]. Raw HRM data were analysed using the HRM software inbuilt in Rotor-Gene Q machine (Qiagen) with the exclusion of No template control (NTC). Two methods were used to classify 27 genotypes. In first method, genotypes were classified based on normalized HRM curves into either control 1 (TU94-2) or control 2 (TW), while remaining as variants depending upon similarity to controls at threshold score >90%. In second method, negative first-derivative melting curves (fluorescence versus temperature plots) were used for predicting melting temperature differences among controls (TU94-2 and TW) and other genotypes and for grouping genotypes. Further, the reliability and synchrony of both above methods were tested by aligning fluorescence data from method 1 with melting temperature data from method 2. Genetic relatedness was also analysed among genotypes using UPGMA to generate the dendrogram. Scoring for each SNP marker was done for three alleles: allele 1 for TU94-2 nucleotide base, allele 2 for TW nucleotide base and allele 3 for variant as presence (1) or absence (0) of respective nucleotide base and data were analysed by software NTSYS-pc version 2.0 (Rohlf, 1998) [270] with Jaccard's similarity coefficients.

2.2.3.5. Phenotyping: Screening for YMD reaction

Recombinant inbred lines developed from the cross TAU-1 x KU96-3 was sown in the filed for YMD screening using infector row method in which susceptible check (LBG17) plants were sown after every 3-5 lines or rows. Mapping population was sown at Trombay loaction

during 2017 and 2018. Yellow mosaic symptoms were scored based on 1-9 scale (Singh et al., 1988) [271] and further classified into four classes: Highly Resistant (HR), Moderately Resistant (MR), Susceptible (S) and Highly Susceptible (HS). Homogeneity or goodness of fit for 1:1 ratio in mapping population was checked by chi-square test.

2.2.3.6 Phenotyping: Screening for bruchid resistance

All the F_2, F_3 plants and 104 RILs developed from the cross population of TU94-2 X Trombay wild were screened for bruchid resistance trait at laboratory condition along with parents.The stock culture of *Callosobruchus maculatus* was maintained on mungbean seeds (*Vigna radiata* (L.)) at constant temperature of 29^{o}C and 60-70 % relative humidity. Incubation was done in a temperature-controlled incubator maintained at 29 ± 1^{o}C.

Bruchid resistance phenotype was evaluated using method described by Dongre et al (1993) [272]. In this method, a pair of newly emerged adults was released in small plastic petri plates containing 50 seeds of each line in 3 replications for 24 h. After incubation period of 24 hr, total number of eggs laid, eggs hatched and number of larvae entered the seeds were counted with the help of microscope. After five days, unhatched eggs remained transparent and hatched eggs from which a larva entered appeared as opaque. Number of adults emerged was routinely checked and recorded. After 50 days, counting for adults emergence was stopped to avoid adult emergence of second generation. Screening data recorded included parameters such as number of eggs laid within 24 h, percent larvae entered, percent adult emerged and developmental period (egg + larva + pupa). The accession with observed longer developmental period for insect, fewer adult emerged and reduced adult size were considered as resistant genotypes.

2.2.3.7. Data analysis for gene tagging

For molecular tagging, segregating markers were scored individually in RILs. For YMD trait, linkage analysis was performed using the software Join-Map version 4.0 (Van Ooijen, 2006) [273] at LOD score of 3. The Kosambi mapping function was used to convert recombination frequency into genetic map distance in centimorgans (cM) (Kosambi, 1944) [274]. For mapping bruchid resistance trait RILs derived from TU94-2 X Trombay wild was used.To identify the position of new SSR markers in reported genetic linkage map of blackgram, genotypic scoring data for all 428 markers were taken from Gupta et al., (2008) [4] along with scoring data of 33 new markers (this study) and combined as single input file for linkage analysis using QTL Ici Mapping ver 4.1. (Wang et al., 2016) [275]. Ordering and rippling of grouped markers were done using 'nnTwoOpt' and 'SAD' (sum of adjacent distances) commands, respectively, for obtaining the shortest map distance. The map distance was expressed in centiMorgan (cM) using the Kosambi (1944) [274] map function. Also all 33 new markers were statistically checked for their association to two bruchid resistance trait related parameters i.e percent adult emergence and development period through one way ANOVA analysis.

2.2.3.8. Validation of SSR amplification and SNP-HRM assay

Successful amplification of target SSR sequence and detection of SNP by HRM were confirmed through re-sequencing of PCR products. Based on the HRM results, five SNPs genotyped as homozygous and another five representative SNPs from each type of HRM melt curves and derivative plots were chosen for sequencing. PCR products of both genotypes were re-sequenced on ABI platform (APS Labs). Forward and reverse (reverse complementary) reads of both genotypes were aligned with the TU94–2 contigs through Clustal Omega online multiple sequence alignment tool.

2.2.4. RNA Extraction

2.2.4.1. From immature seeds

Total RNA was extracted from immature seeds of developing pods using Spectrum ™ Plant Total RNA Kit following manual's instruction as described. A total of 2-3 seeds (each seed from different pod/plant) were crushed in lysis solution/2-ME mixture and vortexed for 30 seconds. Samples were incubated at 56 ˚C for 3-5 minutes. Cellular debris were removed by centrifugation at maximum speed (14,000 rpm) for 3 min. Lysate supernatants were pipetted into a filtration column (blue retainer ring) seated in 2-ml collection tube by positioning the pipette tip at the bottom of the tube but away from the pellet. Cap was closed and centrifuged at maximum speed for 1 min to remove residual debris. Clarified flow-through was saved. of Binding solution (500 µl) was added to the lysate and mixed thoroughly by pipetting at least 5 times. This mixture (700 µl) was added into a binding column (red retainer ring) seated in a 2-ml collection tube. Cap was closed and centifuged at maximum speed for 1 min to bind RNA. Flow-through liquid was discarded and collection tube was tapped (upside down) on a clean absorbent paper to drain the residual liquid. Column was returned to the collection tube and remaninig mixture was pipetted in to the column and centrifugation was repeated. For first column washing – 500 µl of wash solution 1 was used, while for second column washing - 500 µl of wash solution 2 was used. Third column washing was done in a similar fashion as that in second washing. Column was dried by centrifugation at maximum speed for 1 min. Column was carefully removed from column –tube assembly and placed on a new fresh tube. 50 µl of elution solution was added on top centre of the column and left for 1 min as such. Column was centrifuged at maximum speed for 1 minute. The flow-through eluate containing the pure RNA was stored at – 80 °C for downstream processing. For removing traces of genomic-DNA, 5µl each of DNase reaction buffer and DNase enzyme (SIGMA ALDRICH) were added to 50 µl of RNA elute and incubated at room temperature for 15 min.

To stop the reaction of DNase, 5ul of stop solution was added to same tube and incubated at 70 ˚C for 10 min.

2.2.4.2. From leaves

Total RNA was extracted from 100 mg of young leaves using Plant RNA Sure Kit (Qiagen) following kit manual described as follows. The leaf tissue was ground in liquid nitrogen (each leaf from different plant) and transferred to 1.5 ml ependorff tubes. RLC lysis solution/2-ME mixture (500 µl)was added and vortexed for 30 seconds. Transfered the tissue slurry to filtration column (purple retainer ring) seated on 2-ml collection tube by positioning the pipette tip at the bottom of the tube but away from the pellet. Cap was closed and centrifuged at maximum speed for 1 min to remove residual debris. Clarified flow-through was saved. 500 µl of 100% chilled ethanol was added to lysate and mixed thoroughly by pipetting at least 5 times. This mixture (700 µl)was added into a binding column (pink retainer ring) seated on a 2-ml collection tube. Cap was closed and centifuged at maximum speed for 1 min to bind RNA. Flow-through liquid was discarded and collection tube was tapped (upside down) on a clean absorbent paper to drain the residual liquid. Column was returned to the collection tube and remaninig mixture was pipetted in to the column and centrifugation was repeated. For first column washing – 500 µl of wash solution 1 was was used, while for second column washing - 500 µl of wash solution 2 was used. Third column washing was done in a similar fashion as that in second washing. Column was dried by centrifugation at maximum speed for 1 min. Column was carefully removed from column – tube assembly and placed on a new fresh tube. Elution solution (50 µl)was added on top centre of the column and left for 1 min as such. Column was centrifuged at maximum speed for 1 minute. The flow-through eluate containing the pure RNA was stored at – 80 °C for downstream processing. For removing traces of genomic-DNA, 5µl each of DNase reaction buffer and DNase enzyme (SIGMA ALDRICH) were added to 50 µl of RNA elute and

incubated at room temperature for 15 min. To stop the reaction of DNase, 5ul of stop solution was added to same tube and incubated at 70 ˚C for 10 min.

From leaves (Trizol method followed by LiCl precipitation)

About 100 mg of fresh leaf tissue was homogenized in 1ml of Trireagent (Invitrogen) using pestle and mortor. Homogenate was centrifuged at 12000g for 10 min at 2-8 °C and supernatant was transferred to a new tube. To this, 200 µl of chloroform was added, shaken vigorously and allowed to stand for 5-10 min at room temperature. Sample was centrifuged at 12000g for 15 min at 2-8°C. At this stage, top aqueous phase containing RNA was taken carefully with the help of pipette without disturbing middle layer of DNA and lower red phase containing proteins. Iso-propanol (500 µl) was added to this aqueous phase, mixed well and allowed to stand for 5-10 min at room temperature. Sample was centrifuged at 12000g for 10 min at 2-8 ˚C. Supernatant was discarded and pellet was washed with 1ml of 75% ethanol through centrifugation at 7500g for 5 min at 2-8 ˚C. Supernatant was discarded and pellet was air dried in laminar air flow for 5-10 min. To the pellet, 100 µl of nuclease free water was added and tube was kept at room temperature for 10 min for dissolution of RNA pellet. To the RNA elute (50 µl), ~13 µl of 8 M LiCl was added to make a conc. of 2 M and incubated at -80 ˚C for 10 min. Then, this RNA elute was centrifuged at maximum speed of 16000 g for 10 min. Supernatant was decanted carefully without disturbing pellet with the help of pipette. Pellet was washed with 80% and 70% ethanol by centrifugation at maximum speed for 5 min. RNA pellet was air dried in laminar air flow and 50 µl of nuclease free water was added.

2.2.5. cDNA synthesis and Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR)

After RNA extraction and treatment with DNase enzyme, nearly 2 µg of total RNA was used for further processing. The cDNA first strand synthesis was done using a PrimeScript™ RT-PCR Kit (Clontech, USA) and quantitative real time PCR was performed following manufacturer's instructions given in SYBR1 Premix ExTaq™ (TliRNAse H Plus) (Clontech,USA). The PCR amplification was carried out in Rotor-Gene-Q Real-Time PCR System (Qiagen, USA) with the following program: 95˚C for 5 min followed by 35 cycles of 94˚C for 30 sec, 62˚C for 20 sec and 72˚C for 20 sec followed by melting of PCR products from 65˚C to 95˚C for DEGs primer and annealing at 56 ˚C for 30 secs for other gene-specific primers. Quantitative real-time PCR experiments for all gene-specific primers were performed twice with three biological replicates run in triplicate. The relative gene expression levels were calculated by relative quantification (RQ) through the 2-ΔΔCt method (Livak and Schmittgen, 2001) [276].

CHAPTER 3

RESULTS

CHAPTER 3.1

Development of genomic resources for blackgram

This chapter describes the development of genomic resources for blackgram from transcriptome datasets of wild and cultivated blackgram and draft whole genome sequence of mungbean. The wild blackgram studied in this project was Trombay wild (TW) accession INGR10133 which was found growing at Trombay hills by Dongre et. al. (1996) [205] and investigated further by Souframanien et al. (2010) [12].

3.1.1 Transcriptome sequencing, *de-novo* assembly, characterization and annotation of TW

Total RNA was extracted from immature developing seeds of TW (different plants) and cDNA library was prepared.

3.1.1.1 Illumina paired end sequencing and *de novo* assembly of transcriptome

The transcriptome sequencing of Trombay wild (TW) generated 2.97 GB data. This 2.97 GB data comprised of 2,970,339,385 bases and 19,690,124 HQ reads (>20). Sequencing reads in FASTA format were submitted to the NCBI short read archive (SRR 5931432, SRX3091690 and study accession SRP115376). A total of 40178 TCs were obtained from HQ reads assembly with length ranging from 280 to 6445 bp and with an average length of 446 bp (Fig 3.1.1). Contigs with lengths ranging from 200 to 299 bp, 300 to 399 bp, >500 bp and > 1000 bp accounted for 15.9% (6398), 39.7 % (15966), 26.6 % (10694) and 2.2 % (905) of TCs, respectively.

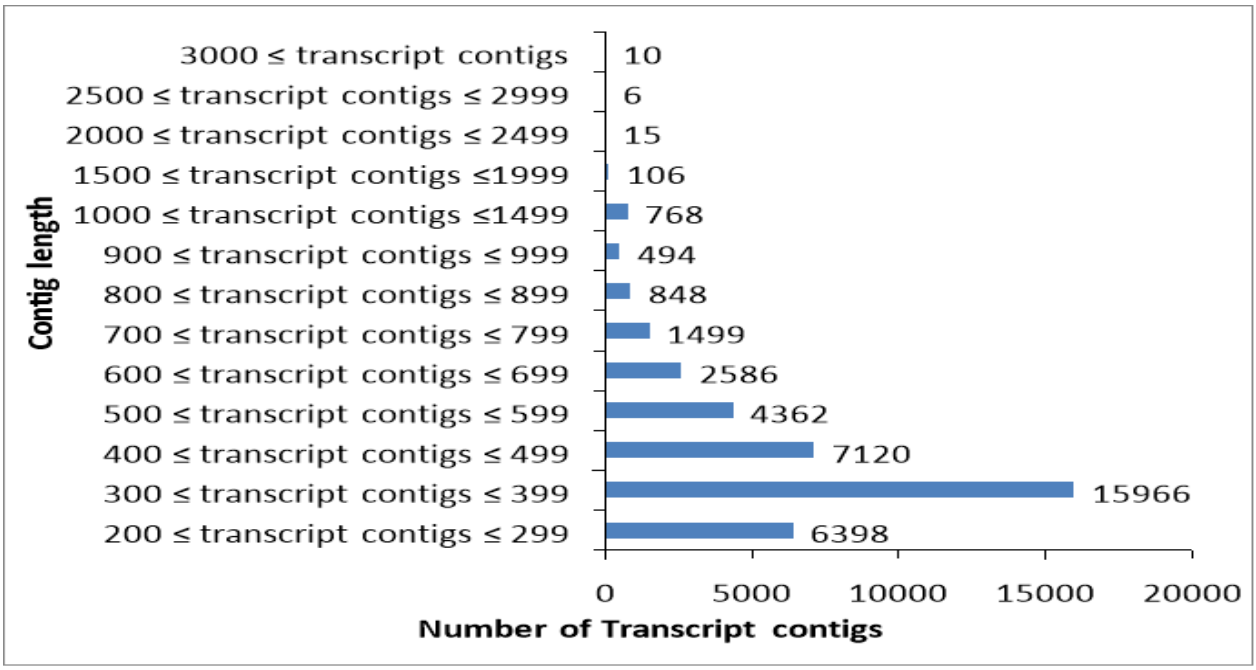


Figure 3.1.1. Distribution of TW transcript contigs based on length.

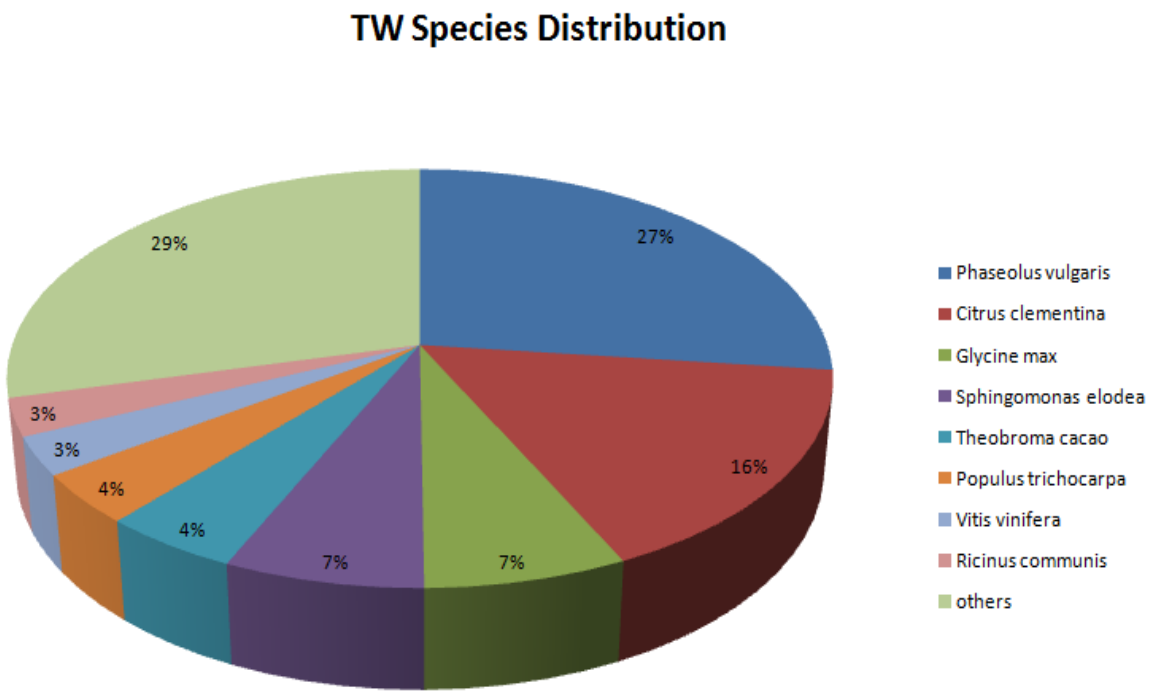


Figure 3.1.2. Wild blackgram (TW) CDS distribution based on homology to different species.

3.1.1.2. Functional annotation

A total of 38,753 (96.4%) CDS were predicted from 40,178 TCs with the help of ORF Finder. After annotation using the NCBI non-redundant (nr) protein database, 28,984 (74.79%) CDS showed significant hits to known proteins and 9769 (25.2%) CDS had no hits in the database. Most of the CDS showed similarity with *Glycine max* (29%), *Phaseolus vulgaris* (27%) and *Citrus clementina* (16%) (Fig 3.1.2). Functional annotation by Blast2GO resulted in classification of 28,106 CDS into gene ontology classes with functional terms. Majority of CDS fall into classes associated with molecular function (11,544, 41.1%), followed by biological process (10,330, 36.8%) and cellular components (6232, 22.2%) (Fig 3.1.3). Bulk of the CDS from molecular function class was assigned to binding and catalytic activities. Within biological process category, CDS majorly lie in cellular processes, metabolic processes followed by establishment of localization (Fig 3.1.4). A total of 140 unique KEGG pathways were predicted for CDS which includes purine metabolism (524), starch and sucrose metabolism (217), glycolysis/gluconeogenesis (183), pyrimidine metabolism (181) and cysteine and methionine metabolism (160).

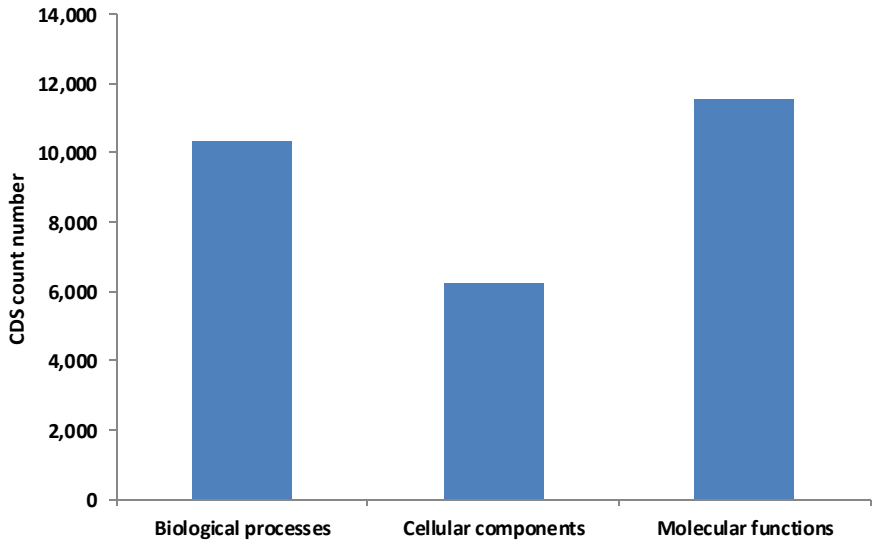


Figure 3.1.3. Wild blackgram CDS distribution into different classes.

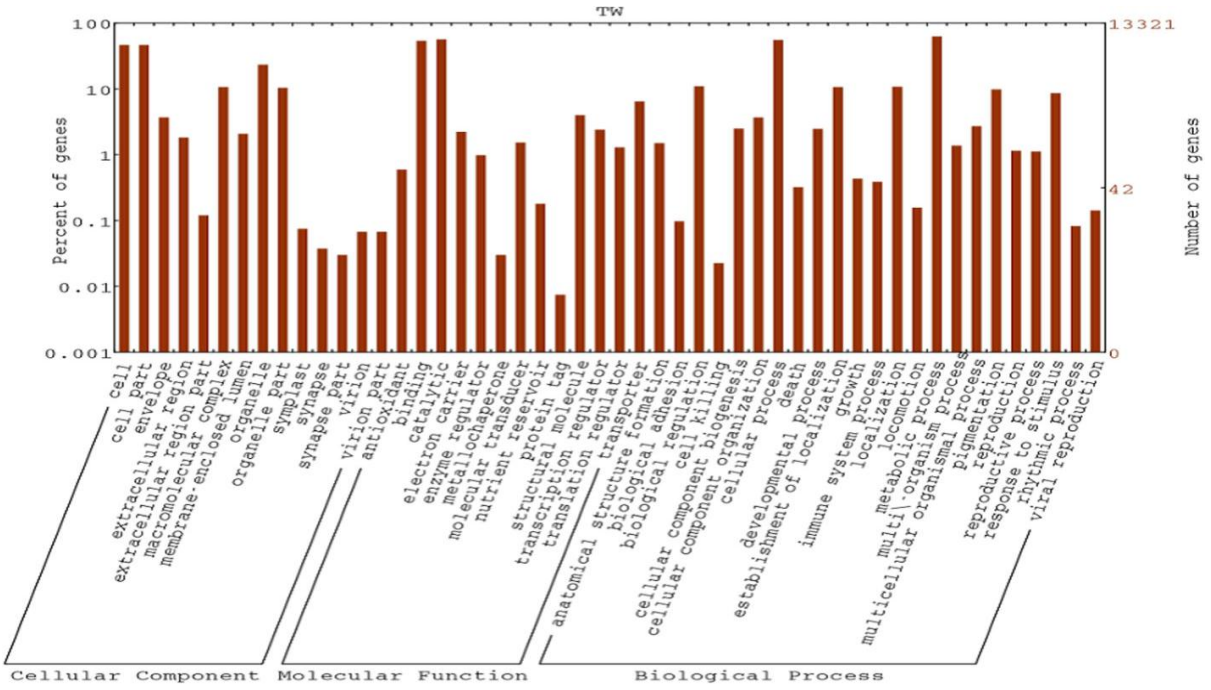


Figure 3.1.4. GO Annotation analysis for wild blackgram CDS using BLAST2GO algorithm.

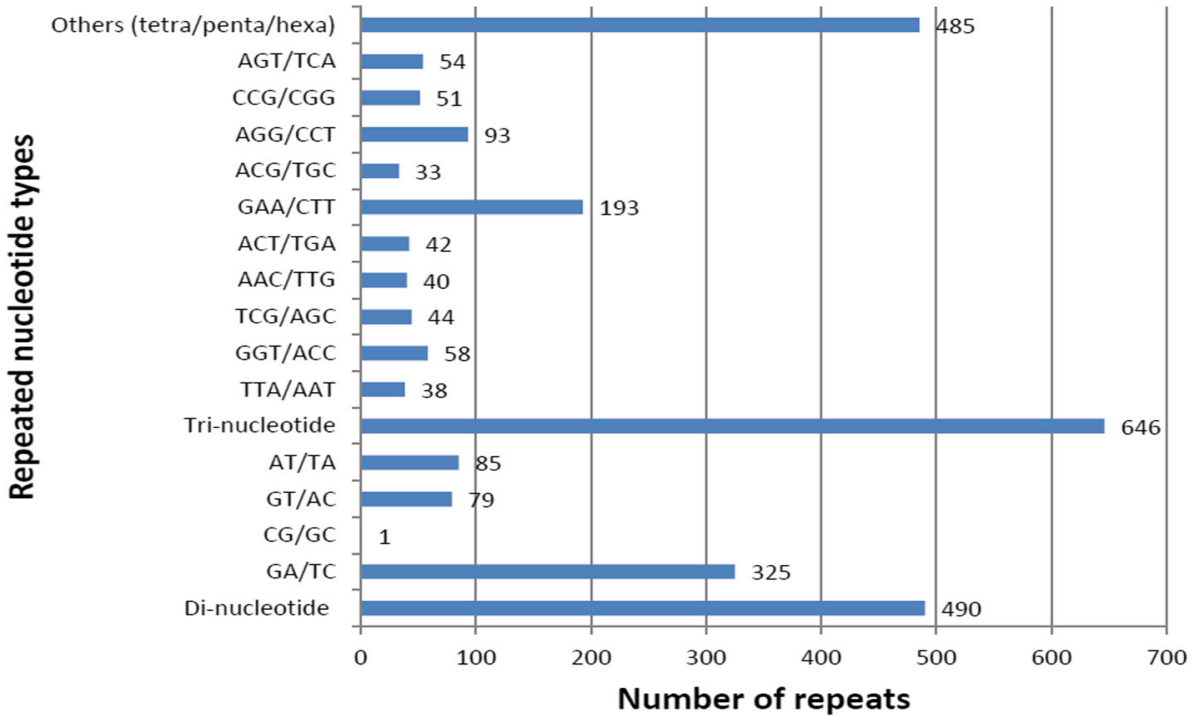


Figure 3.1.5. Simple sequence repeat (SSR) types and distribution in the wild blackgram transcriptome

3.1.1.3. Identification, frequency and distribution of SSRs in blackgram TCs

A total of 1621 SSRs were predicted in 1339 (3.3%) TCs with an average frequency of 1/11.1 kb. Majority of repeats were tri-nucleotides (646,39.9%) followed by di-nucleotides (490,30.2%) and together constituted 70.1% of the total SSRs (Fig 3.1.5). GAA/CTT repeats constituted the largest class (29.9%), followed by AGG/CCT (14.4%) and GGT/ACC repeats (9%). Repeat lengths were observed to vary from 6 to 47 in di-nucleotides, 5–52 in trinucleotides, 4–18 in tetra-nucleotides, 4–30 in penta-nucleotides and 3–10 in hexa-nucleotides. In di-nucleotide SSR class, GA/TC alone constituted 66.3% followed by AT/TA and GT/AC which accounted for 17.3 and 16.1% respectively. while CG/GC occured only once (0.2%) in the transcriptome dataset. SSR motifs found in CDS and UTRs constituted 977 (61.6%) and 610 (38.4%), respectively, of which 19.5 % lie in 5'UTR while 18.9% lie in 3'UTR (Fig 3.1.6). CDS showed predominance of tri-nucleotides (444,44.5%) followed by dinucleotide (242,24.8%) repeats. Thirty four SSR containing contigs showed no ORF.

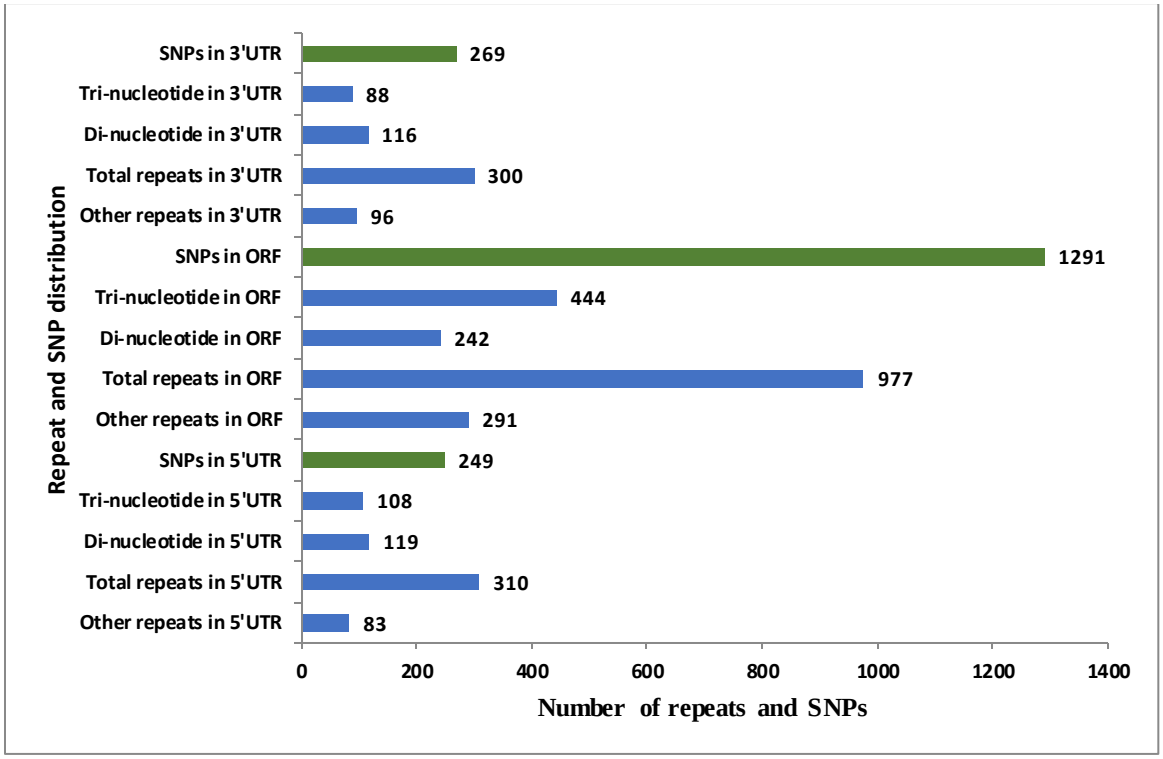


Figure 3.1.6. Frequency and distribution of SSRs and SNPs in coding sequence and untranslated region (UTRs) of blackgram TCs

3.1.1.4. Identification and characterization of SNPs

A total of 1844 SNPs (17 heterozygous and 1828 homozygous) were predicted by aligning TW and TU94-2 HQ reads/TCs with read depth ranging from 20 to 200. Among 1844 SNPs, 1291 (70%) lie in CDS and 518 (28%) are located in untranslated regions (269,14.6% in 3′-UTR and 249,13.5% in 5′-UTR). For 35 SNPs bearing TCs, no ORF was predicted by ORF Finder (Fig 3.1.6). The numbers of transition (Ts) and transversion (Tv) type of SNPs were 1129 (61%) and 716 (39%) respectively, with a Ts/Tv ratio of 1.57 for TW. Among transition type SNPs, A/G (16.4%) type SNP is most abundant followed by T/C (15.3%),G/A (14.7%) and C/T(14.6%) types (Fig 3.1.7). While in transversion substitutions, the most frequent is T/A (5.7%) and G/C (5.7%) followed by A/C (5.1%), A/T (4.9%), C/G (4.8%), C/A (3.9%) and G/T (3.8%).

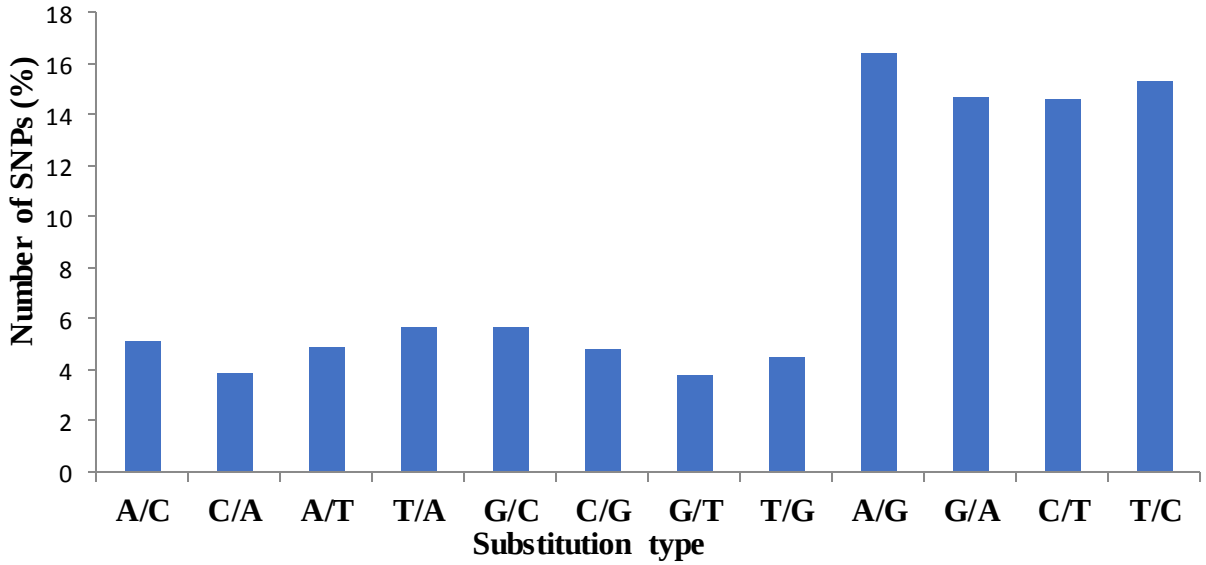


Figure 3.1.7. Classification of blackgram contigs containing SNPs into different substitution types.

3.1.2. Development of Genic-SSR markers from Trombay Wild transcriptome

PCR primers were designed for 1171 SSR motifs out of 1621 SSR motifs and no primers could be designed for 450 SSRs due to inappropriate flanking regions (Table 1 of Annexure I). A total of 100 primer pairs were initially screened (PCR amplification testing) in six

blackgram genotypes (TW, RIL68, Nayagarh, LBG-17, KU96–3, TAU-1). Out of 100 primers screened 58 primers got amplified thus an amplification rate of 58% was achieved. The primers showed a PCR amplification rate of 58% with 31 primers being polymorphic and 19 primers showed null alleles.

3.1.2.1. Genetic diversity analysis in blackgram germplasm

Prescreened 31 SSR primers were used to study the genetic variation in 27 blackgram genotypes yielded 89 alleles with an average of 2.9 alleles/locus. The number of alleles per SSR marker ranged from 1 to 5 with PIC ranging from 0.14 to 0.85 and a mean PIC value of 0.54 (Table 2 of Annexure I). Representative amplification profile of genic-SSR primer (TWSSR62) is shown in Fig 3.1.8. Dendrogram generated based on blackgram genic-SSR markers grouped all 27 genotypes into one major and one minor cluster (Fig 3.1.9). Dendrogram showed highest similarity index between DPU88–31 and IPU02–43 genotypes and lowest between TU94–2 and LBG693 genotypes. The major cluster consisted of 23 genotypes out of 27 with wild accession (TW) appearing distinct from others and the minor cluster comprised of only 4 genotypes.

