Indian J. Genet., 80(2) 179-185 (2020)
DOI: 10.31742/IJGPB.80.2.8

Computational analysis of SNPs and INDELs in cluster bean cultivars involved in multiple trait expression

Sarika Sahu[#], Tanmaya Kumar Sahu[1], Sabri Ghosal[#], Kishor Gaikwad[2] and A. R. Rao[3,*]

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012; [1]ICAR-Indian Agricultural Research Institute, New Delhi 110 012; [#]Amity University, Noida, Uttar Pradesh; [2]ICAR-National Institute of Plant Biotechnology, New Delhi 110 012

(Received: January 2020; Revised: April 2020; Accepted: May 2020)

Abstract

SNPs (Single Nucleotide Polymorphisms) are extensively used in plant breeding programs because of their automation and high precision in allele calling. In the present study, transcriptomes of three cultivars of cluster bean (*Cyamopsis tetragonoloba*), namely, RGC-936, RGC-1066 and M-18 were analysed for the identification of SNPs and Indels using the recently assembled draft genome made available by National Institute of Plant Biotechnology, New Delhi. Besides, a comparison among the identified SNPs and indels of three cultivars was made to mine out the cultivar specific SNPs and indels as well as common markers among the cultivars. In addition, an online database, cbSIR, was developed based on the markers populated from the said cultivars of cluster bean. The results reveal that highest number of SNPs (10279) were present in cultivar RGC-1066 followed by RGC-1066 (9714) and M-18 (7933). The detected SNPs were subjected to functional annotation. In a similar way, Indels were also identified and functionally annotated. Predictions were made based on the involvement of SNP/ Indel possessing genes in the expression of multiple traits such as gum production, auxin transport, disease resistance in the three cultivars of cluster bean.

Keywords: Single nucleotide polymorphism, indels (Insertions/Deletions), next generation sequencing, *in silico* analysis, cluster bean.

Introduction

Acceleration in genetic gain can be achieved through marker-assisted selection (MAS) on the basis of individual genes or through the selection of chromosomal segments at genomic level (Collard et al. 2007). With molecular markers, important genes can be isolated on the basis of their position on the genetic map (Tanksley et al. 1995). Besides, markers can help dissect traits that are controlled by many different factors into their individual components called quantitative trait loci (QTLs), which can be molecularly identified in a subsequent manner (Tanksley et al. 2007). Despite genomic markers like chloroplast marker also play very important role in the identification of seed mediated gene flow and evaluation of risk assessment while transforming transgenic into wild variety of species (Tripathy et al. 2019). Among many types of molecular markers that have been developed in the past, Single Nucleotide Polymorphisms (SNPs) are the most effective markers in plant breeding. SNPs gained much interest in the breeding community (Rafalski et al. 2002) as the variations are caused by transitions (C/T or G/A) or transversions (C/G, C/A, or T/A, T/G), at the same position between individual genomic DNA sequences (Brookes 1999; Trick 2009). Further, SNPs are distributed on intergenic regions, coding regions (exons) and non-coding (introns, 5'UTR, 3'UTR, or exon-intron splice sites) regions of genes (Jehan et al. 2006; Hiremath et al. 2012; Garrido-Cardenas et al. 2018). Majority of the SNPs are biallelic and are tightly linked to or are the actual cause of allelic (phenotypic) variations in traits. SNPs in the coding regions can be categorised into synonymous and nonsynonymous SNPs, with protein sequence being affected by the latter category (Zhao et al. 2019). Besides SNPs, Indels are another type of markers that lay a very important role in exhibiting structural variations on the genome and are being widely distributed throughout the genome of crops like, Arabidopsis, rice, tomato, chickpea (Yang et al. 2014; Lu et al. 2015; Das et al. 2015). The biogenesis of

*Corresponding author's e-mail: ar.rao@icar.gov.in
[3]Present address: Indian Council of Agricultural Research, New Delhi 110 001
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

Indels involves different cellular mechanisms such as movement of transposable elements, replication slippage, unequal crossing-over within the genome (Moghaddam et al. 2014). Further, the *in silico* identification of Indels can easily be detected at the genomic level via bioinformatics approaches at a lower cost and less time. In the recent past, Indels also became ideal markers for breeders in marker assisted crop improvement programs (Jain et al. 2014; Shen et al. 2015).

