Int.J.Curr.Microbiol.App.Sci (2019) *8*(1): 2819-2829

International Journal of Current Microbiology and Applied Sciences
***ISSN: 2319-7706* Volume 8 Number 01 (2019)**
Journal homepage: http://www.ijcmas.com

Review Article https://doi.org/10.20546/ijcmas.2019.801.296

Tools and Resources for SNP Mining in Crop Plants

Saurabh Pandey[1]*, Sunidhi Mishra[2] and Kailash Chandra[3]

[1]National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi-110067, India
[2]Department of Vegetable Sciences, Indira Gandhi Agricultural University Krishak Nagar, Raipur, Chhattisgarh 492012, India
[3]College of Agriculture (S.K.N.A.U. Jobner) Fatehpur Shekhawati, Sikar-332301 (Rajasthan), India

**Corresponding author*

Keywords

Tools, Resources, SNP Mining, Crop Plants

Article Info

Accepted:
17 December 2018
Available Online:
10 January 2019

A B S T R A C T

Molecular genetic markers correspond to highly potent source for the study of plant genomes and the association of inherited phenotypic traits with underneath genetic variation. Single Nucleotide Polymorphism (SNPs) are most abundant form of molecular genetic marker which represents a single nucleotide difference between two individuals at a defined location. Compare to others SNPs are direct sequence variation which offers the precise nature of the allelic variants among different genotypes. Further, it signify recurrent type of genetic polymorphism with high density genome coverage. Advent of Next generation sequencing technology drives the exploration of sequence diversity for various crops. These studies revealed abundance of SNPs in plant systems, with the frequency of 100-300bp per SNP.SNP detection based on EST(expressed sequence tags) sequence data has been performed for crops like maize, barley, tomato and trees like pine and in *Arabidopsis* which is a model plant. Similarly SNP identification based on array analyses has been published for *Arabidopsis*, rice, barley and maize. Amplicon resequencing approach has been utilized for the identification of SNPs in maize, soybean, *Arabidopsis*, rice, tomato, sugarbeet, barley and spruce. There are two sets of data to perform SNP mining one is reference sequence data and other is *de novo* sequence data. This mining for various datasets mainly comprise of subsequent steps: in first step we have to group sequence reads on the basis of their sequence resemblance and confirm identity of reads whether they are covering similar part of genome or they have the same transcript origin. Further we have to align confirm reads and finally identify and categorize sequence variants as probable polymorphic loci/marker. Thus SNP mining can provide better understanding of crops at the gene level, for the detailed analysis of germplasm and eventually for the resourceful management of genetic diversity on a whole genome level inside plant breeding.

Introduction

Molecular markers are the armour of modern plant breeding and its application in crop improvement is now well recognized. Some of the tools that are employed nowadays along with conventional breeding practices are heavily reliant on molecular markers for quick and accurate examination of germplasm, mapping of specific traits and most importantly marker-assisted selection (MAS). These markers can be utilized in parental genotype selection in breeding programs, reduction of linkage drag during backcrossing and highly useful for the selection of phenotypically traits which are hard to establish.

Molecular markers are the integral part of genetics, and utilized for genetic disease associated allele detection, paternity evaluation, forensics and presumption of population history. Moreover, molecular markers are very useful means of genome mapping in every single system, contributing prospective for development of very high-density genetic maps to define associated haplotype blocks or genetic variation of interest. The bulk of variation at the nucleotide level is frequently not observable at the level of phenotype. This type of distinction can be utilized as molecular genetic marker system. There are two prominent candidates for this type of system, simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs).

Among them SNPs correspond to the majority of genetic polymorphism consequently permit the advancement of the premier density of molecular markers (Batley and Edwards, 2007). Prevailing next-generation sequencing (NGS) technologies make available the opportunity of large-scale SNP detection by evaluating whole-genome shotgun sequences of datasets from crop plants with high-quality reference genome sequences. Recently Lai *et al.*, (2015) identified around 4 million intervarietal SNPs in bread wheat. This study also provided insight into the molecular consequences of the evolution and selection that resulted in modern hexaploid wheat. Thus we can say that SNP are future markers, which are presented in the subsequent sections.