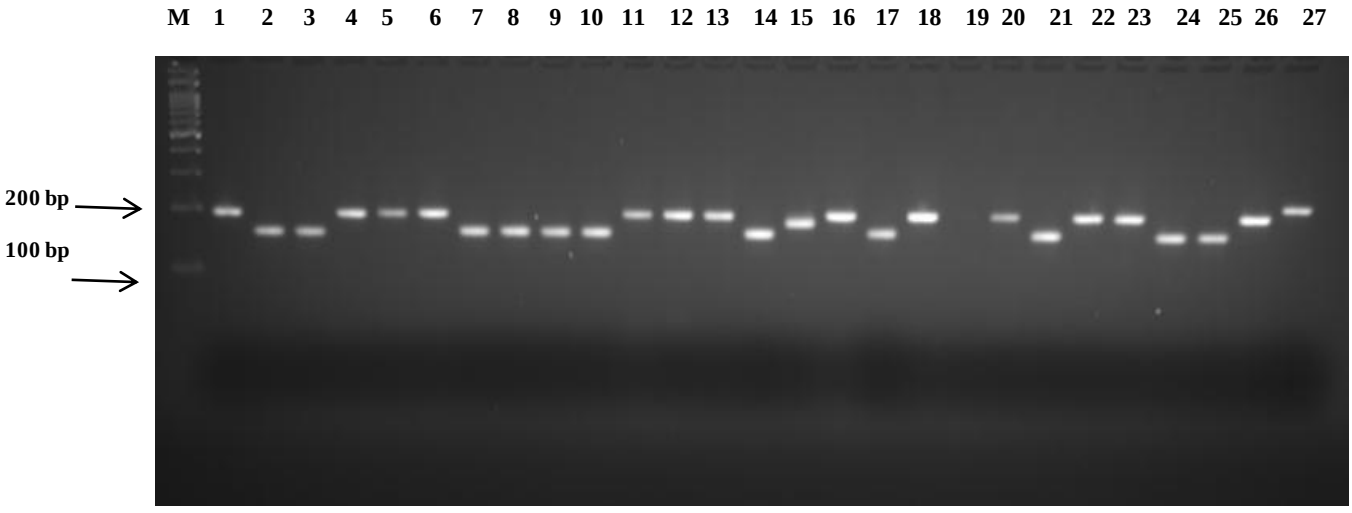


Figure 3.1.8. PCR amplification profile of genic-SSR marker TWSSR62 in 27 blackgram genotypes (Lane 1 to 27 are blackgram genotypes as given in Table 2.1)

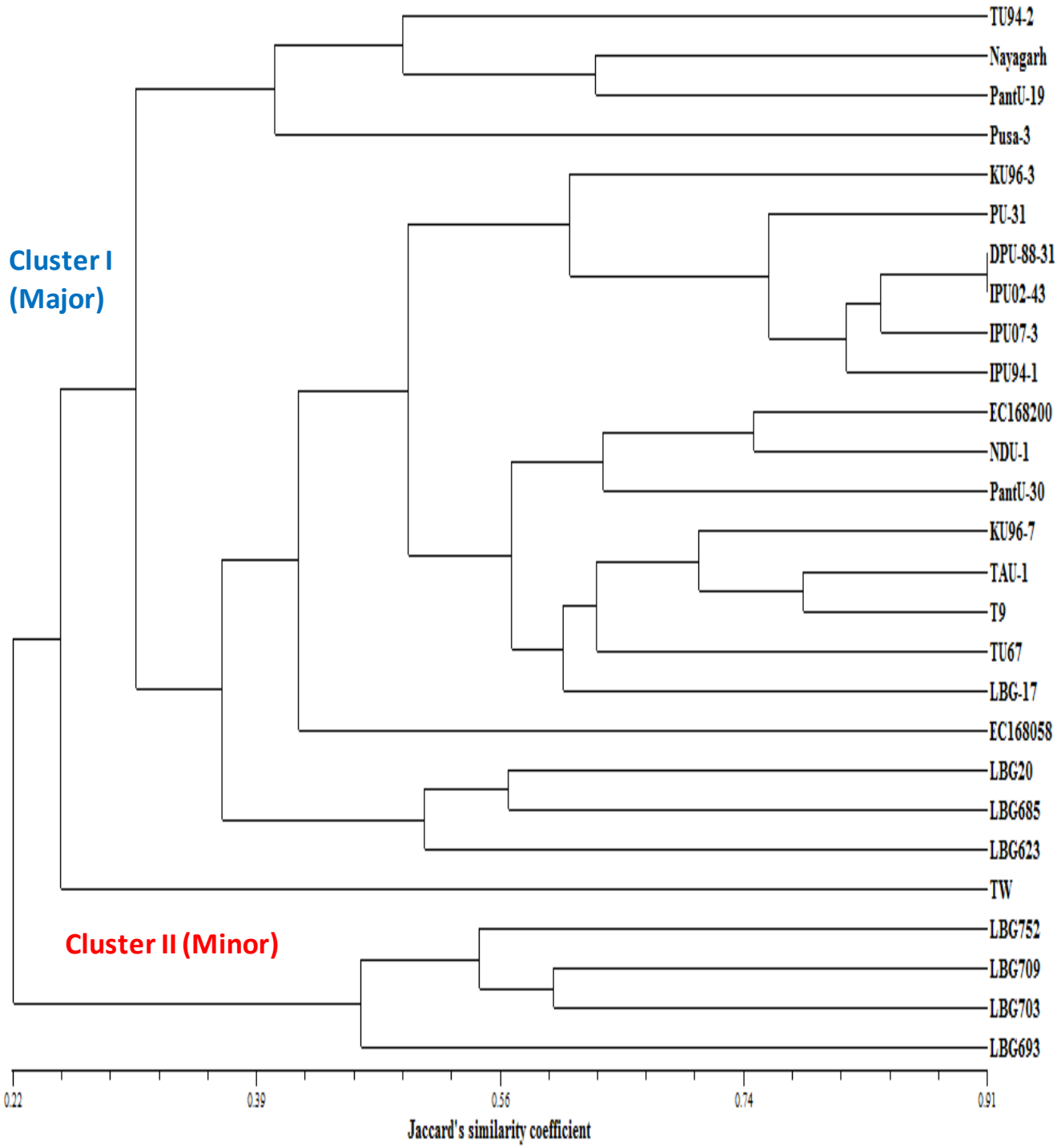


Figure 3.1.9. Dendrogram showing the genetic relationship among 27 blackgram genotypes (accession of genotypes from lane 1 to 27 are given in Table 2.1)

3.1.3. Development of Genic-SNP markers from Trombay wild transcriptome

3.1.3.1. Development of genic-SNP markers

PCR primers were designed for 1749 SNPs out of 1844 SNP loci and no primers could be designed for the remaining 95 SNP loci because their flanking sequences were either too short or the nature of sequence did not fulfil the necessary criteria for primer design (Table 3 of Annexure I). A total of 100 primer pairs were used for screening TW and TU94–2 genomic DNA and an amplification rate of 85% was observed. A total of 79 primers exhibiting monomorphic banding pattern between TW and TU94-2 were used for SNP genotyping by HRM analysis.

3.1.3.2. SNP genotyping by HRM assay

Among 79 SNP primers, 71 showed amplification in PCR-HRM reaction. Raw HRM profiles were filtered for quality check and then analysed through HRM analysis software. A total of 54 primers amplified with a CT value $\leq$ 30 with amplification efficiency > 1.4. Raw HRM profiles were processed to get melt curves, normalized graph and derivative plots with the help of default software. The remaining 17 primers which did not comply with quality check parameters were also analysed. A validation rate of 78.8% (57 SNPs) was obtained for HRM assay, thus confirming transcriptome sequencing results. Single melting transition in normalized HRM melt curves and single narrow peak in derivative plot of validated SNPs (57 SNPs) indicated their homozygous nature. Twelve SNPs showed aberrant curves with multiple melting phases or peaks which suggested heterozygosity. Three SNPs did not show apparent melting temperature difference between both the genotypes. Representative SNPs showing homozygous and heterozygous pattern for normalized HRM curve and derivative plot are shown in Fig 3.1.10. Agarose gel separation of HRM PCR products did not show primer-dimers and non-targeted amplicons (Fig 3.1.11).

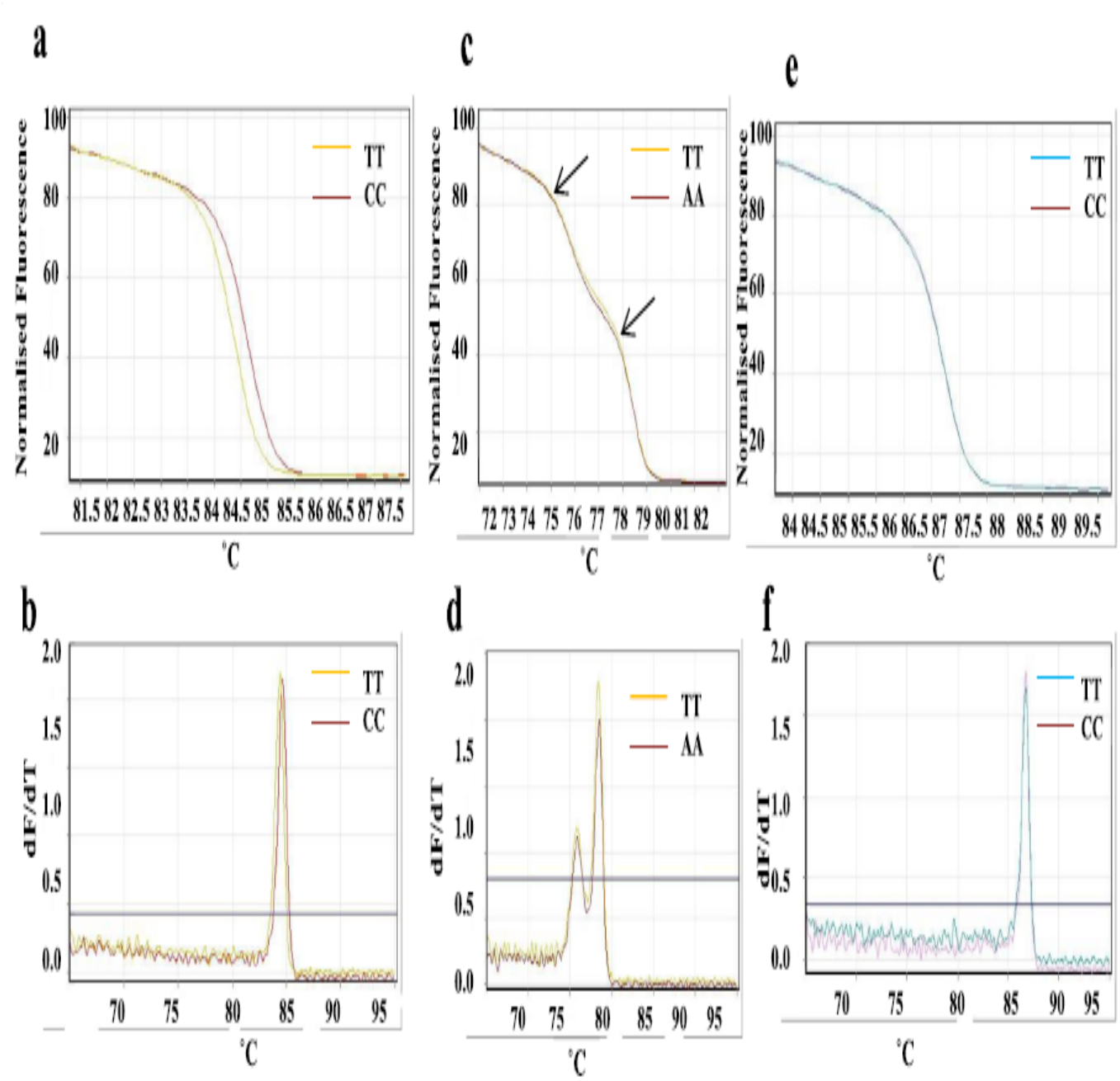


Figure 3.1.10. High resolution melting (HRM) curve profiles of three HRM amplicons. a) normalised plot of two homozygous genotypes showing one melting domain; b) derivative plot of a; c) normalised plot of two genotypes heterozygous at one SNP site (T/A) showing two melting domains (arrows); d) derivative plot of c; e) normalised plot of two genotypes with no apparent SNP site (T/C) showing single curve, f) derivative plot of e.

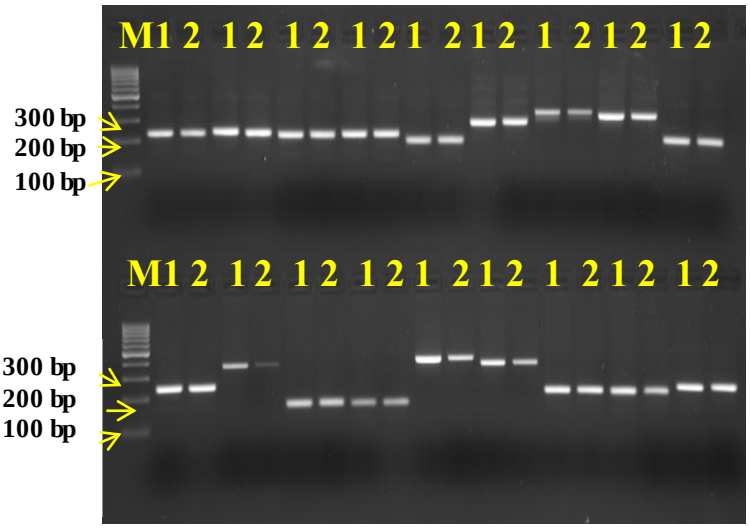


Figure 3.1.11. HRM PCR products of TW (displayed as "1") and TU94-2 (displayed as "2") genotypes for TWSNP primers (I Row-TWSNP82, 84, 85, 86, 88, 89, 91, 92, 93 and II Row-TWSNP94, 95, 96, 97, 98, 99, 100, 101, 102) resolved on 3% agarose gel.

3.1.3.3. Melting temperature difference, genotypes and SNP class

Class I, II and III SNPs were easily distinguished from each other by a considerable shift in T_m between homozygous wild-type and mutant but for class IV SNP, T_m shift was very less. Minimum T_m difference observed for class IV SNP was 0.02. In this study, 31 SNPs could be classified under class I with a T_m difference range of 0.15 to 0.51, 15 SNPs appeared to fall in class II with a T_m difference range of 0.03 to 0.41, 2 and 8 SNPs into class III and IV respectively, with their corresponding T_m difference range of 0.18 to 0.35 and 0.02 to 0.12. Three SNPs (TWSNP 76, 85 and 121) showed no apparent T_m difference. In the present study, one complex amplicon bearing 2 SNPs i.e. TWSNP 50 (T/A and A/G) was genotyped as homozygous with both HRM and RNAseq analysis. In this study, class I and IV SNPs could be easily distinguished from each other due to large T_m difference while other classes are difficult to discern due to overlapping T_m difference.

3.1.3.4. SNP substitution types and its effects on protein function and kinectics

Among HRM validated SNPs, 13 SNPs were non-synonymous missense substitutions, which were expected to have altered function and/or structure of the encoded proteins. These 13 non-synonymous SNPs resulted in amino acid replacements in respective contigs (aminoacid substitution given in parenthesis) which showed homology with cytochrome c-type biogenesis protein CcmE (D to N), U-box domain-containing protein 13 (E to D), cytochrome P450 (N to K), pre-mRNA-splicing factor ATP dependent RNA helicase DEAH2 (K to R), protein phosphatase (V to F), mitochondrial fission 1 protein A-like (A to T), dnaJ protein homolog 1 (M to I), cell division cycle protein 48 homolog (T to I), lariat debranching enzyme (R to Q), 2 uncharacterized proteins (T to A and N to D) and 2 hypothetical proteins (I to V and K to N). I-TASSER results showed changes in protein secondary structure due to these missense mutations. Alteration in secondary structure was predicted in five proteins: U-box domain containing protein 13, pre-mRNA-splicing factor ATP dependent RNA helicase DEAH2, uncharacterized protein LOC108329963 (*Vigna angularis)*, lariat debranching enzyme and cell division cycle protein 48 (Table 4 of Annexure I). For U-box domain-containing protein 13, pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH2 and cell division cycle protein 48, changes in secondary structures did not result in changes in ligand / substrate binding preference or ligand binding sites. But for uncharacterized protein LOC108329963 (*Vigna angularis)* and lariat debranching enzyme, changes in number of helices and strands resulted in alteration in preferred ligand. Helices and strand number remained same with the change for preferred ligand in 6 genes: cytochrome c-type biogenesis protein CcmE, uncharacterized protein LOC108329961 (*Vigna angularis*), cytochrome P450, hypothetical protein LR48_Vigan10g050400 (*Vigna angularis)*, hypothetical protein VIGAN_01393500 (*Vigna angularis* var. *angularis)* and dnaJ protein homolog 1. For example, native form of

cytochrome P450 binds preferably to ligand zinc but the variant form showed binding preference with calcium. Missense mutations with no changes in secondary structure and ligand preference were observed in two genes- protein phosphatase 2C 55 and mitochondrial fission 1 protein A-like. For TWSNP904, amino-acid lysine (K) at 12^{th} position of TU94-2 protein got replaced with threonine (T) in TW protein. I-TASSER results showed no changes in secondary structure due to change from lysine to threonine in peamaclein but alteration in ligand binding preference was seen. In TU94-2, peamaclein protein lysine acted as ligand binding site residue but in TW, threonine (12^{th} position) was not part of ligand binding site. Peamaclein is a novel allergen identified in peach, resistant to digestion and homologous to a potato antimicrobial peptide. TWSNP1304(1), bearing cell division cycle protein 48 contig showed changes in secondary structure with no changes in ligand binding kinectics due to replacement of amino acid threonine (T) (TU94-2 genotype) with isoleucine (I) (TW accession). In the same protein (cell division cycle protein 48), strands were seen at three amino-acid residues (positions 459, 460 and 461) in TU94-2 but coils were seen in corresponding positions in TW protein.

3.1.3.5. SNP genotyping and genetic variation analysis by HRM assay

SNP genotyping and genetic variation analysis using 19 genic-SNPs in 27 blackgram genotypes were performed. Detailed characterization of all 19 genic-SNPs studied in blackgram germplasm is given in Table 5 of Annexure I. All 27 blackgram genotypes were genotyped as homozygous for 18 genic-SNPs and heterozygous for only one SNP (TWSNP1304(1)). Normalized HRM curves and difference plots for TWSNP33 SNP marker showed homozygous pattern (Fig 3.1.12). For TWSNP1304(1) SNP marker, wild accession (TW) and 14 genotypes exhibited heterozygous melt curves (multiple melt phases/peaks in melt curves) while remaining 12 genotypes showed homozygous profile (Single melt phase/peak) (Fig 3.1.13).

Analysis of normalized HRM curves with respect to two controls (TU94-2 and TW) resulted in distribution of all genotypes into 3 groups: group 1: TU94-2, group 2: TW and group 3: Variation (Table 6 of Annexure I). For 2 SNP markers, TWSNP33 (A/C) and TWSNP59 (T/C), all genotypes grouped with control 1 i.e. TU94-2, thus indicating presence of allele A and allele T, respectively. Likewise, for TWSNP73 (C/G) and TWSNP93 (A/G) primer, 24 genotypes grouped with TU94-2, were thus genotyped as C and A for respective SNPs. In contrast, for TWSNP 40a (C/T) and TWSNP859 (T/A), majority of genotypes grouped with control 2 i.e TW and were assigned as C and T genotypes for respective SNPs. Apart from distinguishing genotypes based on fluorescence signals (acquired during PCR product melting), an attempt was made to differentiate genotypes based on differences in melting temperature (T_m) of PCR products. Grouping of genotypes based on melting temperature difference was less effective due to wide distribution of T_m beyond the T_m of both TW and TU94-2. Both methods of differentiating genotypes, method 1 (fluorescence data) and method 2 (melting temperature) were compared through alignment of fluorescence data with melting temperature data. Both methods of analysis showed colinearity and thus, are in agreement with each other (Fig 3.1.14). SNP markers TWSNP33, TWSNP59, TWSNP73 and TWSNP1771 appeared highly efficient as they differentiated all genotypes into 2 controls with no variants found. In contrast, SNP markers TWSNP1304(2), TWSNP4407 and TWSNP904 showed less efficiency in differentiating genotypes as majority of genotypes appeared as variants. Further analysis of genetic closeness among 27 genotypes through UPGMA-based dendrogram resulted in grouping of 26 genotypes into two major clusters with wild accession, TW forming a separate OTU (Fig 3.1.15). The pairwise Jaccard's similarity coefficient ranged from 0.15 to 0.78. Genotypes TU67 and LBG-20, with highest similarity coefficient (0.78) appeared genetically closed. Wild accession (TW) and cultivar TU94-2 were most diverse with very low similarity coefficient of 0.15. Cluster I comprised

of 11 genotypes in three subclusters (Ia, Ib and Ic). The subcluster Ia consisted of 4 genotypes TU94-2, Nayagarh, EC168058 and Ku96-3. The subcluster Ib comprised of genotypes T-9, LBG17, TU67, LBG20 and LBG752. While only 2 genotypes PantU-19 and PantU-30 were grouped under subcluster Ic. Similarly, cluster II consisted of 15 genotypes in 3 subclusters (IIa, IIb and IIc). Sub-clusters IIa, IIb and IIc comprised of 2, 9 and 4 genotypes, respectively (Fig. 3.1.15).

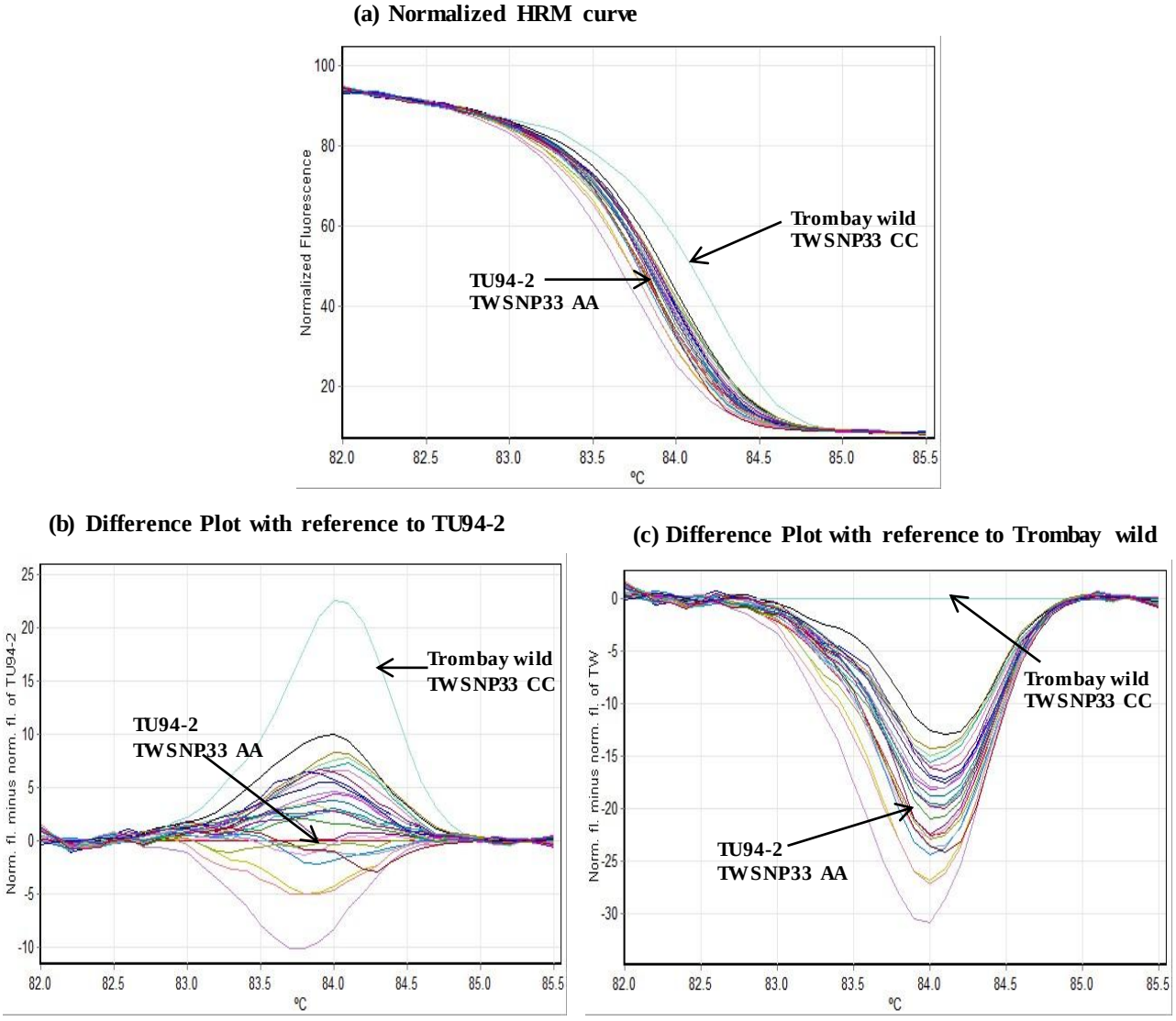


Figure 3.1.12. SNP genotyping with HRM analysis of TWSNP33 A/C functional marker in a set of 27 blackgram genotypes. (a) Normalized HRM curve analysis: all genotypes showed homozygous melt curve pattern, (b) HRM differential plot using"TU94-2" as reference genotype; (c) HRM differential plot using "Trombay wild" as reference genotype.

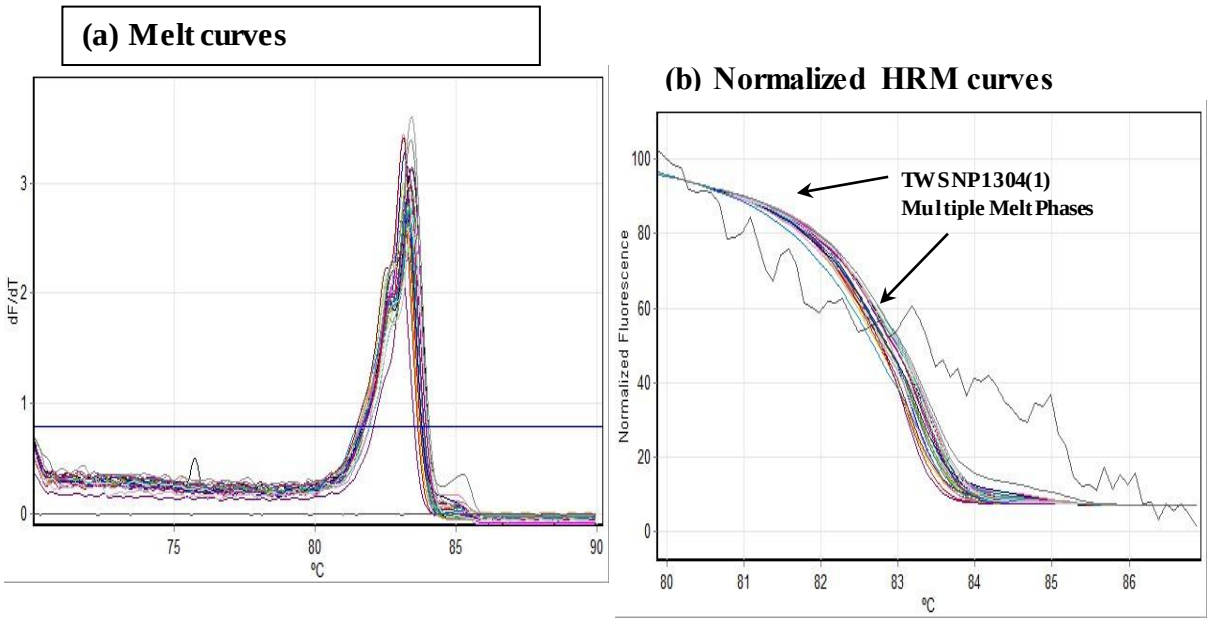


Figure 3.1.13. SNP genotyping with HRM analysis of TWSNP1304(1) SNP in a set of 27 blackgram genotypes. (a) Melt peak plots: multiple melt peaks were observed for some genotypes (b) Normalized HRM curve analysis: Multiple melt phases were observed for some genotypes, thus indicating heterozygous pattern.

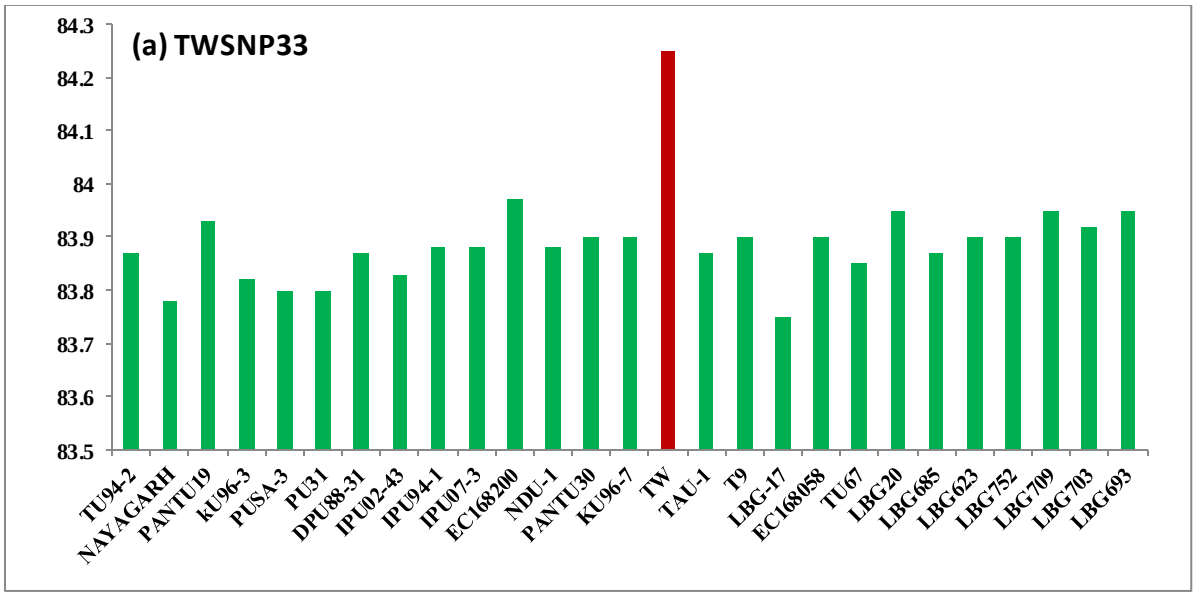


Figure 3.1.14. Graph showing similarity between normalized HRM and melt peak results for TWSNP33 SNP marker. Y-axis denotes melting temperature and X-axis denotes blackgram accessions. Melting temperature results of 27 genotypes were plotted against Y-axis and normalized HRM flourescence results are shown in colours. Green colour : TU94-2 (Cultivar) and Red colour : TW (Wild accession)

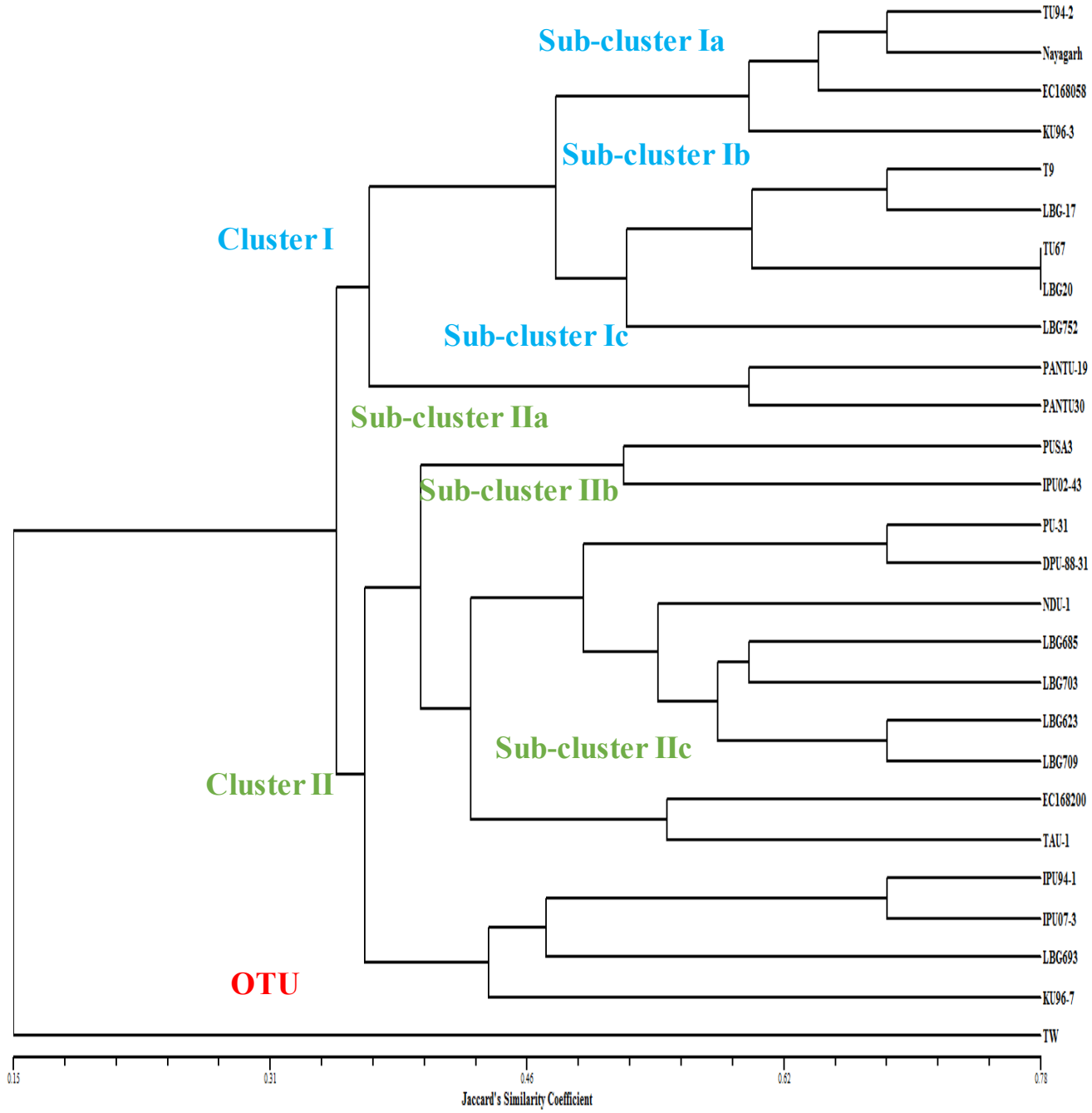


Figure 3.1.15. UPGMA based dendrogram constructed from SNP genotyping of 19 Genic-SNPs through HRM analysis

3.1.3.6. Validation of target SSR motif and SNP amplification through re-sequencing

PCR products amplified using five SSR primers TWSSR 13, 57, 62, 86 and 96 were re-sequenced, which revealed the presence of repeat sequences (GAA)10, (AG)8, (TC)19, (GAT)11 and (TCA)6, respectively. This is similar to the original transcriptome sequence which indicated successful target amplification. HRM PCR products were sequenced to confirm the HRM assay results. Sixteen amplicons were re-sequenced and 9 were found to be homozygous similar to transcriptome sequence data.

3.1.4. Development of Genic-markers from differentially expressed genes (DEGs) of Trombay wild transcriptome

A total of 67 genic-primers were designed for 41 up-regulated DEGs with 5 to 12 fold change and 26 down-regulated DEGs with -5 to -12 fold change in TW with reference to TU94-2 developing seed tissues (Table 7 of Annexure I). A total of 30 primers for up-regulated DEGs and 17 primers for down-regulated DEGs showed amplification in TW and TU94-2 cDNA. Utilization of these primers is described in sub-chapters 3.2 and 3.3.

3.1.5. Development of genic-markers based on Geminivirus-host plant interactions

A total of 177 genic-primers were designed from TU94-2 transcriptome contigs sequences for cellular factors involved in interaction between geminivirus and plant hosts based on literature survey (Table 8 of Annexure I). Out of 177 primers, 110 got amplified in KU96-3, TAU-1 and TU94-2, collectively. Utilization of these primers is described in sub-chapter 3.2.