Cluster bean (*Cyamopsis tetragonoloba* L. Taub.), also known as *guar*, is an important leguminous herb, highly adapted to arid and semi-arid parts of the world requires low inputs and care (Mudgil et al. 2014). It is self pollinated and diploid plant (2n=17) belonging to the tribe Galagae of the Leguminosae family (Ajit et al. 2013). Cluster bean is widely grown in Gujarat, Haryana, Punjab, Rajasthan and Uttar Pradesh (Kumar et al. 2014) and has drawn attention of many researchers because of its gum quality. The guar gum contains ~ 90% galactomannan and has 5-8 times thickening power than that of starch (Kumar et al. 2014). Galactomannan is composed of a β-(1$\rightarrow$4)-linked d-mannose backbone to which single unit α-d-galactosyl residues are attached at O-6 (Daas et al. 2000). The demand of guar gum is significantly increasing in many industries like textile, cosmetics, pharmaceuticals (Mudgil et al. 2014). In conventional guar breeding, usually phenotypic traits were considered as selection criteria for high yield and quality parameters. However, molecular markers offer several advantages over the conventional breeding tools for selection of diverse parents. Pathak et al. (2010) used RAPD (Random amplified polymorphic DNA) marker for the identification of different cultivars of cluster bean. Kuravadi et al. (2014) and Kumar et al (2016) identified and characterized EST-SSR markers in cluster bean. The M-83 cultivar of cluster bean is an improved vegetable variety having glabrous leaves and unbranched stem while RGC-1066 is an improved cultivar that produces more gum (Tanwar et al. 2017). Another cultivar RGC-936 has high seed yield and harvest index (Kumar et al. 2014). However, not many attempts have been made to mine and functionally annotate SNPs/Indels from the transcriptome or whole genome of cluster bean cultivar as well as to study their involvement in multiple trait expression. Moreover, no database or repository on SNPs/Indels of cluster bean is available in public domain.

Therefore, the present study was conducted to (i) mine SNPs and INDELs from the RNAseq data of three cultivars RGC-1066, RGC-936, M-83 (ii) provide functional annotation to the identified SNPs and indels and identify cultivar specific SNPs with multiple trait expression (iii) develop an online repository: cbSIR (http://webapp.cabgrid.res.in/clb_ce/index.php), based on annotated SNPs and indels, for cluster bean breeders and biotechnologists.

Materials and methods

Collection of data and pre-processing

The RNA-seq data of three cultivars of cluster bean RGC-1066, RGC-936, M-83 were downloaded from Sequence Read Archive (SRA) of NCBI, which is a repository of raw sequence reads collected from next generation sequencing technologies like Illumina, 454. The SRA Ids of cultivars RGC-1066, RGC-936 and M-83 are SRR8082057, SRR5428804 and SRR3218523, respectively. Fastq-dump module of SRAtoolkit version 2.10.0 (https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz) was used to convert the SRA files into fastq formatted files. The raw reads of each cultivar was used for the mining of the variants. Unpublished draft genome of cluster bean cultivar RGC-936 from National Institute of Plant Biotechnology (NIPB) New Delhi, India was considered as a reference genome for variant analysis. The quality control and trimming of raw reads were done by using FastQC tool (http://www.bioinformatics.babraham.ac.uk/project) and Trimmomatic tool (Bolger et al. 2014) respectively. The base quality of each read was filtered on the basis of phred score more than 33, opted through standard command given in Trimmomatic tool.

Alignment of reads

The reference genome was first indexed by using alinger tool: bowtie2. Further, the cleaned reads of three cultivars (RGC-1066, RGC-936, M-83) were aligned against the reference genome with a mapper tool: TopHat (Kim et al. 2013). TopHat is a tool for aligning short reads of RNA-seq data to the reference genome using read-mapping algorithm. The mapping percentages between the reads of each cultivar and reference genome were calculated.

Mining of SNPs and Indels

The significantly mapped binary alignment files of all the three cultivars were used for the identification of variants. The SNPs between the reference genome and mapped reads of each cultivar were identified by using the UnifiedGenotyper tool of the GATK (Genome

Analysis Toolkit) version 4.1.2.0. Subsequently, SNPs were filtered on the basis of cut-off quality score > 30 and depth > 10, while Indels were filtered for length greater than 10 base pairs. The substitution of nucleotides like transition and transversion were calculated for each cultivar.