What are SNPs?

In the field of molecular genetics sequence variation at the DNA level are the fundamental needs. SNPs present the eventual form of molecular genetic marker, since basic unit of the inheritance is a nucleotide base, and an SNP characterize a single nucleotide difference among two individuals at a distinct defined location. Basically three forms of sequence variation observed for SNPs: a) transitions (G/A or C/T); b) transversions (C/A, A/T, C/G or T/G); c) small insertions/deletions (indels).

SNPs signifier current type of genetic polymorphism and consequently offer high density information as markers in close proximity to gene/locus of interest. Moreover, this sequence variation can have a key influence on development of organism and responsiveness against surrounding environment.

SNPs present high density markers close to a locus of interest. It distinguishes among allied sequences, together at individual level and among individuals of a population. The occurrence and character of SNPs in plants is commencing to obtain significant interest. Various reports of sequence diversity in recent times have been published for a variety of plant species and these reports confirms the abundance of SNPs within plant systems, with frequency of 100–300 bp per SNP (Appleby *et al.*, 2009).

The nucleotide variants for SNPs at any defined locus could in standard engage four diverse nucleotide variants, but practically they are biallelic in nature. This shortcoming of SNPs is remunerated by the relative abundance of SNPs in contrast with multiallelic markers like SSRs, These SNPs have one more advantage that they are evolutionarily steady, not varying considerably during the course of inheritance among generation. The small mutation speed of SNPs formulates them outstanding markers for learning complex genetic trait variation and as a tool for indulgent genome evolution. Their high density across genome make them suitable for genome mapping projects, and specifically building ultrahigh-density genetic maps, establishment of excellent haplotyping systems for genes/locus of interest and map-based positional cloning. SNPs are utilized regularly in crop breeding programs, for cultivar identification, association with agronomic traits, categorization of genetic resources, genetic diversity analysis and phylogenetic study. Nevertheless, with the development of novel technologies to augment throughput and trim down the cost of SNP assays, beside development of plant genome sequencing projects relevance of SNPs will become more prevalent. Functional classes of SNPs were listed in table 1 as adopted from Mooney *et al* 2005.

Detection methods and techniques used for SNPs

In the sequencing era with vast amount of available data about SNPs, its identification and detection is of utmost importance. Basically there are two methods i. e. *In-vitro* discovery and *in-silico* discovery for the detection of SNPs. *In-vitro* discovery methods comprises of Non sequencing and sequencing based methods and includes Restriction digestion based techniques (RFLPs, CAPs dCAPs), DNA conformation technique (SSCP, DGGE, TGGE), Chip based methods and Target induced local lesions in genome (TILLING) whereas Sequencing based methods comprises of Locus specific PCR amplification, Whole genome shotgun method, Reduced representation shotgun (RRS), Alignment of available genomic sequences and Overlapping regions BACs and PACs. Besides these methods there are various Resequencing methods commercializing now like Pyrosequencing and MALDI-TOF because of their accurate SNP detection (Seal *et al.*, 2014).

There are two sets of data to perform SNP mining one is reference sequence data and other is *de novo* sequence data. This mining for various datasets mainly comprise of subsequent steps: in first step we have to group sequence reads on the basis of their sequence resemblance and confirm identity of reads whether they are covering similar part of genome or they have the same transcript origin. Further we have to align confirm reads and finally identify and categorize sequence variants as probable polymorphic loci/marker.

For the first part where sequence data is acquired from species for which a reference sequence is accessible, a homology search tool is requisite to map the novel sequence reads to the reference set. A global alignment tool or local alignment tool could be utilized, for example Sequence Search and Alignment by Hashing Algorithm (SSAHA) or BLAST be capable of performing this assignment. The other method for generation of reference data is by use of Polymerase Chain Reaction(PCR) product where primers were deliberated for a particular sequence region. Some of the tools such as Mapping and Assembly with Qualities (MAQ), Short Oligonucleotide Alignment Program (SOAP) are utilized for mapping the reference data. For the transcript data it is mapped against a Unigene set, as this result in an ungapped alignment. A spliced alignment