3.1.6. Development of genomic-SSR markers targeting RGAs of mungbean

3.1.6.1. Transferability of mungbean derived SSR markers in blackgram

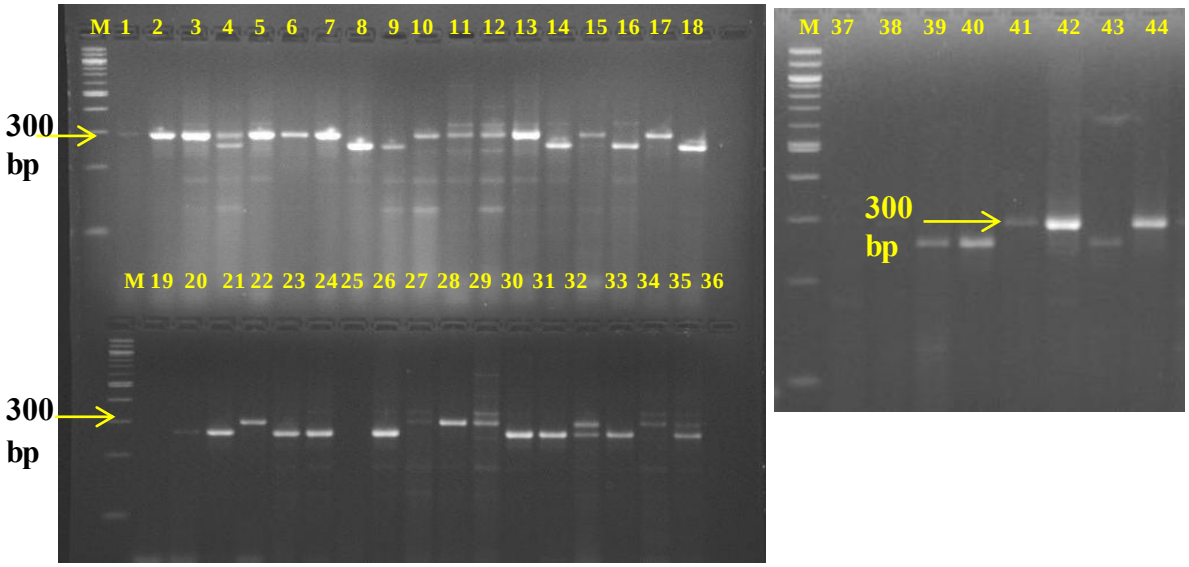
Genomic-SSR markers flanking RGAs found in mungbean scaffolds (accession: PRJNA243847, ID: 243847) were used. Details of all 23 putative resistance proteins with their corresponding homologous mungbean scaffolds resulting from similarity search through tBLASTn algorithm are given in Table 9 of Annexure I. A total of 118 primers were designed for SSRs flanking RGAs lying in mungbean scaffolds (Table 10 of

Annexure I). PCR amplification of mungbean derived primers in blackgram resulted in 70% transferability rate. Thirty randomly selected primers were used for genetic variation analysis in 44 blackgram genotypes differing in disease reaction for YMD and PMD (Table 2.1). These 30 primers designed from different scaffolds of mungbean showed homologies with important resistance genes such as TMV resistance protein N, DNA-damage-repair/toleration protein DRT100, putative disease resistance protein RGA4 and different putative resistance proteins (Table 11 of Annexure I). All 30 SSR primers collectively yielded 90 alleles with an average of 3 alleles/locus. Among 30 primers, 28 show length polymorphism whose PIC ranged from 0.01 to 0.86 with an average of 0.43 (Table 11 of Annexure I) and the remaining 2 primers exhibited 0 PIC value. Di-nucleotide repeats showed PIC values of 0 to 0.86 and for tri-nucleotides 0.11 to 0.67 PIC range was observed. Notably among 30 primers, 24 primers showed presence of null alleles. Representative DNA amplification of blackgram genotypes using mungbean derived resistance gene-SSR primer MRGSSR 118 is shown in Fig 3.1.16. A total of 3 SSR markers MRGSSR12, MRGSSR56 and MRGSSR77 differentiated the YMD resistant and susceptible genotypes. One allele from MRGSSR12 showed amplification in 2 genotypes (EC168200 and IPU02-43) out of 16 resistant and none in the susceptible genotypes studied. Another allele from same primer MRGSSR12 got amplified in 3 (DPU88-31, KU96-7 and IPU07-3) out of 16 resistant genotypes and only 2 out 19 susceptible genotypes. Marker MRGSSR56 showed amplification in 10 out of 16 resistant genotypes and only 3 out of 19 susceptible genotypes. Marker MRGSSR77 marker got amplified in 2 resistant and only in 1 susceptible genotypes studied. Markers MRGSSR12 $((AT)_{13})$, MRGSSR56 $((ATA)_6)$ and MRGSSR77 $((CT)_7)$ sequence lie in mungbean scaffolds JJMO01001477, JJMO01002369 and JJMO01000040 which shared homologies with TMV resistance protein N, DNA damage repair/toleration protein DRT100 and putative disease resistance RPP13-like protein 1, respectively. Interestingly, YMD resistant genotypes

DPU88-31, KU96-7 and Nayagarh appeared to have common genotypes for alleles amplified from markers MRGSSR12 and MRGSSR56. Likewise, Azad-1 YMD resistant genotype showed alleles amplified from both markers MRGSSR12 and MRGSSR77. For PMD trait, 9 markers MRGSSR9, MRGSSR10, MRGSSR12, MRGSSR20, MRGSSR32, MRGSSR51, MRGSSR56, MRGSSR57 and MRGSSR65 differentiated the resistant and susceptible genotypes. Markers MRGSSR9 ($(TTA)_{8,}$ JJMO01000647), MRGSSR10 ($(ATG)_{6,}$ JJMO01001488) and MRGSSR57 ($(TA)_{13,}$ JJMO01000121) got amplified in 9, 5 and 3 out of 12 PMD resistant genotypes and in 3, 2 and 1 out of 7 susceptible genotypes respectively. Interestingly, markers MRGSSR 32 ($(TA)_{9,}$ JJMO01002369) and MRGSSR65 ($(TAA)_{17,}$ JJMO01000122) got amplified in 5 and 1 PMD resistant genotypes, respectively out of 12 resistant genotypes with no amplification in 7 susceptible genotypes. Markers MRGSSR12 and MRGSSR56 both showed amplification in 5 PMD resistant genotypes and 2-3 susceptible genotypes. These marker sequences located in mungbean scaffolds showed homology with TMV resistance protein N, disease resistance protein RPP8, DNA-damage-repair/toleration protein DRT100, probable disease resistance protein At4g33300 and disease resistance RPP13-like protein 4. PMD resistant genotypes PU-31, LBG623 and LBG709 were found to be common for 4 MRGSSR markers (MRGSSR9,12, 20 and 32). MRGSSR12 marker sequence is located at 119560 base-pair (bp) in scaffold JJMO01002369 (482957 bp) (Fig 3.1.17). Upstream to this MRGSSR12 marker sequence, important genes probably present are L-type lectin domain containing receptor kinase VII.2 at 75846 bp, serine/ threonine protein kinase at 63012 bp and variants of TMV resistance protein N at 10788 bp. While downstream of MRGSSR12 sequence, transcription factor bHLH162 at 147106 bp and LRR Extensin like protein 4 at 336926 bp were found. Similarly, marker sequence of MRGSSR56 was found at 190260 bp in scaffold JJMO01001477 (553992 bp) (Fig 3.1.18). In upstream region of MRGSSR56 marker, probable apyrase 7 at 180567 bp and LRR receptor

like serine/ threonine protein kinase GSO1 at 134584 bp were found. While in downstream region of MRGSSR56 marker, plant UBX domain containing protein 10 at 198596 bp, LRR receptor like serine/ threonine protein kinase at 228995 bp and plant intracellular ras-group-related LRR protein 9-like at 446845 bp were present.

Figure 3.1.16. PCR amplification using MRGSSR118 genomic-SSR marker in 44 blackgram genotypes.



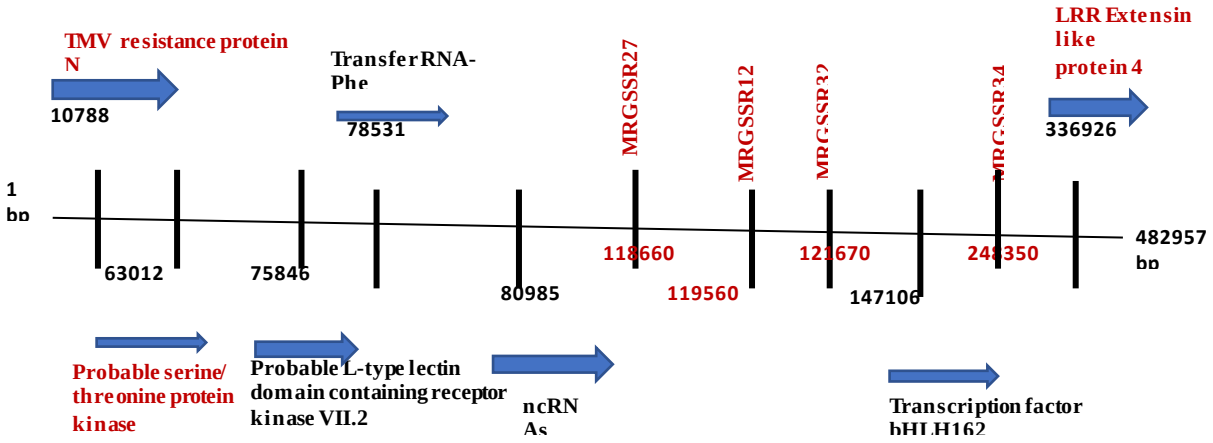
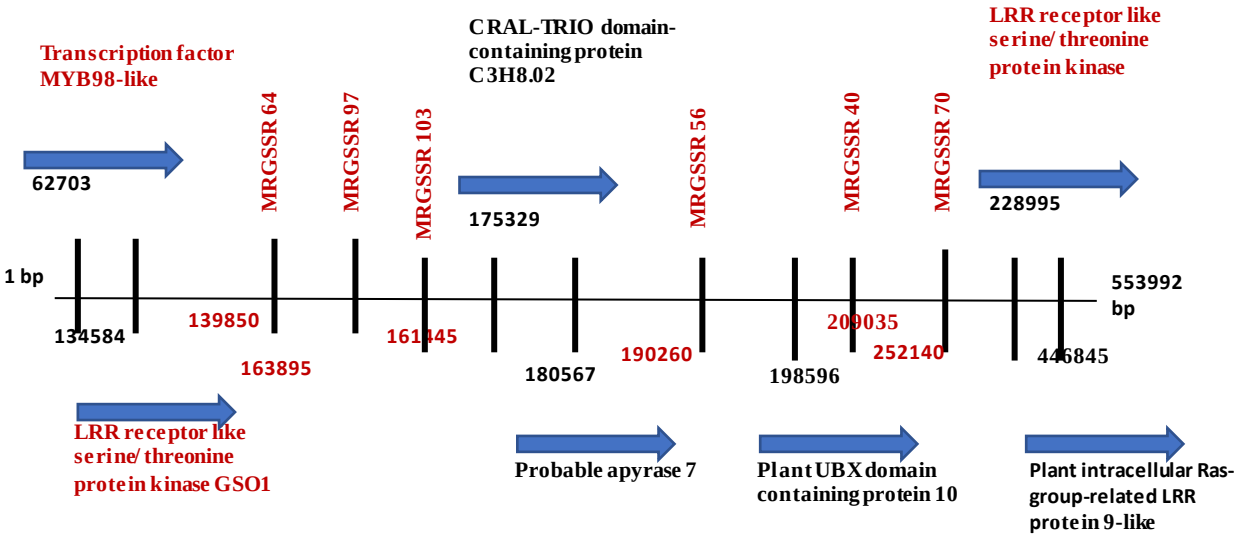


Figure 3.1.17. Physical location of MRGSSR12 primer on their respective mungbean scaffold along with flanking genes.

Figure 3.1.18. Physical location of MRGSSR56 primer on their respective mungbean scaffold along with flanking genes



3.1.6.2. Clustering of blackgram genotypes based on mungbean derived markers

The mungbean derived resistance gene-SSR markers were further utilized for studying genetic variations among 44 blackgram genotypes comprising of 43 cultivars and one wild accession. UPGMA based dendrogram resulted in grouping of 44 genotypes into 7 clusters (Fig 3.1.19). Cluster I comprised of 14 genotypes in 3 sub-clusters (Ia, Ib and Ic). Sub-cluster Ia comprised of 2 genotypes (EC168200-YMD resistant and LBG-17-PMD resistant) and sub-cluster 1b comprised of 5 genotypes of which Pusa-3, LBG 752 and IPU02-043 are resistant to YMD. Sub-cluster Ic consisted of three YMD resistant (DPU88-31, IPU07-3 and KU96-7) and four PMD resistant genotypes (T9, IPU07-3, LBG623 and LBG709). Similarly, cluster II showed grouping of two YMD resistant (KU96-3 and Nayagarh) and two PMD resistant genotypes (LBG20 and KU96-3). Clusters III and IV comprised of YMD resistant genotypes KU96-3, Nayagargh, PLU-1, TU43-1, TU-55-1, PU-31, TU94-2 and Azad-1 along with other blackgram genotypes. Cluster VII displayed grouping of two YMD resistant accessions NDU-1 and PU-19 with a similarity index of 0.27. Accessions LBG693 and IPU02-43 showed the highest similarity coefficient of 0.81. Markers MRGSSR12, 20, 32, 56

and 65 yielded rare/novel alleles, for example MRGSSR56 marker showed occurrence of a different size allele only in two genotypes, wild blackgram (Trombay wild) and NDU-1 and MRGSSR65 marker amplified an allele only in two genotypes LBG17 and IPU02-6 (Fig. 3.1.20).

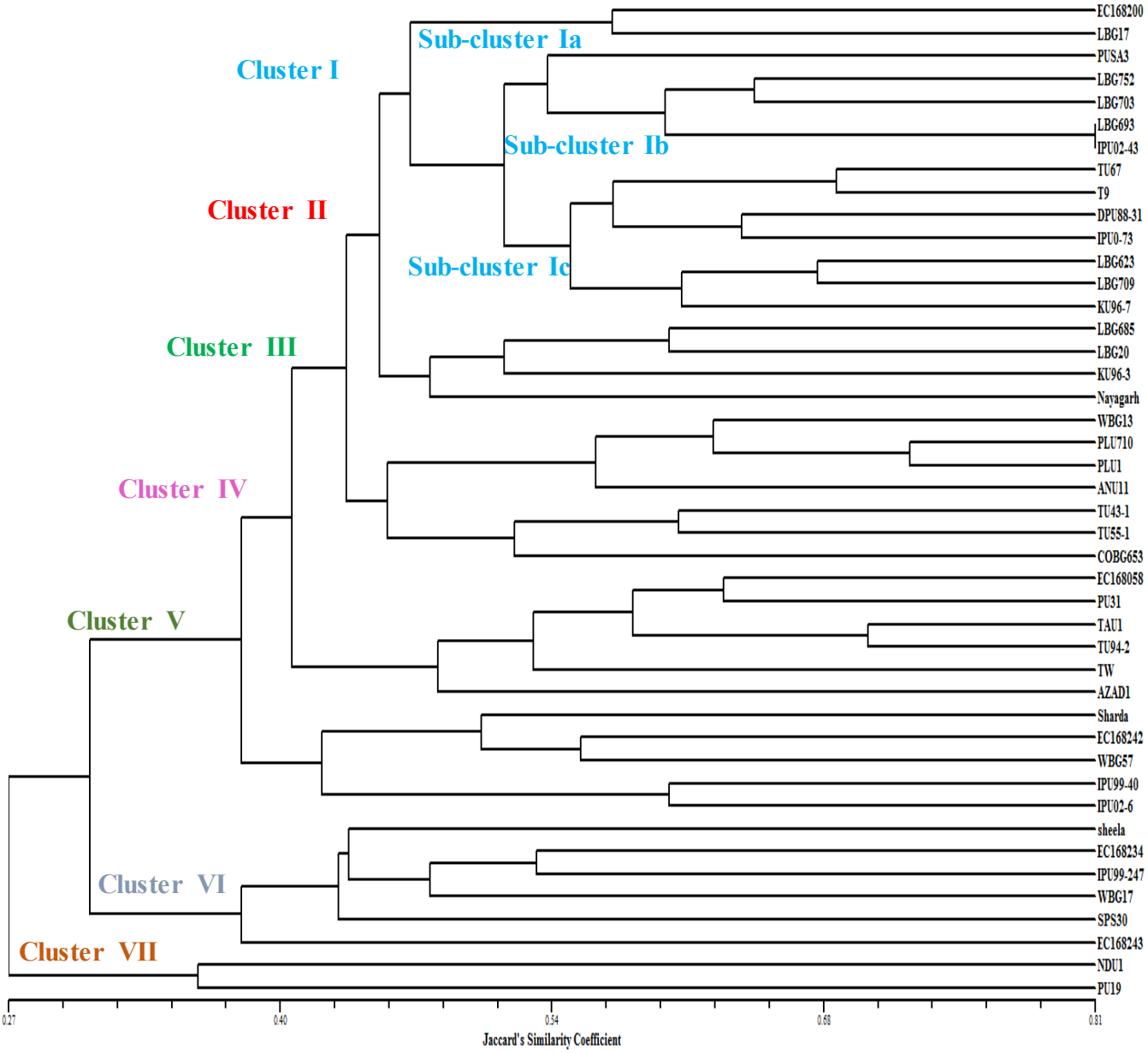


Figure 3.1.19. Dendrogram constructed using Jaccard's similarity coefficient and UPGMA based clustering among 44 blackgram genotypes based on mungbean derived resistance genes-SSR polymorphic markers.

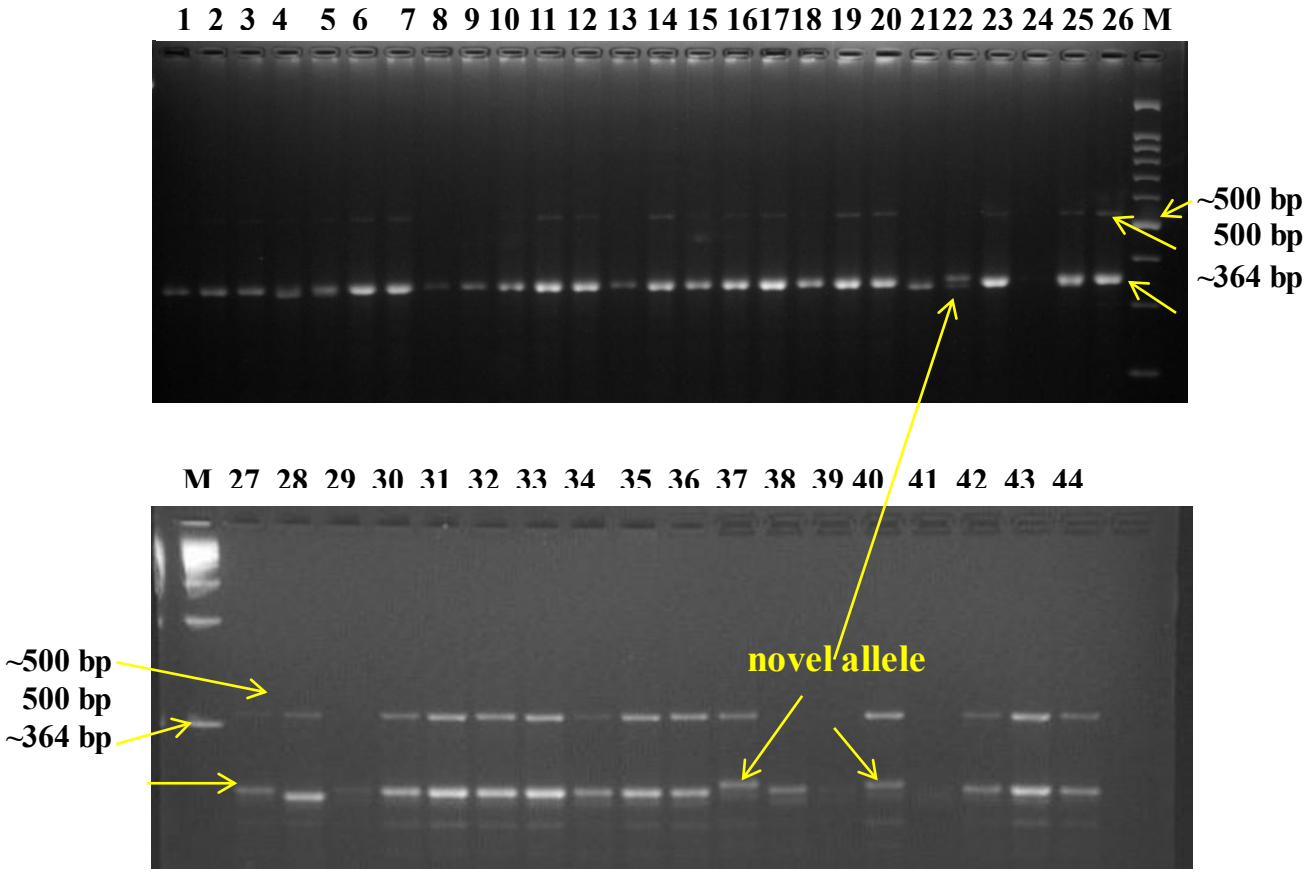


Figure 3.1.20. Agarose gel electrophoresis picture of MRGSSR65 genomic-SSR marker in 44 blackgram genotypes showed PCR amplification of rare/novel allele. (Accessions of genotypes from lane 1-44 are given in Table 2.1)

Chapter 3.2

Development of genic-markers for Yellow mosaic disease resistance trait in *V. mungo* L. Hepper.

3.2. Development of genic-markers for yellow mosaic disease resistance trait in *V. mungo* (L.) Hepper

This chapter is divided into two sub-sections: 3.2.1 Identifying DNA marker(s) linked to yellow mosaic disease resistance trait and 3.2.2 Studying differential gene expression between YMD resistant and susceptible genotypes.

3.2.1. Identifying DNA marker(s) linked to yellow mosaic disease resistance trait

3.2.1.1. Transcriptome characterization of TU94-2 cultivar for R- gene domains

TU94-2 transcriptome dataset consisting of 48,291 contigs were analysed by DRAGO 2 pipeline of PRGDB. A total of 392 contigs were identified to have resistance genes related domains. Among these 392 contigs, 135 contigs harboured only kinase domain, 69 contigs contained kinase and TM (Transmembrane) domains, 44 contigs had NBS domains, 43 contigs contained LRR domains, 34 contigs had TM-LRR domains and 32 contigs contained TM-NBS domains. While contigs with combinations of 3 or more domain types were found to be very few in number (1 to 9). (Fig 3.2.1).

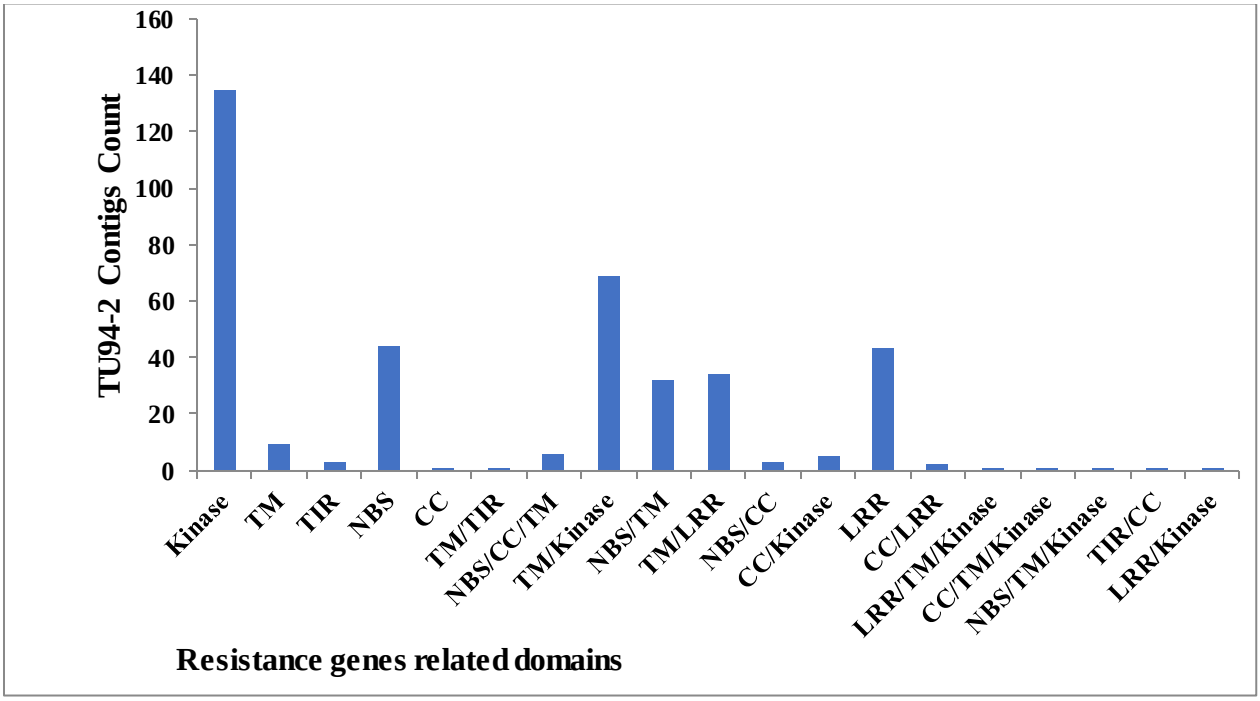


Figure 3.2.1. Transcriptome characterization of TU94-2 genotype for R gene domains through DRAGO 2 pipeline

3.2.1.2. Identification of plant host cellular proteins involved in geminivirus infection cycle

Based on available literature on geminivirus-host interaction, 341 redundant candidate factors such as isoforms, variants of the same gene and same gene sequence from different *Vigna* species or other legume crops were selected. TU94-2 transcriptome contigs were searched for sequence homology with these 341 cellular factors through BLASTn algorithm of NCBI. BLASTn results with TU94-2 contigs showed hits for 123 out of 341 cellular factors and no hits were found for 218 cellular factor sequences. A total of 177 primer-pairs were designed from TU94-2 contigs showing similarities with 123 cellular factors through Primer 3 software (Table 8 of Annexure I).

3.2.1.3. Parental survey, bulk segregant analysis and marker segregation analysis

A total of 610 primer-pairs were screened initially among the parental blackgram cultivars KU96-3 (YMD resistance) and TAU-1 (YMD susceptible). These 610 primers include 315 genic-SSR primers (VmgSSR and TWSSR primer series) from wild and cultivated blackgram transcriptomes (described in sub-chapter 3.1), 118 genomic-SSR primers (MRGSSR primer series) based on putative resistance genes from mungbean genome (described in sub-chapter 3.1) and 177 gene-specific non-SSR primers based on geminivirus-host interaction (YMVF/R series). Primer details of SSR primers derived from cultivated blackgram TU94-2 transcriptome dataset are given in Table 3 of Annexure II (Souframanien and Reddy, 2015) [277] . Out of 610 primers screened, 307 primers got amplified and 153 primers showed length polymorphism between the parents. Bulk segregant analysis of all 153 polymorphic primers in parents, resistant (pooled genomic DNA of 8 highly resistant RILs) and susceptible (pooled genomic DNA of 8 highly susceptible RILs) bulks showed desired segregation pattern for only 3 primers (TWSSR10, VGmSSR86 and VGmSSR547) (Fig. 3.2.2, 3.2.3 and 3.2.4).The segregation patterns of these 3 markers were also confirmed in 218 RILs of F_8 generation of the mapping population from cross TAU-1 x KU96-3.

Figure 3.2.2. Gel electrophoresis pictures of genic-SSR marker VMgSSR547 in bulk segregant analysis and mapping population from cross TAU-1 x KU96-3

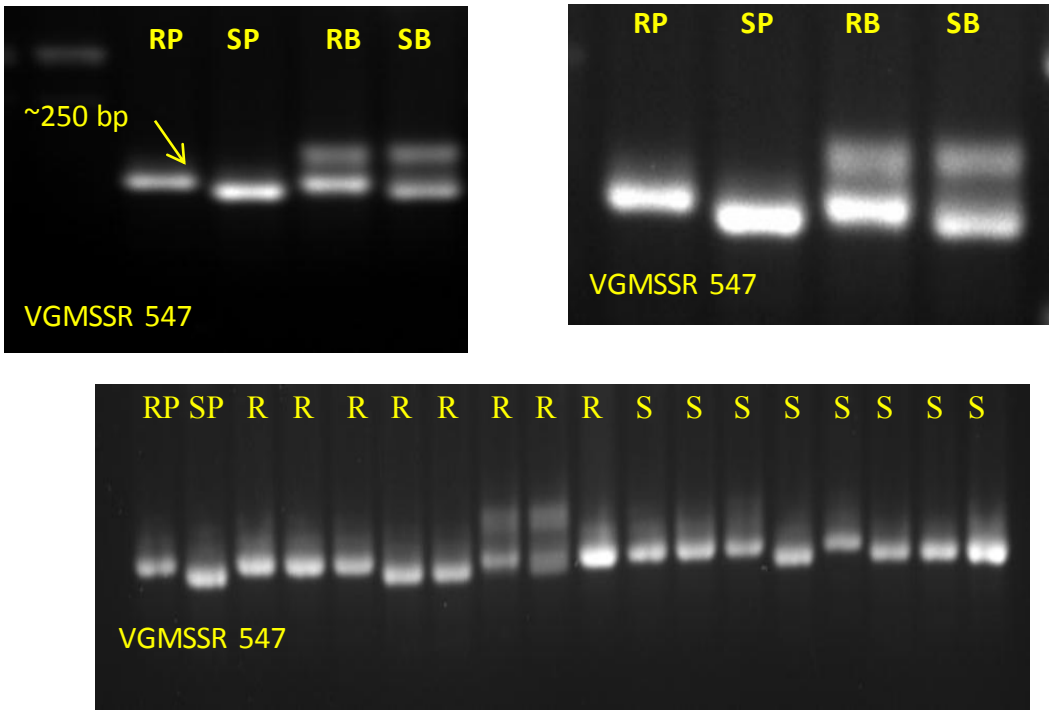
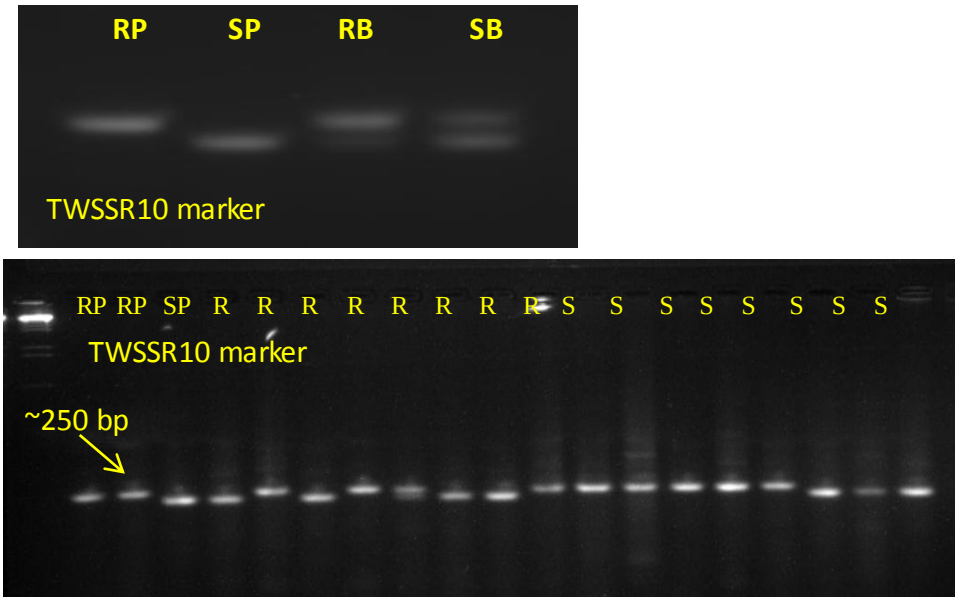


Figure 3.2.3. Gel electrophoresis pictures of genic-SSR marker TWSSR10 in bulk segregant analysis and RIL population from cross TAU-1 x KU96-3



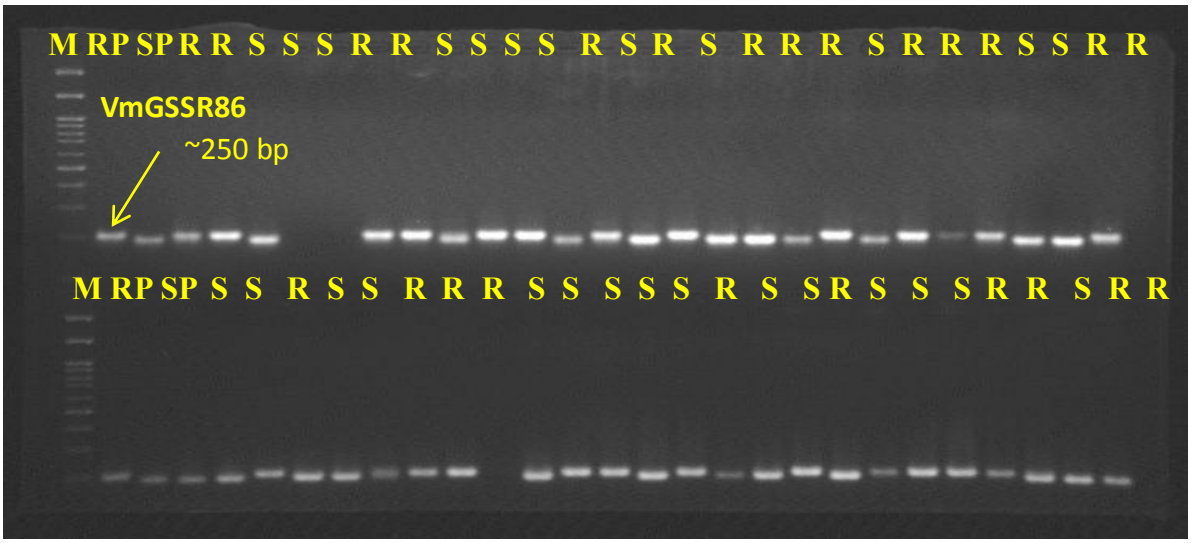


Figure 3.2.4. PAGE and agarose gel electrophoresis pictures of genic-SSR marker VMgSSR86 in bulk segregant analysisand RIL population from cross TAU-1 x KU96-3.

3.2.1.4. Phenotyping: Scoring for YMD reaction and chi square test

Phenotyping for YMD reaction was conducted at Trombay location in 2 different seasons using infector row method. Disease incidence was scored as highly resistant, moderately resistant, susceptible and highly susceptible reactions based on 1-9 scale and finally grouped them into resistant and susceptible class (Fig 3.2.5). In January 2017, out of 204 lines, 90 were resistant and 114were susceptible, while in January 2018, out of 218 lines, 107 were resistant and 111 were susceptible. Chi-square test conducted on mapping population with degree of freedom 1 at p value = 0.05 resulted in $\chi 2$ value of 2.82 (year 2017) and 0.072 (year 2018) (calculated) which were less than table value (3.84) (Table 3.2.1). The goodness of fit resulting from chi-square test reiterated the segregation of mapping population in to 1:1 ratio and thus indicated its suitability for marker development.

Figure 3.2.5. RIL population of F_8 generation from cross KU96-3 x TAU-1 showing yellow mosaic symptoms at Trombay Experimental Field.

Table 3.2.1. Chi-square test for testing 1:1 segregation ratio in RIL mapping population.

F_8 Generation (Year)	Total plants	Yellow mosaic disease				Ratio S:R	χ2 calcualted	χ2 Table value	df	P value
		Observed		Expected						
		R	S	R	S					
KU96-3 x TAU-1 (2017)	204	90	114	102	102	1:1	2.82	3.84	1	0.05
KU96-3 x TAU-1 (2018)	218	107	111	109	109	1:1	0.072	3.84	1	0.05

3.2.1.5. Linkage analysis and marker validation in blackgram germplasm

Analysis of genotypic and phenotypic data for above 3 genic-SSR markers by JOINMAP software version 4.0 (Van Ooijen, 2006) [273] at LOD score 3 with Kosambi mapping function showed linkage of 1 genic-SSR marker TWSSR10, to YMD resistance trait at a

distance of 36.7 cM (Fig 3.2.6). The 2 genic-SSR markers (TWSSR10 and VGmSSR547) along with the previously reported CEDG180 marker were used for validation in 27 diverse blackgram genotypes (12 Resistant and 15 Susceptible genotypes) with known disease reactions to YMD (12 Resistant:PantU-31, IPU07-3, TU-67, PantU-19, KU96-7, IPU02-43, DPU88-31, TU 94-2, Nayagarh, EC-168200, Pusa-3 and KU96-3 and 15 susceptible: *V. mungo* var. *silvestris*, TAU-1, COBG653, LBG709, LBG611, LBG20, LBG17, T9, LBG623, LBG752, LBG703, LBG685, EC168058, EC168235 and IPU94-1) and resistance allele frequency was calculated. The resistance allele frequency for marker VmgSSR547 was 6/13 for resistant and 2/12 for susceptible genotypes while for marker TWSSR10, resistance allele frequency was 10/13 for resistant and 4/12 for susceptible genotypes (Fig 3.2.7). For YMD reported SSR marker, CEDG180 (from adzukibean), resistance allele frequency was 6/13 for resistant and 7/12 for susceptible genotypes (Fig 3.2.8). The resistance allele frequencies for both genic-SSR markers were similar to allele frequency observed for reported YMD marker CEDG180. Third genic-SSR marker, VmgSSR86 did not show amplification in all 27 blackgram genotypes and thus, resistance allele frequency could not be calculated.

Figure 3.2.6 Linkage analysis showed mapping of one genic-SSR marker (TWSSR10) at a distance of 36.7 cM away from YMD resistance trait on blackgram.

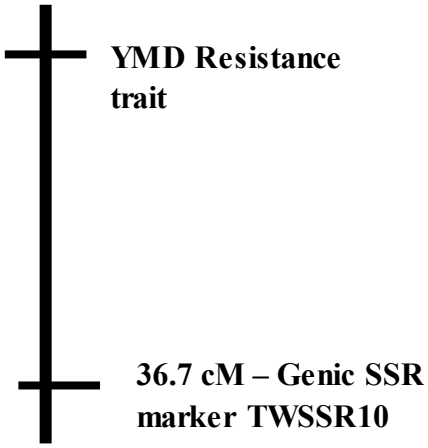


Figure 3.2.7. Gel electrophoresis picture showing segregation pattern of YMD resistance linked markers TWSSR10 and VMgSSR547 in different blackgram genotypes.

