Identification and characterization of SNPs in released, landrace and wild accessions of mungbean (Vigna radiata (L.) Wilczek) using whole genome re-sequencing

Pooja Bangar1,2 · Neetu Tyagi1 · Bhavana Tiwari1 · Sanjay Kumar1 · Paramananda Barman1 · Ratna Kumari1 · Ambika Gaikwad1 · K. V. Bhat1 · Ashok Chaudhury2

Abstract
Mungbean [Vigna radiata (L.) R. Wilczek var. radiata] is vital grain legume having nutritional and socio-economic importance, especially in the developing countries. We performed whole-genome re-sequencing of three accessions representing the wild progenitor species, released and landrace of mungbean to identify SNPs with relevance to genetic relationships analysis. Approximately 9.3 million raw reads were obtained using Ion Torrent PGM™ platform and more than 92% of the reads were mapped to the reference mungbean genome. We identified a total of 233,799 single-nucleotide polymorphisms in relation to the reference genome (SNPs: 103,341 in wild, 93,078 in released and 37,380 in landrace accessions) and 9544 insertions and deletions (InDels: 4742 in wild, 3608 in released and 1194 in landrace accessions) in the coding and non-coding regions. In all accessions, genomic variants were unevenly distributed within and across the mungbean chromosomes. Among these 5339, 4739 and 1795 SNPs were non-synonymous in 815, 790 and 317 genes of wild, released and landrace accessions, respectively. These polymorphisms might contribute to the variation in important pathways of genes for abiotic and biotic stress tolerance and important agronomic traits, such as seed dormancy, flowering time and seed size in mungbean. Among the randomly selected SNPs, a selected subset was validated using Sanger sequencing technique. The genomic variations among mungbean wild, released and landrace accessions constitute a powerful tool to support genetic research and molecular breeding of mungbean.

Keywords Domestication · Insertion/deletion · Mungbean · Single-nucleotide polymorphism · Whole-genome re-sequencing
Introduction

Mungbean is a diploid (2n = 22), self-pollinated and fast-growing (< 60 days) legume belonging to the subfamily Papilionoid of the family Fabaceae (Van et al. 2013). It is distributed mainly in Asia, Australia, West Indies, South and North America and tropical and subtropical Africa (Gowda et al. 2015). Archaeological evidence and morphological diversity studies suggested that mungbean was domesticated in the Indian subcontinent about 3500 years ago and later spread to other Asian and African countries amid the initial domestication process (Vishnu-Mittre 1974; Fuller and Harvey 2006; Lambriades and Godwin 2007). Vigna radiata var. subalobata (Roxb. Verdcourt) is presumed to be the originator of mungbean and widely distributed in the tropics of the Indian subcontinent, northern Australia to western Africa and Papua New Guinea (Smartt 1990; Tomooka et al. 2002). Among legumes, mungbean is an excellent and easily digestible source of protein for predominantly vegetarian populations in India. The raw and mature seeds are rich in nutrients including carbohydrates, protein, fiber, minerals, antioxidants like flavonoids (Quercetin-3-O-glucoside) and phenolics (Guo et al. 2012). In addition, mungbean roots fix atmospheric nitrogen via symbiotic relationship with nitrogen-fixing rhizobia (Chankaew et al. 2011; Allito et al. 2015). The raw and mature seeds are rich in nutrients including carbohydrates, protein, fiber, minerals, antioxidants like flavonoids (Quercetin-3-O-glucoside) and phenolics (Guo et al. 2012). In addition, mungbean roots fix atmospheric nitrogen via symbiotic relationship with nitrogen-fixing rhizobia (Chankaew et al. 2011; Allito et al. 2015). Thus, this crop is valuable both economically as well as nutritionally and supports sustainable agriculture (Graham et al. 2003; Bangar et al. 2018). However, despite being a valuable crop, overall production of mungbean is hampered by several factors, such as narrow genetic base, biotic and abiotic stresses (Singh et al. 2015). Morphophysiological and biochemical response of mungbean [Vigna radiata (L.) Wilczek] varieties at different developmental stages under drought stress was recently reported by Bangar et al. (2019).

Molecular marker development is critical for crop improvement programs (Bangar et al. 2018). Among the molecular markers, SSR (simple sequence repeats) has been used mostly for assessing genetic diversity in mungbean (Chen et al. 2015; Kaur et al. 2016). However, availability of only a limited number of polymorphic SSR markers in mungbean has hindered its use in marker-assisted breeding programs (Gwag et al. 2010; Shrivastava et al. 2014). Attempts were also made to use SSR markers from other closely related legumes, like moth bean, adzuki bean, cowpea and rice bean, in diversity analyses (Jingade et al. 2014; Shivakumar et al. 2017), which were not sufficient to meet the basic requirements in terms of polymorphisms. SNPs (single-nucleotide polymorphisms) representing single-nucleotide changes at a specific locus are considered as the most abundant and ultimate form of genetic variation (Mammadov et al. 2012). Among the genomic variations, SNPs including InDels (insertions and deletions) have gained importance due to their wide genomic distribution with chromosome-specific positions, co-dominant inheritance, amenability to automated genotyping and high reproducibility (Ganal et al. 2009; Mammadov et al. 2012; Huq et al. 2016). The high density of SNPs has resulted in the increased use of SNPs as the markers for various applications, such as accurate variety identification, diversity analysis, marker-assisted and genomic selection, association and QTL analysis and seed-purity testing (Van et al. 2013; Kang et al. 2014; Huq et al. 2016).