tool will be used for mapping to genomic data where such datasets are not available. The mapped data is then aligned at the position of the new sequence read on the reference. Pairwise or multiple alignments can be utilized for the assessment of base structure on each position and the resultant SNP recognition. Software tools like CAP3 and Phrap (http://www.pharp.org) are extensively utilized for accumulating the sequences to contigs. The sequence variants at each position are symbolized by multiple reads. More sequence reads obtainable in a species demonstrating a certain genomic region boost the chances of finding a polymorphism. A sequence variant (allele) can also be eminent from a sequencing error when it is established by multiple reads. Higher the number of reads per allele, higher is the probability of it being a true polymorphism.

For the *de novo* sequence data set the alignment of the sequence data that fit in to the same region of the genome, particular assembly tools are employed to split up the input datasets that are not assembled as contigs. For large number of reads more time will be required. Various tools specific for initial sequence fragment segregation into homologous groups which are again decomposed into clusters of unique origin for example Teraclu (http://www.timelogic.com), TGICL and d2cluster have been devised. After the clustering step, each cluster requires to be processed to align all reads within the cluster. All nucleotides from diverse reads at the identical position on the gene or genome are aligned and can be easily compared. If some fragments cannot be correctly aligned, they do not belong to a single cluster and are split into a second cluster. After individual reads have been clustered into aligned homologous groups, the final step of polymorphism recognition is identifying variations in the alignment and applying a scoring scheme.

Plant SNP databases and identification tools

There are various computational databases and approaches were developed for the discovery of new SNPs. There are more than ten diverse methods are accessible for SNP genotyping. In recent times, a range of online and freely accessible databases and tools have been devised for the recognition of SNPs in genomic sequences.

dbSNP

It is a public domain archive for a large compilation of simple genetic polymorphisms.

It includes SNPs, retrotransposons, small scale insertions and deletions, STR's etc. Each dbSNP entry includes frequency of polymorphism along with method of experiment and the protocols used to study the variations, etc.

POLYMORPH website

Dedicated database for Arabidopsis genome and we can question the SNP by diverse methods such as coding SNP by region, gene, SNPs flanked between accessions, SNPs by allele occurrence, etc. The assay progress menu having tools for CAPS marker searches, primer design and tool to create assay progress format and also information about repetitive sequences can be establish in repetitive menu.

Barley SNP database

It contains information on SNP polymorphisms from genes linked with abiotic stress in eight cultivars of barley. This database comprises 1717 studied contigs of barley, 1479 deliberate primers, and 1505 contigs and a precious SNP resource for barley genetics.

SNiPlay (http://sniplay.cirad.fr/, http://banana-genome.cirad.fr/sniplay)

It is web-based user pleasant interface web-based tool for polymorphism discovery. It has the SNP analysis pipeline comprising novel SNP detection tools as well as tools to calculate various types of statistical index. When the input file is provided in fasta, alignment format SNiPlay detects SNPs and insertion/deletion events and to dig out allelic information for each polymorphic position. There are a range of input file formats are accessible and the suite distinguish polymorphisms from the input file provided.

The SNiPlay database at present comprises data related to 4 Vitis projects and one *Coffea* project. The 'project overview' option summarises the data for each project. The Coffea transcriptome containing 5229 amplicons, 5201 genes, 59776 SNPs and also density of polymorphisms in the whole sequence.

AutoSNPdb

It comprises data in relation to contigs, SNPs and indels observed in Barley, Brassica, Rice and Wheat. The nucleotides are marked in four diverse colours and consensus sequences are also exhibit in in a different way and the SNPs are marked with different colours.

It also shows Uniref and GenBank annotation, Gene Ontology (GO) terms and reference genome location on the chromosomes. The SNPs are given a co-segregation score, which is a measure of the number of SNPs in the alignment that shares the same pattern of polymorphism between aligned sequences.

Maritime pine SNP database contains information of SNPs in the ESTs in pine trees mined by three programs namely Phred, Phrap and PolyBayes (http://www.pharp.org).