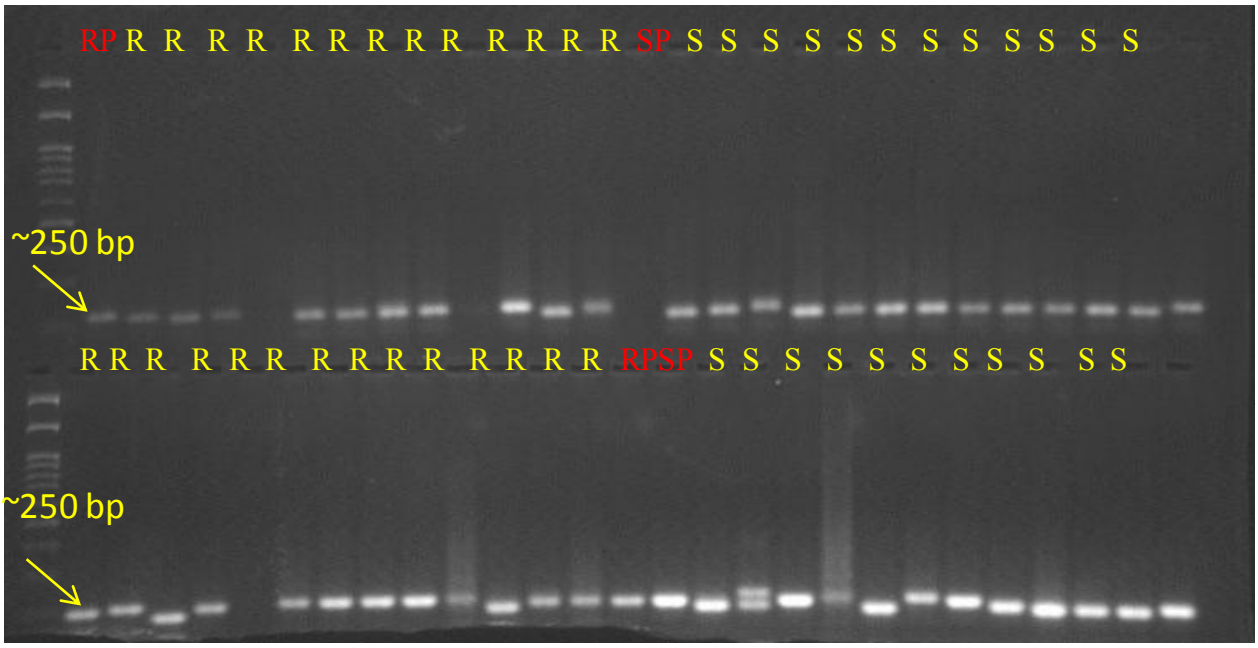
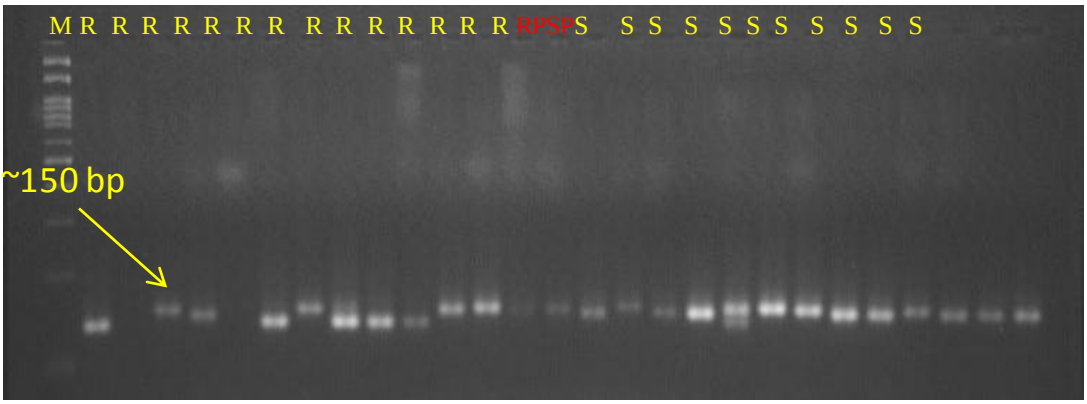


Figure 3.2.8. Gel electrophoresis picture showing segregation pattern of reported YMD resistance linked marker CEDG180 in different blackgram genotypes



3.2.1.6. Functional annotation and localization of markers on mungbean genome

To know the significance of three markers found for YMD in this study, marker sequence i.e transcript contigs were analysed for sequence similarity with NCBI nr database and flanking genomic regions (upstream and downstream) of mungbean genome was analysed for defense related genes.NCBI BLAST results showed homolgy of TW contig 2139 (TWSSR10 marker)

with the transcription factor TCP9, TU94-2 contig 1833 (VmgSSR86 marker) sequence with formin-like protein 2 and TU94-2 contig 14292 (VmgSSR547 marker) sequence with DDB1- and CUL4-associated factor 8. Also, localization of these marker sequences (contigs) on mungbean genome through BLASTn and visualization through genome browser showed presence of R genes in upstream and downstream regions. Marker VmgSSR86 sequence was located at LOC106768783 (6390K-6400K) and TMV resistance protein N was found at distance of 8980k-8984k on mungbean genome. Marker VmgSSR547 sequence was located at LOC106772191 (44,410K-44,416K) and enhanced disease resistance 2 was found at LOC106770618 at a distance of 37,732-37,739K. TWSSR10 marker sequence was located at LOC106779364 (15,300-16,900) but no resistance gene was found in proximate genomic regions (Table 3.2.2).

Table 3.2.2. Functional annotations and localization of genic-SSR marker sequences on mungbean genome

Marker – Genic- SSR	TU94-2/TW transcript contig number	NCBI Annotation (*Vigna radiata* var. *radiata*)	Function	Location on *V. radiata* genome	Flanking resistance gene(s)
TWSSR10	2139 (TW Contig)	Transcription factor TCP9	Salicylic acid synthesis) Wang et al., 2015) [278]	LOC106779364 , 15,300-16,900 (TCP9 transcription factor)	Not found on intact chromosome
VGMSSR86	1833 (TU94-2 Contig)	Formin-like protein 2	Has been hypothesized for role in pathogen perception	LOC106768783 , 6390K-6400K (Formin)	8980k-8984k (TMV Resistance N) Tobacco mosaic virus
VGMSSR547	14292 (TU94-2 Contig)	DDB1- and CUL4-associated factor 8	E3 ligases associated proteins involved in development	LOC106772191 , 44,410K-44,416K (DDB1 and CUL4 associated factor 8)	LOC106770618 , 37,732-37,739K (Enhanced disease resistance 2)

3.2.2. Gene expression study between YMD resistance and susceptible genotypes

To reveal the underlying molecular mechanism and to explore the key cellular factors controlling or regulating resistance to YMD in KU96-3 cultivar, expression pattern of key cellular proteins were analysed by quantitative RT-PCR. The relative expression of cellular factors studied were selected based on their key roles in geminivirus infection cycle (Literature survey) and DEGs related to defense/stress identified from developing seeds transcriptome of YMD resistant cultivar TU94-2.

3.2.2.1 Identification of DEGs related to defense responses from transcriptome of TU94-2 immature seeds

TU94-2 transcriptome analysis showed basal level up-regulation of genes encoding pathogen recognition receptors (PRRs) and LRR-containing proteins such as LRR receptor-like serine/threonine-protein kinase RKF3, receptor-like protein kinase FERONIA, LRR containing protein DDB_G0290503, two F-box/LRR-repeat proteins (protein At4g29420 and protein 4) and four uncharacterized transmembrane proteins (transmembrane protein 205, transmembrane 9 superfamily member 2, transmembrane protein 230 and transmembrane protein PM19L). Interestingly, transcripts of 3 autophagy related proteins ATG 3, 5 and 8C, serine/threonine-protein kinase ATG1t and chloroplast located lipoxygenase (LOX) involved in biosynthesis of jasmonic acid were found up-regulated. Noteworthy components of PTI and ETI signaling including calcium-dependent protein kinase 28 (CPK28), mitogen-activated protein kinase kinase 5, mitogen-activated protein kinase kinase kinase NPK1, tobamovirus multiplication protein 1 (TOM1) were found to be over-expressed. A number of transcription factors associated with pathogenesis were also found upregulated which belonged to families such as MYB, NAC and WD-repeat domains containing transcription factors. Several transcripts of DnaJ, heat shock chaperones, ubiquitin activating, conjugating enzymes, ligases, SEC interacting proteins and proteasomes were observed to be highly expressed. In addition, another susceptibility factor found to be enriched in TU94-2 cultivar

was DOWNY MILDEW RESISTANCE 6 (DMR6). DEGs involved in defense responses to geminivirus and found up-regulated in TU94-2 developing seeds are described in Table 1 of Annexure II and DEGs involved in defense responses to other phytopathogens are given in Table 2 of Annexure II.

3.2.2.2. Relative expression of genes upon YMV invasion

The blackgram cultivars KU96-3 (YMD resistant) and TAU-1 (YMD susceptible) were grown under controlled lab conditions (control) and in experimental field (test). Young leaves from control plants with no yellow mosaic symptoms and from test plants with yellow mosaic patches were used for RNA extraction as shown in Fig 3.2.9. The relative expression of 16 genes that were selected on the basis of their significant role in geminivirus infection cycle in plant host and their high expression profile in YMD resistant TU94-2 developing seeds transcriptome were analysed by qRT-PCR. Gene expression pattern experiments showed significant upregulation of some genes in resistant cultivar compared to susceptible cultivar (TAU-1) under YMV infection which throws light on the molecular mechanism of disease reaction. Out of 16 genes analysed, 10 genes were upregulated and 6 genes were downregulated in KU96-3 cultivar as compared to susceptible cultivar TAU-1 (Table 3.2.3). Graphical representation of expression pattern of 16 genes analysed by qRT-PCR are given in Fig 3.2.10.1, Fig 3.2.10.2 and Fig 3.2.10.3. Transcription factors MYB1R1 and WD repeat-containing protein DWA2 showed moderate fold changes of 1.46 and 2.29, respectively. The 70 kDa DNA-binding subunit of replication protein A (RPA) showed upregulation in KU96-3 cultivar with a fold change of 2.4. RPA is a single stranded DNA binding protein playing a key role in replication. Putative leucine-rich repeat-containing protein DDB and Tobamovirus multiplication protein 1-like were upregulated in KU96-3 cultivar with 3.27 and 2.39 fold changes, respectively. Transcripts encoding leucine-rich repeat receptor-like protein kinase was found to be highly expressed in KU96-3 with fold change of 27.9. Autophagy-related

protein 8C and argonaute 1 (AGO1) transcripts were also upregulated in KU96-3 with 7.99 and 8.5 fold changes. Notably, transcripts encoding eukaryotic translation initiation factor 4B3 and importin subunit alpha-2-like were found to be very highly upregulated in KU96-3 with fold changes of 1112 and 1844 respectively. On the other hand, transcripts for 6 genes were found downregulated in KU96-3,which included DNA repair protein RAD4 (0.81 FC), allene oxide cyclase 4 (0.17 FC), SUPPRESSOR OF GENE SILENCING 3 (0.6 FC), RNA-directed DNA methylation 3 (0.53 FC), RNA-directed DNA methylation 4 (0.4 FC) and eukaryotic translation initiation factor 4E (eIF4E) (0.75 FC).

Figure 3.2.9. KU96-3 and TAU-1 plants grown under lab conditions with no yellow mosaic symptoms and in experimental field showing yellow mosaic symptoms.

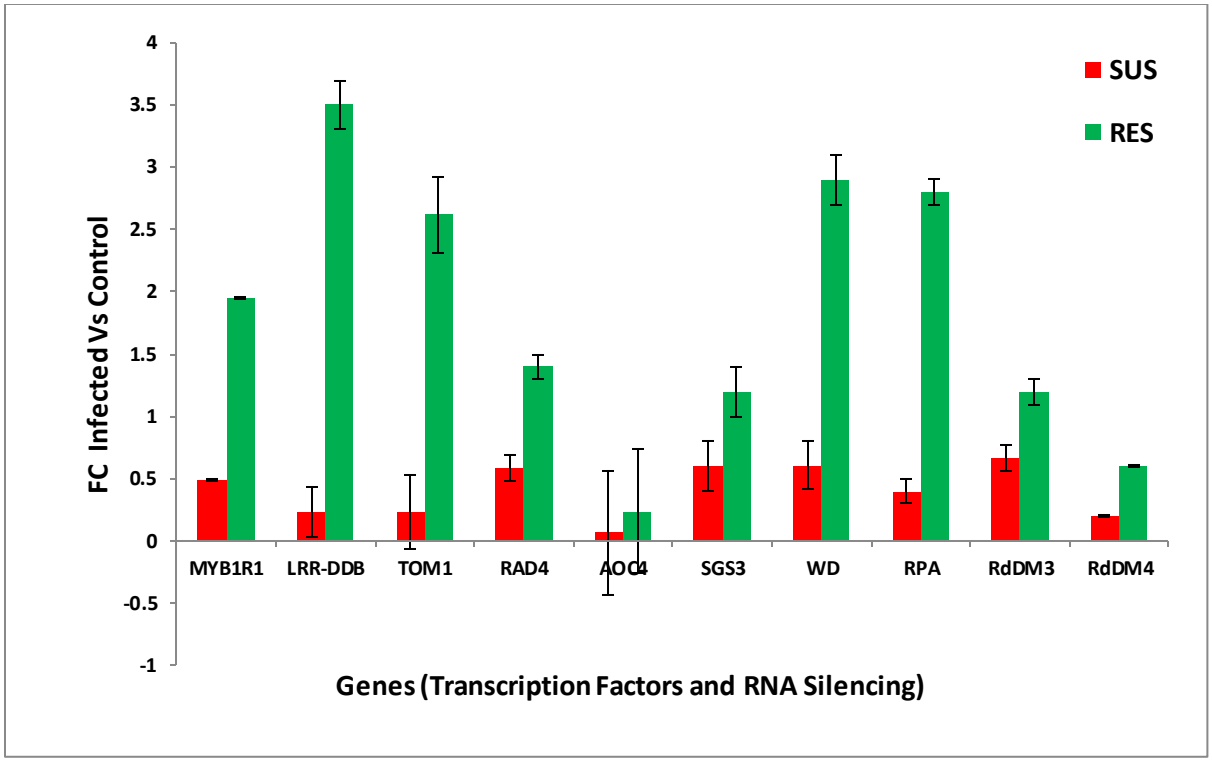


Figure 3.2.10.1. Gene expression profiles of defense related transcription factors and other gene transcripts analysed by qRT-PCR

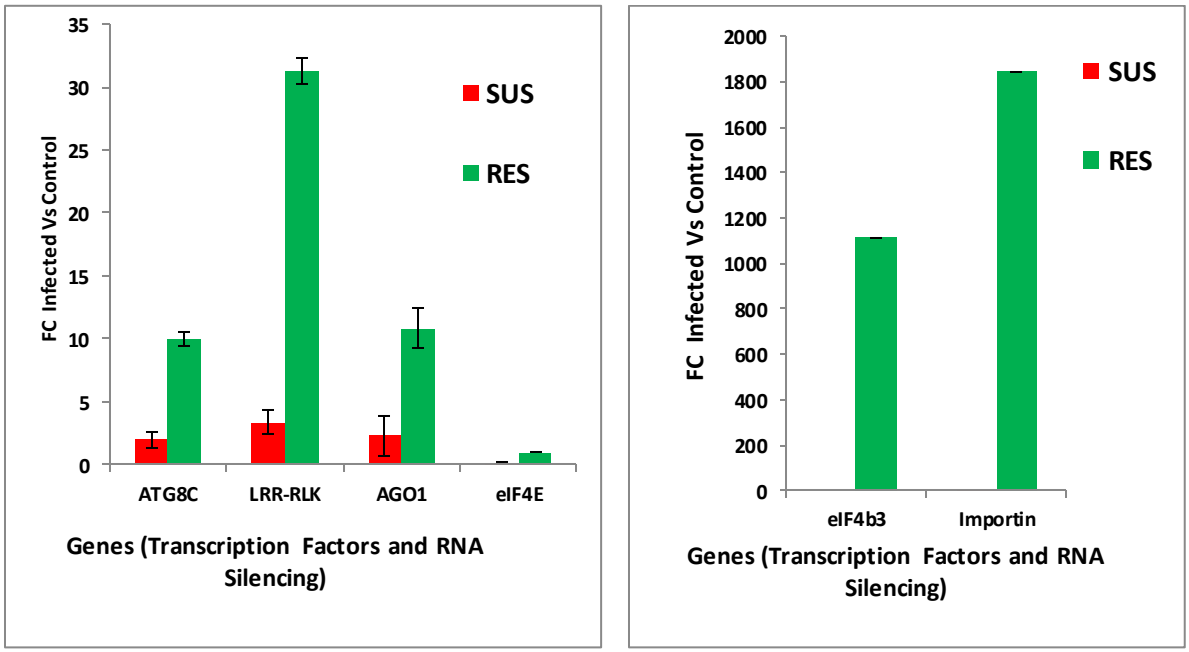


Figure 3.2.10.2. Gene expression profiles of RNA silencing and autophagy related gene analysed by qRT-PCR

Figure 3.2.10.3. Gene expression profiles of transcription factor 4b3 and importin transporter analysed by qRT-PCR

Table 3.2.3. Gene expression profiles of 16 genes analysed upon YMV invasion by qRT-PCR

Primer code	Gene abbr.	NCBI Annotation	Resistance (R) FC	Susceptible (S) FC	R Vs S FC
Q2001	MYB1R1	Transcription factor MYB1R1	1.95	0.49	1.46
Q7450	LRR-DDB	Putative leucine-rich repeat-containing protein DDB	3.5	0.23	3.27
Q383	TOM1	Tobamovirus multiplication protein 1-like	2.62	0.23	2.39
YMV29	RAD4	DNA repair protein RAD4	1.4	0.59	0.81
YMV66	AOC4	Allene oxide cyclase 4, chloroplastic-like (AOC1)	0.24	0.07	0.17
YMV92	SGS3	Suppressor of gene silencing 3	1.2	0.6	0.6
TUCDS41	WD	WD repeat-containing protein DWA2	2.9	0.61	2.29
YMV10	RPA	Replication protein A 70 kDa DNA-binding subunit	2.8	0.4	2.4
YMV122	RdDM3	RNA-directed DNA methylation 3	1.2	0.67	0.53
YMV22	RdDM4	RNA-directed DNA methylation 4	0.6	0.2	0.4
Q9754	ATG8C	Autophagy-related protein 8C	9.99	2	7.99
TWCDS69	LRR-RLK	Probable leucine-rich repeat receptor-like protein kinase	31.3	3.4	27.9
YMV68	AGO1	Argonaute 1 (AGO1)	10.8	2.3	8.5
YMV69	eIF4E	Eukaryotic translation initiation factor 4E (eIF4E)	0.97	0.22	0.75
TWCDS70	eIF4b3	Eukaryotic translation initiation factor 4B3	1112.8	0.0003	1112.7
YMV86	Importin	Importin subunit alpha-2-like	1845	0.6	1844.4

Moreover, based on upregulation of defense related DEGs in YMD resistant TU94-2 seed transcriptome (YMV non-infected immature seeds) and upregulated genes in YMD resistant KU96-3 (YMV infected young leaves), a hypothesized pathway describing the basal and induced expression of defense related genes is proposed which gives a preliminary idea of the molecular mechanism of YMD resistance in KU96-3 and TU94-2 cultivars (Fig 3.2.11).

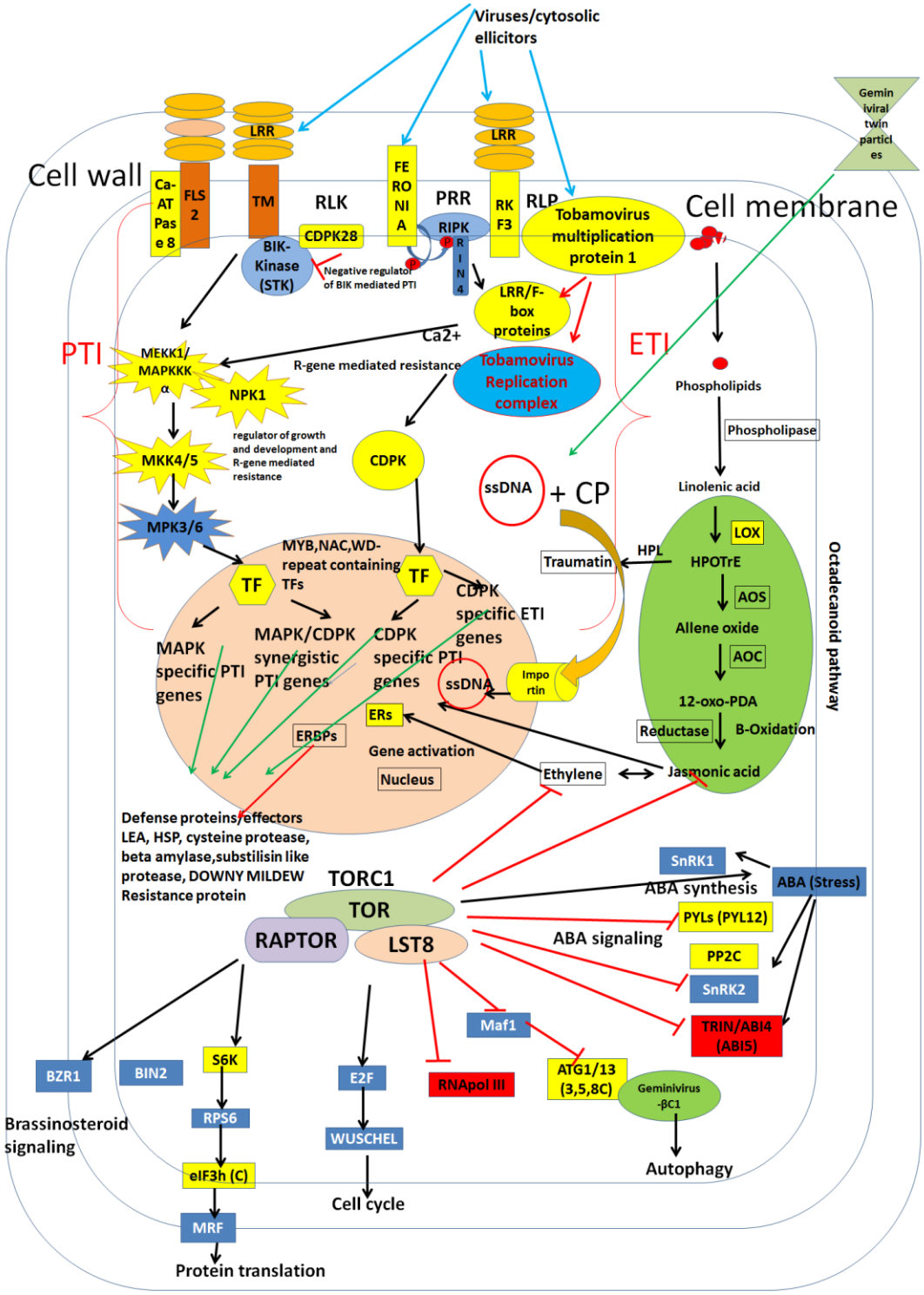
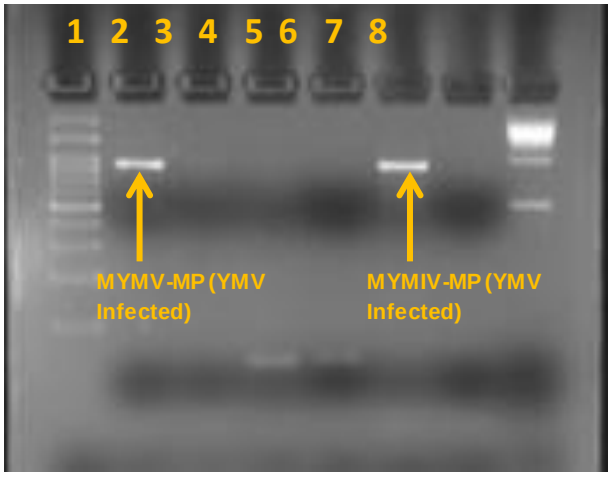


Figure 3.2.11. Proposed hypothetical pathway based on upregulation of defense related DEGs in YMD resistant TU94-2 seed transcriptome (YMV non-infected immature seeds) and upregulated genes in YMD resistant KU96-3 (YMV infected young leaves) at basal and induced stages, respectively.

3.2.2.3. Identification of YMV strain prevalent at Trombay

Total DNA isolated from field grown YMV infected KU96-3 and TAU-1 (test) leaves and lab grown uninfected KU96-3 and TAU-1 (control) leaves were PCR tested for different viral genes such as coat protein and movement protein of MYMV and MYMIV strains. PCR amplification of MYMV and MYMIV movement protein genes indicated mixed infection and hence, prevalence of both these YMV strains at Trombay (Fig 3.2.12).

Figure 3.2.12. Gel electrophoresis image showing PCR testing for MYMV and MYMIV viral genes in YMV infected and noninfected leaves



LANES

1.100b.p ladder
2. MYMV-MP (I)= +
3. MYMV-MP(UI)= -
4. CP (I)= -
5. CP (UI)= -
6. MYMIV-MP (I)= +
7. MYMIV-MP (UI)= -
8. 500b.p ladder

Chapter 3.3

Development of genic-markers for bruchid resistance trait in *V. mungo* var. *silvestris.*

3.3. Development of genic-markers for bruchid resistance trait in *V. mungo* var. *silvestris*

This chapter is divided into two sub-sections: 3.3.1Studying differential gene expression between bruchid resistant and susceptible genotypes and 3.3.2 Identifying DNA marker(s) linked to bruchid resistance trait.

3.3.1. Gene expression study between bruchid resistant and susceptible genotypes

Storage pest, bruchids caused huge loss to blackgram production under poor storage conditions. In blackgram, very few resistance sources are reported but have not yet been studied. Trombay wild urd (INGR10133) (TW), is a wild accession of blackgram which is a native of Trombay region. In TW, the resistance trait is under the control of two dominant duplicate genes with larval antibiosis as a resistance mechanism. The resistance factors and or the mechanism were found to be expressed constitutively. Antibiosis resulted in reduced survival, longer developmental period (88 days as compared with 34 days on TU 94-2) and reduced body weight of *C. maculatus* (Dongre et al., 1996; Souframanien and Gopalakrishna, 2007) [205, 16] when fed with dried TW seeds. Moreover, biochemical basis of resistance in TW is not known till date. In this sub-chapter, molecular mechanism and biochemical basis of resistance in TW operating at basal level were studied through RNA sequencing and expression pattern analysis for important genes in non-infested developing seeds of TW. We also validated the RNA-Seq results through quantitative real-time PCR (qRT-PCR) method. This study enlightened the transcriptional differences related to innate immune system of wild and cultivated blackgram against bruchids.

3.3.1.1. Identification, annotation and reannotation of DEGs

A total of 2306 DEGs (CDS) including 1116 upregulated and 1190 downregulated were identified from TW and TU94-2 transcriptomes. DEGs identified using wild and cultivated blackgram transcriptome datasets have been previously annotated with reference to green plant database. This annotation resulted in majority of CDS (DEGs) remaining

uncharacterized. In the present study, 682 DEGs with considerable fold change values ranging from 3 to 12 (264 up-regulated DEGs) and -3 to -12 (418 down-regulated DEGs) were again annotated with reference to NCBI nr database and previously uncharacterized CDS were assigned with functions. Heat map showed the expression pattern of top 100 DEGs of TW and TU94-2 genotypes (Fig 3.3.1).

3.3.1.2. Transcriptome characterization of Trombay wild

The RNASeq results showed up-regulation of genes related to pests/pathogen perception, defense responses, resistance factors and other significant cellular processes. The up-regulated transcripts related to stimuli perception in TW transcriptome included three LRR-RLKs (leucine-rich repeat receptor-like protein kinase); leucine-rich repeat receptor-like protein kinase At1g68400, LRR receptor kinase BAK1 or somatic embryogenesis receptor kinase 1-like (SERK) and one each of LRR receptor-like serine/threonine-protein kinase ERECTA, RPK; receptor serine/threonine-protein kinase-like protein At1g28390, L-type lectin-domain containing receptor kinase S and transmembrane protein 87B. Four transcripts related to signal transduction were found up-regulated of which 2 were serine/threonine-protein kinases: Serine/threonine-protein kinase STY8 isoform C, probable serine/threonine-protein kinase At4g35230 and 2 were calcium related proteins: Calmodulin-like protein 1 (CML-1) and calnexin homolog isoform X1. Genes encoding several ribosomal proteins, translation initiation factors, carbohydrate metabolism genes, tubulin chains and myosin binding proteins, cell wall proteins such as glycosyltransferases, arabinogalactans and expansins were highly expressed. Stress/defense related genes that were highly expressed included heat shock proteins, stress response proteins, DnaJ homologs, chaperons, acid phosphatases, 7S globulins, vicilins, thaumatin like protein (TLP), miraculin and thioredoxin. Some of these are well reported effectors against pathogens and pests.

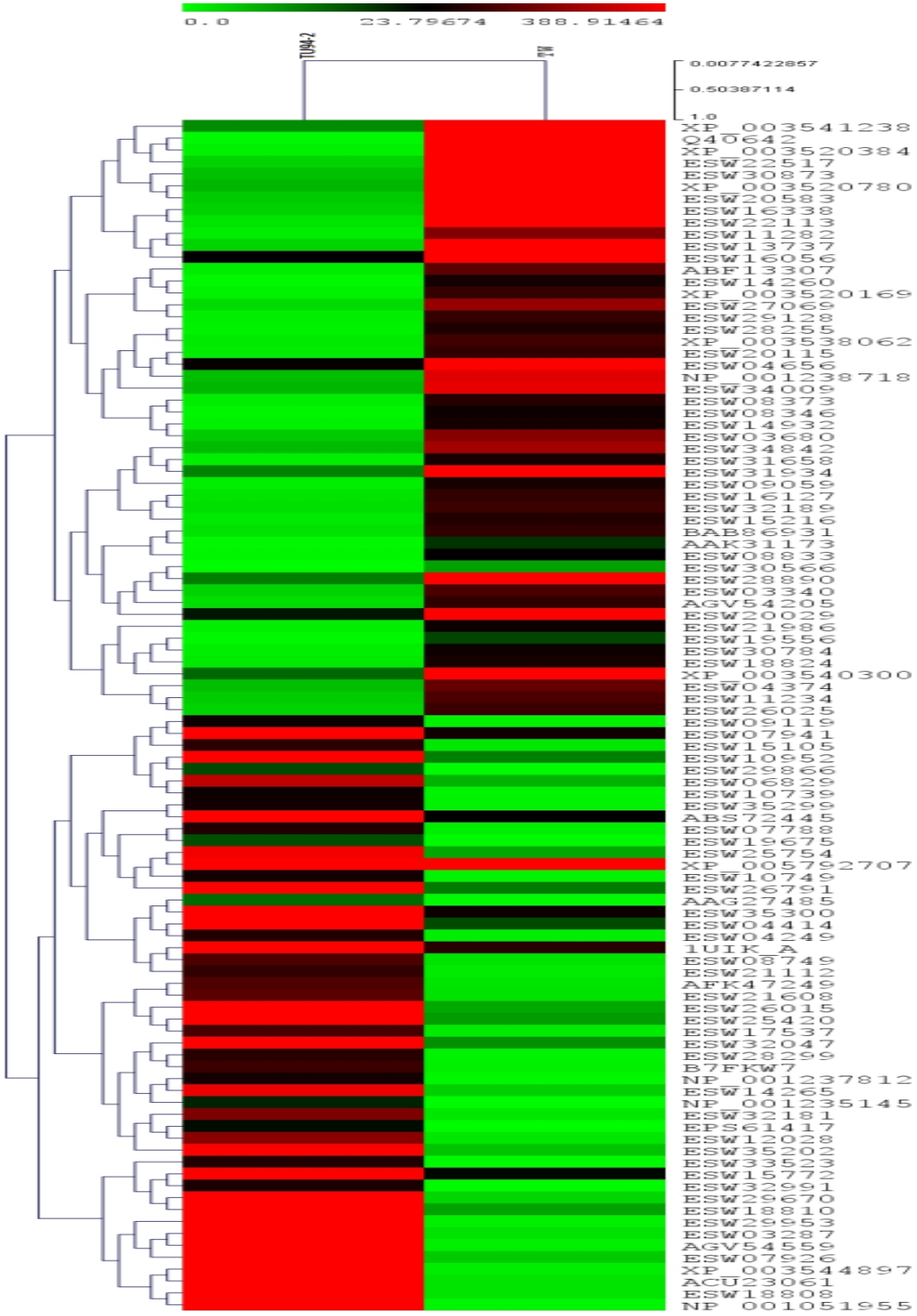


Figure 3.3.1. Heat Map showing differential expression pattern of top 100 coding sequences (CDS) of TU94-2 cultivar and Trombay wild (TW) blackgram. Green colour denotes down-regulation and red colour denotes up-regulation of CDS.

Augmented levels of transcripts of abscisic acid receptor PYL9, protein phosphatase 2C (PP2C) and abscisic acid insensitive 5 (ABI5) were found. Interestingly, hydroperoxide lyase (HPL, CYP74B) which is a cytochrome P450 present in chloroplasts was found upregulated in TW. Transcription factors that were found up-regulated are ethylene-responsive transcription factor ERF061, zinc finger A20 and AN1 domain-containing stress-associated protein 5, transcription factor LHW, transcription termination factor MTEF18, RING-H2 finger protein ATL8 and DNA-directed RNA polymerase III subunit RPC7. DEGs related to pest resistance that were found upregulated in TW developing seeds are described in Table 1 of Annexure II. DEGs involved in defense responses to other phytopathogens are given in Table 1 of Annexure III.

3.3.1.3. Validation of DEGs using quantitative real-time PCR

Approximately, 1µg of total RNA from developing seeds of wild and cultivated genotypes grown under controlled conditions of non-oviposition, non-infestation by bruchids were taken for downstream processing. The elongation factor EF 1α gene was used as a reference gene or internal control for normalisation in real-time PCR reactions. Details of primers used in this study are given in Table 2 of Annexure III. Expression patterns of 27 genes were analysed by realtime PCR. Of which, 22 DEGs (15 up-regulated and 7 down-regulated genes) were from the RNA-seq dataset of blackgram developing seeds and 5 genes coding for putative factors reported for imparting resistance to pests were based on literature survey. The trend of gene expression pattern for all 22 genes (selected from RNASeq)was same in both the qRT-PCR and RNA Sequencing and hence, the DEGs were validated successfully (Fig 3.3.2.1 and Fig 3.3.2.2). TW transcript encoding acid phosphatase showed higher (1.3 fold) expression level as compared to cultivar (Fig 3.3.3.1). The universal stress protein PHOS32 showed 2.5 fold increased expression in TW than cultivar (Fig 3.3.3.2). A leucine-rich repeat receptor like kinase gene (LRR-RLK) showed enhanced expression (1.4 fold change) in TW

in comparison to cultivar. Likewise, increase in fold changes were also observed in TW as compared to TU94-2 cultivar for following transcripts: golgin subfamily A member 6-like protein 6, multiple organellar RNA editing factor 2, RING-H2 finger protein ATL8-like, ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1, protein RETICULATA-RELATED 1 (chloroplastic), EID1-like F-box protein 2, prostatic spermine-binding protein, protein PNS1, glycosyltransferase BC10, gibberellin receptor GID1, geranylgeranyl pyrophosphate synthase 7 (chloroplastic) and 50S ribosomal protein 5 (chloroplastic).

Figure 3.3.2.1. Expression pattern of 15 up-regulated genes as examined by qRT-PCR. For validation, log of fold change values as calculated by ΔΔCt method in the RT-PCR and FPKM values (log FPKM TW/TU94-2) in the RNASeq were used for each gene.

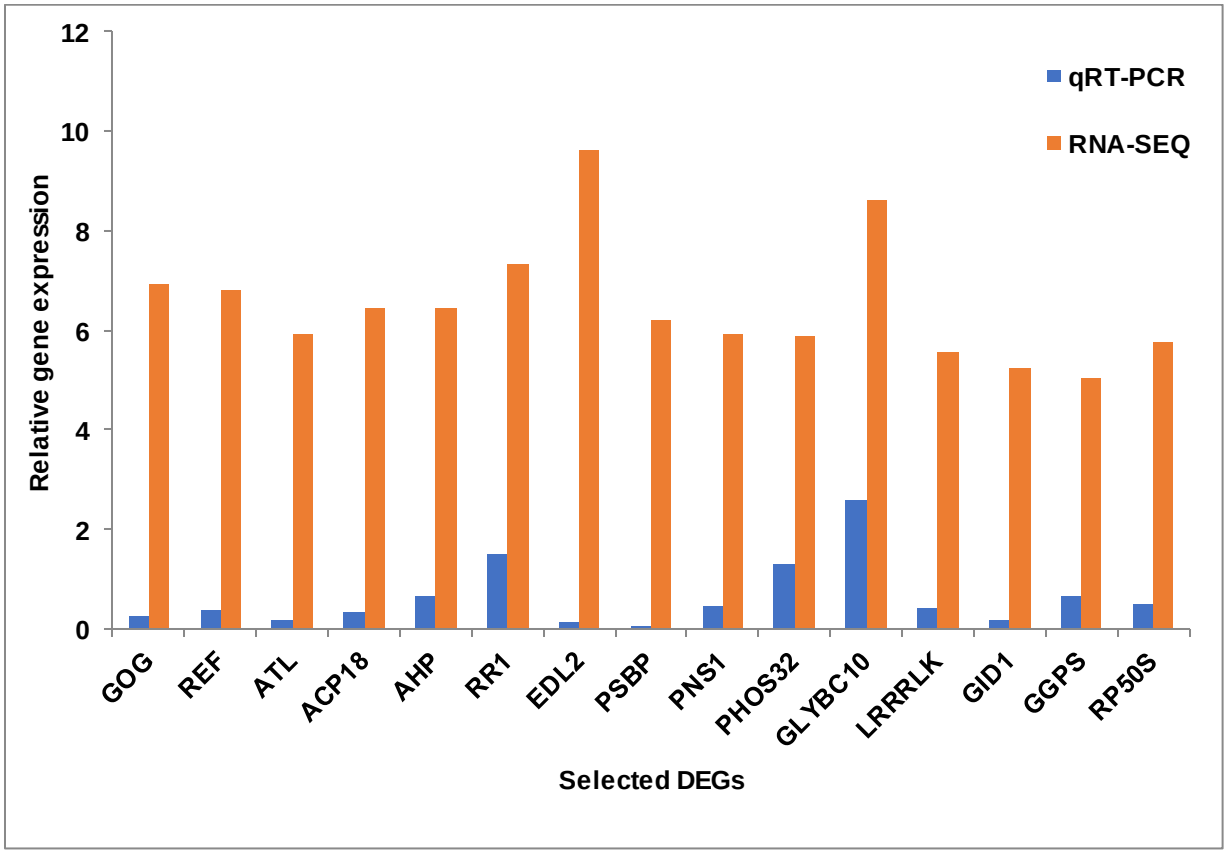


Figure 3.3.2.2. Expression pattern of 7 down-regulated genes as examined by qRT-PCR. For validation, log of fold change values as calculated by ΔΔCt method in the qRT-PCR and FPKM values (log FPKM TW/TU94-2) in the RNA-Seq were used for each gene.