Presence of a narrow genetic base has necessitated the use of wild relatives for crop improvement programs (Zhang et al. 2017a, b; Muñoz et al. 2017). Further, the rapid development of the next-generation sequencing (NGS) technologies has enabled genome-wide discovery of SNPs for characterization of genetic differences among varieties (Lam et al. 2010; Zhang et al. 2014; Wei et al. 2016; Valliyodan et al. 2016). Whole-genome sequencing has facilitated generation of genome-wide markers in several legumes, such as lotus, soybean, pigeon pea, etc. (Sato et al. 2008; Schmutz et al. 2010; Varshney et al. 2012, 2013). Using whole-genome platform, genetic linkage mapping, domestication and evolution process were explained in wild and cultivated mungbean accessions (Isemura et al. 2012; Van et al. 2013; Kang et al. 2014; Schafleitner et al. 2016). The rapid development of NGS technologies along with the accessibility of complete mungbean genome facilitates an opportunity for the advance mungbean genomic research.

In this study, we sequenced whole genomes of a wild accession (Vigna radiata var. subalobata), a released variety (SML-668) and a landrace (MCV-1) of mungbean using Ion Torrent Personal Genome Machine™ (PGM™) platform. We analysed the genetic divergence of the wild accession and the domesticates after comparing the patterns of genetic variation among all accessions. The SNP polymorphisms observed in this study will enable the researchers to utilize genome-assisted breeding programs for utilization of genetic variations in wild and landraces in mungbean improvement.

Materials and methods

Sample preparation

The materials used in this study were maintained at the National Bureau of Plant Genetic Resources (NBPGR), Pusa, New Delhi. Genomic DNA was isolated from the fresh, young leaf material of the released variety SML-668 (selected from AVRDC Line, NM94, Punjab Agricultural University, Ludhiana, Punjab), wild Vigna radiata var.
sublobata (BBYD-2711) and landrace MCV-1 (collected from Pune, Maharashtra) using Qiagen DNeasy Plant Mini Kit (Cat no. 69106). The quantity, as well as quality of isolated DNA, was analyzed with the Bioanalyzer 2100 (Agilent Technologies, Singapore) and Qubit 2.0 Fluorometer (Invitrogen Corporation) using the Qubit dsDNA BR Assay (Molecular probe, Life Technologies, Invitrogen Division).

### Library preparation and sequencing

The DNA libraries were prepared by the Ion Shear Plus reagent kit (Life Technologies, USA) for each sample according to the user guide (Cat. no. 4471248, Revision C.0) with minor modifications (shearing time 90 s). First, emulsion PCR of prepared libraries was performed using the Ion OneTouch 2 (OT2) followed by the clonal amplification of DNA [according to Ion PGM™ Template OT2 400 Kit Cat. no. 4479878, Revision 3.0]. The prepared sample library was loaded on the Ion Torrent PGM machine (Life Technologies) using Ion PGM Sequencing 318v2 chip.

### Sequence read mapping

CLC Genomics Workbench (version 8.0.1) was used for the sequence mapping. The following criteria were used for trimming the raw sequence: adaptor sequences, short reads (< 101 bp), low-quality sequences (Phred quality score Q > 29), and ambiguous nucleotides (N > 2) were removed. The resultant high-quality, filtered reads were mapped onto the mungbean reference genome using default parameters. The recent version of mungbean [Vigna radiata var. radiata (assembly Vradiatora_ver6), RefSeq assembly accession: GCF_000741045.1] retrieved from NCBI was used as a reference (https://www.ncbi.nlm.nih.gov). For further analysis, the mapping output was changed to BAM format.

### Identification and analysis of SNPs, InDels and structural variants

SNPs and InDels were identified from the mapping output files using the ‘Fixed Ploidy Variant Detection’ tool of the CLC Genomics Workbench. The following criteria were used for identification of SNP and InDels: Ploidy of sample 2, required variant probability ≥ 90%, average quality of the SNP base > 29, minimum coverage 10, and minimum frequency 20%. Default values were used for all filters, such as general filters, noise, quality, direction, and position. Pyro-variants from the raw data were removed using the technology-specific filters. To remove false-positive SNPs from the data, SNPs showing homopolymer and 100% non-reference allele in output were filtered out. Further variant call format (VCF) files of SNPs and InDels of each accession were generated. Structural Variants tool of CLC Genomics Workbench was used to identify large and complex structural variants, such as insertions, deletions, inversions, translocations and tandem duplications, in read mapping using default parameters. Further, SAMtools (v0.1.16) software was used to measure genome coverage (Li et al. 2009).

### Structural and functional annotations of SNPs and InDels

Genetic variant annotation and effect prediction toolbox SnpEff 4.3 k (https://snp eff.sourceforge.net/) was used to annotate and classify the genomic variants (SNPs and InDels) using the VCF files as input (Cingolani et al. 2012). For this, SnpEff Vigna radiata database was created using the V. radiata_ver6 genome annotation file (gff) and genome sequence. Using this database, both HTML and text files were obtained as an output for all the accessions. Genomic variants were annotated based on their impact (high, moderate, low and modifier) and functional class (synonymous and non-synonymous substitutions) in the output. The variants were also located based on their genomic regions, such as downstream, upstream, exon, intron, intragenic and intergenic region, transcript, 3’and 5’un-translated regions (UTRs). The detected synonymous and non-synonymous SNPs and InDels of all accessions were visualized on the nuclear chromosomes using Circos Visualization tool (version 0.69) (Krzywinski et al. 2009). Further SNP variant containing genes were BLAST (cut-off value 1e-5) searched against NR database, and Blast2Go was used for functional analysis of variants. Gene ontology (GO) category was assigned to each variant according to the biological process (BP), cellular component (CC) and molecular function (MF) and the results were visualized by WEGO 2.0 (Web Gene Ontology Annotations Plot) (Ye et al. 2006). Finally, high- and moderate-effect genomic variants were mapped to the KOBAS 3.0 version database for biological pathways analysis. The conserved domains in the proteins were predicted against the Pfam database (release 27.0) at the BLAST threshold 1e−5 (Wu et al. 2006).