Functional annotation and categorization of SNPs

The unpublished reference genome of cluster bean was searched against the viridiplantae database of UniProt for homology search. The filtered hits were submitted to UniProt (https://www.UniProt.org/uploadlists/) to retrieve the gene names and functions. The scaffolds having SNPs were annotated from each cultivar. Augustus tool (Hoff et al. 2013) was used for the gene prediction of reference genome. Further, presence of SNPs in exons, introns and CDS regions were extracted using in-house shell script from the annotated reference genome.

Development of the online repository

The information on the SNPs and INDELs identified from this study were organized in the form of an online repository named as "Cluster bean SNPs and INDELs Repository-cbSIR". The front-end of the repository was designed using HTML and PHP 5.4.7 to facilitate the retrieval of information. Whereas the back-end database was created in the interface of MySQL. The data model constitutes both the conceptual and logical data models. The database schema includes ten entities where the relations between the entities are suitably defined using the primary and foreign keys. The search options are also provided in the information system. Further, the search interface was made interactive and user friendly. Besides, JavaScript was used for client-side applications such as form-validation, redirection, and other customizations.

Results and discussion

Pre-processing of RNA-seq data of cluster bean cultivars

The RNA-seq data of cluster bean cultivars, namely, RGC-1066, RGC-936, M-83 were checked for quality. The high quality data were subjected to Trimmomatic for the removal of adapters. The cleaned reads of each cultivar were mapped against the reference genome with TopHat aligner tool and mapping percentages of each cultivar are found to be 69.3%, 60.03% and 64.4% in M-83, RGC-1066 and RGC-936 cultivars, respectively.

In silico identification and structural annotation of SNPs

The schematic pipeline used for identification of SNPs and Indels are shown in Fig. 1. The UnifiedGenotyper tool of GATK (https://gatk.broadinstitute.org/hc/en-us/sections/360007279572-4-1-2-0) was used for calling variants from each of the binary alignment mapped (bam) files of each cultivar. Total number of SNPs in M-83, RGC-1066 and RGC-936 are found to be 69906, 70886 and 46699 respectively. The ratio of transitions to transversions (Ts/Tv) in M-83, RGC-936 and RGC-1066 are observed as 0.96, 0.92 and 1.01, respectively. The parameters for identification of significant SNPs were set as quality score: 30 and depth: >10. The filtered SNPs in M-83, RGC-1066 and RGC-936 cultivars are now 7933, 9714 and 10279 respectively. The ratio of Ts/Tv is observed as 1.5 for M-83, 1.2 for

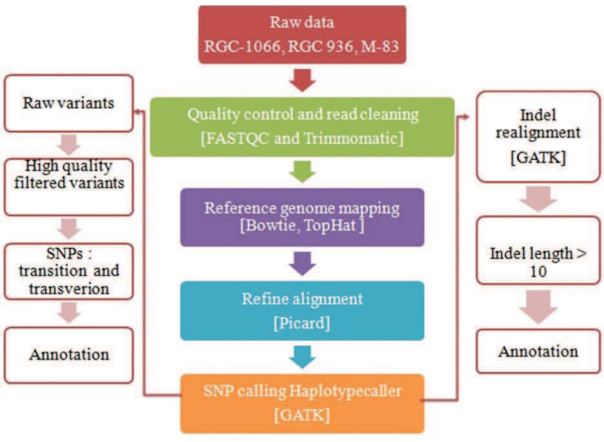

Fig. 1. Schematic pipeline for the identification and characterization of SNPs and indels

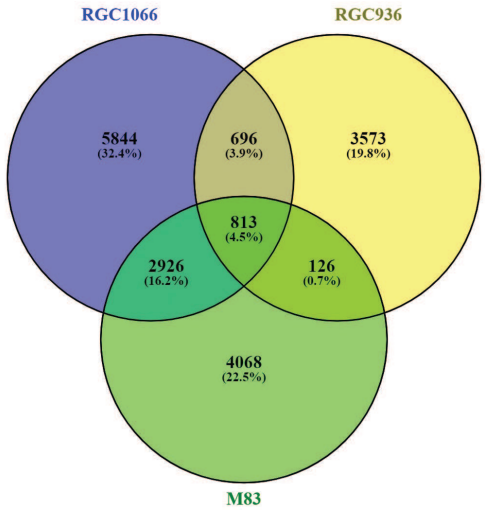

Fig. 2. Venn diagram showing the distribution of SNPs in the cultivars of cluster bean