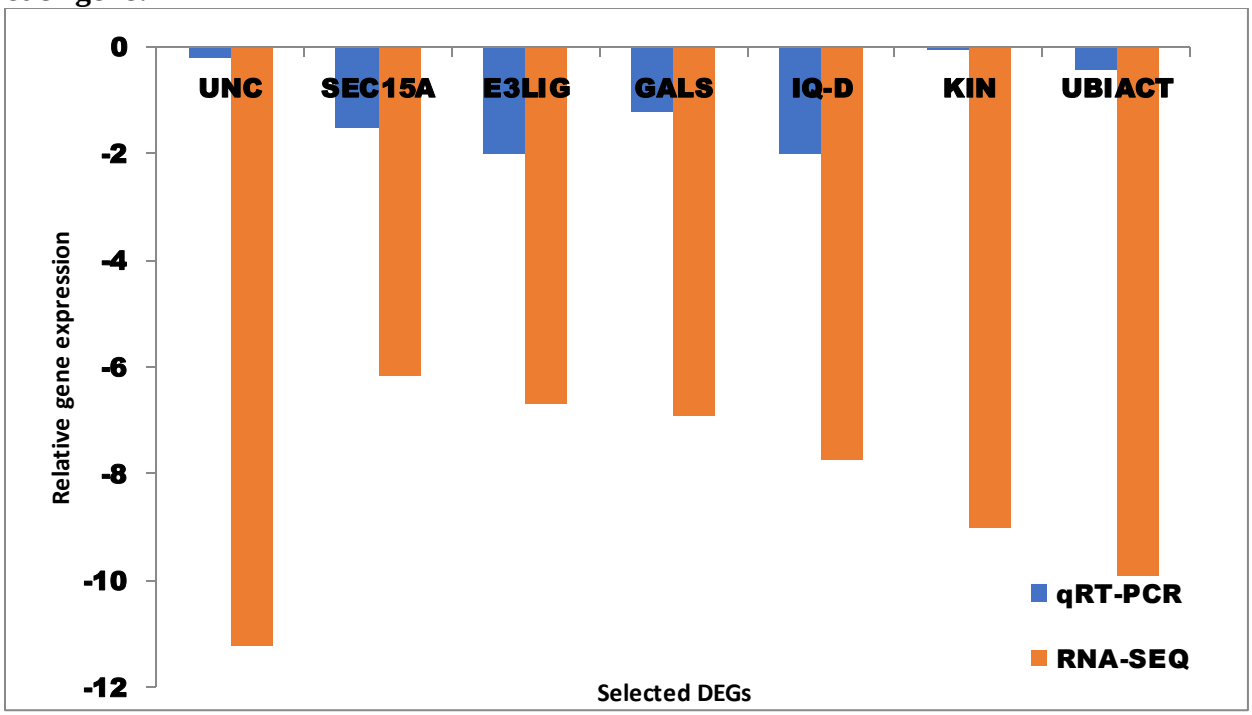


Figure 3.3.3.1. Expression levels of 12 DEGs analysed in non-infested TW developing seeds as examined by qRT-PCR. The expression levels were normalized with the help of EF1*α* gene of blackgram. Expression levels were calculated by ΔΔCt method in the RT-PCR but not converted to log values. Expansion of abbreviations of genes are given in Table 2 of Annexure III. Bars represent mean ± standard deviation.

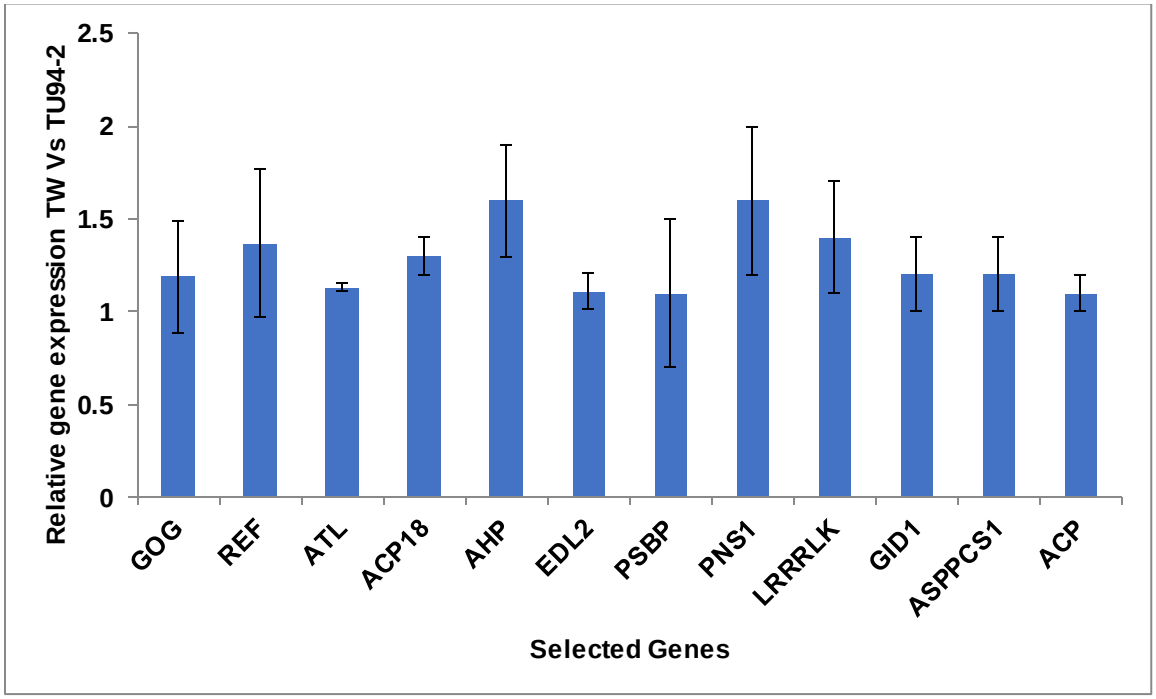
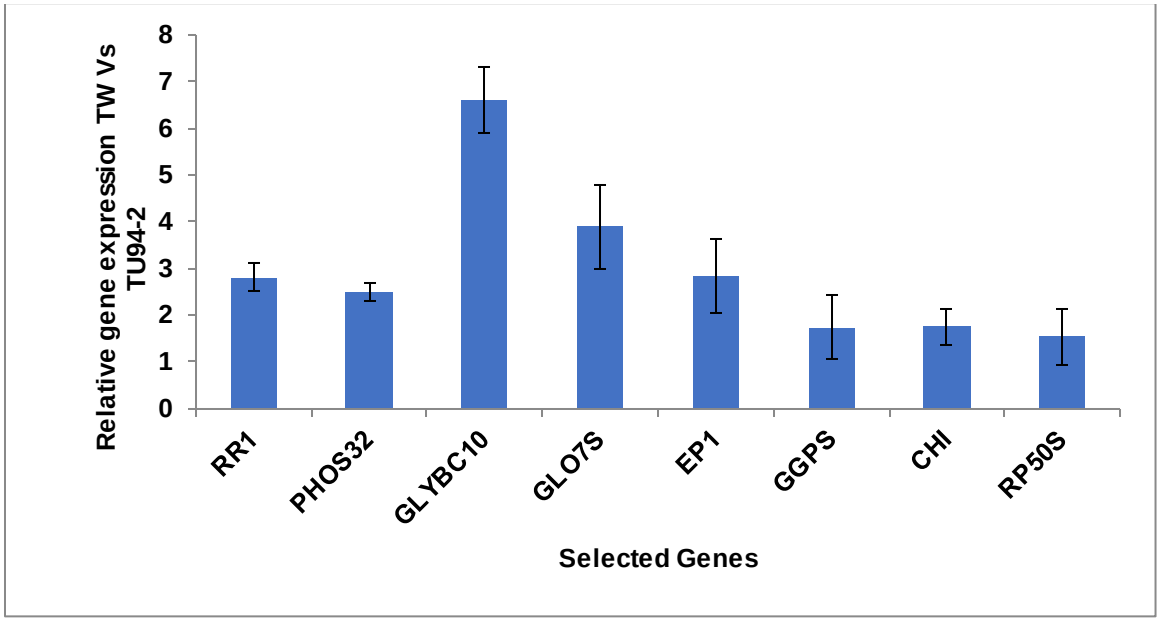


Figure 3.3.3.2. Expression levels of 8 DEGs analysed in TW non-infested developing seeds as examined by qRT-PCR. The expression levels were normalized with the help of EF1*α* gene of blackgram. Expression levels were calculated by by ΔΔCt method in the RT-PCR but not converted to log values. Expansion of abbreviations of genes are given in Table 2 of Annexure III. Bars represent mean ± standard deviation.



Moreover, based on RNA sequencing and qRT-PCR data of TW developing seeds, a hypothetical pathway has been proposed describing basal upregulation of bruchid and other defense related DEGs in TW (resistant) seed transcriptome (bruchid non-infested immature seeds) (Fig 3.3.4).

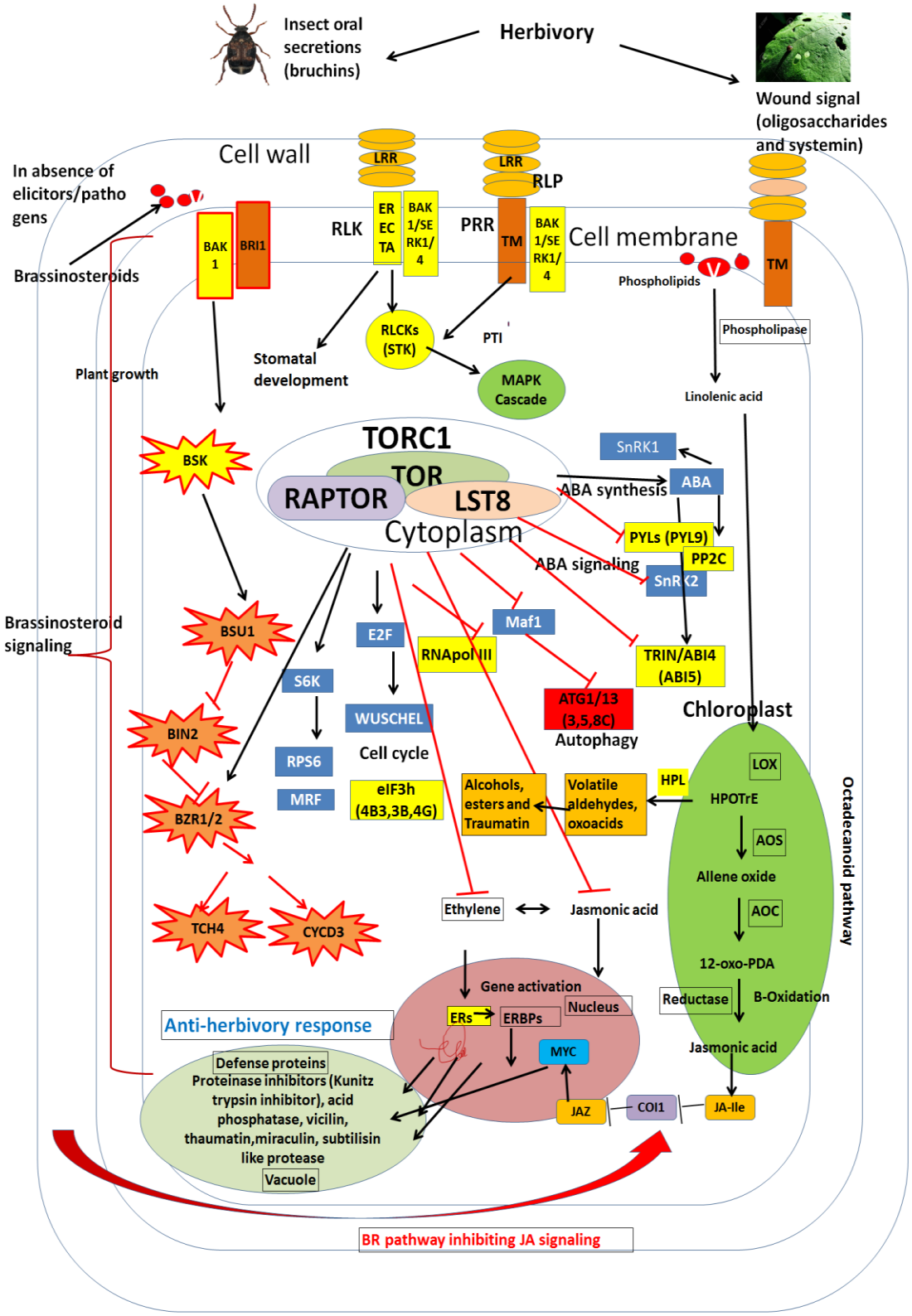


Figure 3.3.4. Proposed hypothetical pathway based on constitutive upregulation of bruchid related and other defense related DEGs in resistant TW as identified by RNA sequencing and validated by qRT-PCR.

3.3.2. Development of genic-markers for bruchid resistance trait in Trombay wild

3.3.2.1. SSR Analysis

Parental survey (TW and TU94-2) with 434 SSR primers resulted in PCR amplification of 235 primers, of which 106 primers showed length polymorphism on resolving through agarose gels. Polymorphic primers (106) comprised of 48 VMgSSR, 34 TWSSR and 24 MRGSSR primers. A total of 57 randomly selected polymorphic primers were screened in 8 bruchid resistant and 8 bruchid susceptible RILs. Representative amplification profile of genic-SSR primer TWSSR68 resolved on agarose gel is shown in Fig 3.3.5. Further, 33 primers which showed desired amplification pattern in 8 bruchid resistant and 8 bruchid susceptible RILs were analysed over the entire mapping population of 104 RILs (TU94-2 x TW). Amplification pattern of genic-SSR marker TWSSR66 in RIL mapping population resolved through agarose gel electrophoresis is shown in Fig 3.3.6.

Figure 3.3.5. Amplification pattern of genic-SSR marker TWSSR68 in 8 bruchid resistant and 8 bruchid susceptible RILs along with parents TW and TU94-2.

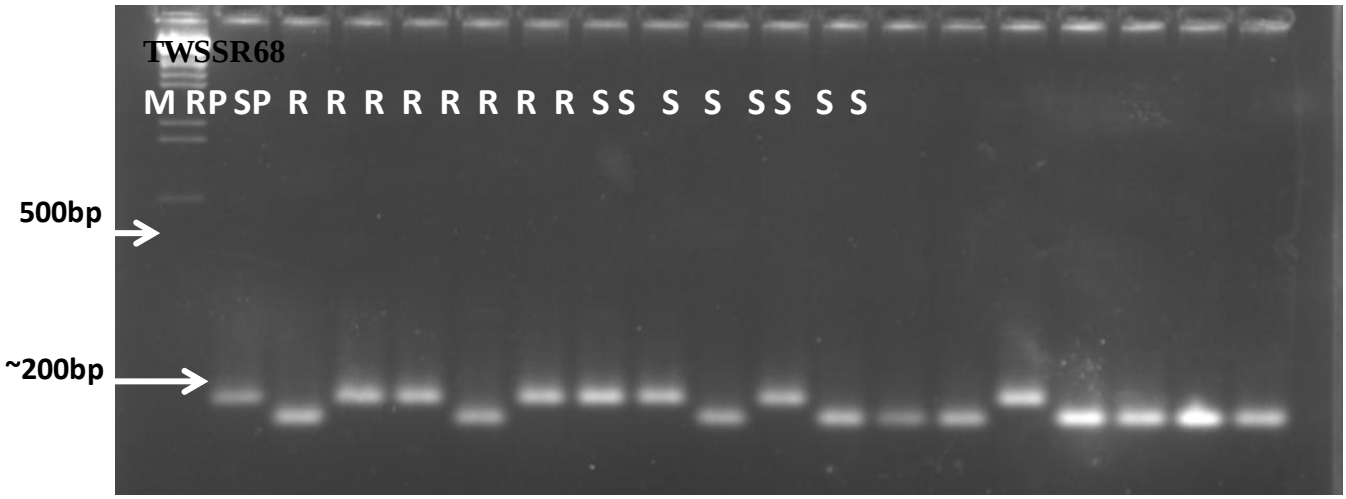
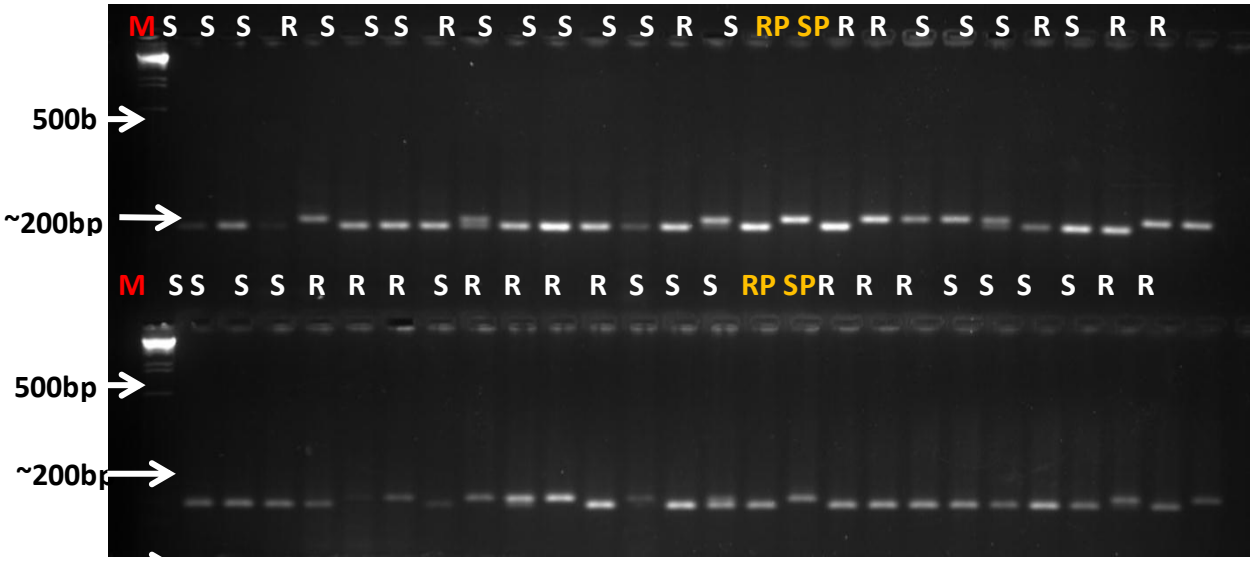


Figure 3.3.6. Amplification pattern of genic-SSR marker TWSSR66 in RIL mapping population of the cross TU94-2 x TW.



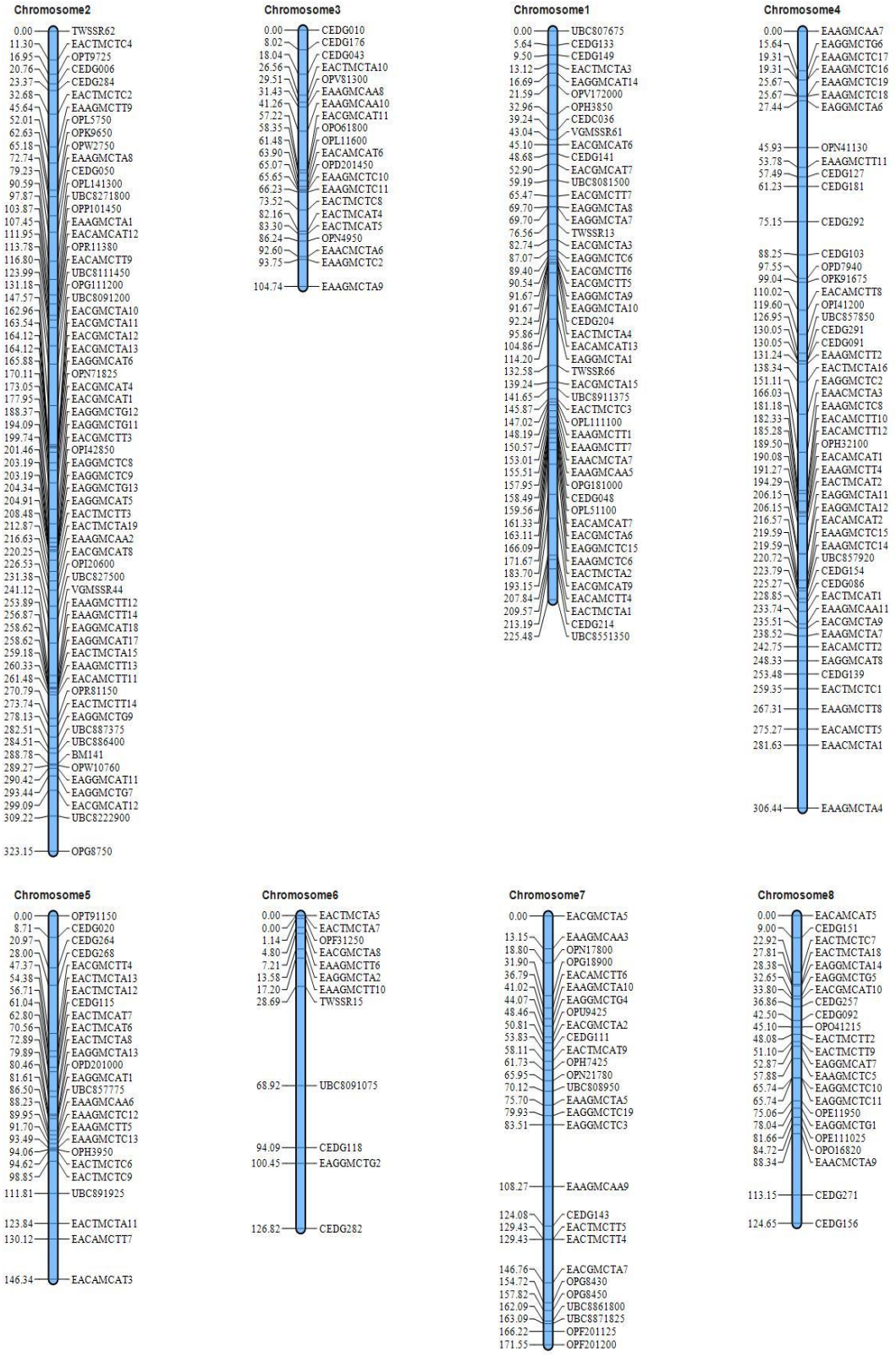
3.3.2.2. Re-construction and comparative analysis of genetic linkage maps

To integrate the newly developed markers from present study into the earlier reported map of blackgram, the genetic linkage map was re-constructed through QTL ICI mapping at LOD score of 5. New linkage map was built using a total of 461 markers, 428 from being earlier an study (Gupta et al., 2008) [4] and 33 new markers from the present study. Out of 33 new markers, 11 were mapped on to the re-constructed linkage map (Fig 3.3.7). Comparison between reported and newly re-constructed linkage map showed high similarity or conservation of map features with very few differences (Table 3.3.1). Among 11 newly placed markers on re-constructed map, 2 markers were mapped on LG2, 3 markers on LG1, 1 marker each on LG 6 and LG 9 and 4 markers were located on LG10 (Fig 3.3.7). The re-constructed linkage map was 2326.23 cM in length with linkage group (LG) 3 being the shortest (104.74) and LG 10 appeared as longest (410.07) (Table 3.3.1). LG6 consisted of

only 12 markers and LG10 showed highest marker numbers of 90. The average length of the LGs was 211.48 cM with a marker density of 0.2 markers per cM and 22 markers do not get mapped.

The reconstructed map was larger (2326.23 cM) as compared to the previous map (865.1 cM) and both the maps showed LG3 as shortest and LG10 as longest linkage groups. The sequence of molecular markers in the previous linkage map constructed using 428 markers by joinmap software (Gupta et al., 2008) [4] and linkage map reconstructed using 461 markers (428 previous markers + 33 new markers) by ICIM software were similar except for some discrepancies at few marker positions. In re-constructed map, the sequence of reported markers flanking the bruchid resistance QTLs was the same with previous map except for Cmrdp1.2 and Cmrdp1.3 in LG1 and LG2 respectively. For QTL Cmrdp1.2, the reported flanking markers (EACG/MCTA-15 and EAGG/MCTA-1) were located near the terminal in the previous map as compared to the center of LG1 in the re-constructed map (Fig 3.3.8). In case of QTL Cmrdp1.3, unlike the previous map, both the reported flanking markers ($OPI20_{600}$ and $OPL14_{1300}$) of LG2 were not placed together in the reconstructed map. The reported flanking marker $OPL14\text{-}_{1300}$ was located at a different position albeit on same LG2 (Fig3.3.8).

Figure 3.3.7. Reconstructed ICIM analysis based genetic linkage map of blackgram using 428 markers (Gupta et al., 2008) [4] and 33 new genic-SSR markers (this study).



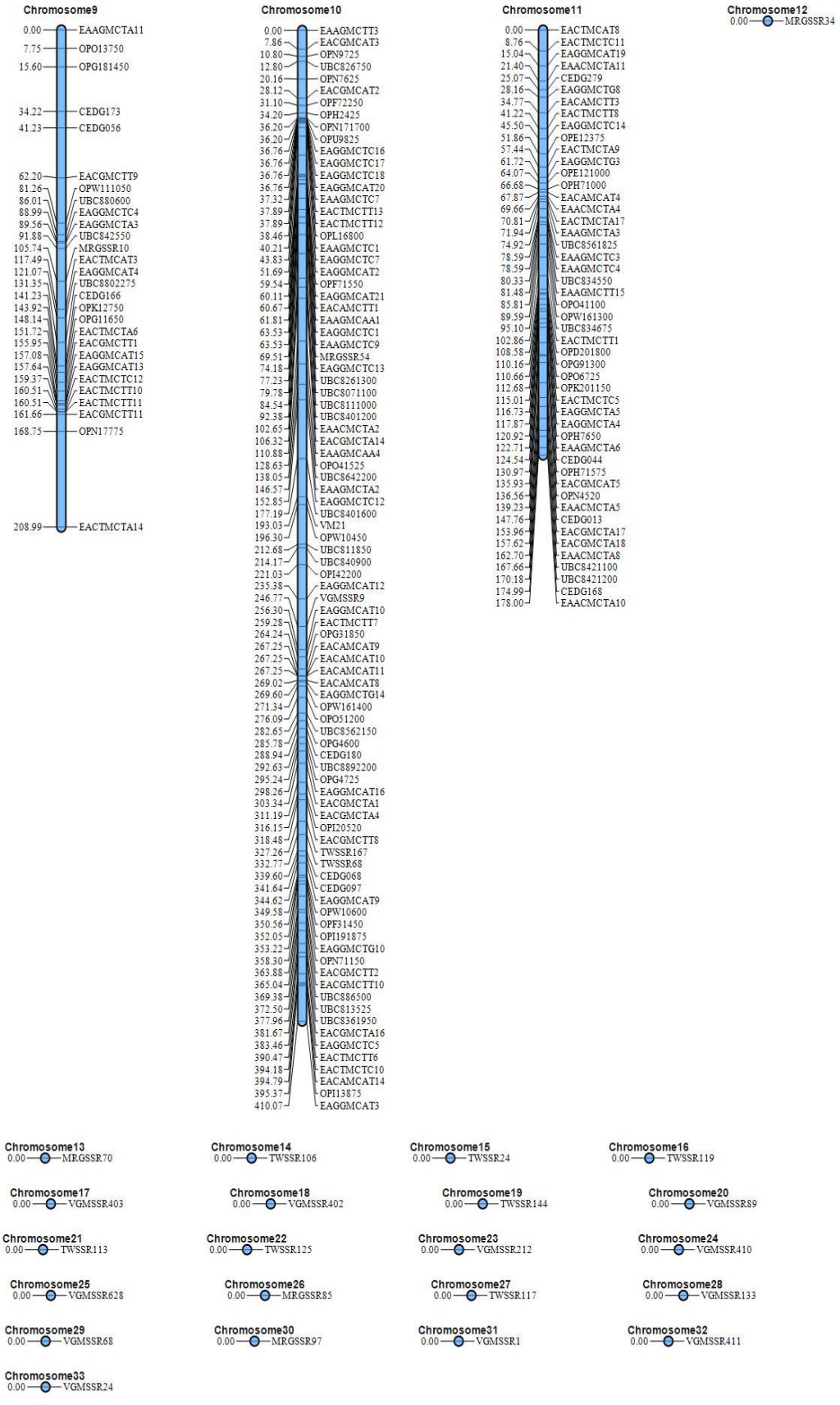
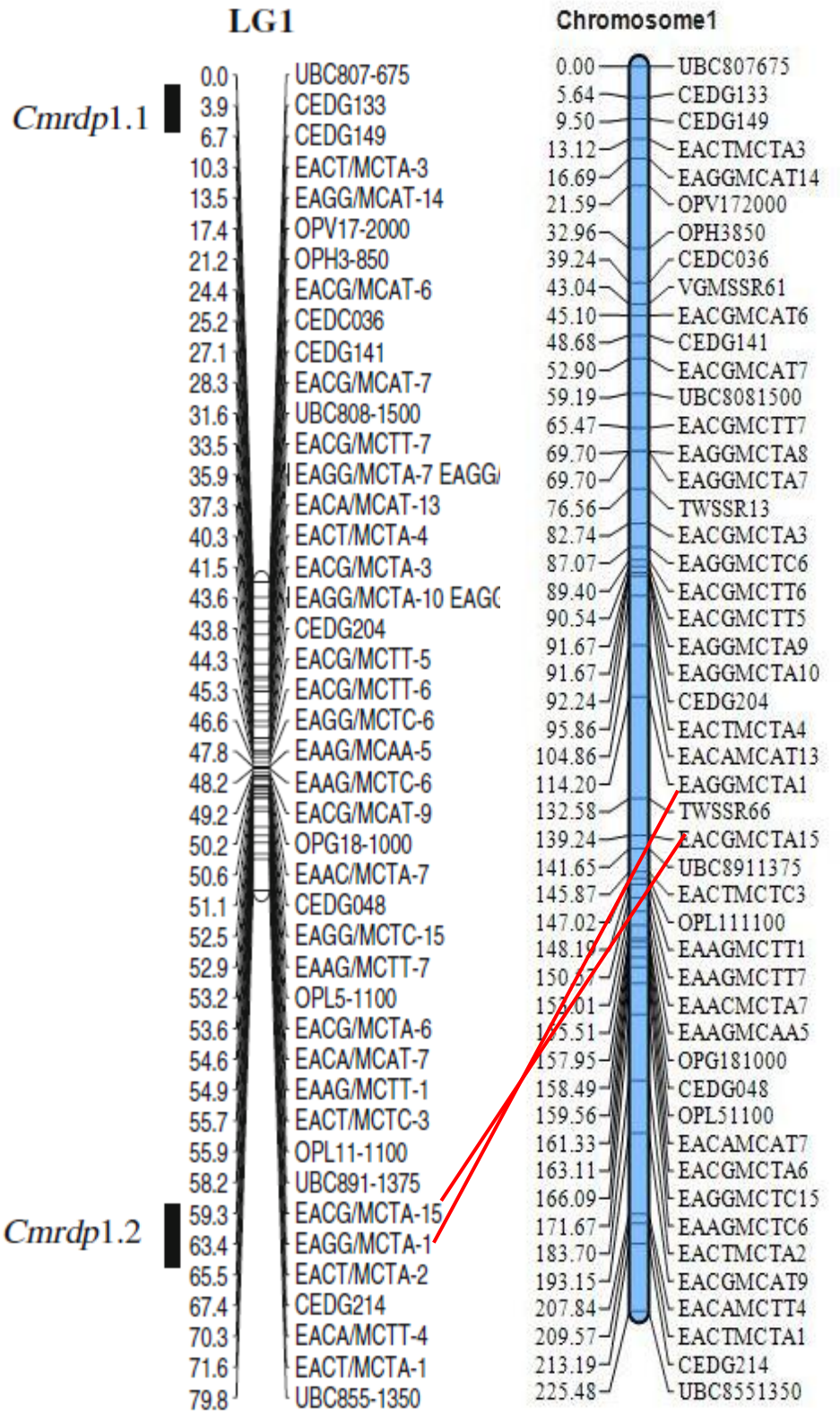

Chromosome9
0.00 EAAGMCTA11
7.75 OPO13750
15.60 OPG181450
34.22 CEDG173
41.23 CEDG056
62.20 EACGMCTT9
81.26 OPW111050
86.01 UBC880600
88.99 EAGGMCTC4
89.56 EAGGMCTA3
91.88 UBC842550
105.74 MRGSSR10
117.49 EACTMCAT3
121.07 EAGGMCAT4
131.35 UBC8802275
141.23 CEDG166
143.92 OPK12750
148.14 OPG11650
151.72 EACTMCTA6
155.95 EACGMCTT1
157.08 EAGGMCAT15
157.64 EAGGMCAT13
159.37 EACTMCTC12
160.51 EACTMCTT10
160.51 EACTMCTT11
161.66 EACGMCTT11
168.75 OPN17775
208.99 EACTMCTA14
Chromosome10
0.00 EAAGMCTT3
7.86 EACGMCAT3
10.80 OPN9725
12.80 UBC826750
20.16 OPN7625
28.12 EACGMCAT2
31.10 OPF72250
34.20 OPH2425
36.20 OPN171700
36.20 OPU9825
36.76 EAGGMCTC16
36.76 EAGGMCTC17
36.76 EAGGMCTC18
36.76 EAGGMCAT20
37.32 EAAGMCTC7
37.89 EACTMCTT13
37.89 EACTMCTT12
38.46 OPL16800
40.21 EAAGMCTC1
43.83 EAGGMCTC7
51.69 EAGGMCAT2
59.54 OPF71550
60.11 EAGGMCAT21
60.67 EACAMCTT1
61.81 EAAGMCAA1
63.53 EAGGMCTC1
63.53 EAAGMCTC9
69.51 MRGSSR54
74.18 EAGGMCTC13
77.23 UBC8261300
79.78 UBC8071100
84.54 UBC8111000
92.38 UBC8401200
102.65 EAACMCTA2
106.32 EACGMCTA14
110.88 EAAGMCAA4
128.63 OPO41525
138.05 UBC8642200
146.57 EAAGMCTA2
152.85 EAGGMCTC12
177.19 UBC8401600
193.03 VM21
196.30 OPW10450
212.68 UBC811850
214.17 UBC840900
221.03 OPI42200
235.38 EAGGMCAT12
246.77 VGMSSR9
256.30 EAGGMCAT10
259.28 EACTMCTT7
264.24 OPG31850
267.25 EACAMCAT9
267.25 EACAMCAT10
267.25 EACAMCAT11
269.02 EACAMCAT8
269.60 EAGGMCTG14
271.34 OPW161400
276.09 OPO51200
282.65 UBC8562150
285.78 OPG4600
288.94 CEDG180
292.63 UBC8892200
295.24 OPG4725
298.26 EAGGMCAT16
303.34 EACGMCTA1
311.19 EACGMCTA4
316.15 OPI20520
318.48 EACGMCTT8
327.26 TWSSR167
332.77 TWSSR68
339.60 CEDG068
341.64 CEDG097
344.62 EAGGMCAT9
349.58 OPW10600
350.56 OPF31450
352.05 OPI191875
353.22 EAGGMCTG10
358.30 OPN71150
363.88 EACGMCTT2
365.04 EACGMCTT10
369.38 UBC886500
372.50 UBC813525
377.96 UBC8361950
381.67 EACGMCTA16
383.46 EAGGMCTC5
390.47 EACTMCTT6
394.18 EACTMCTC10
394.79 EACAMCAT14
395.37 OPI13875
410.07 EAGGMCAT3
Chromosome11
0.00 EACTMCAT8
8.76 EACTMCTC11
15.04 EAGGMCAT19
21.40 EAACMCTA11
25.07 CEDG279
28.16 EAGGMCTG8
34.77 EACAMCTT3
41.22 EACTMCTT8
45.50 EAGGMCTC14
51.86 OPE12375
57.44 EACTMCTA9
61.72 EAGGMCTG3
64.07 OPE121000
66.68 OPH71000
67.87 EACAMCAT4
69.66 EAACMCTA4
70.81 EACTMCTA17
71.94 EAAGMCTA3
74.92 UBC8561825
78.59 EAAGMCTC3
78.59 EAAGMCTC4
80.33 UBC834550
81.48 EAAGMCTT15
85.81 OPO41100
89.59 OPW161300
95.10 UBC834675
102.86 EACTMCTT1
108.58 OPD201800
110.16 OPG91300
110.66 OPO6725
112.68 OPK201150
115.01 EACTMCTC5
116.73 EAGGMCTA5
117.87 EAGGMCTA4
120.92 OPH7650
122.71 EAAGMCTA6
124.54 CEDG044
130.97 OPH71575
135.93 EACGMCAT5
136.56 OPN4520
139.23 EAACMCTA5
147.76 CEDG013
153.96 EACGMCTA17
157.62 EACGMCTA18
162.70 EAACMCTA8
167.66 UBC8421100
170.18 UBC8421200
174.99 CEDG168
178.00 EAACMCTA10
Chromosome12
0.00 MRGSSR34
Chromosome13
0.00 MRGSSR70
Chromosome14
0.00 TWSSR106
Chromosome15
0.00 TWSSR24
Chromosome16
0.00 TWSSR119
Chromosome17
0.00 VGMSSR403
Chromosome18
0.00 VGMSSR402
Chromosome19
0.00 TWSSR144
Chromosome20
0.00 VGMSSR89
Chromosome21
0.00 TWSSR113
Chromosome22
0.00 TWSSR125
Chromosome23
0.00 VGMSSR212
Chromosome24
0.00 VGMSSR410
Chromosome25
0.00 VGMSSR628
Chromosome26
0.00 MRGSSR85
Chromosome27
0.00 TWSSR117
Chromosome28
0.00 VGMSSR133
Chromosome29
0.00 VGMSSR68
Chromosome30
0.00 MRGSSR97
Chromosome31
0.00 VGMSSR1
Chromosome32
0.00 VGMSSR411
Chromosome33
0.00 VGMSSR24

Table 3.3.1. Comparative analysis of markers associated with bruchid resistance QTL in the previous genetic linkage map (Gupta et al., 2008, Souframanien et al., 2010) [4,12] and reconstructed genetic linkage map (present study) of blackgram.