### SNPs validation

Two SNPs per chromosome were selected randomly for the experimental validation of identified SNPs between the varieties. Primer pairs were designed from the 250 bp flanking sequences of the selected SNPs using Primer 3 (https://bioinfo.ut.ee/primer3-0.4.0/). Polymerase chain reaction (PCR) amplification was done in a thermal cycler (Eppendorf) with 25 µl reaction mixture under temperature conditions of 94 °C for 6 min, followed by 35 PCR cycles inclusive of denaturation for 1 min at 94 °C, primer annealing for 1 min at 54 °C and primer elongation at 72 °C for 1 min, followed by a step of extension at 72 °C for 10 min. Only
those amplified PCR products which are single and clear distinct band on an agarose gel were purified and sequenced by the Sanger sequencing method. The high-quality sequences of the accessions were aligned with the reference genome and analyzed for SNP variation by Jalview software version 2.10.1 (Waterhouse et al. 2009).

Results

Whole-genome sequencing and mapping

Whole-genome sequencing was performed with Ion Torrent PGM™ sequencer using 318 v2 chip followed by a quality check of sequences with NGS core tools in CLC Genomics Workbench. A total of 3.3 million, 3.4 million and 2.6 million raw reads were obtained from the wild accession, released variety, and landrace, respectively. After removing low-quality reads, high-quality trimmed reads were mapped to the reference genome (Vigna radiata var. radiata) individually. Approximately 91–92% reads of all accessions were mapped with 52–60% coverage on the reference genome. Data of high-quality trimmed reads, mapped alignments and GC% of all three re-sequenced accessions are presented in Table 1.

Identification, characterization and chromosomal distribution of genomic variants

SNPs and InDels were identified in all accessions against the reference genome individually. We identified 103,341 SNPs (1 variant every 3230 bases) in wild accession, 93,078 SNPs (1 variant every 3585 bases) in released variety and 37,380 SNPs (1 variant every 8927 bases) in landrace in comparison with the reference genome. The number of transitions (Ts) and transversion (Tv) identified in wild accession was 66.46% and 33.54%, respectively, with Ts/Tv ratio of 1.98. Similarly, the released variety has 66.83% transition and 33.17% transversion with Ts/Tv ratio 2.01 and landrace had 66.20% transitions and 33.80% transversion with Ts/Tv ratio 1.95. Thus, the overall frequency of transitions was found to be higher than transversion in all the accessions (Supplementary Fig. S1). Transition frequency of A/G was more than C/T, while T/C was higher as compared to G/A among the accessions. However, the transversion frequency of A/T and C/G was higher as compared to T/A and G/C. Likewise, on individual comparison to the reference genome, 4742 InDels (2594 insertions and 2148 deletions), 3608 InDels (2,019 insertions and 1589 deletions) and 1194 InDels (693 insertions and 501 deletions) were identified in wild, released, and landrace accession, respectively. The length of insertions and deletions was 1–29 bp in wild accession, 1–36 bp in released variety and 1–24 bp in landrace (Supplementary Fig. S2). Most of the InDels were found to be less than 5 bp in length. Among the InDels detected, single-base InDel was found at a higher rate (68.9%, 71.8%, and 75.3%) in all three accessions of Vigna. The longest deletion of 36 bp was found in released variety and the longest insertion of 26 bp was found in wild accession. InDels which are greater than 1 kb in length were considered as structural variations. Wild accession harbors a maximum number of structural variations (5043) followed by released variety (3686) and landrace (3533). Complex structural variations, such as tandem duplication, replacement, cross-mapped breakpoints, identified are presented in Supplementary Table S1. Among these, maximum and minimum numbers of genomic variations were found in wild accession and landrace, respectively.

For analyzing the chromosomal distribution of genomic variations, the latest version of mungbean genome was used as a reference. The SNPs and InDels detected from each accession are shown in the Vigna radiata chromosomes using Circos Visualization tool with 100 kb sliding window size (Fig. 1). Maximum and minimum number of genomic variations was found on chromosome 8 and chromosome 3, respectively (Table 2). However, among all the accessions, the highest density of SNPs was found in chromosome 8 and chromosome 3, respectively (Table 2). However, among all the accessions, the highest density of SNPs was found in chromosome 1 (44.9/100 kb) and lowest in chromosome 3 (19.6/100 kb). Distribution of SNPs and InDels showed a positive correlation with the length of chromosomes in all the accessions. However, it was observed that SNPs and InDels distribution was not uniform within and across the chromosomes, as regions without SNPs/InDels were also noticed. Further to identify SNP hotspots, criteria of ≥ 400 SNPs per 100 kb were considered for all the accessions. In the wild accession, five hotspots were identified in chromosomes 1, 2, 9 and 10,
while in the landrace and released variety only two hotspots were found on chromosome 9.