RGC-936 and 1.4 for RGC-1066 after filtering the significant SNPs. However, the number of common SNPs among all cultivars is 813 (Fig. 2). The presence of SNPs on various genomic regions like introns, CDS (coding sequence) were given in Table 1.

Table 1. Cultivar-wise SNP distribution on CDS and intron as well as InDels distribution

Cultivar	Transitions			Transversions			Indels		
	CDS	Intron	Total	CDS	Intron	Total	Insertions	Deletions	Total
M-83	1762	503	2265	1048	363	1411	877	316	1193
RGC-936	1876	775	2651	1070	1243	2313	1295	645	4604
RGC-1066	1587	534	2121	1153	678	1831	3927	677	1940

Prediction and structural annotation of Indels

Indels are the common forms of polymorphism corresponding to the addition or removal of base pairs in the DNA sequence of an organism. They are the second most common type of genomic variants and are the most common type of structural variants on genome (Lin et al. 2017). The GATK tool is an efficient tool to find the indels of size less than 50 base pairs. The Indels of length > 10 are only reported here, as the Indel length less than 10 base pairs are commonly appeared to be an outcome of slipped strand replication (Moghaddam et al. 2014). A total of 1193, 1940 and 4604 Indels are found in M-83, RGC-1066 and RGC-936 cultivars respectively. Here, the number insertions are found to be greater than the number of deletions in each cultivar (Table 1). Also, the length distribution of insertion markers is given in Table 2. Besides, the number of insertion markers falling in the range of 10-20 base pairs are highly abundant in each cultivar. However, the length of insertion markers greater than 40 are almost found to be nil in the cultivar RGC-1066.

Table 2. Distribution of insertion markers on the basis of nucleotide length

Range	M-83	RGC-1066	RGC-936
10-20	617	1283	2702
20-30	181	6	819
30-40	60	5	306
>40	19	1	100

Functional annotation of SNPs

The transcripts having SNPs were annotated with blastx program against viridiplantae database (https://www.UniPrott.org/UniPrott/?query=reviewed:yes%20taxonomy:33090). Around 18% (1173) of SNPs present on the transcripts were annotated in M-83 cultivar which is an improved vegetable variety. Among the identified SNPs, 1173 SNPs are falling in the exonic region. It was found that three SNPs falling in the exonic regions are participating in disease resistant function while few SNPs are participating in auxin transport regulation. In case of RGC-936, a high yielding cultivar, total of 14% of SNPs were annotated. The annotated SNPs falling in exons are 1039, respectively. Whereas a total of 66% SNPs present on the transcripts of RGC-1066 were annotated and among them 57% fall in the exonic region. The GO terms were obtained from the Agrigo (http://bioinfo.cau.edu.cn/agriGO/index.php) and further classified by WEGO tool (Ye et al. 2018) that was used for the classification of GO terms into biological process, cellular component and molecular function (Fig. 3). The GO terms in biological processes reveal that the genes having SNPs are highly involved in abiotic stresses and regulation of cellular processes. In the classified group of cellular component, the major categories of transcripts having SNPs were found in cell part, cell organelles and intrinsic component of membrane. Under the group of Molecular function, genes bearing SNPs were participating in carbohydrates derivative binding, organic cyclic compound binding and heterocyclic compound binding. SNPs which are specific to RGC-1066 cultivars are mainly participating in the transcription factor, stress responsive protein, UDP-galactose, Beta-galactosidase and glucomannan-4-beta mannosyl transferase. The last three enzymes are mainly found to be participating in the synthesis of guar gum (Daas et al. 2000). Most of the SNPs of RGC-936 cultivar are involved in transcription factor and few among them are found to be participating in production of galactomannan. Further, the SNPs of improved vegetable cultivar: M-83 are participating in auxin transport protein, disease resistance and galactosyltransferase.