Linkage Group	Length (cM) (Previous study, 428 markers)	Length (cM) (This study, 428 + 33 = 461 markers)	QTL name	Reported flanking markers from previous map (Souframanien et al., 2010) [12]	Map distance (cM) of new genic-SSR markers from reported QTLs markers on reconstructed genetic linkage map
LG-1	79.8	225.48	Cmrdp1.1 Cmrdp1.2	CEDG133 and CEDG149 EACG/MCTA-15 and EAGG/MCTA-1	VMgSSR61 (33.54cM) away from CEDG133 and CEDG149, TWSSR13 (67.06cM) away from CEDG133 and CEDG149, TWSSR66 (6.66 cM) from EACG/MCTA-15 and (18.38 cM) from EAGG/MCTA-1
LG-2	68.5	323.15	Cmrdp1.3 Cmrdp1.4 Cmrdp1.5	OPL14-1300 and OPI20-600 OPR1-1380 and EACA/MCAT-12 EAAG/MCTA-1 and UBC827-1800	VMgSSR44 (14.59 cM) from OPI20-600, TWSSR62 (111.95 cM) from OPR1-1380 and EACA/MCAT-12 (Cmrdp1.4), TWSSR62 (97.97 cM) from EAAG/MCTA-1 and UBC827-1800 (Cmrdp1.5)
LG-3	44.9	104.74	Cmrae1.1	EACT/MCTC-8 and EACT/MCAT-5	
LG-4	99.6	306.44	Cmrae1.2	CEDG086 and CEDG154	
LG-5	107.8	146.34			
LG-6	61.6	126.82			
LG-7	69.0	171.55			
LG-8	81.1	124.65			
LG-9	76.0	208.99			
LG-10	115.0	410.07	Cmrdp1.6	EACG/MCTA-14 and EAAG/MCAA-4	MRGSSR54 (36.81cM) and VMgSSR9 (135.89cM) away from reported flanking markers (EACG/MCTA-14 and EAAG/MCAA-4)
LG-11	61.8	178			
Total	865.1	2326.23			

Figure 3.3.8.1. Comparison of the marker position between previously reported and new re-constructed genetic linkage maps of blackgram (LG1).



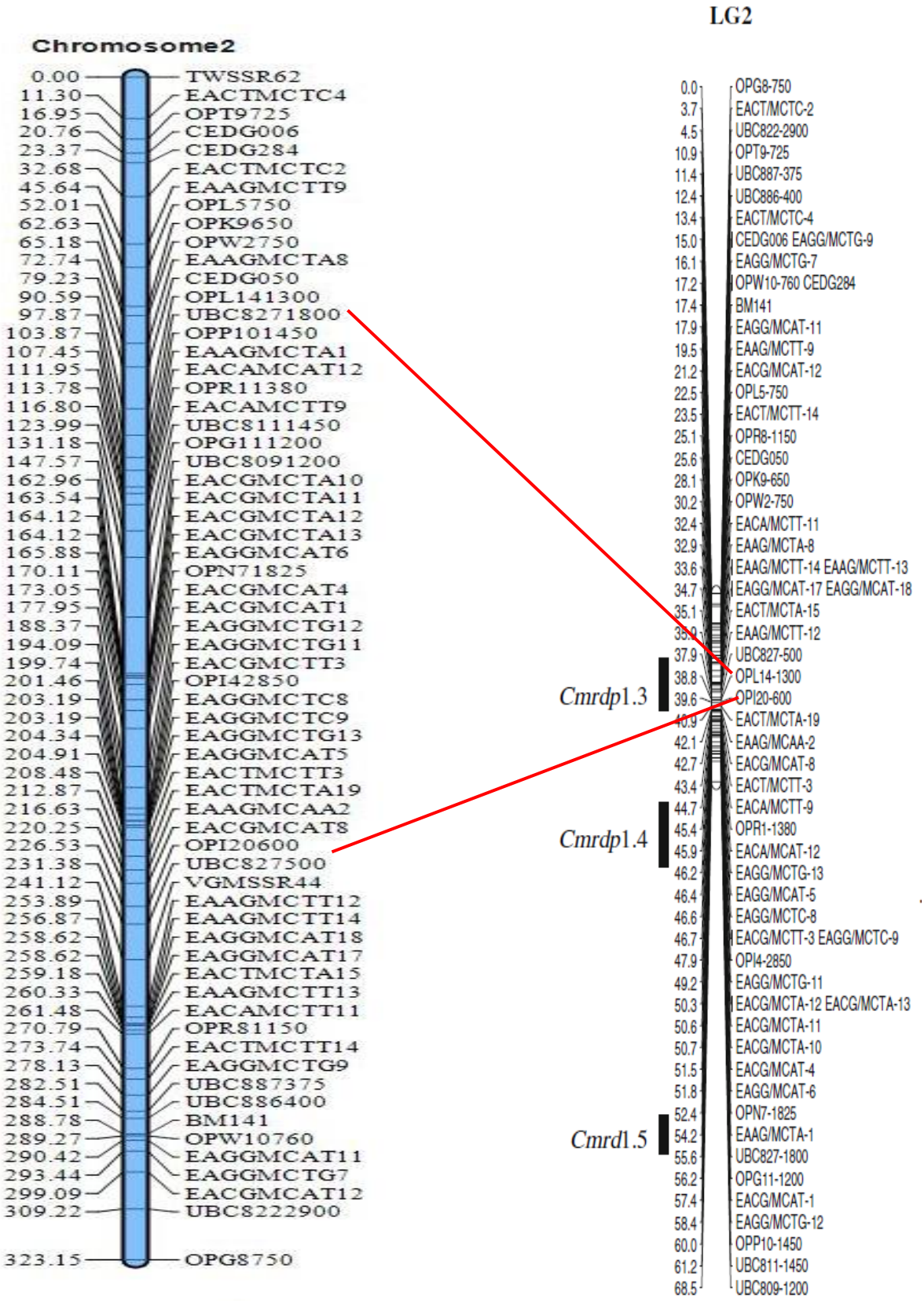


Figure 3.3.8.2. Comparison of the marker position between previously reported and new re-constructed genetic linkage maps of blackgram (LG2).

3.3.2.3. Markers mapped in proximity to bruchid resistance QTLs

In respect of the bruchid resistance QTL Cmrdp1.1 on LG1, two markers VMgSSR61 and TWSSR13 were mapped at positions 33.54 cM and 67.06 cM away from prior reported flanking markers CEDG133 and CEDG149, respectively. With regards to another QTL Cmrdp1.2 on same LG1, one marker TWSSR66 was positioned within the region of 25.04 cM spanning the previously reported flanking markers EACG/MCTA-15 and EAGG/MCTA-1 (Fig 3.3.9). As regards to QTL Cmrdp1.3 on LG2, a marker VMgSSR44 was identified within the flanking markers $OPI20_{600}$ and EAAGMCTT12. On the same LG2, a marker TWSSR62 was detected that was positioned at 111.95 cM and 97.97 cM away from the reported QTLs Cmrdp1.4 (flanked by markers OPR_{11380}, EACA/MCAT-12) and Cmrd1.5 (flanked by markers EAAG/MCTA-1, UBC827-1800), respectively(Fig 3.3.9). On LG10 two markers MRGSSR54 and VMgSSR9 were identified on either side of the QTL Cmrdp 1.6 that were located 36.81 cM and 135.89 cM away from the reported flanking markers EACG/MCTA-14 and EAAG/MCAA-4, respectively (Fig3.3.9). No new markers could be located in the vicinity of the QTLs on LG3 and LG4.

3.3.2.4. Validation of markers in blackgram germplasm

A total of 15 new markers (both mapped and unmapped on linkage map) were used for validating the association of markers with bruchid resistance in 27 blackgram genotypes with known resistance reaction. Among the 27 blackgram genotypes, 26 were susceptible to bruchid infestation and only 1 genotype i.e wild accession of blackgram (TW) possessed resistance to bruchids. Three genic-SSR markers TWSSR14, TWSSR68 and TWSSR87 were identified that were 100% efficient in discriminating the resistant TW from rest of the susceptible genotypes (Fig 3.3.10). Genic-SSR markers TWSSR13 and TWSSR16 were able to differentiate the TW resistant genotype from all but one (PantU-30) susceptible genotypes (Fig 3.3.11).

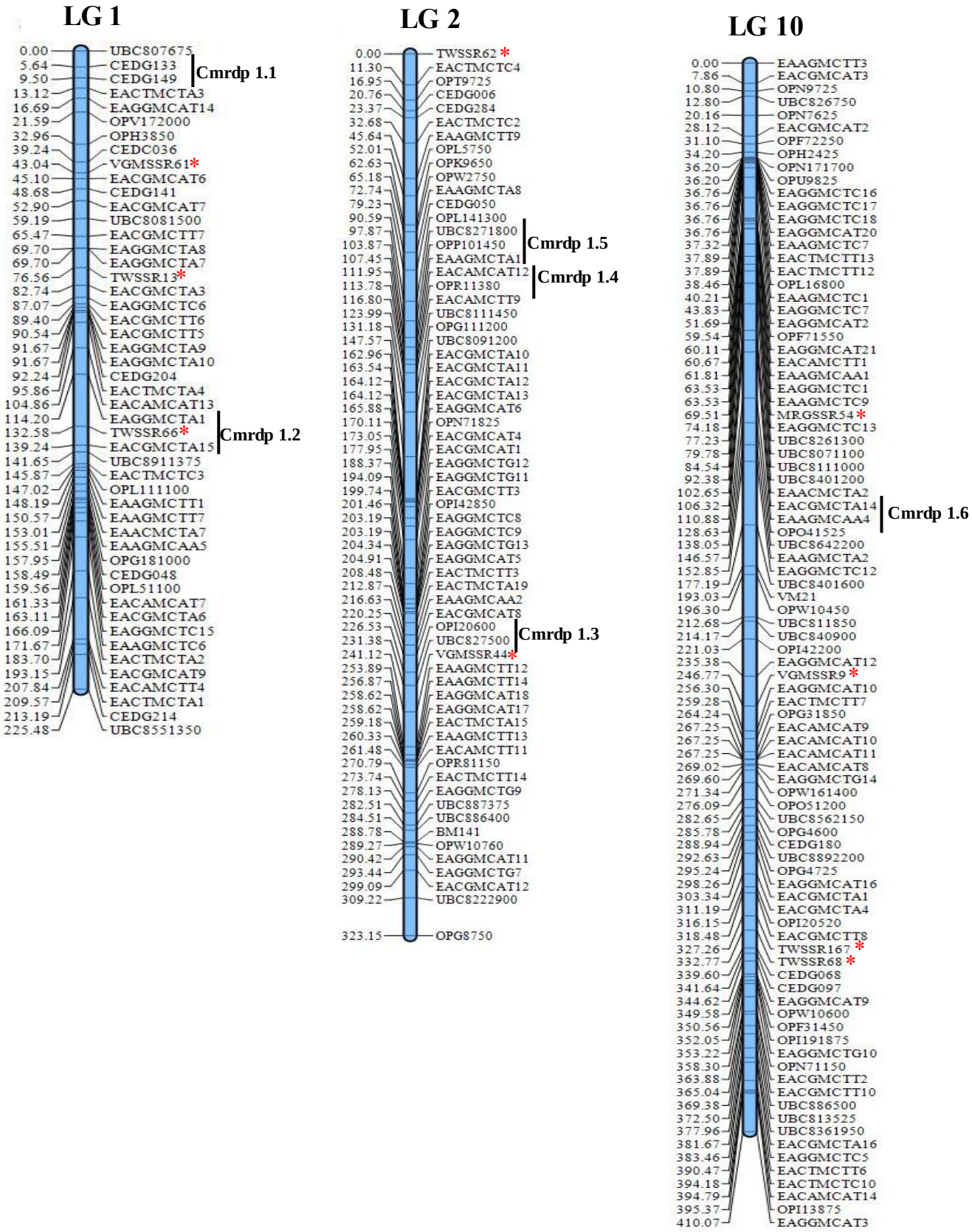


Figure 3.3.9. Reconstructed genetic linkage map of blackgram showing the location of newly identified SSR markers and QTLs associated with bruchid resistance. Asterik sign denotes new genic-SSR markers and one genomic-SSR marker.

In addition, the marker TWSSR13 also was potent in differentiating another resistant genotype (RIL68) from the susceptible genotypes. Even though the marker TWSSR62 was located close to bruchid resistance QTL, it was not as efficient as the other markers as it could differentiate the TW resistant from only 15 susceptible genotypes.

Figure 3.3.10. Validation of genic-SSR marker TWSSR14 in 27 blackgram germplasm consisting of 26 bruchid susceptible (Lanes 1-14, 16-27) and 1 bruchid resistant (TW) wild (Lane 15) accessions. (Lane M: 100bp ladder)

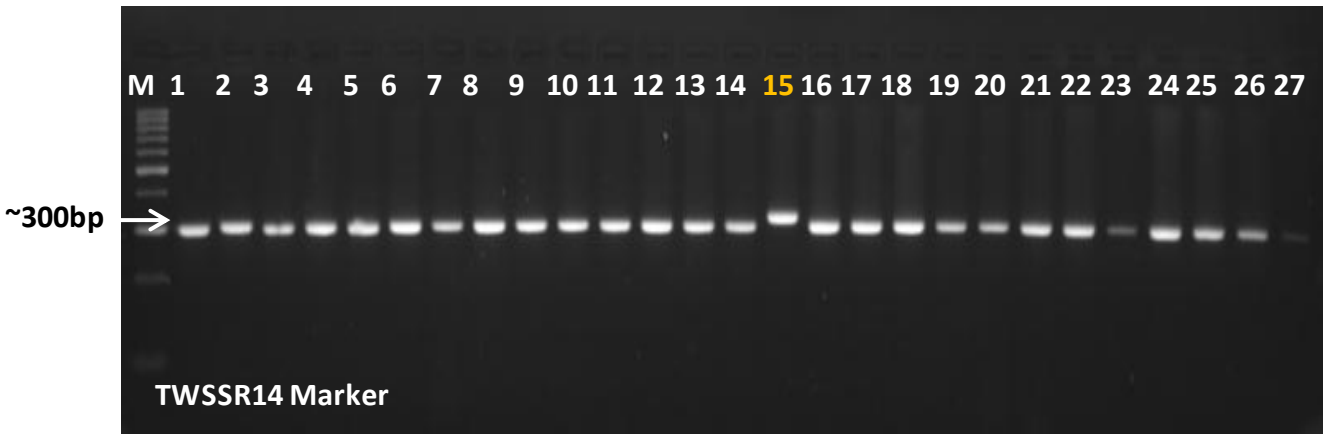
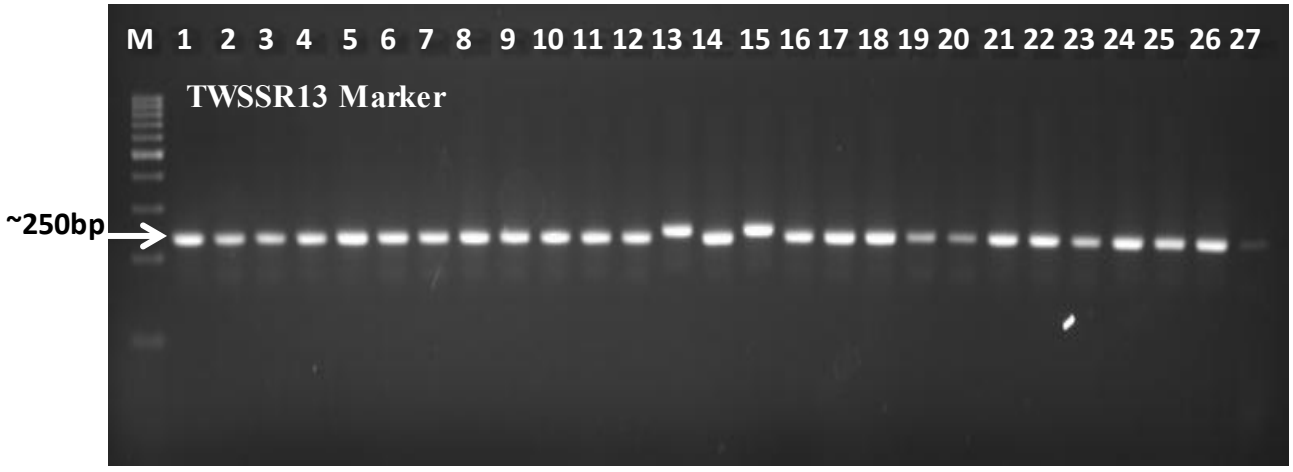


Figure 3.3.11. Validation of genic-SSR marker TWSSR13 in blackgram germplasm. (Lane M: 100bp marker, Lane 13,15: bruchid resistant and rest of the lanes represent susceptible genotypes).



3.3.2.5. Identification of candidate resistance genes

The candidate genes related to bruchid resistance and associated with the newly developed markers were identified by searching for sequence homology of genic-SSR marker sequences with adzuki bean genome. Homology search through BLASTN and BLASTX showed presence of LRR-proteins and other resistance related genes on adzuki bean genome in proximity to marker sequences (Table 3.3.2). Genic-SSR marker TWSSR167 (LG10) shared homology with vicilin-like seed storage protein and was located on chromosome 6 (Chr06: 4,096,533..4,099,730) of adzuki bean (Fig 3.3.12). Although TWSSR14 marker sequence shared similarity with homeobox-leucine zipper protein ATHB-13 and was located on chromosome 1 of adzuki bean (LOC108323167 and 12,922,508..12,925,097), important resistance genes were located in the flanking regions of this marker sequence. The resistance genes located in the vicinity included beta-D-glucanase, endochitinase, TMV resistance protein N-like and LRR repeat protein and receptors (Fig 3.3.13). Interestingly, the flanking regions of TWSSR15 marker on adzuki bean genome showed presence of significant defense and resistance related genes such as wound-induced protein-like (LOC108328764), Thaumatin-like protein (LOC108327281), galactosidase (LOC108328830) and LRR receptor-like protein kinase (Fig 3.3.14). This TWSSR15 marker sequence showed homology with vacuolar-sorting receptor 1-like and was found positioned at LOC108329520 on chromsome 3 of adzukibean. Graphical representation of TWSSR87 marker's location on azukibean is given in Fig 3.3.15. Genomic-SSR marker MRGSSR54 was located close (36.81cM) to bruchid resistance QTL Cmrdp 1.6 on LG10. This marker sequence was found harboured in mungbean scaffold JJMO01001680 which shared homology with putative disease resistance protein At4g11170 having TIR, LRR and NB-ARC domains.

Table 3.3.2. Genic-SSR markers with their annotation, position on blackgram linkage map and adzuki bean genome.

Markers	Annotation (NCBI BLAST)	Blackgram linkage map	Position of SSR marker sequence on Azuki bean genome	Accession and NCBI gene Id
TWSSR167	Vicilin-like seed storage protein At2g18540	LG10	Chr06:LOC108334794 4,096,533..4,099,730	NC_030642.1 108334794
TWSSR13	Sugar carrier protein C	LG1	Chr01: LOC108320676 7,809,808..7,813,062	NC_030637.1 108320676
TWSSR14	Homeobox-leucine zipper protein ATHB-13	Un-mapped	Chr01: LOC108323167 12,922,508..12,925,097	NC_030637.1 108323167
TWSSR15	Vacuolar-sorting receptor 1-like	LG6	Chr03: LOC108329520 9,530,319..9,536,040	NC_030639.1 108329520
TWSSR16	Small nuclear ribonucleoprotein SmD3b-like	Un-mapped	Chr08: LOC108340453 22,852,256..22,871,904	NC_030644.1 108340453
TWSSR62	Transcription initiation factor IIF subunit beta	LG2	Chr09: LOC108343208 3,225,407..3,231,508	NC_030645.1 108343208
TWSSR66	Uncharacterized sequence	LG1	LOC108347851 472,148..474,033	NW_016114844.1 108347851
TWSSR68	UPF0496 protein At2g18630-like	LG10	Chr06: LOC108334559 3,947,383..3,949,484	NC_030642.1 108334559
TWSSR87	Histidine kinase 4 (*Vigna radiata*)	Un-mapped	Chr01: LOC106770782 4,438,961..4,447,494	NC_028351.1 106770782
VMgSSR61	Enhancer of mRNA-decapping protein 4-like	LG1	LOC108320978 60,128….69,128	NW_016115001.1 108320978
VMgSSR9	Palmitoyl-acyl carrier protein thioesterase, chloroplastic	LG10	Chr01: LOC108320727 35,036,858..35,041,318	-
VMgSSR44	Uncharacterized protein LOC108331185	LG2	Chr04: LOC108331185 6,809,424..6,809,822	NC_030640.1 108331185
MRGSSR10	JJMO01001488 - TMV resistance protein N (Mungbean)	LG9	JJMO01001488 mungbean scaffold	Raizada and Souframanien unpublished data
MRGSSR54	JJMO01001680 - Protein SUPPRESSOR OF npr1-1, CONSTITUTIVE 1 (Mungbean)	LG10	JJMO01001680 mungbean scaffold	Raizada and Souframanien unpublished data

Figure 3.3.12. Graphical representation of genic-SSR marker's position on adzuki bean genome. The TWSSR167 marker sequence showed homology with vicilin-like seed storage protein and was found located on chromosome 6 of adzuki bean genome.

Figure 3.3.13. Graphical representation of TWSSR14 genic-SSR marker's position on adzuki bean genome. The TWSSR14 marker sequence showed homology with homeobox-leucine zipper protein ATHB-13 and was found located on chromosome 1 of adzuki bean genome.

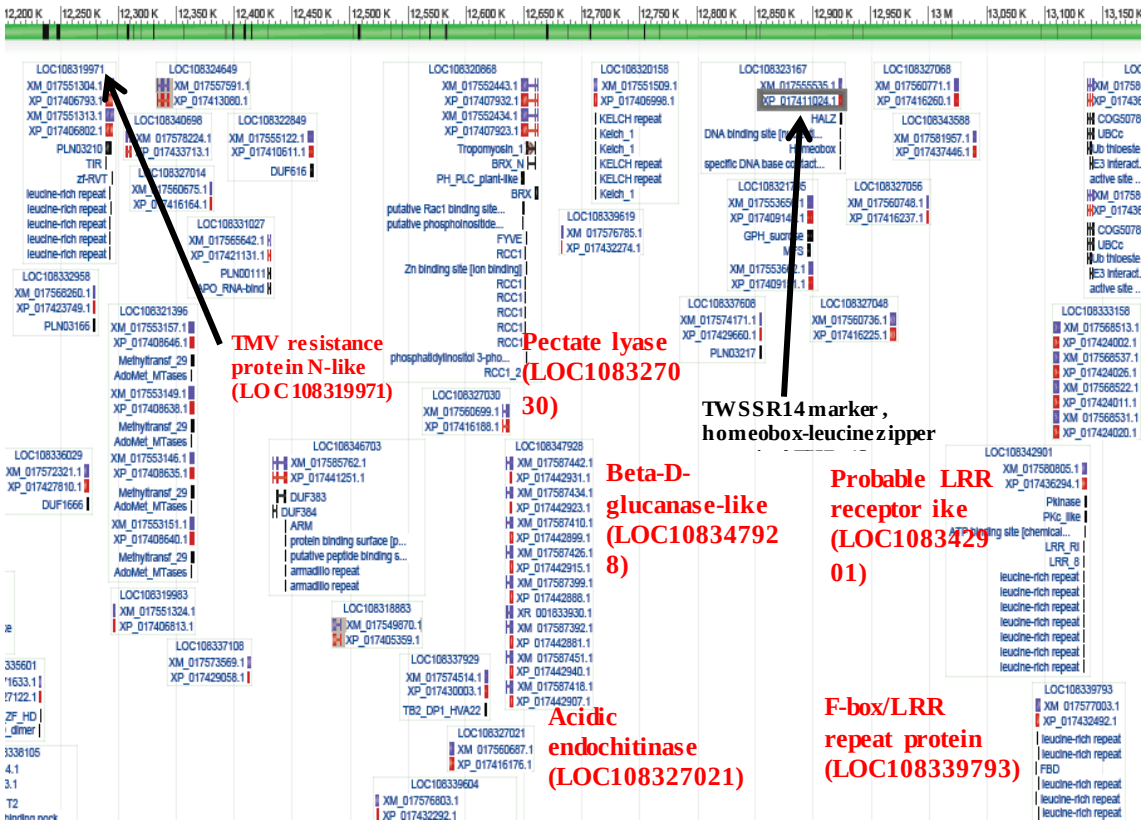


Figure 3.3.14. Graphical representation of TWSSR15 genic-SSR marker's position on adzuki bean genome. The TWSSR15 marker sequence showed homology with vacuolar sorting receptor protein and was found located on chromosome 3 of adzuki bean genome.

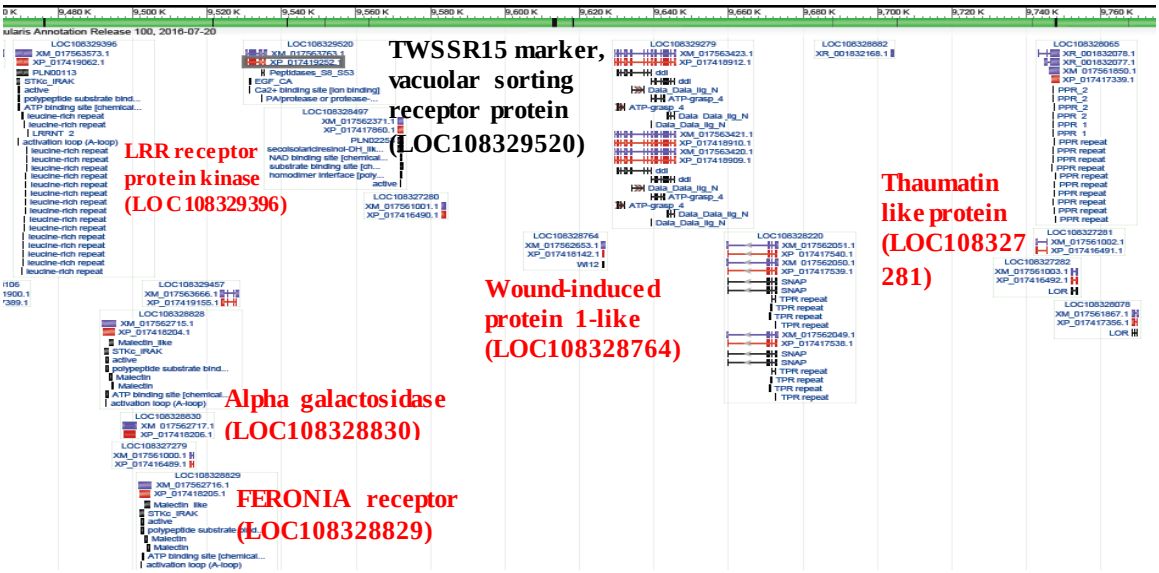
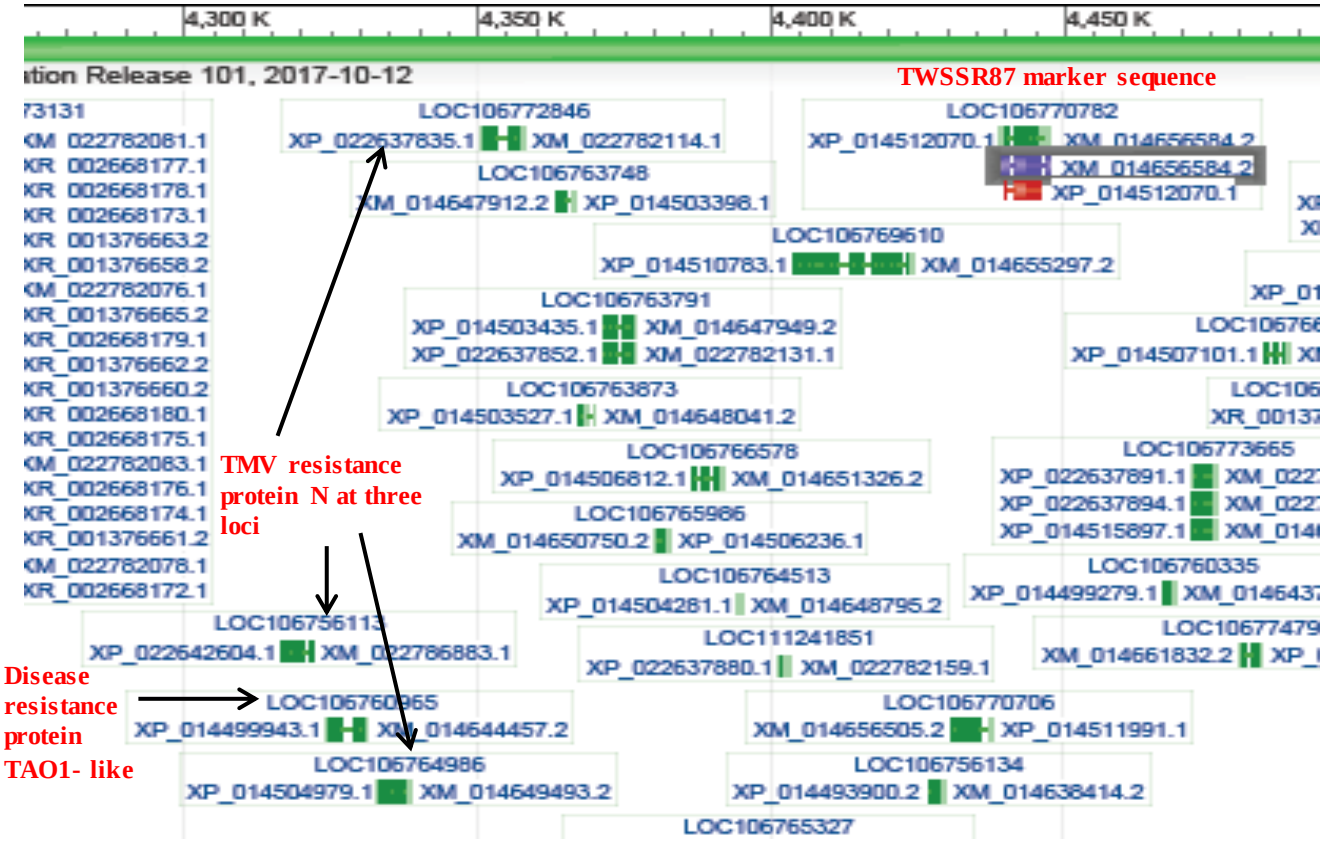


Figure 3.3.15. Graphical representation of TWSSR87 genic-SSR marker's position on mungbean genome. The TWSSR87 marker sequence showed homology with histidine kinase 4 and was found located on chromosome 1 of mungbean genome.



3.3.2.6. Detection of putative SSR marker associated with quantitative trait locus (QTLs).

One-way ANOVA performed on the mean of the groups formed on the basis of individual segregation pattern of 33 SSR markers, resulted in the identification of 30 SSR markers putatively associated with 2 quantitative traits studied. The results of one – way ANOVA analyses on the marker- phenotype association are given in Table 3.3.3 and Table 3.3.4 for percent adult emergence and development period respectively. A simple regression analysis was performed keeping phenotypic values as dependent variable and the individual segregation pattern of SSR marker loci as independent variable.

One way ANOVA analysis showed association of 29 markers with trait percent adult emergence and 30 markers with another trait development period. While regression analysis showed 19 markers (out of 29 markers) and 17 markers (out of 30 markers) associated with trait percent adult emergence and development period respectively. Out of 33 markers, 29 markers found associated with both traits through one-way ANOVA. A total of 12 markers were found associated with both traits through regression analysis.

Table 3.3.3. SSR marker associated with percent adult emergence identified through one-way ANOVA and simple regression.

One- way ANOVA			Simple regression		
Marker	**F value**	**Probability**	**Marker**	**R^2 value**	**Probability**
TWSSR62	19.3	0	-	-	-
TWSSR 68	4.57	0.03	TWSSR 68	0.039	0.057
TWSSR 15	14.4	0	-	-	-
TWSSR 66	11.1	0.001	TWSSR 66	0.007	0.407
MRGSSR 34	11.1	0.001	MRGSSR 34	0.028	0.108
TWSSR 167	8.73	0.004	TWSSR 167	0.035	0.07
VGMSSR 9	8.73	0.004	-	-	-
MRGSSR 10	4.06	0.047	MRGSSR 10	0.024	0.136

VGMSSR 44	14.2	0	-	-	-
TWSSR 24	6.04	0.015	-	-	-
VGMSSR 402	7.54	0.007	VGMSSR 402	0.015	0.245
TWSSR 106	12.1	0	TWSSR 106	0.007	0.419
-	-	-	MRGSSR 54	0.077	0.008
TWSSR 113	11.4	0.001	-	-	-
-	-	-	TWSSR 119	0.036	0.104
VGMSSR 403	10.8	0.001	VGMSSR 403	0.006	0.425
VGMSSR 89	10.2	0.002	-	-	-
MRGSSR 70	4.85	0.03	-	-	-
VGMSSR 68	6.18	0.015	-	-	-
TWSSR 144	8.73	0.004	-	-	-
VGMSSR 1	16.9	0	VGMSSR 1	0.035	0.069
VGMSSR 24	9.73	0.003	VGMSSR 24	0.009	0.413
VGMSSR 628	11.4	0.001	VGMSSR 628	0.025	0.159
VGMSSR 411	16.9	0	VGMSSR 411	0.009	0.349
MRGSSR 97	18.9	0	MRGSSR 97	0.007	0.437
VGMSSR 133	9.92	0.002	VGMSSR 133	0.065	0.028
TWSSR 117	10.2	0.002	TWSSR 117	0.011	0.324
VGMSSR 212	17.7	0	VGMSSR 212	0.098	0.002
VGMSSR 410	16.7	0	-	-	-
MRGSSR 85	11.9	0	MRGSSR 85	0.022	0.149
TWSSR 125	11.5	0.001	-	-	-

Table 3.3.4. SSR marker associated with development period of bruchid pest identified through one-way ANOVA and simple regression.

One- way ANOVA			Simple regression		
Marker	**F value**	**Probability**	**Marker**	**R^2 value**	**Probability**
TWSSR62	18.2	0	TWSSR62	0.01	0.33
TWSSR 68	6.26	0.014	-	-	-
TWSSR 15	14.2	0	-	-	-
TWSSR 66	11	0.001	TWSSR 66	0.017	0.216
MRGSSR 34	11	0.001	MRGSSR 34	0.035	0.073
TWSSR 167	9.24	0.003	TWSSR 167	0.017	0.211
-	-	-	TWSSR 13	0.02	0.168
VGMSSR 9	9.24	0.003	-	-	-
MRGSSR 10	5.35	0.023	MRGSSR 10	0.007	0.409
VGMSSR 44	12.8	0	-	-	-
TWSSR 24	7.49	0.007	TWSSR 24	0.013	0.27
VGMSSR 402	10	0.002	-	-	-
TWSSR 106	22.9	0	-	-	-
MRGSSR 54	5.19	0.025	-	-	-
TWSSR 113	14.1	0	-	-	-
VGMSSR 403	12.1	0	VGMSSR 403	0.001	0.477
VGMSSR 89	12.9	0	-	-	-
MRGSSR 70	5.57	0.02	MRGSSR 70	0.019	0.185
VGMSSR 68	6.04	0.016	-	-	-
TWSSR 144	9.24	0.003	-	-	-
VGMSSR 1	27.7	0	VGMSSR 1	0.035	0.07
VGMSSR 24	9.4	0.003	VGMSSR 24	0.019	0.226
VGMSSR 628	10	0.002	VGMSSR 628	0.019	0.213

VGMSSR 411	14.7	0	VGMSSR 411	0.019	0.188
MRGSSR 97	15.9	0	MRGSSR 97	0.05	0.034
VGMSSR 133	7.86	0.006	-	-	-
TWSSR 117	19.9	0	TWSSR 117	0.021	0.169
VGMSSR 212	27.4	0	VGMSSR 212	0.061	0.016
VGMSSR 410	30.8	0	-	-	-
MRGSSR 85	22.8	0	-	-	-
TWSSR 125	10.8	0.002	TWSSR 125	0.017	0.241

CHAPTER 4

DISCUSSION

This chapter is divided into 3 sub-chapters with sub-sections described as follows.