The selected SNPs were validated by their gel amplification in all the accessions. Table 3 lists the primer descriptions used for the verification of the selected SNPs along with their chromosomal reference positions and genomic locations. The results of in silico discovery of 16 SNPs out of selected 22 SNPs in mungbean were verified by the Sanger sequencing method followed by aligning the accession sequences using Jalview software (Fig. 2).
| SNP ID   | Chromosome | Reference position | Reference V. sublobata | MCV-1     | SML-668    | Variant Type     | Forward (F) primer (5′-3′) | Reverse (R) primer (5′-3′) | Product size | Tm (F) | Tm (R) | GC% (F) | GC% (R) |
|---------|------------|--------------------|------------------------|-----------|------------|----------------|---------------------------|---------------------------|--------------|--------|--------|---------|---------|
| 1       | NC_028351.1 | 4,139,677          | C                      | T         | A          | Upstream gene  | TACGAAAAG GAGGGC ATATG   | AATTTATGC TGGCTC CAAGA   | 179          | 54.92  | 54.98  | 45      | 40      |
| 2       | NC_028351.1 | 22,089,531         | T                      | C          | G          | Intergenic     | ACTGTTGTTG GAGGCC TAGCT  | ACCCAGTTT CAACCC AGGAA   | 167          | 58.94  | 59.07  | 50      | 50      |
| 3       | NC_028352.1 | 111,585            | A                      | T          | G          | Missense       | TACCGGTAC AGCTGA GAATA  | TGCGATATT CAAGTC TGGAG   | 190          | 55.09  | 54.98  | 45      | 45      |
| 4       | NC_028352.1 | 20,606,865         | A                      | C          | T          | Intergenic     | CATCCCAACA ACTAGC CTAGG | TAGTACACA CCGAGA GATGT   | 227          | 55.23  | 55.04  | 45      | 45      |
| 5       | NC_028353.1 | 7,339,845          | G                      | T          | C          | Missense       | AACCTCCTC TGAATG ACAAC  | TCGGCCCA TTGGCT ACCAT    | 190          | 54.98  | 54.77  | 45      | 40      |
| 6       | NC_028353.1 | 12,661,878         | G                      | A          | T          | Intergenic     | ATTTTCCCA ACTTAG AGGCC  | TAAGGCCCT TCAAGA GTTTC   | 153          | 55.29  | 54.98  | 45      | 45      |
| 7       | NC_028354.1 | 1,017,819          | T                      | A          | C          | Intergenic     | TCGGATGAC AGAGTT AGACT  | AACCTCGT ACTTTG ACCCTC  | 215          | 55.03  | 54.96  | 45      | 45      |
| 8       | NC_028354.1 | 13,487,393         | T                      | A          | C          | Upstream gene  | GACAGTTGCA CAGAGA AACCT  | GAAGTCTCTC TGCTAC GACC   | 173          | 54.94  | 55.03  | 45      | 45      |
| 9       | NC_028355.1 | 5,770,737          | T                      | C          | A          | Intergenic     | ATGTCACAACC AAATCC AAACG | ACTTTTGGCT CCTGGT ACTAG  | 198          | 54.77  | 55.31  | 40      | 52.6    |
| 10      | NC_028355.1 | 31,958,930         | C                      | A          | T          | Intergenic     | GTGCTTGTCT TCAACC CATAC  | TCTCTTCAT TCGAGC AAGGC  | 220          | 55.17  | 55.01  | 45      | 45      |
| 11      | NC_028356.1 | 23,575,348         | A                      | C          | T          | Upstream gene  | AAGTCTACG GCCCTCA TACAT  | TCAGAC AAGGC GGCGACTCT AAAGCT TGGTTG | 211          | 55.04  | 55.08  | 45      | 45      |
| 12      | NC_028356.1 | 29,215,660         | T                      | C          | G          | Intergenic     | AAATCTGGCC TTGAAG TCATC  | TCATCGAC TTGGT ACAGC    | 193          | 55.05  | 55.19  | 40      | 45      |
| 13      | NC_028357.1 | 9,767,300          | A                      | C          | T          | Upstream gene  | CTGTTGCTTT TTTGTT TCAGA  | TCATCAAGC TTGTTC CTGAA   | 187          | 55.14  | 54.88  | 45      | 40      |
Table 3 (continued)

| SNP ID | Chromosome | Reference position | Reference V. sublobata | MCV-1 | SML-668 | Variant Type | Forward (F) primer (5'-3') | Reverse (R) primer (5'-3') | Product size T_m (F) | T_m (R) | GC% (F) | GC% (R) |
|--------|------------|-------------------|---------------------|-------|---------|--------------|--------------------------|--------------------------|----------------------|------|--------|--------|
| 14     | NC_028357.1| 19,212,708        | T                   | C     | G       | C            | Intergenic               | TGCCTGATTT GAAGAT GTAGA GCAACAAA AGTCAC AAACAA | 194     | 54.52  | 54.91  | 45     | 40     |
| 15     | NC_028358.1| 19,533,745        | A                   | C     | G       | A/T          | Upstream gene            | GCAACAAA AGTCAC AAACAA TGTTCGTGT GTCAT | 178     | 55.51  | 54.96  | 40     | 40     |
| 16     | NC_028358.1| 29,706,229        | G                   | A     | T       | A            | Synonymous               | GGCACCACG TATTAT AGTGCTA CAGACCAAT AAGGACAA ACCCA ACTTC | 206     | 54.97  | 54.99  | 45     | 45     |
| 17     | NC_028359.1| 16,146,976        | A                   | G     | T       | G            | Intergenic               | CCAATCAGT GAGTC AAGGC CAGGATGAA AGGACAA ACCCA CAGACCAAT AAGGACAA ACCCA ACTTC | 179     | 54.81  | 54.07  | 40     | 40     |
| 18     | NC_028359.1| 17,596,728        | C                   | A     | G       | G            | Intergenic               | GGCATATGAGT CTTCAT GGTGTC CAGGAATGAA AGGACAA ACCCA CAGGATGAA AGGACAA ACCCA ACTTC | 207     | 55.13  | 55.01  | 45     | 45     |
| 19     | NC_028360.1| 12,538,607        | A                   | G     | T       | -            | Upstream gene            | GGGATCTCCC TTAGGAG TGATGTC AAGGACAA AGGACAA ACCCA CAGGATGAA AGGACAA ACCCA ACTTC | 190     | 54.97  | 54.91  | 50     | 45     |
| 20     | NC_028360.1| 2,851,951         | A                   | -     | T       | T            | Intergenic               | GGGCTAATA AGCTAC CTGTG TTTCTGCA CCTAGA ACTCC | 167     | 55.17  | 54.99  | 45     | 45     |
| 21     | NC_028361.1| 12,573,184        | T                   | C     | A       | -            | Intergenic               | ATTTGCGGAT CTCAAC GAGTGC AAGGACGTT GTCAT ATGGT | 188     | 55.88  | 55.06  | 40     | 40     |
| 22     | NC_028361.1| 13,099,471        | T                   | G     | G       | A            | Intergenic               | ATGTAGGAG GAGAGA GTGCC CAGGATGAA AGGACAA ACCCA CAGGATGAA AGGACAA ACCCA ACTTC | 201     | 55.12  | 55.04  | 50     | 45     |
Structural and functional annotations of genomic variants