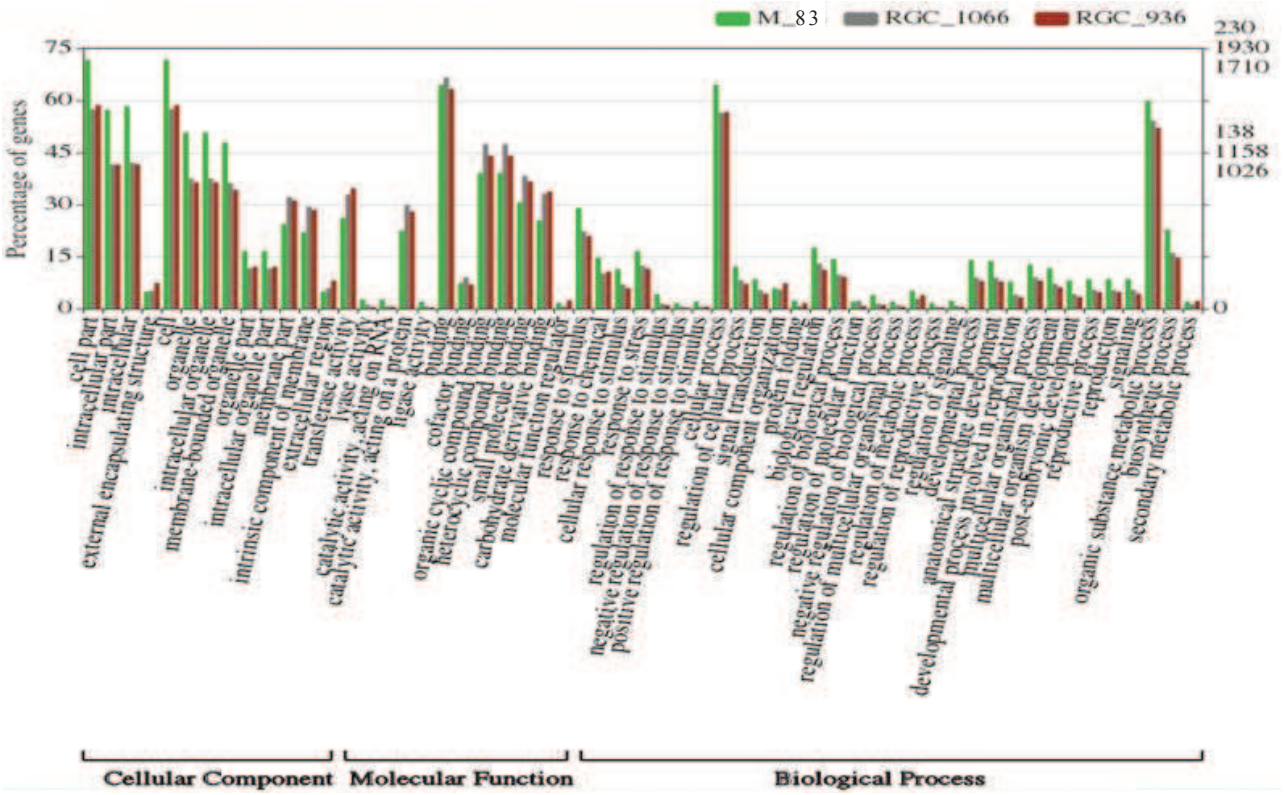

Fig. 3. Classification of GO terms of transcripts possessing SNPs from each cultivar: M-83, RGC-1066 and RGC-936 in green, grey and red color respectively

Fig. 4. Homepage of cbSIR database

Functional annotation of INDELs

InDels having a length greater than 10 base pairs were annotated for each cultivar. The annotated Indels in M-13, RGC-936 and RGC-1066 are 24, 24 and 27 respectively. Among annotated Indels, insertion markers were mostly participating in more activities as compared to deletions. However, merely 5, 4 and 2 deletion markers were annotated in RGC-1066, M-83 and RGC-936 respectively. These Indels markers are found to participate in different functions like transcription factor, chloroplast gene (ycf2), heat shock protein. Das et al. (2015) revealed that indel markers-containing candidate gene regulating flowering and maturity in chickpea.