ESTree DB

It comprises SNP description of peach tree and almond. SNPs in this database can be availed directly in numerically ordered reports (i.e. SNP report 1, SNP report 2, etc.), or via a search engine that allows cluster recovery by SNP number, EST name or EST GI number. It facilitates users to view the contig alignments, cluster information and co-segregation score. Local BLAST search can also be performed on ESTree nucleic; SNP or protein DB.

Plant Markers

It comprises markers for over 50 different plant species. Plant Markers is a genetic marker database that contains a inclusive pool of predicted molecular markers. It utilizess the SNIPER algorithm, SNPs are selected and is validated for both plants and animals. Panzea (Canaran *et al.,* 2008) project explains the genetic construction of complex traits in maize and teosinte. It illustrates the domestication traits and agronomic traits like flowering time, plant height and kernel quality. The database design is based on the Genomic Diversity and Phenotype Data Model (GDPDM). With the marker search utility of Panzea, one can get the markers utilised in this project, i.e. the marker SNPs, HapMap-SNPs, SSR, sequencing, indels, CAPS, Isozyme and other markers in each of the chromosomes. It also displays the sequence and location of the markers in the sequences. By utilizing Polymorphism search, one can obtain the genotypes of Panzea SNP or SSR markers that have been assayed on a pair of maize or teosinte inbred lines of choice.

GABI 'Genomanalyse im biologischen System Pflanze'

It is the name of a large combined network of diverse plant genomic research projects. The objective of Genome Analysis of the Plant

Biological System (GabiPD) is to assemble, incorporates, analyse and envisage primary information from GABI projects. GabiPD represents a repository and investigation platform for a wide array of heterogeneous data from high-throughput experiments in several plant species. Data from different 'omics' fronts (e.g. genomics, transcriptomics, proteomics and metabolomics) originating from 14 different model or crop species are incorporated here. The Green Cards (http://www.gabipd.org/database/cgi-bin/ GreenCards.pl.cgi) allow visualisation of annotations and displaying protein domains and gene structure. It gives interactive gene maps from potato and barley and protein 2D gels from *A. thaliana* and *Brassica napus*.

TreeSNPs

TreeSNPs is based on PostgreSQL database and Ruby on Rails and its basis code can be attained and customized freely. To this database, users can also incorporate the data that is generated as a result of their work from genes families from well-studied species, recognition of gene homologs, construction of primer sets covering different gene regions, PCR results, sequencing results and SNP discovery. The main benefit of TreeSNPs over existing systems is its ability to integrate data from any project involved in SNP identification from PCR-amplified sequenced fragments. SNPs were recognized by amplification of 1kb fragments by PCR from genomic DNA of various strains, following sequencing of the PCR products, and alignment of those sequences.

BGI-RIS database

Dedicated database comprises of assembling contigs and anchoring contigs and scaffolds onto rice chromosome based mapped genetic markers and BAC-based physical maps of sequence contigs of *Beijing indica* and *Syngenta japonica*. RePS (repeat – masked Phrap with scaffolding) (http://ftp.genome. washington.edu/cgi-bin/RepeatMasker) is a program that explicitly identifies repetitive sequences and tagging repeat sequences using K-mer repeats from the shotgun sequencing data and removes them prior to assembly.

IRIS

It is the execution of the International Crop Information System (ICIS), it includes germplasm pedigrees, field evaluations, structural and functional genomic data and environmental data.

Orygenes DB

It is a tool to study rice reverse genetics. Using genome browser (Gbrowse), a web-based application for displaying genomic annotations and other features, Orygenes searches in all the chromosomes and shows the position of the markers whether it is in intron, exon, 3'UTR or in the promoter regions and shows orientations of markers, i.e. in reverse or forward direction. It also gives link to the Gramene Marker database where particulars of the markers positions are available. This database holds basic information about different markers used for genetic mapping. Other than Orygenes DB, Rice Tos17 insertion mutant database (http://tos.nias.affrc.go.jp/) contains information for Tos17, T-DNA and D's in the sequence of rice.