*Vigna radiata* annotated reference genome was used to evaluate the impact and possible effects of the SNPs and InDels identified in all accessions using the SnpEff software. SNPs and InDels were classified on the basis of their impacts on gene expression into four classes, i.e. high effect (modify splice sites, frame-shift, start lost and stop gained/lost), moderate effect (causes non-synonymous substitutions), modifier (intergenic, intragenic, intron, upstream, downstream and UTR variants) and low effect (causes synonymous substitutions and start gained). In all accessions, a major proportion of SNPs were modifiers (95–96%), followed by moderate (2.2–2.6%), low impact (1.5–1.9%) and smallest subset of high-effect SNPs (0.17–0.23%). Similarly, InDels showed around 97% modifier genomic variants in the three accessions. InDels belonging to high-impact variants ranged from 1.9 to 2.3% according to genotypes. Approximately 0.15–0.26% and 0.30–0.67% genomic variants are of low and moderate type effects, respectively. Impacts of genomic variants are shown in Supplementary Table S2.

Structural annotation showed that SNPs were abundant in the intergenic region (38.36–41.78%) followed by downstream region (19.93–20.95%) and upstream region (17.18–19.68%) in all the accessions. In the structural component of the gene, a significant number of SNPs were found in intron region (5.25–5.93%) followed by exon region (4.84–5.06%) and minimum SNPs were observed in 3’UTR (0.13–0.26%) and 5’UTR (0.06–0.12%) in all the accessions (Fig. 3). Likewise, maximum numbers of InDels were found in the intergenic region (34.22–38.89%) followed by downstream (20.45–22.37%) and upstream (19.02–21.09%) regions. Genic regions contained a significant number of InDels in introns (6.48–8.47%) and exons (2.87–3.57%) and a minimum number in 3’UTRs (0.13–0.36%) and 5’UTRs (0.07–0.17%) in the accessions (Fig. 4). It was observed that the majority of genomic variations were located in...
non-coding regions and low diversity was present in coding regions. Further, SNP variants were functionally evaluated in terms of mis-sense, nonsense and silent SNPs. Among the protein-coding regions, wild accession showed 3191 synonymous SNPs in 570 genes and 5339 non-synonymous SNPs in 815 genes. The non-synonymous SNPs included mis-sense substitution in 4991 (77.44%) SNPs in 620 genes and 348 (5.4%) nonsense SNPs showed substitution by premature stop codon introduction in 195 genes. The released variety showed 2,882 synonymous SNPs in 528 genes and 4739 non-synonymous SNPs in 790 genes. The non-synonymous SNPs included 4442 (77.31%) mis-sense SNPs in 610 genes and 297 (5.17%) nonsense SNPs in 180 genes. Similarly, the landrace showed 1105 synonymous SNPs in 228 genes and 1795 non-synonymous SNPs in 317 genes. The non-synonymous SNPs in the landrace included 1701 (78.49%) mis-sense SNPs in 256 genes and 94 (4.3%) nonsense SNPs.

We further analysed variants which cause a change in the amino acid in the coding part of gene, thereby play an important role in the functional variations. For this, we mapped the high- and moderate-effect genomic variants to KOBAS database and all variants were mapped to 178 unique KEGG pathways. Functional annotation of high- and moderate-effect SNPs gene revealed their maximum correspondence to signal transduction (average of 15.58%) followed by carbohydrate metabolism (an average of 9.74%), global and overview maps (average of 8.08%) with minimal correspondence to membrane transport pathways (an average of 0.5%) (Supplementary Fig. S3). Further, GO analysis depicted a significant representation of GO terms in the genes associated with biological processes (cellular process, 54.05%), followed by cellular components (cell, 27.02%), and molecular functions (catalytic activity, 18.91%) (Fig. 5).

**Comparative genomic variations in the wild, released and landrace accessions**

We observed that wild-specific SNP alleles (39.15%) were more abundant than the released variety of specific alleles (33.11%) and landrace specific alleles (8.89%). On comparison of the wild accession with landrace, we found 6072 unique SNPs in 340 loci in wild accession and 749 unique SNPs in 74 loci in landrace in addition to 134 common loci in the protein-coding regions of the wild accession and the landrace (Supplementary Table S3). Likewise,
by comparison of wild accession with released variety, we found 4007 unique SNPs in 245 loci in wild accession and 3635 unique SNPs in 237 loci in released variety in addition to 229 common loci in protein-coding regions of wild accession and released variety (Supplementary Table S4). Further, between released and landrace varieties, we found 4696 unique SNPs in 317 loci in released variety and 534 unique SNPs in 59 loci in landrace in addition to 149 common loci in both released and landrace varieties (Supplementary Table S5).