4.1 Development of genomic resources in blackgram.

4.2 Development of genic-markers and gene expression study in *V. mungo* L. Hepper for yellow mosaic disease resistance trait.

4.3 Development of genic-markers and gene expression study in *V. mungo* var. *silvestris* for bruchid resistance trait.

4.1. Development of genomic resources in blackgram

The Next Generation Sequencing (NGS) technologies especially RNA-Seq combined with bioinformatics analysis has great advantages on non-model species. RNA-Seq approach is a high throughput technology for examining the fine structure of a transcriptome that allows gene discovery and molecular marker development. The transcriptome sequence dataset of wild blackgram developed in the present study could serve as a reference transcriptome for future blackgram molecular research. It would aid development of gene-based markers and studying the effects of biotic, abiotic stresses and specific mutations on phenotypes (Thompson et al., 2009) [279]. Transcriptome data obtained in the present study from immature developing seeds, would aid in understanding the gene expression pattern underlying the seed development and pave way for incorporating desirable improvements in seed quality and yield. Molecular analysis of seed development and other seed parameters (quality and yield) were studied by transcriptome sequencing (Weber et al., 2005) [280]. In addition, the wild accession of blackgram, *Vigna mungo* var. *silvestris* as a source of bruchid resistance genes is highly pertinent to the blackgram breeding programme. Transcriptome sequencing and comparative analysis of cultivated (Souframanien and Reddy, 2015) [277] and wild blackgram species (this study) would facilitate exploration of the genetic differences between both species and comprehension of other aspects of genome evolution. Transcriptome sequencing of cultivated and wild species have been reported for legumes and other crops such as soybean (Libault et al., 2010) [281] chickpea (Garg et al., 2011) [282],

pigeonpea (Dutta et al., 2011) [283], mungbean (Chen et al., 2015) [284] and peanut (Zhang et al., 2012) [285].

4.1.1. Characterization of wild blackgram transcriptome

The present study was undertaken with the prime objective of generating transcriptome dataset of blackgram wild accession (TW) and its further utilization in development of molecular markers associated with important traits. The TW transcriptome dataset yielded 40,178 TCs which is less than that of 48,291 in cultivated blackgram (TU94-2). This difference in number of TCs could be attributed to the small seed size of TW, differences in gene expression patterns, or due to introgression or loss of genes in the cultivated blackgram during domestication and technical errors during sample processing. Most of the transcriptomic features such as contig length, SSR type and class etc of TW transcriptome compared with that of TU94-2 transcriptome. The lengths of contigs which varied from 280 to 6445 bp with an average of 446 bp in TW are comparable to that in cultivated blackgram (Souframanien and Reddy, 2015) [277]. In contrast to the similarities between TW and TU94-2 trancriptome features, few differences were observed were. For example in TW accession, CDS showed 27%, 16% and 7% similarities with common bean, *Citrus clementina* and *Glycine max*, respectively, whereas in case of TU94–2, CDS showed 50% similarity with common bean followed by 11% with *Glycine max* (Souframanien and Reddy, 2015) [277]. The homologies observed between wild blackgram, cultivated blackgram and common bean are obvious as they all belong to sub-tribe *Phaseolinae*. Similar kind of inter-generic gene conservations have been reported among other legumes such as field pea, faba bean, lentil with *Glycine max* (Kaur et al., 2012) [286] and chickpea with *Medicago*, soybean and common bean (Kudapa et al., 2014) [287]. The search for homology in the datatbase did not yield any hits in about 25.20% of CDS from TW transcriptome. This could be attributed to novel splicing variants, novel or rare genes, DEGs, crop-specific genes and contigs with

less conserved regions (3′-UTRs; C-termini or 3′sequences) (Garg et al., 2011) [282]. Since, TW is a wild species and thus, more genetically diverse, percentage of functional annotation of CDS was more in TW (72%) compared to TU94-2 (65%). Similarly, more number of KEGG pathways were identified in TW (140) of which, 524 CDS were associated with purine metabolism which are involved in ammonia assimilation and detoxification in specialized tissues, primary nitrogen metabolism in tropical legumes, nucleotide biosynthesis and degradation in cotyledons and embryonic axes of blackgram (Souframanien and Reddy, 2015) [277].

4.1.2. Frequency and distribution of genic-SSRs

In this study, a total of 1621 SSRs were discovered in 3.3% of the transcripts with a frequency of one SSR per 11.1 kb which is comparable to that in cultivated blackgram (Souframanien and Reddy, 2015) [277] but higher than in mungbean (3.3 kb) (Gupta et al., 2014) [288], pigeonpea (8.4 kb) (Dutta et al., 2011) [283] and soybean (7.4 kb) (Cardle et al., 2000) [289]. The less number of SSRs obtained in wild blackgram transcriptome could be owing to the reduced number of transcripts (40,178) compared to the cultivated blackgram (48,291). The SSR frequency is known to generally vary among plant species and can be assigned to a variety of reasons such as genomic composition, SSR search criteria, size of the transcriptome dataset, mining tools and sequence redundancy (Morgante et al., 2002) [290]. Present study showed predominance of tri-nucleotides in wild transcriptome which is expected to evade frame-shift mutations. Nontrimeric SSRs get suppressed in coding regions (Wakeley, 1994) [291] as observed in mungbean (Gupta et al., 2014) [288] and common bean (Blair et al., 2011) [292]. Wild transcriptome showed abundancy of di-nucleotides (30.2%) and tri-nucleotides AAG/CTT (29.9%) nearly similar however tri-nucleotides were found to be predominant in cultivated blackgram (Souframanien and Reddy, 2015) [277], cultivated peanut (Liang et al., 2009) [293], chickpea (Garg et al., 2011) [282], mungbean

(Chen et al., 2015) [284] and pigeonpea (Dutta et al., 2011) [283]. Among di-nucleotides, AG/CT were found to be the most common motif present in wild blackgram commensurate with other legumes like chickpea (71.1%) and cultivated blackgram (59.7%) (Souframanien and Reddy, 2015; Garg et al., 2011) [277,282]. Bioinformatic analysis showed almost equal distribution of di-nucleotide repeats in UTRs (235, 48%) and CDS (242, 49.4%), while tri-nucleotides repeats (444, 68.7%) were predominantly found in CDS. As reported in cultivated blackgram (Souframanien and Reddy, 2015) [277], 5'UTRs in wild blackgram also accommodated more triplets than 3'UTRs. Occurrence of SSRs in genic or expressed regions (coding regions and UTRs) and in genomic regions including introns is a non-random phenomenon with strong biasness and varies among taxonomic groups. SSR distribution depends upon relative frequencies of replication slippage, point mutation, selection pressure and environmental stresses that increase mutation events in SSRs (Trifonov, 2004) [294]. Development and application of genic-SSR markers from transcriptome data have been reported in many crop species including legumes such as blackgram (Souframanien and Reddy, 2015) [277], mungbean (Gupta et al., 2014) [288], cowpea (Gupta and Gopalakrishna, 2010) [295], pigeonpea (Dutta et al., 2011) [283], common bean (Blair et al., 2011) [292] and chickpea (Jhanwar et al., 2012) [296]. In the present study, the genic-SSR markers developed showed an amplification rate of 58% which is acceptable and could be due to sequencing errors or occurrence of SNPs causing disruption in priming sites and large PCR amplicons could be due to intervening introns as primers were designed from expressed genomic regions but screened with genomic DNA.

4.1.3. Genetic diversity study in blackgram germplasm

Genetic diversity analysis among 27 blackgram genotypes with 31 genic-SSR markers amplified a total of 89 alleles. The PIC value ranging from 0.14 to 0.85 with an average of 0.54 in the present study is higher than that reported in cultivated blackgram but is

comparable with that in other legumes including pigeonpea (Dutta et al., 2011) [283]. PIC value depends on number of samples taken, accession source and marker scoring skills. High PIC values of these genic-SSR markers from wild transcriptome indicate their higher efficacy and utility. In the present study 19 SSR primers showed null alleles in 27 different blackgram genotypes which could be due to primer site variation or unrecognized intron splice sites disrupting priming sites or presence of large introns between the primers, resulting in a too large product or in extreme cases, failed amplification (Souframanien and Reddy, 2015) [277]. This is anticipated due to the diverse genetic nature of wild from cultivated species as primers were designed from wild species and screened in 27 genotypes comprising of cultivars, varieties, landraces etc.

4.1.4. Frequency and distribution of genic-SNPs

Recent trend showed extensive utilization of EST-SNPs for high-resolution genotyping, construction of genetic maps, and genome-wide association studies in several crops (Jhanwar et al., 2012) [296]. In the present study, 1845 SNPs were identified in 40,178 TCs of TW which is comparable to that reported in chickpea (Jhanwar et al., 2012) [296], soybean (Yadav et al., 2015) [297] and mungbean (Van et al., 2013) [298]. Transition to transversion ratio (Ts/Tv ratio) of 1.6 calculated for TW species in the present study was higher in comparison to maize (Ching et al., 2002) [299]. The Ts/Tv ratio helps in understanding the process of molecular evolution (Sablok et al., 2011) [262] and in checking the quality of SNP calls (DePristo et al., 2011) [300]. However, transition bias over transversion is universal (Rosenberg et al., 2003) [301] and is caused partially by cytosine methylation (Shen et al., 1994) [302]. A high Ts/Tv ratio also generally indicates more accuracy (Liu et al., 2012) [303] in SNP call. Therefore, the high transition bias in the present study, may reflect either high methylation levels in the blackgram genome or good accuracy of SNP calling.

4.1.5. High Resolution Melting assay

HRM is a very efficient method for genotyping SNP/SSR/indels and could be used for many purposes such as plant cultivar identification, genetic mapping, QTL analysis, pathogenic species diagnosis, and gene discovery. There are various reports of its applications in plants such as alfalfa (Han et al., 2011) [304], almond (Wu et al., 2008) [267] and potato (De Koeyer et al., 2010) [305]. In this study, we attempted to validate SNPs that were identified from transcriptome sequencing of wild and cultivated blackgram with the help of a less expensive HRM assay. The success rate of 78.87% obtained for validation of SNPs by HRM assay in the present study is similar to that reported in other crops such as 91% in alfalfa (Li et al., 2012) [306]. Validation rate of HRM assay is a highly variable parameter influenced by several factors including amplicon size, number of SNPs genotyped and failed PCR amplification. But validation rate and thus, assay high-throughputness can be improved by standardization of kit components, primer concentrations, template DNA concentrations, in-silico detection of amplicons and primer-pairs for secondary structures. In the present study, discrepancies were observed in 12 SNPs which were genotyped as heterozygous by HRM assay due to melting pattern characteristic of heterozygosity, but identified as homozygous by RNASeq. Appearance of multiple melt domains in HRM curve is not only because of heteroduplex formation from heterozygous genotypes, but also could be due to sequencing and experimental errors, secondary structures, characteristics of DNA, concentrations of the ions and the volume of the solution in the assay system and sensitivity or specificity of HRM machine (Wu et al., 2008) [267]. Hence, these 12 SNPs were assigned as putative homozygotes. Similar discrepancy in genotyping through HRM assay as also reported in alfalfa (Li et al., 2012) [306]. HRM assay performed in the present study able to differentiate A/T variation (8 TWSNPs) of class I SNP which is the most difficult SNP class to be resolved by melting analysis (Liew et al., 2004) [268]. Total 17 SNPs which did not

qualify quality control parameters were analysed further by HRM assay. Among these, 10 SNPs showed single melting phase and thus, genotyped as homozygous by HRM assay which were also predicted as homozygous by RNASeq. Whereas remaining 7 SNPs, although identified as homozygous by RNASeq, showed multiple melt domains in HRM assay and thus, appeared as heterozygous. This showed the accurate genotyping of SNPs even with poor quality melt curves thus, suggest robustness of HRM assay observed in this study.

4.1.6. Functional significance of SNPs validated by HRM assay in different genes

Among HRM validated SNPs, 13 SNP loci were non-synonymous and exhibited missense mutations. These SNP bearing contigs (TU94-2 cultivar) showed homology with physiological genes such as cytochrome c-type biogenesis protein, U-box domain containing protein 13, cytochrome P450, pre-mRNA-splicing factor, ATP-dependent RNA helicase DEAH2, protein phosphatase, mitochondrial fission 1 protein A-like, dnaJ protein homolog 1 and cell division cycle protein. Thus, genotyping for these functional SNPs differentiated the wild accession from TU94–2 and other cultivars. Also, non-synonymous SNPs in physiological genes that alter the structure and function of encoded proteins could produce favourable alleles providing adaptive and evolutionary advantages (Gailing et al., 2009) [269]. For example, in this study a non-synonymous SNP in U-box domain-containing protein 13 could explain the tolerance of plant towards various abiotic stresses such as cold, heat and salt, because U-box containing genes played role in abiotic stresses (Liu et al., 2012) [303]. Similarly, pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH2 is involved in RNA processing and associated with early flowering (Herr et al., 2006) [307]. Cytochrome P450s are stress-responsive genes and involved in the biosynthetic pathways of plant allelochemicals such as insect toxins and repellents (Schuler, 1996) [308]. Therefore, a non-synonymous SNP altering secondary structure and thus, catalytic activity could be responsible for difference in susceptibility to various insects found in crop germplasm.

Likewise, HSP40/Dnaj protein is involved in viral pathogenesis such as for Potato virus Y (Hofius et al., 2007) [309] and Tomato spotted wilt virus (Soellick et al., 2000) [310]. Therefore, presence of a non-synonymous SNP could serve as a molecular marker linked to these virus resistance genes. Protein phosphatase 2c is involved in abscisic acid (ABA) signal transduction and thus, presence of a non-synonymous SNP may be responsible for modulating plant growth, seed dormancy and stomatal closure (Meyer et al., 1994) [311]. Similarly, cell division cycle CDC48 protein controls the turnover of immune receptors and mediates the degradation of viral proteins in plant cells (Begue et al., 2008) [312].

4.1.7. Genetic variation analysis using genic-SNPs through HRM assay

SNPs validated by HRM assay in TW and TU94-2 genotypes were further screened in different blackgram genotypes for analysing genetic variations using functional SNPs (genic-SNPs). SNP genotyping by HRM assay using 19 genic-SNPs were successfully demonstrated in discriminating the intra-specific accessions of blackgram along with establishing the phylogenic relationships. Development of SNP derived functional markers from transcriptomes have also been reported in other plants (Guo et al., 2017) [313]. SNP markers are the third-generation molecular markers with wider applications (Ganal et al., 2009) [314]. Conventional SNP screening approaches such as cleaved–amplified polymorphic sequencing, single-strand conformation polymorphism, denaturing high performance liquid chromatography, gene chip technology etc., are less accurate, time consuming and are of low resolution (Glavač and Dean, 1993) [315]. HRM is an efficient, rapid, highly sensitive and economical method for SNP screening with no post-PCR processing and contamination as compared to SnaPshot (Mehta et al., 2017) [316]. Analysing molecular genetics of germplasm using genic-SSR and SNP markers will help in exploring and understanding of minute genetic differences controlling desired traits, such as yield potential, resistance against pathogen infections and tolerance to abiotic stress. UPGMA

based dendrogram constructed from SNP genotyping data showed grouping of genotypes with common parents and same YMD and/or PMD reactions. Grouping of cultivars with common parents is anticipated due to genetic closeness. Three genotypes, IPU02-43, IPU07-3 and PantU-31 were grouped in cluster II along with their common parent DPU88-31. Likewise, in the same cluster II, Pusa3, DPU88-31 and NDU-1 having one common parent (T9) were placed together. Within cluster I, TU94-2 and KU96-3 both resistant to YMD and PMD are placed together in sub-cluster Ia, 6 genotypes all resistant to PMD except TU-67 (unknown reaction to PMD) grouped in sub-cluster Ib and sub-cluster Ic is comprised of two genotypes of which both are YMD resistant (PantU-19 and PantU-30) and PantU-19 alone is also resistant to PMD. In cluster II, YMD resistant Pusa-3 and IPU02-43 are grouped together in sub-cluster Ia. Clustering of YMD and PMD resistant genotypes based on genic-SNPs could help in genotype selection for breeding purposes. Moreover, grouping of genotypes together with common parents indicates the narrow gene pool or lack of genetic variation in blackgram cultivated germplasm. The dendrogram clustering, as reiterated by SNP genotyping by HRM assay, showed majority of blackgram genotypes to share allelic pattern (genic-SNPs) with TU94-2 cultivar for most of the SNP markers. Similar UPGMA clustering patterns of genotypes were reported in other genetic diversity studies in blackgram based on RAPD, ISSR, SSR and AFLP markers (Souframanien and Gopalakrishna, 2004; Gupta and Gopalakrishna, 2009; Souframanien and Gopalakrishna, 2009) [60, 61, 317]. Thus, applicability of genic-SNPs for studying genetic diversity in blackgram germplasm has been demonstrated and also serves as a theoretical basis for genetic improvement through exploitation of elite and novel/rare alleles. In addition, the significance of genic-SNPs could be perceived from the identification of a SNP differentiating the lone bruchid resistant wild accession of blackgram TW from the rest of the susceptible genotypes. The SNP marker TWSNP33 grouped all genotypes with TU94-2 leaving alone TW that formed a separate

OTU. This SNP bearing contig coded for UDP-glycosyltransferase 89B1protein. In plants, UDP-glycosyltransferases (UGTs) enzymes catalysed glucosylation of a diverse array of aglycones including plant hormones, secondary metabolites involved in stress and defense responses, and xenobiotics such as herbicides (Li et al., 2001) [318]. Moreover, 99 UGTs have been employed for molecular phylogenetic analysis in *Arabidopsis* (*L*i et al., 2001) [318]. The contigs containing the markers TWSNP 61 and TWSNP 73 and coding for molecular chaperone regulator 7 (BAG family) and WRKY transcription factor 17 respectively, grouped majority of the genotypes with TU94-2. BAG molecular chaperone regulator 7 control programmed cell death and unfolded protein response (UPR) during heat and cold tolerance (Williams et al., 2010) [319] and WRKY TF 17 acts as a negative regulator of defense signaling (Journot-Catalino et al., 2006) [320]. Notably, the TWSNP 59 marker grouped all genotypes with TU94-2 and none with TW. This marker sequence containing contig codes for casein kinase 1-like protein 3 which is involved in blue light responses (e.g. hypocotyl elongation and flowering) by phosphorylating CRY2 to reduce its stability (Tan et al., 2013) [321]. Besides serving as genetic diversity tools, these SNP markers could be associated with wild specific traits (novel/rare alleles) and thus, be used as functional markers for discriminating desired genotypes from other genotypes (Ganopoulos et al., 2013) [322]. The SNP markers TWSNP40a and TWSNP859 coding for cell wall-associated hydrolase and heat shock cognate 70 kDa protein 2-like respectively, grouped most of the genotypes with TW. SNP effects on protein secondary structure and functions due to missense substitutions were also studied. The I-TASSER results showed that the SNP (TWSNP904) present in peamaclein protein caused alteration in ligand binding kinetics resulting in different alleles in wild and cultivated blackgram. Peamaclein is an allergen from peach which showed homology with a potato antimicrobial peptide. In Pinus, the non-

synonymous SNP in the coding region of drought responsive genes was found to be responsible for greater adaptability to various abiotic stresses (Eveno et al., 2008) [323].

4.1.8. Genetic variation analysis using mungbean derived genomic-SSRs flanking RGAs in mungbean genome scaffolds

The cross-transferability of SSR markers from a crop to other related species lacking in genomic resources is an economical way of developing SSR markers in such crops (Souframanien et al., 2017) [24]. Several reports have claimed successful transferability of SSR markers to related species (Souframanien and Gopalakrishna, 2009; Gupta et al., 2014) [317, 288]. In this study, we screened the genomic-SSR markers associated with RGAs derived from mungbean genome for their transferability in blackgram. The transferability rate achieved was 70% which is high in comparison to other similar reports such as 50% for cowpea unigene-SSR markers (Souframanien et al., 2017) [24] and 68% collectively for adzukibean, commonbean, cowpea and mungbean derived markers in blackgram (Souframanien and Gopalakrishna, 2009) [317]. The transferability rate for markers depends on genetic closeness among the species and conservation of PCR primer binding sites flanking the SSR motifs (Souframanien et al., 2017) [24]. All transferable markers did not show null alleles in this study which however, is commonly observed in other similar studies like in all cowpea unigene-SSR markers (Souframanien et al., 2017) [24]. High transferability rate and less frequent null alleles observed in this study could be because of genome derived markers that lack problems inherent with genic-markers such as disrupted priming sites due to intronic splice sites, large introns etc. Another reason could be due to genetic relatedness between blackgram and mungbean. Genetic variation analysis using 30 SSR markers in 44 blackgram genotypes yielded 90 alleles with PIC value varing from 0 to 0.86 (average PIC: 0.43). This is comparable with other genomic SSR markers in other *Vigna* species (Souframanien and Gopalakrishna, 2009) [317] and thus, demonstrate the

utility of these resistance genes based genomic-SSR markers in blackgram. MRGSSR12 marker for SSR with $(AT)_{13}$ motif and MRGSSR110 for SSR with $(AT)_7$ with showed high PIC values of 0.86 and 0.83, respectively. These highly polymorphic markers were derived from scaffolds homologous with TMV resistance protein N and putative disease resistance protein At4g11170. Markers amplifying di-nucleotide motifs showed high PIC value which is consistent with earlier reports in cowpea (Souframanien and Gopalakrishna, 2009) [317] and soybean (Hisano et al., 2007) [324]. The 3 markers, MRGSSR1 for $(AC)_7$, MRGSSR82 for $(TA)_{32}$ and MRGSSR84 for $(AG)_9$ repeat motifs showed very low PIC values of 0.01, 0 and 0, respectively. These markers were derived from the mungbean scaffolds having homologies with protein suppressor of npr1-1, constitutive 1 and DNA-damage-repair/toleration protein DRT100. Low PIC values indicate lack of size variation in these 3 repeat motifs which might be due to their location in coding sequence of resistance genes (Richard and Dujon, 1997) [325] or within regulatory regions or in introns and are under strong selection pressure to be kept in a proper size range (Li et al., 2004) [326]. Significance of variable repeat motifs depends on its location whether in coding sequences or regulatory regions where they could cause a frame shift, a fluctuation of gene expression, inactivation of gene activity, and/or a change of function, and eventually phenotypic changes (Li et al., 2004) [326]. In this study, 3 SSR primers (MRGSSR12, 56 and 77) differentiated YMD resistant and susceptible genotypes while 9 SSR primers (MRGSSR9, 10, 12, 20, 32, 51, 56, 57 and 65) were able to differentiate PMD resistant and susceptible genotypes. Similarly, RGA derived SSR and ISSR markers were used to distinguish YMD resistant and susceptible genotypes in blackgram (Souframanien and Gopalakrishna, 2009) [317]. In the present study, null alleles were frequently observed with mungbean derived SSR markers in YMD resistant genotypes. This could be due to genotypes involved in this study that might have less number of functional variants of RGAs or altered priming sites, thus resulting in no

amplification. The high amplification rate and more allelic variation in susceptible genotypes may be due to the pseudogenes which are very frequent among NBS/LRR-like genes and are expected to display much more polymorphism than functional genes (Calenge et al., 2005) [327]. MRGSSR56 marker sequence from scaffold JJMO01001477 showed homology with DNA-damage-repair/toleration protein DRT100 having 9 LRR repeats which is able to complement bacterial recA mutations but native function is not known. Marker MRGSSR12 was designed from mungbean scaffold bearing TMV resistance protein N. TMV resistance protein N is a disease resistance protein having one TIR, one NB-ARC domain and six LRR repeats which trigger a defense system in plant against pathogens through direct or indirect interaction with avirulence protein (The Uniprot Consortium, https://doi.org/10.1093/nar/gkw1099). MRGSSR12 marker showed amplification of one allele in 2 YMD and 2 PMD resistant genotypes and none in the susceptible genotypes. Likewise, marker MRGSSR65 showed amplification of single allele in one PMD resistant genotype with no amplification in any of the susceptible genotypes. These results suggest the presence /identification of novel or rare alleles of the resistance genes by some MRGSSR markers in wild blackgram (Trombay wild), NDU-1, LBG17 and IPU02-6. Occurrence of similar rare alleles in YMD resistant genotypes has been previously reported in blackgram (Gupta et al., 2015) [328]. The significance of flanking genes found near to markers MRGSSR12 and MRGSSR56 sequences in mungbean scaffold is as follows: L-type lectin-domain containing receptor kinase VII.2 has 3 molecular functions i.e ATP binding, carbohydrate binding and serine threonine kinase activity which is known to be involved in signalling cascades. Transcription factor bHLH162 is involved in defense response to fungus and regulation of transcription by RNA polymerase II. LRR receptor like serine/ threonine protein kinase GSO1 together with GSO2 played key role in several physiological processes such as cell wall organization, regulation of cell division, root development and response to

wounding. Probable apyrase 7 functions in anther dehiscence and pollen exine formation and exhibited inducible expression pattern due to wounding and drought stress. LRR receptor like serine/threonine protein kinase plays role in signaling during pathogen recognition and activation of plant defense mechanisms (Afzal et al., 2008) [329]. UPGMA based dendrogram using mungbean derived genomic-SSR markers resulted in distribution of 44 blackgram genotypes into 7 clusters. The genetic closeness for some of the cultivars can be explained due to common parents in their pedigree. For example, genotypes IPU07-3 and IPU02-43 although from different crosses, DPU88-31 x PDU-1 and DPU88-31 x DUR-1, respectively, had one parent common and were grouped together in cluster I. Also, both genotypes IPU07-3 and IPU02-43 are resistant to YMD and PMD which might suggest sharing of resistance genes. DPU88-31 was grouped with one of its parent T9 in sub-cluster Ic of cluster I. Likewise, genotypes Pusa3 and DPU88-31 both were grouped together in cluster I along with their one shared parent, T9. In this study, grouping of blackgram genotypes with similar disease reaction was observed in cluster I which has also been reported earlier in blackgram (Souframanien and Gopalakrishna, 2009) [317]. Grouping of individuals based on disease reaction observed in this study could be due to housing of resistance genes on the same scaffold. For example, TMV resistance protein N, Protein suppressor of npr1-1, constitutive 1, putative late blight resistance protein homolog R1B-8 and putative disease resistance protein At4g11170 exhibited homology with the same scaffold JJMO01000125. Clustering of RGAs was also reported in other crops such as rice (Monosi et al., 2004) [330] and tomato (Dickinson et al., 1993) [331]. These transferable genomic-SSR markers would be a valuable resource for blackgram genetic analysis because resistance genes-based SSR marker polymorphism would represent the variation present in the resistance sources of blackgram genotypes. Since, these SSR markers were designed from mungbean wgs scaffolds that exhibited homology with resistance proteins and could

be lying within CDS, UTRs or regulatory regions of resistance genes, would offer an opportunity to investigate the consequences of SSR polymorphism on gene functions and regulation. These SSR markers would be helpful in the selection of appropriate genotypes for breeding cultivars with multiple stress tolerance.

4.2. Development of genic-markers and gene expression study for Yellow mosaic disease resistance trait in *V. mungo* L. Hepper

4.2.1. Development of genic-markers linked to YMD resistance trait

KU96-3 is an elite high yielding variety of blackgram specific for northern regions and possess resistance to YMV infection. No molecular markers linked to YMD resistance trait has been reported in this variety. Parental survey with 610 primers resulted in amplification in only 307 primers. The low amplification rate could be attributed to the genic nature of primers as primers were designed from transcriptome sequence dataset which constituting the expressed regions of genome. Low amplification turnover could be due to intervening introns and nucleotide variations present at primer annealing sites as primers were screened with genomic DNA of blackgram. Out of 307 primers showing amplifiction, only 153 showed size variation. The lack of polymorphism is comprehensible as both the parents are genetically closely related cultivars lacking genetic variations and the inherent conserved nature of expressed regions of genomes i.e. transcriptomes. It also could be due to the type of SSRs targeted for primer designing as transcriptome of both wild and cultivated blackgram both showed predominance of di and tri-nucleotides SSRs which by default possessed less extent of size variation (Dubey et al. 2019) [332]. Resolving of PCR products having narrow allelic size differences on agarose gels becomes difficult resulting in no apparent observable size variation and constraining the number of polymorphic primers. Therefore, screening with 610 genic-SSR primers resulted in only 3 with desired amplification pattern in the initial bulks and parents of the BSA. Linkage analysis with these 3 SSR markers resulted in

mapping of only one SSR marker TWSSR10 at a distance of 36.7 cM from YMD resistance trait while rest two markers did not get mapped. Moreover, validation of these 3 markers in germplasm showed congruency with the previously reported YMD marker CEDG180. Gene function annotation of these marker sequences using NCBI nr database resulted in homology of TWSSR10 marker sequence (TW contig 2139) with transcription factor TCP9 that positively regulates isochorismate synthase 1 of Salicylic acid synthesis pathway (Wang et al., 2015) [278] and VMgSSR86 marker sequence (TU94-2 Contig 1833) with formin-like protein 2 that has been hypothesized for role in pathogen perception. Formins bind with plasma membrane-localized phosphoinisotides (van Gisbergen et al., 2012) [333], and also with Rho-guanosine triphophatases (GTPases) (Bechtold et al., 2014) [334]. Both phosphoinisotides and Rho-GTPase have been known to play important roles in plant-pathogen-associated processes, including pathogen entry and immune activation (Hung et al., 2014; Kawano et al., 2014) [335, 336]. Moreover, when VMgSSR86 and VMgSSR547 marker sequences were mapped on mungbean genome, the resistance genes TMV Resistance gene N and enhanced disease resistance 2, respectively were found in the flanking regions. Therefore, with the above results it could be suggested that, these 3 genic-SSR markers though not directly linked to YMD resistance gene(s), may be involved in defense response pathway or the QTLs associated with YMD resistance trait. Hence, it could be inferred that the 610 primers deployed for screening with low amplification rate of only 50% were insufficient for mapping the desired trait necessitating the need to deploy more number of primers targeting the SSRs.

4.2.2. Gene expression study for YMD resistance trait

To get an insight into the basis of molecular mechanism of YMD resistance, transcriptome of YMD resistant TU94-2 cultivar at seed developing stage and expression patterns of specific genes in YMD resistant KU96-3 cultivar upon YMV challenge were studied.

4.2.2.1. Transcriptome characterization of TU94-2 cultivated blackgram: Defense system in response to viruses and other phytopathogens

RNA sequencing results of cultivar TU94-2 showed up-regulation of defense related genes. RNA-Seq showed up-regulation of FERONIA (FER) that serves as a receptor for a unique peptide ligand, RALF1 (Rapid Alkalinization Factor 1) and play role in effector-triggered immunity (ETI) through the RALF1–FER–RIPK signaling module that may intersect with the RIPK–RIN4 (RPM1-induced protein kinase - RPM1-interacting protein 4) pathway (Liao et al., 2017) [337]. Several uncharacterized RLKs and LRR containing proteins were found up-regulated in TU94-2 developing seeds that may be candidate R genes and may be involved in the perception of geminivirus and other pathogens: TYLCV-encoded C4 protein interacts with BARELY ANY MERISTEM (BAM) 1 and 2, while BCTV-C4 and *cabbage leaf curl virus*-NSP interact with LRR receptor like kinase (Rosas-Diaz et al., 2018; Piroux et al., 2007) [87,88] . Mitogen-activated protein kinase kinase 5 is the component of innate immune MAP kinase signaling cascade (MEKK1, MKK4/MKK5 and MPK3/MPK6) downstream of bacterial flagellin receptor FLS2 and plays role in hydrogen peroxide generation during hypersensitive response-like cell death (Ren et al., 2002) [338]. Transcript of mitogen-activated protein kinase kinase kinase NPK1 was found over-expressed which plays role in the NACK-PQR (NPK1-NQK1/MEK1-NRK1) MAP kinase signaling pathway controlling resistance gene-mediated responses such as the N-mediated resistance to tobamovirus (TMV) and the Rx-mediated hypersensitive response (HR) to *potato virus X* (PVX) (Jin et al., 2002) [339]. Plants employ both RNA silencing and autophagy as antiviral defense strategies during geminivirus infection for silencing of viral transcripts and degradation of viral virulence factors, respectively (Rodríguez-Negrete et al., 2009; Choi et al., 2018) [112,130]. The TU94-2 transcriptome showed up-regulation of three autophagy related proteins ATG 3, 5 and 8C and one serine/threonine-protein kinase ATG1t that may be

involved in interaction with geminiviruses as observed for autophagy-related NbATG8f protein with the *Cotton leaf curl Multan virus* CLCuMuB-βC1 protein (Yakupjan et al., 2017) [340]. The RNA-Seq data showed enriched transcript of Tobamovirus multiplication protein 1, which is a susceptibility factor and necessary for intracellular multiplication of tobamovirus (Yamanaka et al., 2002) [341] but its overexpression leads to increased accumulation of the membrane-bound forms and decreased accumulation of the soluble forms, thus, inhibiting tobamovirus multiplication (Hagiwara-Komoda et al., 2008) [342]. The family of TFs with the most members represented by DEGs was ERFs, followed by the zinc finger CCCH containing protein, MYB, WRKY, NAC and WD-repeat families which regulates several jasmonate and ethylene responsive defense genes under pathogen attack (Huang et al., 2015) [343] as reported in *G. arboreum* defense against CLCuD (Naqvi et al., 2017) [344]. Transcript for ERF 9 was up-regulated which binds to the GCC-box pathogenesis-related promoter element under stress (Ohta et al., 2001) [345] and negatively regulates defense against necrotrophic fungi (Maruyama et al., 2013) [346]. Gene encoding TIFY 10A was over-expressed which is a repressor of jasmonate responses and gets induced by wounding, jasmonate application and hervibory (Chung and Howe, 2009) [347]. Geminivirus infection also induces the expression of a DNA-binding protein TIFY4B that acts as a geminiviral resistance factor and the interaction of CabLCV and TGMV TrAPs with TIFY4B inhibits its potential role in cell cycle arrest (Chung and Sunter, 2014) [94]. Transcripts for β−1,3-glucanase, DnaJ, heat shock chaperones and callose synthase were observed up-regulated which might hinder cell to cell movement of viral particles as observed in β−1,3-glucanase interaction with TGB2 protein of Potato Virus X (PVX) (Stange, 2006) [348]. In TU94-2, several transcripts of components for ubiquitin proteasome system (UPS) were found up-regulated that are known to target virus proteins for degradation as defense strategy (Mandadi and Scholthof, 2013) [140] as evident in SUMO-

conjugating enzyme 1 (SCE1) interacting with geminiviral Rep protein (Sánchez-Durán et al., 2011) [142] and ubiquitin-conjugating (UBC) enzyme (SlUBC3) interacting with *Cotton leaf curl Multan virus* (CLCuMV) βC1 protein (Eini et al., 2009) [349]. This suggests that ubiquitin mediated proteolysis could be a defense strategy against symptom development. Gene encoding lipoxygenase (LOX) was found up-regulated that is known to be involved in jasmonic acid (JA) synthesis which gets induced by wounding, hervibory and pathogen invasion followed by induction of genes encoding proteinase inhibitors, flavonoid biosynthesis (chalcone synthase and phenylalanine ammonia lyase), sesquiterpenoid biosynthesis (hydroxymethylglutaryl CoA reductase), thionin (antifungal protein), and osmotin (antifungal protein). Interestingly, another gene coding for immunity suppressor found up-regulated was DOWNY MILDEW RESISTANCE 6 encoding a salicylate-5hydroxylase that converts salicylic acid (SA) to 2,3-dihydroxybenzoic acid (2,3-DHBA) (Zhang et al., 2017) [350]. It hitherto, negatively regulates defense related genes (e.g. PR-1, PR-2, and PR-5) and is required for susceptibility to the downy mildew pathogens *Hyaloperonospora arabidopsidis*, *Pseudomonas syringae* pv. *tomato* DC3000 and oomycete *Phytophthora capsica* (De Toledo et al., 2016) [351]. Above RNA-Seq results suggest that in cultivated blackgram developing seeds, majority of up-regulated cellular components function in defense and stress responses. such as upregulation of PTI and ETI components, tobamovirus multiplication protein, downy mildew resistance protein and LOX enzyme. As observed for TW, different gene expression profiles in TU94-2 imparts enhanced tolerance to diverse range of pathogens for which TW might be susceptible and thus making TU94-2 different from TW.