Other Rice marker databases

It consists of Rice Mutant Database (RMD) which is an archive of T-DNA insertion mutant's information. The National Institute of Agrobiological Sciences in Tsukuba, Japan has produced high quality genetic maps of rice with 3000 markers, 30,000 full length cDNA, expression profiles, Tos17, D's, and this data

is publicly available. A unification tool Rice PIPELINE dynamically collects and collates data from various databases like KOME, INE, RED, Tos17 and PLACE. The function of Rice PIPELINE is to provide unique scientific resource that pools publicly available rice genome data for search by clone sequence, Genbank accession id or other keywords. A marker search in the Rice PIPELINE database leads directly to the data available on specific DNA clones and with the help of this information gene function can be identified. INE (Sakata *et al.,* 2000) is a database that integrates the genetic map, physical map and sequencing information of the rice genome. Knowledge based Oryzae Molecular Biological Encyclopedia (KOME) (Kikuchi *et al.,* 2003) is a database of full length cDNAs of 28,469 unique genes of rice. In this database full sequencing, nucleotide analysis, amino acid analysis, GO classification and digital mapping on the genome sequences of *Indica* and *Japonica* cultivars could be performed. It also searches InterPro by keyword and a lot of information about SNPs can be obtained such as GO identification of proteins,etc.

Grain genes

It is a database for triticeae, oats and sugarcane. This database have information about the cytogenetic maps, genomic probes, nucleotide sequences, genes, alleles, QTLs, pedigrees of cultivars, germplasms, pathogens and pathologies(Carollo *et al.,* 2005).

The legume information system database

It utilizes the CMAP software and informs about the maps such as the SNP, SSR, STS, telomere, RFLP's, RAPD's of the *Glycine max, Lotus japonicus, Medicago truncatula, Arachis hypogaea, Cajanus cajan, Cicer arietunum, Medicago sativa, Phaseolus vulgaris, Pisum sativum, Trifolium repens* and *Vigna unguiculata*.

Eucalyptus SNP Database (EUSNPDB)

It is a eucalyptus SNPs database. SNPs are mined from the EST libraries of Eucalyptus from nucleotide sequences collected from GenBank.

EST sequences are retrieved from the dbEST containing 35,320 sequences from mesophyll leaves, differentiating xylem, flower, shoot apex, woody tissue and root. Shannon Index is used to estimate the distribution of SNPs/Indels. A total 33,466 SNP sites and 5874 indel polymorphisms in 26,026 ESTs analysed (Singh *et al.,* 2011).

Table.1 SNP functional classes

Functional class	Short form	Description
Coding SNPs	cSNP	Positions that fall within the coding regions of genes
Regulatory SNPs	rSNP	Positions that fall in regulatory regions of genes
Synonymous SNPs	sSNP	Positions in exons that do not change the codon to substitute an amino acid
Non-synonymous SNPs	nsSNP	Positions that incur an amino acid substitution
Intronic SNPs	iSNP	Positions that fall within introns

Table.2 List of Crops and no. of SNPs detected in it

CropOP		Citation	No.of SNPs detected
Alfalfa		Li *et al.*, 2012	872.384
Cotton		Bayers *et al.*, 2012 Zhu *et al.*, 2014	151,712 40,503
Peanut		Khera *et al.*, 2013 Zhou *et al.*, 2014	8,486 1,765
Potato		Uitdewilligen *et al.*, 2013	42,625
Rapeseed		Trick *et al.*, 2009 Hu *et al.*, 2012 Huang *et al.*, 2013	41,593 655 892,803
Wheat		Allen, 2013 Cavanagh *et al.*, 2013	10,251 25,454
White Clover		Nagy *et al.*,2013	208,854