We observed multiple variations in the high- and moderate-effect SNPs associated with the disease-resistance pathways, receptor-like protein kinases and agronomic traits in all the accessions. Further, we found 568 SNPs in 46 protein kinase loci in wild accession, 161 SNPs in 14 loci in landrace and 575 SNPs in 32 protein kinase loci in released variety. Similarly, 764 SNPs were found in 19 disease-resistance loci in wild accession, 440 SNPs in 12 loci in landrace and 931 SNPs in 28 loci in released variety. SNPs associated with agronomic characteristics, such as seed size, seed color, seed dormancy and flowering timing among the wild, landrace and released accessions, were compared (Supplementary Table S6). A total of 169 SNPs in wild accession and 63 SNPs in released variety were in the callose synthase gene, 29 SNPs were detected in the pentatricopeptide repeat. In addition, genes with an important role in regulating plant growth and flower development, such as FAR1-related 5-like, F-box, and agamous-like MADS-box protein, were observed to have SNPs in all accessions analysed. In this study, we found unique loci for agamous-like MADS-box protein (NC_028358.1, LOC106760530) in the wild accession only. However, high-impact SNP was found in the lowest number, but they had a direct effect on the functionality of important genes related to the signaling pathways, growth and developmental traits of plants either by premature stop/start codon and or by changing gene splicing sites. We found a total of 347 SNPs in 203 loci in wild accession, 290 SNPs in 179 loci in released variety and 101 SNPs in 64 loci in landrace. By Pfam domain analysis of non-synonymous SNPs, 68.92%, 70.39%, and 79.76% transposon domains were found in released variety, wild accession, and landrace, respectively (Supplementary Table S7). The proportion of long-terminal repeats (LTR) (class I transposon) was the highest (~20%), while MuDr family transposase (class II transposon) proportion was the lowest (~0.5%) in released variety and landrace accessions.

Discussion

Whole-genome re-sequencing is currently the most convenient approach for the genome-wide SNP identification and trait discovery in various crop plants including mungbean (Lam et al. 2010; Jain et al. 2014; Jiao et al. 2016). Genomic variations, such as SNPs, are linked to the majority of important traits used for numerous applications, such as in genetic mapping, quantitative trait locus (QTL) analysis and genotypic diversification and evolutionary studies of plant species (Jain et al. 2014; Cleverenger et al. 2015). In the present study, genomic variants were identified and compared between wild accession (Vigna radiata var. sublobata), and released variety (SML-668) as well as landrace (MCV-1) of mungbean using NGS platform. In our data, 92% of the reads were mapped onto the reference genome indicating good quality of the data generated in this experiment. We observed high genetic variation in wild accession among all the accessions analysed indicating that these could be used as a source of new allelic variations in the mungbean germplasm improvement programs. Among the observed SNPs, a large proportion of the SNPs were detected in both wild and released mungbean variety, suggesting that most of the polymorphism observed in the modern released variety of mungbean came from their wild progenitors (Fig. 6). Also, the non-synonymous-to-synonymous ratio indicated a high substitution rates in mungbean wild, released and landrace accessions compared to the reports in other plants, such as rice (1.2) and soybean (1.47) (Kenneth et al. 2009; Valliyodan et al. 2016). The results of the present study indicate the need to further analyze the diversity in mungbean genome so that more useful genomic variations can be discovered for use in marker-assisted breeding in this important pulse crop.

Analysis for chromosomal-level distribution of the variants revealed that SNP variations were dispersed unevenly...
across the chromosomes but positively correlate with the length of chromosomes. It was observed that high- and low-density regions of genomic variants are present in the chromosomes. Such uneven dispersal of variants has also been identified in other plant species, such as in tomato, rice, mei, and chickpea (Corrado et al. 2013; Jain et al. 2014; Zhang et al. 2017a, b; Kujur et al. 2015). The high-density or hotspot regions were proposed to be the result of high rate of mutation in the genes residing in those regions (Lercher and Hurst 2002). In this study, high-density regions of SNPs were found on chromosomes 1, 2, 9 and 10 in wild accession. These chromosomal regions have been reported to harbor QTLs related to important agronomic traits, such as seed dormancy, seed size, pod dehiscence, growth habit and flowering time, that might have been used in the domestication of modern mungbean cultivars (Isemura et al. 2012; Li et al. 2014). On the other hand, the low-density regions represent a reduction in the genetic diversity. Therefore, knowledge of genomic variants residing in the high- and low-variant density regions provides a better breeding approach for crop improvement (Jain et al. 2014). We observed that SNPs identified exhibited transition bias as the numbers of transition were higher than transversion, which is comparable with the previous studies (Jain et al. 2014; Zhang et al. 2014). This might be related to the high methylation frequency of C to T changes. Structural annotation of all SNPs and InDels indicated that these were present in only 5% of the coding regions and the rest were located in intergenic and non-coding regions of mungbean genome. Similar results have also been reported in other species, such as in melon, strawberry, soybean, and tomato (Natarajan et al. 2016; Corrado et al. 2013; Hawkins et al. 2016). One possible explanation for the low occurrence of such variants could be that the non-synonymous SNPs in coding regions can lead to a functional change of gene expression and might affect plant growth and development, thus resulting in positive or negative effects of selection process during the course of evolution. It had reported that InDels of shorter length cause a more deleterious effect on gene expression than longer InDels in all crops (Choi et al. 2012). Further, transposon elements (TEs) have played a key role in the course of plant evolution, genome expansion, and gene regulation by causing various kinds of genetic variations (Lisch 2013). We observed more transposon elements in wild relatives than released variety. Similar findings of the transposon elements have been reported in other studies (Wang et al. 2013; Kang et al. 2014) which also substantiate the present conclusions.

A significant number of high- and moderate-effect SNPs were identified in the genes related to plant development, signal transduction, flowering, and carbohydrate metabolism. The SNPs in the gene loci, protein kinases loci, leucine-rich receptors, serine/threonine protein kinases and cysteine-rich protein kinases were present in higher frequencies in wild accession than the released variety which is comparable to the earlier studies (Li et al. 2014; Zhang et al. 2017a, b). KEGG pathway analysis of the genomic variants showed enrichment of variants in genes required for carbohydrate metabolism and signal transduction pathways in mungbean (Hudson et al. 1999; de Folter et al. 2005; Jung et al. 2012). Since both these pathways are known for their role in plant abiotic stress tolerance, understanding implications of these genomic variations are likely to provide in-depth information about the positive selection for abiotic stress tolerance during the domestication and selection of mungbean (Brozynska et al. 2016; Pandey et al. 2016).