Repository

As no online database on SNPs/Indels of cluster bean is available in public domain, the same was developed under the present study. The database cbSIR can

useful for further crop improvement program of cluster bean. The basic information on cluster bean and about the repository are provided in the website of cbSIR. The identified SNPs of M-83, RGC-1066, RGC-936 cultivars are given in cbSIR. The indels identified from the considered cultivars of cluster bean are also given in the cbSIR with the retrieval as well as search facilities. Moreover, the facilities like export of information to different formats (.pdf, .xls, .doc, .xml, .csv), quick search having position, reference allele, alternate allele, depth(DP), quality, genotype (heterozygous, homozygous) filtering criteria are given to the end-users for easy retrieval of information. In addition, facility for fetching information on cultivar specific SNPs as well as SNPs common in all cultivars are provided to the end-users.

The high-throughput sequencing technology has proved to be an effective approach for transcriptome profiling. The *in silico* identification of SNPs and indels markers from transcriptome data is cost-effective and less time consuming. The cultivar specific markers like SNPs and Indels can be used in future for crop improvement program of cluster bean. The SNPs present on gene regulating glucomannan synthesis in RGC-1066 cultivars can be used in future as a candidate gene, will help to identify candidate gene in genomic assisted breeding. The online database on SNPs and indels of cluster bean is given at http://webapp.cabgrid.res.in/clb_ce/index.php for end-users.

Authors' contribution

Conceptualization of research (SS, ARR); Designing of the experiments (SS, ARR); Contribution of experimental materials (Public domain); Execution of field/lab experiments and data collection (SS, KG); Analysis of data and interpretation (SS, ARR, TKS); Preparation of manuscript (SS, ARR, TKS, KG, SG).

Declaration

The authors declare no conflict of interest.

Acknowledgements

This study was supported by the grants ICAR-Consortia Research Platform on Genomics (CRP-Genomics/IX/2017) received from Indian Council of Agricultural Research (ICAR). Authors acknowledged the computational support received from ICAR-IASRI and SS acknowledged the support received from Amity University, Noida, Uttar Pradesh while pursuing her Ph.D. degree programme. This study was financially supported by the grant ICAR-Consortia Research Platform on Genomics (CRP-Genomics/1X/2017).

References

Ajit P., Yengkokpam P., More, V., Ghorpade B. and Swamy M. K. 2013. Molecular characterization of cluster bean (*Cyamopsis tetragonoloba*) cultivars using PCR-based molecular markers. Int. J. Adv. Biotechnol. Res., **4**: 158-166.

Bolger A M., M. Lohse and B. Usadel. 2014. Trimmomatic: A flexible read trimming tool for illumina NGS data. Bioinformatics, **30**(15): 2114-2120.

Brookes A. J. 1999. The essence of SNPs. Gene., **234**(2): 177-86.

Collard B. C. Y. and Mackill D. J. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philos T R Soc B., **363**: 557-572.

Daas P. J. H., Schols H. A. and De Jongh H. H. J. 2000. On the galactosyl distribution of commercial galactomannans. Carbohydr. Res., **329**: 609-619.

Das S, Upadhyaya H.D., Srivastava R., Bajaj D., Gowda C. L., Sharma S., Singh S., Tyagi A. K. and Parida S. K. 2015. Genome-wide insertion–deletion (InDel) marker discovery and genotyping for genomics-assisted breeding applications in chickpea. DNA Res., **22**(5): 377-86.

Garrido-Cardenas J. A., Mesa-Valle C. and Manzano-Agugliaro F. 2018. Trends in plant research using molecular markers. Planta, **247**(3): 543-57.

Hiremath P. J., Kumar A., Penmetsa R. V., Farmer A., Schlueter J. A., Chamarthi S. K., Whaley A. M. Carrasquilla-Garcia N., Gaur P. M. and Upadhyaya H. D. 2012. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol J., **10**(6): 716-32.

Hoff K. J. and Stanke M. 2013. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res., **41**(W1): W123-8.

Jain M., Moharana K. C., Shankar R., Kumari R. and Garg R. 2014. Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. Plant Biotechnol. J., **12**: 253-264.

Jehan T. and Lakhanpaul S. 2006. Single nucleotide polymorphism (SNP)–methods and applications in plant genetics: a review, Ind. J. Biotech., **5**(4): 435-459.

Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R. and Salzberg S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions

and gene fusions. Genome Biol., **14**(4): R36.

Kumar S., Parekh M. J., Patel C. B., Zala H. N., Sharma R., Kulkarni K. S., Fougat R. S., Bhatt R. K. and Sakure A. A. 2016. Development and validation of EST-derived SSR markers and diversity analysis in cluster bean (*Cyamopsis tetragonoloba*). J. Plant. Biochem. Biot., **25**(3): 263-9.

Kumar V. and Kaushik V. K. 2014. Growth substances response to cluster bean [*Cyamopsis tetragonoloba* (L.)] growth parameters. Agric. Sustain. Dev., **2**: 11-13.

Kuravadi N. A., Tiwari P. B., Tanwar U. K., Tripathi S. K., Dhugga K. S., Gill K. S. and Randhawa G. S. 2014. Identification and characterization of EST-SSR markers in cluster bean (*Cyamopsis* spp.). Crop Sci., **54**(3): 1097-1102.

Lin M., Whitmire S., Chen J., Farrel A., Shi X. and Guo J. T. 2017. Effects of short indels on protein structure and function in human genomes. Sci. Rep., **7**(1): 9313.

Lü Y., Cui X., Li R., Huang P., Zong J., Yao D., Li G., Zhang D. and Yuan Z. 2015. Development of genome wide insertion/deletion markers in rice based on graphic pipeline platform. J. Integr. Plant Biol., **57**(11): 980-91.

Moghaddam S. M., Song Q., Mamidi S., Schmutz J., Lee R., Cregan P., Osorno J. M. and McClean P. E. 2014. Developing market class specific InDel markers from next generation sequence data in *Phaseolus vulgaris* L. Front. Plant Sci., **5**: 185.

Mudgil D., Barak S. and Khatkar B. S. 2014. Guar gum: Processing, properties and food applications - A Review. J. Food Sci. Tech., **51**: 409-418.

Pathak R., Singh S. K., Singh M. and Henry A. 2010. Molecular assessment of genetic diversity in cluster bean (*Cyamopsis tetragonoloba*) genotypes. J. Genet., **89**: 243-6.

Rafalski A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant. Biol., **5**(2): 94-100.

Shen X., Liu Z. Q., Mocoeur A., Xia Y. and Jing H. C. 2015. PAV markers in Sorghum bicolour: genome pattern, affected genes and pathways, and genetic linkage map construction. Theor. Appl. Genet., **128**: 623-637.

Tanksley S. D. and Fulton T. M. 2007. Dissecting quantitative trait variation—examples from the tomato. Euphytica, **154**(3): 365-70.

Tanksley S. D., Ganal M. W. and Martin G. B. 1995. Chromosome landing: A paradigm for map-based gene cloning in plants with large genomes. Trends in Genet., **11**(2): 63-8.

Tanwar, U. K., Vikas P. and Gursharn S. R. 2017. RNA-Seq of Guar (*Cyamopsis tetragonoloba* L. Taub.) Leaves: De Novo Transcriptome Assembly, Functional Annotation and Development of Genomic Resources. Front. Plant Sci., **8**: 91.

Trick M., Long Y., Meng J. and Bancroft I. 2009. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotech. J., **7**: 334-346.

Tripathy Kabita, Singh B., Misra G. and Singh Nagendra K. 2019. Identification distribution and comparative analysis of microsatellites in the chloroplast genome of *Oriza* species. Indian J. Genet., **79**(3): 536-544. DOI: 10.31742/IJGPB.79.3.2.

Yang J., Wang Y., Shen H. and Yang W. 2014. In Silico identification and experimental validation of Insertion–Deletion polymorphisms in tomato genome. DNA Res., **21**(4): 429-38.

Ye J., Zhang Y., Cui H., Liu J., Wu, Y., Cheng Y., Xu H., Huang X., Li S., Zhou A. and Zhang X. 2018. WEGO 2.0: A web tool for analyzing and plotting GO annotations, update. Nucleic Acids Res., **46**(W1): W71-W75.

Zhao Y., Wang K., Wang W. L., Yin T. T., Dong W. Q. and Xu C. J. 2019. A high-throughput SNP discovery strategy for RNA-seq data. BMC Genomics, **20**(1): 160.