Table.3 Validated SNPs in various crops

Crop	Gene	function	Trait associated with SNP	Utility of SNP
Barley (Hordeum vulgare)	B-Amylase gene (Shu and Ramussen, 2014)	Degradation of starch	Enzyme thermostability	To select barley seedling carying superior allele of B-Amylase
Wild Barley (*H. spontaneum*)	Dhn1 & Dhn5 (Dehydrin) Karami *et al.*, 2013	Adaptive response of plant to environmental stress	Resistance to water stress	For water stress adaptation
Rice (*Oryza sativa*)	Wx (waxy) gene Sd-1(semi-dwarfing) gene (Yang *et al.*, 2014)	1) Control amylose synthesis by coding starch synthase enzyme 2) Dwarfism	1) Amylose content 2) Dwarfism	1) For development of new cultivar 2) Selection of sd-1 in breeding programme
Wheat (*Triticum aestivum*)	Pin b (Puroindolin b) Rht 1 & Rht 2 gene	Thicken the coat Dwarfism	1) Grain hardiness 2) Dwarfism	Breeding program
Soybean (*Glycine max*)	Rhg 1& Rhg 4	Soyabean Cyst nematode resistance allele	SCN resistance	Breeding programme
Onion (*Allium cepa*)	SNP allele in Plastosome	Responsible for CMS	Cytoplasmic male sterility and fertility	For development of CMS lines
Musturd (*Brassica juncea*)	FAE 1 gene	Fatty acid elongase	Erucic acid content	Breeding programme

QualitySNP tool

It utilizes a haplotype-based strategy to detect reliable synonymous and non-synonymous SNPs from public EST data without the requirement of trace/quality files or genomic sequence data. It uses an algorithm called CAP3 for clustering and alignment and three more filters for the recognition of reliable SNPs. In initial filtering level it screens possible SNPs and identifies variations between or within different genotypes and passes the screened information to next filter, where it utilizes a haplotype-based approach to perceive consistent SNPs.

Clustering is performed and possible false SNPs are recognized within second level. In third and final filtering procedure it calculates a confidence score for each SNP based upon sequence redundancy and quality. Non-synonymous SNPs are consequently identified by detecting open reading frames of consensus sequences (contigs) with SNPs.

In conclusion, SNP mining can provide better understanding of crops at the gene level, for the detailed analysis of germplasm and eventually for the proficient management of genetic diversity contained by plant breeding on a whole genome level. Several identified SNPs in respective crops and their validation were listed in table 2 and 3.

While a variety of methods area cross-the-board for SNP detection are easily accessible and the pace with which SNPs can be recognized in major crop plants are still increasing with the involvement of the next-generation sequencing techniques, various constraints will be there to be deal with before large-scale SNP genotyping will be utilized in major crop plants as frequent tool for reasons such as association genetics and plant breeding.

The genomic sequences of major crop plants will be on hand in the near future, the focus should be positioned in the discovery of SNPs in as many genes as possible and the parallel investigation of many diverse lines. It is very likely that by the use of amplicon resequencing or sequence capture techniques in combination with the next-generation sequencing technologies, we will have within a few years identified SNPs and haplotypes in almost all of the 30 000–60 000 genes of a crop plant. Moreover, emphasis should be put onto the recognition of SNPs within the actual diversity range of breeding material and validated SNPs should be mapped exactly in large segregating populations. At present, large-scale SNP analysis in many crop plants is still based on individual SNPs. In the future this needs to shift toward haplotype-specific SNPs for more efficient association studies as it is done in human genome analysis. This needs major gene and intergenic regions to be sequenced in many individuals. With haplotype-based SNPs identified in the gene repertoire of a crop plant, the door will be open toward the detailed analysis of germplasm, efficient association studies of SNP markers with traits, and eventually the efficient management of genetic diversity within plant breeding on a whole genome level.

References

Allen, A.M., 2013, Discovery and development of exome-based, codominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Pl. Biotechnol. J.*, 11: 279–295.

Appleby, N., Edwards, D. and Batley, J., 2009, New technologies for ultra-high throughput genotyping in plants. *Methods in Mol. Biol.*, 513:19–39.

Byers, R.L., Harker, D.B., Yourstone, S.M., Maughan, P.J., and Udall, J.A., 2012.

Development and mapping of SNP assays in allotetraploid cotton. *Theor. Appl. Genet*, 124: 1201–1214.

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L. and Akhunova, A., 2013, Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. USA*, 110:8057–8062.

Chuang, L., Yang, C., Tsui, K., Cheng, Y., Chang, P., Wen, C. and Chang, H., 2008, Restriction enzyme mining for SNPs in genomes. *Anticancer Res.*, 28: 2001-2008

Edwards, D., Forster, J.W., Chagné, D., and Batley, J. 2007, What are SNPs, in Association Mapping in Plants. *Springer*, 41–52.