Conclusion

Overall, in this study, a substantial number of genomic variations, i.e. SNPs and InDels, were detected using Ion torrent PGM™ platform. Most of the modern mungbean released varieties are suffering from the high degree of homogeneity (Lakhanpaul et al. 2000). These identified genomic variations provided significant information on functional and structural diversity among wild, released and landrace accessions of mungbean. Further, this study brings into focus the significance of wild species diversity in mungbean breeding programs. The integrated analysis of high-effect and non-synonymous genomic variations with QTLs could be used to discover and clone genes related to the important agronomic traits. In addition, medium- or low-density SNP arrays could be constructed with a set of SNPs having high agronomic values for specific germplasm screening and molecular breeding programs. Thus, the genomic variations (SNPs and InDels) generated in this study will add beneficial genomic resources that can be efficiently exploited for the future mungbean improvement programs.

Acknowledgements The authors thank the Indian Council of Agricultural Research, New Delhi and, Director, National Bureau of Plant Genetic Resources for the facilities provided for the study. The first author, PB acknowledges the DBT Fellowship from The Department of Biotechnology, Ministry of Science & Technology, Government of India, New Delhi for the Ph. D program. The funding for the wet laboratory analysis from CRP on Genomics project of the Indian Council of Agricultural Research, New Delhi is gratefully acknowledged.

Compliance with ethical standards

Conflict of interests Authors declare no conflict of interest.

References

Allito BB, Nana EM, Alemneh AA (2015) Rhizobia strain and legume genome interaction effects on nitrogen fixation and yield of grain legume: a review. Mol Soil Biol 20:6
Bangar P, Chaudhury A, Umdale S, Kumari R, Tiwari B, Kumar S, Gaikwad AB, Bhat KV (2018) Detection and characterization of polymorphic simple sequence repeats markers for the analysis of genetic diversity in Indian mungbean \textit{(Vigna radiata)} (L.) Wilczek. Indian J Genet Plant Breed 78:111–117

Bangar P, Chaudhury A, Tiwari B, Kumar S, Kumari R, Bhat KV (2019) Morphophysiological and biochemical response of mungbean \textit{(Vigna radiata)} (L.) Wilczek\textregistered{} varieties at different developmental stages under drought stress. Turk J Biol 43:58–69

Brozynska M, Furtado A, Henry RJ (2016) Genomics of crop wild relatives: expanding the gene pool for crop improvement. Plant Biotechnol J 14:970–1085

Chankaew S, Somta P, Sorajjapinun W, Srinives P (2011) Quantitative trait loci mapping of Cercospora leaf spot resistance in mungbean, \textit{Vigna radiata} (L.) Wilczek. Mol Breed 28:255–264

Chen H, Qiao L, Wang L, Wang S, Blair MW, Cheng X (2015) Assessment of genetic diversity and population structure of mung bean (\textit{Vigna radiata}) germplasm using EST-based and genomic SSR markers. Gene 566:175–183

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and Indels. PLoS ONE 7:e46688

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnPEFF: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6:80–92

Clevenberk J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA (2015) Single nucleotide polymorphism identification in polyploids: A Review, Example, and Recommendations. Mol Plant 8:831–846

Corrado G, Pifanelli P, Caramante M, Coppola M, Rosa Rao R (2013) SNP genotyping reveals genetic diversity between cultivated landraces and contemporary varieties of tomato. BMC Genomics 14:835

de Foter S et al (2005) Comprehensive interaction map of the Arabidopsis MADs-box transcription factors. Plant Cell 17:1424–1433

Fuller DQ, Harvey EL (2006) The archaeobotany of Indian pulses: identification, processing and evidence for cultivation. Environm Archaeol 11:219–246

Ganal MW, Allmann T, Roder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Gowda CL, Laxmipathi SK, Chaturvedi PM, Gaur CV, Kumar S, Jukanti AK (2015) Pulses research and development strategies for India. 17–33

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. Plant Physiol 131:872–877

Guo X, Li T, Tang K, Liu RH (2012) Effect of germination on phytochemical profiles and antioxidant activity of mungbean sprouts (\textit{Vigna radiata}). J Agric Food Chem 60:11050–11055

Gwag JG, Dixit A, Park YJ, Ma KH, Kwon SJ, Cho GT, Lee GA, Lee SY, Kang HK, Lee SH (2010) Assessment of genetic diversity and population structure in mungbean. Genes Genom 32:299–308

Hawkins C, Caruana J, Schiksnis E, Liu Z (2016) Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. Sci Rep 6:29017

Hudson M, Ringli C, Boylan MT, Quail PH (1999) The FAR1 locus encodes a novel nuclear protein specific to phytochrome A signaling. Genes Dev 13:2017–2027

Huq MA, Akter S, Nou S, Kim HT, Jung YJ, Kang KK (2016) Identification of functional SNPs in genes and their effects on plant phenotypes. J Plant Biotechnol 43:1–11

Isemura T, Kaga A, Tabata S, Somta P, Srinives P, Shimizu T, Jo U, Vaughan DA, Tomooka N (2012) Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean \textit{(Vigna radiata)}. PLoS ONE 7:e41304

Jain M, Moharanah KC, Shankar R, Kumari R, Garg R (2014) Genome wide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. Plant Biotech J 12:253–264

Jiao K, Li X, Guo W, Yuan X, Cui X, Chen X (2016) Genome resequencing of two accessions and fine mapping the locus of lobed leaflet margins in mungbean. Mol Breed 36:128