4.2.2.2. Gene expression study in YMD resistant KU96-3 cultivar through qRT-PCR

A total of 16 host genes significant in geminivirus infection cycle selected based on literature survey and TU94-2 transcriptome dataset, were analysed for their expression pattern in YMD

resistant and susceptible genotypes. Functions of these genes could be understood from the following discussion. Transcription factor MYB1R1 binds selectively to the DNA sequence 5'-(GA)GATAA-3' and gets induced due to high salinity, drought and ABA followed by regulation of drought-responsive genes (Shin et al., 2011) [352]. MYB transcription factors are master regulators of cellular responses and are known for their involvement in plant development, secondary metabolism, hormone signaling, biotic and abiotic stress responses as observed in rice and *Arabidopsis* (Katiyar et al., 2012) [353] and in cotton against CLCuD (Naqvi et al., 2017) [344]. WD repeat-containing protein DWA2 is a component of the CUL4-RBX1-DDB1-DWA1/DWA2 E3 ubiquitin-protein ligase complex as the substrate recognition module that acts as negative regulator in abscisic acid (ABA) signaling through degradation of ABI5 (Lee et al., 2010) [354]. Putative leucine-rich repeat-containing protein DDB showed 3.27 fold change which suggests its role in defense signaling as leucine-rich repeat proteins are frequently found involved in immune signaling activation against geminiviruses (Kundu et al., 2019) [15]. Tobamovirus replication proteins interact with many host cellular proteins (Liu et al., 2005) [355], with themselves (Goregaoker et al., 2001) [356], with the tRNA-like structure of the viral genomic RNA (Osman and Buck, 2003) [357], and with small RNA duplexes (Csorba et al., 2007; Hagiwara-Komoda et al., 2008) [358,342]. Among interacting host cellular proteins, Tobamovirus multiplication protein 1 (TOM1) is a seven-pass transmembrane protein localized on both vacuolar membranes and unidentified membranes of higher density (Hagiwara et al., 2003) [359] and acts as a susceptibility factor necessary for intracellular multiplication of Tobamovirus. Simultaneous knockout of TOM1 and its homolog TOM3 in *Arabidopsis thaliana* (Yamanaka et al., 2002) [341] and of orthologs of TOM1 and TOM3 in *Nicotiana tabacum* (Asano et al., 2005) [360] strongly inhibited Tobamovirus multiplication. TOM1 was reported for its interaction with the Hel domain of Tobamovirus replication proteins (Yamanaka et al., 2002) [341]. TOM1 is

an integral component of Tobamovirus replication complexes and aids recruiting of Tobamovirus replication proteins on membranes. Although TOM1 is required for Tobamovirus replication complex assembly, its overexpression leads to increased accumulation of the membrane-bound forms of the replication proteins and decreased accumulation of the soluble (nonmembrane-bound) ones thus inhibiting Tobamovirus multiplication (Hagiwara-Komoda et al., 2008) [342] probably by its interference with the functions mediated by other interaction partners. Likewise, mutations in viral or host factors that increase the affinity to the partner can also lead to above similar results, as with the case of TLJ mutants. As a result, mutations in TOM1 leads to reduced efficiency of intracellular multiplication of Tobamoviruses (e.g. crucifer strain TMV-Cg) due to reduced accumulation of viral coat protein (CP) and reduced amplification of TMV-related RNAs (UniProt). Therefore, for successful establishment of Tobamovirus replication, Tobamovirus replication proteins have evolved to have optimal (i.e., not too strong but not too weak) affinity to each of the interacting partners in their natural host cells that harbor given concentration of the host factors of given amino acid sequences. As described above, Tobamovirus replication proteins participate in both viral RNA synthesis and suppression of PTGS. Several reports demonstrated that TOM1 overexpression decreases the activity of Tobamovirus replication proteins to suppress PTGS (Hagiwara-Komoda et al., 2008) [342]. Several studies reported the involvement of leucine-rich repeat receptor kinases in defense response against viruses including geminiviruses (Reitz *et al.,* 2015) [361]. Plants employ both RNA silencing and autophagy as antiviral defense during geminivirus infection (Rodríguez-Negrete et al., 2009; Muthamilarasan and Prasad, 2013; Choi et al., 2018) [112,82,130]. RNA silencing leads to degradation of viral transcripts, while autophagy related genes result in degradation of viral virulence factors. In *N. benthamiana*, a key autophagy-related NbATG8f protein, interacts with the CLCuMuB-βC1 protein and leads to its degradation (Haxim et al., 2017) [362] and

similar degradations of viral proteins were also observed in other studies (Clavel et al., 2017) [363]. Likewise, argonaute 1 is a component of RNA silencing RISC complex which is involved in loading siRNA to the RISC complex for degrading viral transcripts, spreading of silencing signals for establishing systemic antiviral immunity and symptom recovery (Ghoshal and Sanfacon, 2015) [364]. Importin is a cytoplasmic-nuclear transporter involved in transport of proteins from cytoplasm to nucleus and reported for its interaction with coat protein of mungbean yellow mosaic virus for transport of viral particles to the nucleus (Guerra-Peraza et al., 2005) [146]. Although the mechanism is not known but upregulation of importin was found to be associated with increased resistance to geminivirus which is counter-intuitive. Similarly, high expression of eukaryotic translation initiation factor 4B3 in KU96-3 upon YMV invasion is not reported in literature but the upregulation of this translation factor may be required by resistant host to synthesize defense related proteins to combat the viral invader. DNA repair protein RAD4 is involved in repairing of DNA upon UV exposure and is known to interact with geminivirus ToLCNDV-Ren (AC3) protein (Pradhan et al., 2017) [365]. AOC1 is involved in jasmonic acid biosynthesis which gets repressed upon CaLCuV infection (Ascencio-Ibanez et al., 2008) [366], may be as a consequence of the inverse regulation between jasmonate and salicylic acid signaling pathways, as SA pathway gets activated during geminivirus-host interaction. In *Arabidopsis thaliana*, efficient RNA silencing requires RDR6 and its dsRNA-binding partner, Suppressor of Gene Silencing 3 (SGS3), to amplify viral siRNAs for mounting effective defense response against virus infection (Wang et al., 2010; Wang et al., 2011) [105, 111]. Also, studies on virus induced gene silencing (VIGS) triggered by *Cabbage leaf curl virus* (CaLCuV), showed conversion of geminivirus-derived transcripts into dsRNAs by RDR6 and SGS3 (Muangsan et al., 2004) [367]. RNA-directed DNA methylation is a process of transcriptional gene silencing which is a major defense response against geminiviruses

(Raja et al., 2008) [46]. Eukaryotic translation initiation factor 4E are recessive resistance genes of host plants against geminiviruses whose products are essential for the virus multiplication and infection cycle (Machado et al., 2017) [368]. Downregulation of AOC in resistance background upon YMV invasion is expected due to negative regulation of SA pathway which gets induced upon viral infection. However, RAD4, SGS3, RdRM, and eIF4E were observed to be upregulated in resistant genotypes upon geminivirus attack in several crops, but in this study, their downregulation suggests low transcript counts which could be due to the inactive state of these genes at that particular initial stage of YMV infection /or leaf sampling.

4.3. Development of genic-markers and gene expression study for bruchid resistance trait in *V. mungo* var. *silvestris*

4.3.1. Transcriptome characterization of wild blackgram: Innate immune system in response to bruchid pests and other phytopathogens

For elucidation of an effective defense response by plants against pathogens, timely perception of pathogens by plasmamembrane localized or cytosolic receptors is required (Morillo and Tax, 2006) [369]. RNASeq results of TW showed up-regulation of membrane receptors BAK1, SERK1, ERECTA and lectin domain receptors which regulates PAMP triggered immunity. For example BAK1 is involved in immunity to diverse RNA viruses (Nicaise and Candresse, 2017) [370]. Recently, SNP based high-density linkage map showed association of lectin receptor kinase and chitinase with bruchid resistance trait in wild blackgram accession TC2210 (Somta et al., 2019) [5]. Transcriptomics of oviposited developing seeds of mild tolerant blackgram, speculated that the receptor serine/threonine kinase (RSTK) is responsible for the perception of elicitors of bruchids (bruchins) (Baruah et al., 2017) [14]. Serine/threonine-protein kinase At4g35230/ BR-signaling kinase 1 (BSK1) is a downstream signaling molecule for receptor kinase BRI1 which positively regulates

brassinosteroid signaling and also plant immunity (Tang et al., 2008) [371]. These up-regulated receptor kinases in TW developing seeds transcriptome could be involved in bruchid resistance. Among differentially expressed transcription factors (TFs), most abundant were the ethylene response factors (ERFs) and the Tri-helix transcription factors. ERFs control expression of PR genes, ET, SA, and JA inducible genes (Gutterson and Reuber, 2004) [372]. The bHLH transcription factor is involved in immunity to viruses such as TYLCV and CLCuV (Wang et al., 2015; Naqvi et al., 2017) [278,344] . Interestingly, TW RNASeq showed upregulated transcripts for hydroperoxide lyase (HPL, CYP74B) which is involved in biosynthesis of jasmonic acid and green leaf volatiles that are deterrents to insects/pests (Rustgi et al., 2019) [373]. Besides HPL, transcripts encoding several anti-insect and anti-pathogenic compounds were found over-expressed in TW at basal level. These are acid phosphatase, Thaumatin like-proteins (TLPs), trypsin inhibitor/miraculin, vicilin and 7s globulin. TLPs are the PR family 5 (PR5) proteins are known to get induced by pathogen/pest attack (Van Loon et al., 2006) [374]. Proteinase inhibitors (PIs) are generally present in seeds and serve as defense compounds against insects, nematodes, viruses, bacteria and fungi pathogens. Trypsins control the developmental processes of insects such as molting and synthesis of neuropeptides and thus, the actions of trypsin inhibitors lead to growth and developmental retardation of the insect larvae (Shukle et al., 1985) [375]. Likewise, numerous studies have been conducted on analysing detrimental effects of legume vicilins, 7S globulins and acid phosphatases on insect's development especially in *C. maculatus* (Liu et al., 2005) [355] and also on pathogens and nematodes (Jakobek and Lindgren, 2002) [376]. Hence, RNA-Seq results of wild blackgram developing seeds suggest that over-expression of above mentioned anti-insect transcripts may be imparting resistance to brcuhids in TW. Also differential expression of several other defense effectors in TW make it different from cultivated blackgram with regard to tolerance to specific pests/pathogens.

4.3.2. Development of genic-markers linked to bruchid resistance trait in wild blackgram

Despite its significance, blackgram is lagging behind from other pulses in molecular genetics research. Major reason behind this, is the lack of genomic resources in blackgram. However, bruchid resistance QTLs has been mapped on genetic linkage map of wild blackgram (acc:INGR10133) but the reported flanking markers comprising of AFLP, ISSR and RAPD, are not practical for regular screening. Therefore, in this study, we have developed genic-SSR markers for bruchid resistance QTLs/trait and also mapped these markers on reported genetic linkage map of blackgram through re-construction of linkage map using previous reported 428 markers (Gupta et al., 2008) [4] along with 33 new SSR markers (from this study). Besides the known advantages associated with SSR markers, the markers developed in this study were designed from expressed regions of genome (Transcriptome) which have scope for direct gene tagging and identification of candidate resistance genes. Genome of blackgram had not been sequenced /published during the course of this study, yet but the genetic closeness among *Vigna* species (Kang et al., 2014) [266] allowed the use adzuki bean genome to identify the candidate genes for bruchid resistance QTLs. Annotation of markers with reference to adzuki bean or mungbean genome and validation of markers in blackgram germplasm resulted in identification of defense related genes which could be candidate genes related to bruchid resistance QTLs/trait. Both reconstructed map of this study and previous reported map showed colinearity with few discrepancies in order of the markers. This might be due to different softwares used for construction of genetic linkage map. A total of 8 bruchid resistance QTLs on 5 LGs (1,2,3,4 and 10) had been mapped on previous linkage map of blackgram (Souframanien et al., 2010) [12]. Newly developed marker TWSSR13 was located on LG1 near to the QTL Cmrdp1.1. On adzuki bean genome (Chromosome, chr: 1), the flanking regions of this marker sequence showed presence of defense related genes

such as transport inhibitor response 1-like protein TIR1 (contain LRRs) (Navarro et al., 2006) [377] and MYB domain containing binding protein (Tsuda and Somssich, 2015) [378]. Comparative analysis showed presence of QTL Brm3.2 (LG1) for bruchid resistance from *V. umbellata* on chromosome 1 of adzuki bean (Somta et al., 2019) [5]. Therefore, these observations suggest that blackgram QTL Cmrdp1.1 (LG1) and *V. umbellata* QTL Brm3.2 (LG1) both on adzuki bean chr 1 might be same QTL. In this study, an un-mapped marker TWSSR14 was found located at LOC108323167 on chr 1 of adzuki bean. Flanking regions of this marker sequence showed presence of beta-D-glucanase, endochitinase, TMV resistance protein N-like, LRR repeat protein and receptors. β-1-3-glucanases (PR-2) have been found to degrade pathogen carbohydrates to restrict microbial growth and also to make available PAMPs to plant PRRs for evoking defense response against leaf rust in wheat (Anguelova-Merhar et al., 2001) [379]. Chitinases (PR-3, 4, 8, and 11) degrade fungal cell walls, inhibit fungal spore germination (Broekaert et al., 1988) [380] and have been recently found to be involved in bruchid resistance in blackgram (Somta et al., 2019) [5]. TMV resistance protein N is a disease resistance protein that restricts pathogen by inducing hypersensitive response (Dinesh and Baker, 2000) [381]. NB-LRRs are characteristic domains of R proteins involved in pathogen perception (Caplan et al., 2008) [382]. Therefore, these genes could be candidate genes associated with bruchid resistance QTL on LG1 of blackgram. Remarkably, TWSSR167 marker was mapped on LG10 at a distance of 216 cM from Cmrdp1.6 QTL and shared homology with vicilin-like seed storage protein. Legume vicilins bind to chitin-containing structures of insect midgut such as *C. maculatus* and *Z. subfasciatus* (Firmino et al., 1996) [383] and thus affect larval development of these insects (Yunes et al., 1998) [384]. Detrimental effects of vicilins on pests (*C. maculatus)* have been studied in legumes (Yunes et al., 1998) [384]. Therefore, it could be postulated that the bruchid resistance due QTL Cmrdp1.6 in TW could be because of anti-insect vicilin proteins.

Apart from markers for reported QTLs, TWSSR15 marker in this study was mapped on LG6 with no associated QTL (Souframanien et al., 2010) [12]. But recently, on the SNP based linkage map of blackgram, bruchid resistance QTLs (qVmunBr6.1 and qVmunBr6.1) were identified on LG6 and also same QTLs on chr 8 of both adzuki bean and mungbean (Somta et al., 2019) [5]. The flanking regions of TWSSR15 marker on chr 3 of adzukibean showed presence of defense related proteins such as wound-induced protein-like (LOC108328764), thaumatin-like protein (LOC108327281), galactosidase (LOC108328830) and LRR receptor-like protein kinase. Moreover, chr 3 of adzuki bean (*Vigna nepalensis*) bear QTL Brm1.3.1 (Somta et al., 2008) [227]. These results indicated the presence of new QTL/QTLs on LG6 of TW (wild species studied in this project) that were not identified previously in the less saturated linkage map (428 markers, Souframanien et al., 2010) [12]. However, the association of TWSSR15 marker either with qVmunBr6.1 or Brm1.3.1 needs further investigation. TLPs are PR5 proteins reported to affect cucumber mosaic virus (CMV) multiplication and its spread in *N. tabacum* (Kim et al., 2005) [385] and also interfere in expression of some PIs to restrict pests (Singh et al., 2013) [386]. In rough lemon, miraculin-like protein (RlemMLP2- thaumatin motif) exhibited both antifungal and anti-insect activities (trypsin inhibitor) (Tsukuda et al., 2006) [387]. In potato, two cysteine-rich isoforms of wound-induced protein (Win1 and Win2) got induced upon mechanical wounding and shared homology with chitinase (Stanford et al., 1989) [388]. LRR receptor like protein kinase up-regulated in TW may play role in pest perception as has also been speculated in bruchid egg-induced transcript dynamics in blackgram (Baruah et al., 2017) [14]. Marker TWSSR16 sequence was found located on chr 8 (LOC108340453) of adzuki bean genome with flanking regions showing predominance of uncharacterized proteins, secondary metabolism related genes and apoptosis-inducing factor. Comparative analysis showed presence of QTLs qVmunBr6.1 and qVmunBr6.1 from blackgram LG6 on chr 8 of both adzuki bean and

mungbean (Somta et al., 2019) [5]. The marker VMgSSR9 on LG10 was mapped 135.89 cM away from QTL Cmrdp1.6 in blackgram. DnaJ and glucan endo beta-glucosidase 8-like genes were found at regions flanking to LOC108320727 on chr 1 of adzuki bean. Comparative analysis showed QTL Brm3.2 from *V. umbellata* (LG1) to be located on chr1 of adzuki bean. These results indicate that the QTLs Cmrdp1.6 of blackgram and Brm3.2 of *V. umbellata* located on the same adzuki bean chr 1 could be one and the same. The markers TWSSR62, TWSSR66, TWSSR68, VMgSSR61 and VMgSSR9 were mapped near to QTLs with no defense related genes being found in the regions flanking for these markers on adzuki bean genome. This could be due to incomplete annotation of genomic regions of adzuki bean. Combining the results of germplasm validation and the mapping of markers close to QTLs suggest that the markers TWSSR13 and TWSSR68 are candidate genic-SSR markers linked to bruchid resistance QTLs. Notably genomic-SSR marker MRGSSR54 from mungbean scaffold got mapped very close (36.81cM) to QTL Cmrdp 1.6 on LG10. This mungbean scaffold (JJMO01001680) showed homology with putative disease resistance protein At4g11170 having TIR, LRR and NB-ARC domains and thus, could be a candidate gene for bruchid resistance. Also, in the present study ANOVA and simple regression analysis of all 33 markers showed association of 29 markers and 12 markers, respectively for both traits percent adult emergence and development period. Altogether, the results from mapping of markers near to QTLs, their annotations on adzuki bean genome and their validation in blackgram germplasm combinedly suggest that the markers TWSSR13, TWSSR14, TWSSR16, TWSSR68, TWSSR87, TWSSR167 are potentially linked to bruchid resistance QTLs/trait and could be exploited in breeding programmes for marker-assisted selection of genotypes.

Chapter 5

Summary and future perspective

Pulses are one of the most important dietary components of human diet. These are rich sources of proteins and thus, are important for vegetarian population in fulfilling their protein needs. Among the different pulses, blackgram (*Vigna mungo*) is one of the most popular pulse, (also known as king of pulses) and contains easily digestible non-flatulence proteins. Currently, blackgram is gaining international significance and is being grown in countries such as Thailand, Japan etc., for various purposes. Besides being nutritional, it is also recognized for its medicinal significance for being antihyperglycemic (treatment of type 2 diabetes) , anti-cardiovascular and anti-cancerous. Blackgram is extensively grown in Asian countries, of which India is the largest producer of blackgram. Blackgram production is very much affected by biotic and abiotic stresses. Among biotic constraints, yellow mosaic disease (YMD) and pests (bruchid) attack are the most common, causing a huge loss to standing crop and stored seeds, respectively. YMD is caused by a geminivirus and transmitted by whitefly. While bruchids (*Callosobruchus* species) are storage pests causing damage to stored seeds. Chemical control methods are not economical and have concerns associated with food and environment safety. Thus, host-plant resistance (HPR) can be a viable option for durable protection from pathogens and pests. Very limited molecular research has been conducted in blackgram which included development of 3 genetic linkage maps, few YMD linked markers, identification and mapping of bruchid resistance associated QTLs and few transcriptomic studies on YMV invasion and bruchid egg deposition. These preliminary studies laid a foundation for further research and future directions are needed in blackgram to accelerate blackgram varietal development programmes. Further, additional constraints in blackgram improvement programmes include lack of genome sequence, genotype specific YMD resistance linked markers, difficult to use bruchid QTLs markers, lack of bruchid resistance sources and numerous virus strains breaking resistance of elite cultivars. Hence complete and clear understanding of molecular mechanism for resistance

along with the development of genomic resources are needed for accelerating blackgram breeding programmes. Wild progenitor of blackgram (*Vigna mungo* var. *silvestris*), Trombay wild, TW (INGR10133) is a resistance source for bruchid infestation (*C. maculatus*) with antibiosis mechanism. Till date, no genic-markers were reported for blackgram and thus in the present work, TW is molecularly characterized through RNA sequencing and a preliminary idea on YMD resistance mechanism is also established.

Genomic resources for blackgram were developed in the form of transcriptome dataset and markers were derived from the same. The transcriptome sequencing of Trombay wild (TW) generated 2.97 GB data accounted by 2,970,339,385 bases corresponding to 19,690,124 HQ reads (SRR 5931432, SRX3091690). After assembly of HQ reads, 40,178 transcript contigs (TCs) were obtained which were further analyzed by bioinformatics tools. A total of 38,753 CDS were predicted from 40,178 TCs by ORF Finder that showed similarities with *Phaseolus vulgaris, Glycine max* and *Citrus clementina* when annotated by BLASTx algorithm (NCBI), GO and KEGG. A total of 1621 simple sequence repeats (SSRs) were identified in 1339 TCs with a frequency of one SSR per 11.1 kb. PCR primer pairs were designed for 1171 SSR loci. An amplification rate of 58% was achieved. HQ reads alignment of TW and TU94-2 blackgram resulted in identification of 1844 SNPs. Positional distribution showed the presence of 1291 SNPs in CDS and 518 in UTRs. The transition (Ts) type SNPs were 1129 and transversion (Tv) type SNPs were 716 with a Ts/Tv ratio of 1.57 for TW and the most abundant SNP being of A/G type. PCR primers were designed for 1749 SNP loci. HRM assay was standarized in blackgram for SNP genotyping with a validation rate of 78.8%. SNPs were also classified into 4 classes based on Tm difference in HRM assay and non-synonymous SNPs were analysed *insilico* for predicting SNP effect on protein kinetics by I-TASSER. To the best of our knowledge, this is the first report on transcriptome characterization, markers development and standardization of HRM assay for

SNP genotyping of wild accession of blackgram. Besides these markers, genomic-SSR primers flanking RGAs on mungbean scaffolds were designed from mungbean genome that showed 70% transferability in blackgram. In addition, SSR and SNP markers developed in this study (31 SSR, 19 SNP and 30 mungbean SSR) were deployed in genetic variation analysis of blackgram germplasm. Host cellular factors involved in geminivirus infection cycle were targeted for marker development. A total of 177 genic-primers based on cellular proteins were designed from TU94-2 blackgram contigs and tested for their amplification in 4 blackgram genotypes by PCR.

Previously reported YMD linked molecular markers in blackgram are not tightly linked and are genotype specific. No marker has been reported for the YMD resistant cultivar KU96-3 grown in northern region. For targeting specific genes related to YMD resistance trait, several strategies were deployed. Characterization of YMD resistant TU94-2 transcriptome (48,291 contigs) for resistance gene analogues using DRAGO 2 pipeline resulted in identification of 392 contigs containing R gene domains. A total of 341 host cellular factors with redundancy (including isoforms) involved in geminivirus infection cycle (based on literature survey) were used for designing 177 gene specific primers. This is the first report on targetting cellular factors for YMD linked marker development in blackgram. In initial screening (Parental survey), a total of 610 primers designed in this study were used, resulting in 50% PCR amplification. Out of 153 polymorphic markers, only 3 markers (TWSSR10, VGMSSR86 and VGMSSR547) showed desired amplification pattern in BSA followed by screening in 218 RILs of mapping population. Phenotyping for YMD reaction was conducted at Trombay in 2 different seasons using infector row method and RILs showed segregation in 1:1 ratio that was confirmed by chi-square test. Linkage analysis using JOINMAP software showed linkage of 1 genic-SSR marker (TWSSR10) with YMD resistance trait at a distance of 36.7 cM. Validation results in blackgram germplasm for these

genic-SSR markers were similar to that observed for previously reported YMD marker CEDG180. Gene function annotation for these marker sequences using NCBI nr database showed homology of TWSSR10 (TW contig 2139) and VMgSSR86 (TU94-2 Contig 1833) marker sequences, respectively with transcription factor TCP9 that positively regulates isochorismate synthase 1 of salicylic acid synthesis pathway and with formin-like protein 2 that has been hypothesized for role in pathogen perception. Formins binds plasma membrane-localized phosphoinisotides and Rho-guanosine triphophatases (GTPases) which are involved in plant-pathogen-associated processes, including pathogen entry and immune activation. Also, on mungbean genome, resistance genes TMV Resistance gene N and enhanced disease resistance 2 were found flanking to VMgSSR86 and VMgSSR547 marker sequences, respectively. These results suggest that, these 3 genic-SSR markers may not be directly linked to YMD resistance gene(s) but may be involved in defense response pathway and associated with QTLs for YMD resistance trait. To get an insight into the molecular mechanism operating in YMD resistant KU96-3 cultivar in response to YMV invasion, expression pattern of 16 genes with significant roles in geminivirus infection cycle were analysed by qRT-PCR. In resistant background under infected condition, 10 genes showed upregulation which included transcription factor MYB1R1 and WD repeat-containing protein DWA2, replication protein A (RPA), LRR-containing protein DDB, Tobamovirus multiplication protein 1, LRR-RLK, autophagy-related protein 8C, argonaute 1 (AGO1), eukaryotic translation initiation factor 4B3 and Importin subunit alpha-2. Among these, autophagy-related protein 8C, argonaute 1 (AGO1) and importin subunit alpha-2 are the most important ones which indicate that the resistance or defense response could be due to activation of autophagosome mediated degradation of viral proteins, RNA silencing of target viral transcripts or genome and importin mediating virus particles movement in host plant. While transcripts for 6 genes which were otherwise reported to be upregulated in other

studies, were found downregulated in this study in resistant background that included DNA repair protein RAD4, Allene oxide cyclase 4, SUPPRESSOR OF GENE SILENCING 3, RNA-directed DNA methylation 3, RNA-directed DNA methylation 4 and Eukaryotic translation initiation factor 4E (eIF4E). The low expression of these genes may be due to the fact that the study was conducted at early stage of infection while the genes could be activated at later stages of infection.

Bruchid pests pose a huge troll on blackgram seeds during storage. No bruchid resistant variety has been reported in blackgram till date. Trombay wild urd (INGR10133) (TW) is a well known source of bruchid resistance in blackgram native to trombay hills and used in the present study. As of now, it is known that in TW, the resistance trait is under the control of two dominant duplicate genes and the QTLs associated with bruchid resistance have been identified and mapped. The resistance mechanism has been reported to be larval antibiosis which is constitutively expressed as dried seeds showed resistance to bruchids infestation. But biochemical basis of resistance and markers linked to QTLs which can be of practical use have not been reported. RNA sequencing results for TW developing seeds revealed upregulation of bruchid resistance and other defense related genes such as LRR-RLKs, acid phosphatase, 7S globulins, vicilins, thaumatin, miraculin, thioredoxin, hydroperoxide lyase (HPL, CYP74B). These upregulated defense related genes have been reported to confer resistance to bruchids in other studies and few of them have been validated successfully by qRT-PCR in this study. For bruchid resistance linked marker development, of the 434 SSR primers screened in parental survey, 235 and 106 primers showed amplification and polymorphism (48 VMgSSR, 34 TWSSR and 24 MRGSSR), respectively. A total of 33 polymorphic primers were used for screening the mapping population of 104 RILs. Reconstruction of genetic linkage map using genotypic data of these 33 new genic markers and 428 previously reported markers of earlier study (Gupta et al., 2008) [4] resulted in mapping of 11 markers in the reconstructed map. Comparative analysis of

reconstructed map and previously reported map showed high similarity between both with very few differences. The sequences of molecular markers in both the linkage maps were similar except at few locations. New genic-SSR markers that were found located near to the QTLs are VMgSSR61 and TWSSR13 for QTL Cmrdp1.1, TWSSR66 for QTL Cmrdp1.2, VMgSSR44 for QTL Cmrdp1.3, TWSSR62 for QTL Cmrdp1.4 and MRGSSR54 and VMgSSR9 markers for the QTL Cmrdp 1.6. Marker validation in germplasm showed 100% discrimination of resistant TW from all susceptible genotypes for markers TWSSR14, TWSSR68 and TWSSR87. SSR markers TWSSR13 and TWSSR16 showed >90% discrimination ability. To identify the candidate genes for resistance, genic-SSR marker sequences were searched for homology with adzuki bean genome. The marker TWSSR167 showed homology with vicilin-like seed storage protein which is reported for conferring resistance to bruchids. Although marker TWSSR14 sequence showed homology with homeobox-leucine zipper protein ATHB-13 which is not related to bruchid resistance, but defence related genes were found in upstream and downstream regions to this marker sequence in adzukibean genome such as beta-D-glucanase, endochitinase, TMV resistance protein N-like and LRR repeat protein and receptors. Similarly, TWSSR15 marker showed similarity with vacuolar-sorting receptor 1-like, but the flanking regions on adzukibean genome showed presence of defense related proteins such as wound-induced protein-like, thaumatin-like protein, galactosidase and LRR receptor-like protein kinase.

Based on these studies on genomic resources development, molecular markers and resistance mechanism for YMD and bruchid resistance, the following important conclusions were drawn:

From wild accession of blackgram, transcriptome dataset was generated and 3350 genic-markers were developed.

1. Protocol for SNP genotyping by HRM assay was standardized in blackgram.
2. A total of 2305 DEGs were identified by RNA sequencing of wild and cultivated blackgram immature seeds and RNAseq results were validated by qRT-PCR.
3. One genic-SSR marker (TWSSR10) derived from wild transcriptome was found linked to YMD trait at a distance of 36.7 cM.
4. In YMD resistant KU96-3 cultivar, RNA silencing and autophagy based molecular resistance mechanism has been speculated.
5. A total of 4 new genic-SSR markers were mapped in proximities to bruchid resistance QTLs in reconstructed genetic linkage map of blackgram.
6. RNA sequencing of wild blackgram developing seeds and gene expression analysis (qRT-PCR) showed anti-insect compounds such as acid phosphatase, vicilin, trypsin inhibitor and miraculin to be responsible for bruchid resistance in wild blackgram.

Future perspectives

1. Transcriptome dataset and genic-markers developed in this study will aid in accelerating molecular research in blackgram and is expected to enhance efficiency of marker assisted selection (MAS) in blackgram breeding programme.
2. RNAseq analysis resulted in identification of DEGs related to YMD and bruchid resistance whose function could be studied by transgenics.
3. The genic-markers developed for bruchid resistance could be useful for MAS.
4. Nucleotide variations present in DEGs related to YMD and bruchid resistance can be exploited for marker development.

This is to certify that all the changes suggested by the examiners have been incorporated into the final version of the thesis entitled 'Development of genic-markers for yellow mosaic virus and bruchid resistance traits in black gram (*Vigna mungo* (L.) Hepper)' submitted to the Dean-academic, HBNI in partial fulfillment of requirements for the Degree of Doctor of Philosophy of Homi Bhabha National Institute.

Signature of the Student

Endorsement by the Thesis Supervisor:

(I certify that the changes suggested by the reviewers are incorporated in the thesis before final submission)

Signature (with date) 15/2/2024.

सोफ्रमणीयन ज. / Dr. SOUFRAMANIEN J.
वैज्ञानिक अधिकारी / Scientific Officer
नाभिकीय कृषी एवं जैव प्रौद्योगिकी प्रभाग
Nuclear Agriculture & Biotechnology Division
भाभा परमाणु अनुसंधान केंद्र / Bhabha Atomic Research Centre
ट्रांबे, मुंबई-400085 / Trombay, Mumbai-400085.

Name: Dr. Jegadeesan Souframanien

Designation: Scientific Officer-G

Department/Centre: Nuclear Agriculture And Biotechnology Division

Name of the CI/OCC: Bhabha Atomic Research Centre, Mumbai

Thesis Highlight

Name of the Student: Avi Raizada

Name of the CI/OCC: Bhabha Atomic Research Center **Enrolment No.: LIFE01201404006**

Thesis Title: Development of genic-markers for yellow mosaic virus and bruchid resistance traits in black gram (*Vigna mungo* (L.) Hepper)

Discipline: Life Sciences **Sub-Area of Discipline: Plant Molecular Genetics**

Date of viva voce: 18/12/2021

Black gram (*Vigna mungo* L. *Hepper*) is an important pulse crop that constitutes major dietary component for population of Asian countries. Its production is majorly affected by yellow mosaic disease (YMD) and bruchid pest infestation. Black gram improvement programmes are getting hampered due to lack of genomic resources which further limits the molecular research. To date, very few YMD linked DNA markers with limited transcriptomic studies were reported and efficient markers and basis of resistance for bruchids attack are not yet worked out except one study on transcript dynamics upon bruchid oviposition on developing seeds. Here, large genomic resources for black gram were developed. For this, transcriptome dataset of wild black gram (TW) was generated by NGS technology (RNA-Seq) and used for developing genic-markers. Also transcriptome dataset was analysed for the presence of structural elements such as CDS, ORF, SSR, SNP through several bioinformatics tools and databases such as NCBI nr protein database, Blast2GO, KEGG database. A total of 1621 SSRs and 1844 SNPs were identified. PCR primers were designed for 1171 SSR and 1749 SNP loci and an amplification rate of 58% and 85% were achieved. SNP genotyping was performed with a validation rate of 78.8% by HRM assay. SSR and SNP markers were also used to study genetic variations in 27 different black gram genotypes. Besides genic-markers, 118 genomic SSR-markers were developed from mungbean genome scaffolds that harboured RGAs and a transferability rate of 70% was observed in 44 different black gram genotypes. For YMD resistance trait, marker TWSSR10 was mapped at a distance of 36.7 cM that showed homology with the transcription factor TCP9 (salicylic acid biosynthesis). RNA-Seq of TW and TU94-2 developing seeds showed upregulation of insecticidal and autophagy related transcripts that were validated by qRT-PCR. Gene expression studies in KU96-3 (YMD resistant) and TAU-1 (YMD susceptible) genotypes under YMV infected conditions showed high expression of autophagy related genes and argonaute. New genic-SSR markers in close proximity to bruchid resistance QTLs were developed and mapped on reconstructed linkage map. In summary the generation of huge genomic resources, a preliminary idea on YMD resistance mechanism and biochemical basis of bruchid resistance could be helpful in YMD and bruchid resistance breeding programmes.

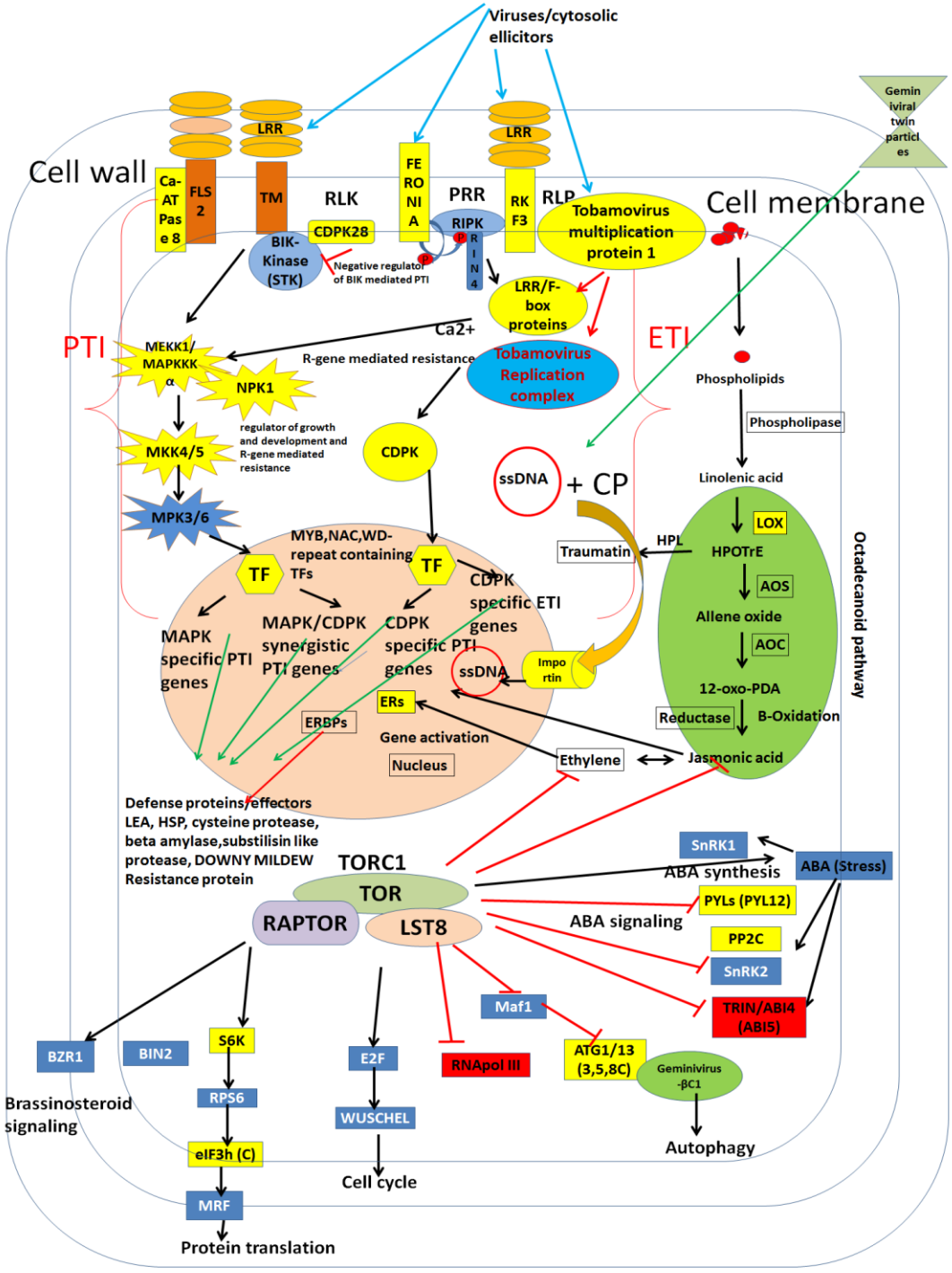


Figure 1. Proposed hypothetical pathway based on upregulation of defense related DEGs in YMD resistant TU94-2 seed transcriptome (YMV non-infected immature seeds) and upregulated genes in YMD resistant KU96-3 (YMV infected young leaves) at basal and induced stages, respectively.