Ganal, M., Thomas, A. and Marion, R., 2009, SNP identification in crop plants.*Curr. Op. in Pl. Biol.*, 12:211 –217.

Hu, Z., Huang, S., Sun, M., Wang, H., and Hua, W., 2012, Development and application of single nucleotide polymorphism markers in the polyploid *Brassica napus* by 454 sequencing of expressed sequence tags. *Plant Breed*.,131:293–299.

Huang, S., Deng, L., Guan, M., Li, J., Lu, K., Wang, H., Fu, D., Mason, A.S., Liu, S., and Hua, W. 2013, Identification of genome-wide single nucleotide polymorphisms in allopolyploid crop Brassica napus. *BMC Genomics*14:717.

Karami, A., Shahbazi, M., Niknam, V., Shobbar, Z., Tafreshi, R., Abedini, R. and Mabood, H.2013. Expression analysis of dehydrin multigene family across tolerant and susceptible barley (*Hordeum vulgare* L.) genotypes in response to terminal drought stress. *Acta Physiol. Planta*. 35(7):2289–97.

Khera, P., Upadhyaya, H.D., Pandey, M.K., Roorkiwal, M., Sriswathi, M., Janila, P., Guo, Y., Mckain, M.R., Nagy, E.D. and Knapp, S.J. 2013 Single nucleotide polymorphism–based genetic diversity in the reference set of peanut (*Arachis spp*.) by developing and applying cost-effective competitive allele specific polymerase chain reaction genotyping assays. *Pl. Genome*6:1–11.

Kwak, P. and Chen, X., 2003, Detection of Single Nucleotide Polymorphisms. *Curr. Issues Mol. Biol.*, 5: 43-60

Lai, K., Michał, L., Lee, H., Berkman, P., Bayer, P., Visendi, P., Ruperao, P., Fitzgerald, T., Zander, M., Chan, C., Manoli, S., Stiller, J., Batley, J. and Edwards D., 2015, Identification and characterization of more than 4 million intervarietal SNPs across the group 7 chromosomes of bread wheat. *Pl. Biotech. J.*, 13: 97–104.

Li, X., Acharya, A., Farmer, A.D., Crow, J.A., Bharti, A.K., Kramer, R.S., Wei, Y., Han, Y., Gou, J. and May, G.D., 2012, Prevalence of single nucleotide polymorphism among 27 diverse alfalfa genotypes as assessed by transcriptome sequencing. *BMC Genomics*, 13:568.

Mooney, S., 2005, Bioinformatics Approaches and resources for single nucleotide polymorphism functional analysis. *Brief., Bioinfo.*,6:44-56.

Nagy, I., Barth, S., Mehenni-Ciz, J., Abberton, M.T., and Milbourne, D., 2013, A hybrid next generation transcript sequencing-based approach to identify allelic and homeolog-specific single nucleotide polymorphisms in allotetraploid white clover. *BMC Genomics*, 14:100.

Seal, A., Gupta, A., Mahalaxmi, M., Aykkal, R., Singh, T.R. and Arunachalam, V., 2014, Tools, resources and databases for SNPs and indels in sequences: a review. *Int. J. Bioinfo. Res. and App.*,10(3): 264–296.

Shu, X. and Ramussen, S. (2014) Quantification of amylose, amylopectin, and β-glucan in search for genes controlling the three major quality traits in barley by genome-wide association studies. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2014.00197.

Uitdewilligen, J., Wolters, A., D'Hoop, B., Borm, T.J.A., Visser, R.G.F., and Eck, H., 2013, A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One*, 8:e62355.

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., Wang, G., Luo, Q., Zhang, Q., Liu, Q. and Xiong, L. (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Comm.*, 5: 5087.

How to cite this article:

Saurabh Pandey, Sunidhi Mishra and Kailash Chandra. 2019. Tools and Resources for SNP Mining in Crop Plants. *Int.J.Curr.Microbiol.App.Sci.* 8(01): 2819-2829.
doi: https://doi.org/10.20546/ijcmas.2019.801.296