Jingade P, Bhosale LV, Sanjayaraj JA, Rajanna R, Jain M, Ravikumar RL (2014) Characterization of microsatellite markers, their transferability to orphan legumes and use in determination of genetic diversity among chickpea \textit{(Cicer arietinum)} \textit{L.} cultivars. J Crop Sci Biotechnol 17:191–199

Jung CH, Wong CE, Singh MB, Bhalla PL (2012) Comparative genomic analysis of soybean flowering genes. PLoS ONE 7:e38250

Kang YJ et al (2014) Genome sequence of mungbean and insights into evolution within \textit{Vigna species}. Nat Commun 5:5443

Kaur G, Joshi A, Jain D, Choudhary R, Vyas D (2016) Diversity analysis of green gram \textit{(Vigna radiata)} (L.) Wilczek\textregistered{} through morphological and molecular markers. Turk J Agric For 40:229–240

Kenneth L et al (2009) Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106(30):12273–12278

Krzywinski M et al (2009) Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645

Kujur A et al (2015) Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. Front Plant Sci 6:162

Lakhanpal S, Bhat KV, Chadha S (2000) Random amplified polymorphic DNA analysis in Indian mungbean \textit{(Vigna radiata)} (L.) Wilczek\textregistered{} cultivars. Genetica 109:227–234

Lam HM et al (2010) Re-sequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42:1053–1059

Lambriides CJ, Godwin I (2007) Mungbean. In: Kole C (ed) Genome Mapping and Molecular Breeding in Plants Pulses, Sugar and Tuber Crops. Springer, Berlin, pp 69–903

Lercher MJ, Hurst LD (2002) Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet 18:337–340

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) 1000 genome project data processing subgroup The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li YH, Reif JC, Jackson SA, Ma YS, Chang RZ, Qiu LJ (2014) Detecting SNPs underlying domestication-related traits in soybean. BMC Plant Biol 14:251

Lisch D (2013) How important are transposons for plant evolution? Nat Rev Genet 14:49–61

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. Int J Plant Genom 2012:728398. \url{https://doi.org/10.1155/2012/728398}

Muñoz N, Liu A, Kan L, Li MW, Lam HM (2017) Potential uses of wild germplasms of grain legumes for crop improvement. Int J Mol Sci 18:328

Natarajan S, Kim HT, Thamilarasan SK, Veerappan K, Park JI, Nou IS (2016) Whole genome re-sequencing and characterization of powdery mildew disease-associated allelic variation in melon. PLoS ONE 11:e0157524

Pandey MK et al (2016) Emerging genomic tools for legume breeding: current status and future prospects front. Plant Sci 7:455

Sato S et al (2008) Genome structure of the legume \textit{Lotus japonicus}. Plant Cell 17:191–199

Schaeferlein R et al (2016) BMC Plant Biol 16:159
Schmutz J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183
Shivakumar MS, Ramesh S, Rao AM, Udaykumar HR, Keerthi CM (2017) Cross legume species/genera transferability of SSR markers and their Utility in assessing polymorphism among advanced breeding lines in Dolichos Bean (Lablab purpureus L.). Int J Curr Microbiol App Sci 6:656–668
Shrivastava D, Verma P, Bhatia S (2014) Expanding the repertoire of microsatellite markers for polymorphism studies in Indian accessions of mung bean (Vigna radiata L. Wilczek). Mol Biol Rep 41:5669–5680
Singh AK, Singh SS, Prakash V, Kumar SA, Dwivedi SK (2015) Pulses production in India: present status, bottleneck and way forward. J AgriSearch 2:75–83
Smartt J (1990) Grain legumes: evolution and genetic resources. Cambridge University Press, UK
Tomooka N, Vaughan DA, Moss H, Maxted N (2002) The Asian Vigna: genus Vigna subgenus Ceratotropis genetic resources. Kluwer Academic Publishers, Dordrecht
Valliyodan B et al (2016) Landscape of genomic diversity and trait discovery in soybean. Sci Rep 6:23598
Van K, Kang YJ, Han KS, Lee YH, Gwag JG, Moon JK, Lee SH (2013) Genome-wide SNP discovery in mungbean by Illumina HiSeq. Theor Appl Genet 126:2017–2027
Varshney RK et al (2012) Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89
Varshney RK et al (2013) Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. Nat Biotechnol 31:240–246
Vishnu-Mittre B (1974) Palaeobotanical evidence in India. In: Hutchinson J (ed) Evolutionary Studies in World Crops: Diversity and Change in the Indian Sub-continent. Cambridge University Press, Cambridge, pp 3–30
Wang X, Weigel D, Smith LM (2013) Transposon variants and their effects on gene expression in Arabidopsis. PLoS Genet 9:e1003255
Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191
Wei X, Zhu X, Yu J, Wang L, Zhang Y, Li D, Zhou R, Zhang X (2016) Identification of sesame genomic variations from genome comparison of landrace and variety. Front Plant Sci 7:1169
Wu J, Mao X, Cai T, Luo J, Wei L (2006a) KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Res 34(2):W720–W724
Wu J, Mao X, Cai T, Luo J, Wei L (2006b) KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Res 34(2):W720–W724
Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34(2):293–297
Zhang H, Mittal N, Larry J, Barazani LO, Song BH (2017a) Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. Evol Appl 10:5–24
Zhang J, Zhao K, Hou D, Cai J, Zhang Q, Cheng T, Pan H, Yang W (2017b) Genome-wide discovery of DNA polymorphisms in Mei (Prunus mume Sieb. et Zucc.), an ornamental woody plant, with contrasting tree architecture and their functional relevance for weeping trait. Plant Mol Biol Rep 35:37–46
Zhang S, Chen W, Xin L, Gao Z, Hou Y, Yu X, Zhang Z, Qu S (2014) Genomic variants of genes associated with three horticultural traits in apple revealed by genome re-sequencing. Hort Res 1:14045