GENOME WIDE SNP DISCOVERY IN JAPONICA PGMS RICE VIA GBS, A STRIKING VENTURE IN EXPLOITATION OF GENOMIC DIVERSITY

Imdad Ullah Zaid[1], Saima Noor[1], Mudassar Mushtaq[1], Muhammad Amir Zia[1], Kainat Bibi[2], Sabahat Noor[1], Ghulam Muhammad Ali[1], Aitezaz Ali Asad Shahani[3] and Zaheer Abbas[1]*

1National Institute for Genomics and Advanced Biotechnology (NIGAB), National Agricultural Research Centre, Islamabad 45500, Pakistan

2Crop Sciences Institute, National Agricultural Research Centre, Islamabad 45500, Pakistan

3Key laboratory of crop sciences and plant breeding genetics, college of agriculture, Yanbian University, Yanji, Jilin, China

Abstract

Analysis of genomic variations using DNA sequencing technologies has imparted significance to crop genetics and breeding. In the present study, a photo sensitive genic male sterile (PGMS) line of a Japonica rice has been sequenced by Genotyping by Sequencing (GBS) method of next generation sequencing (NGS) for detecting genome wide single nucleotide polymorphisms (SNPs). The optimized protocol of GBS revealed 57.2% of uniquely mapped short reads onto the Nipponbare reference rice genome. The uniquely mapped reads of 200bp were further imputed in TASSAL-GBS software for identification of total 1894 genome-wide SNPs. The identified SNPs within the PGMS genome was scattered unevenly across and within the 12 chromosomes of rice. The annotation of identified SNPs confirmed the presence of 963 SNPs within the intragenic spaces along with 140 and 765 SNPs in the intron and exon regions, respectively. Also, the presence of 498 non-synonymous SNPs in the sequenced genome inferred significance for future breeding applications. The outcome of our study implied that 1004S could be elite breeding germplasm for enhancing the genetic diversity of two-line hybrid rice system. Moreover, the identified SNPs concluded to be potential resource for molecular breeding and genome-wide variations on hybrid performance.

Key words; GBS, PGMS, SNP discovery, annotation, diversity.

1. Introduction

Mining of genetic diversities in crop genomes has a broad spectrum of applications in breeding experiments. Genetic diversity through molecular markers has a remarkable insight to the processing, propagation and elucidation of genetic information. Contribution of molecular markers in plant breeding has vast range of genetic resource conversation and their proper management (Zeid 2003; Bertini *et al*. 2006; Perseguini *et al*. 2011; Skuza *et al*. 2013; Dao *et al*. 2014; Kumbhar *et al*. 2015). Molecular markers provide an efficient genetic platform for accurate assessment and characterizations of genetic resources of crop germplasm. Plant breeders use molecular markers to screen the genetic basis of germplasm and manipulate their genetic diversities at locus/gene level to provide useful information of quantity and distribution of genetic diversity within and among breeding populations (Mondini *et al*. 2009).

The concept of molecular markers in plant breeding has been used since ages; it offered significance over phenotypic observations. Tagging alleles and genes authorizing quantitative traits is the optimum usage of molecular markers (Wu and Tanksley 1993; Rahman *et al*. 2008, Fu *et al*. 2010; Khaled *et al*. 2015; Qin *et al*. 2015). Also, it extended the significance in locating disease resistance locus in many plant species (Michelmore *et al*. 1991; Borovkova *et al*. 1995; Michelmore 1995; Huang *et al*. 1997; Ragimekula *et al*. 2013; Ashkani *et al*. 2015). However, these electrophoretically separated DNA markers distinguish the genetic polymorphisms in a breeding population and consumed several months of genotyping. For genome-wide scanning of large breeding population in small period of time with low cost and more efficient and reliable for exploiting the genomic variations. NGS is the most efficient and time saving method for scanning a breeding population and exploiting genetic polymorphism, with low cost and high throughput genotyping.

NGS is the phenomenal invention soaring the magnificent horizon of genome-wide DNA polymorphisms (Nielsen *et al*. 2011; Kumar *et al*. 2012; Trick *et al*. 2012). With the advent of NGS technology, multiple genomes (animals and plants) have been sequenced for marker identifications, such as for SNPs, insertion and deletions (InDels) (Feltus *et al*. 2004; Van Tassell *et al*. 2008; Wiedmann *et al*. 2008; Amaral *et al*. 2011; Subbaiyan *et al*. 2012; Huang *et al*. 2013; Kim *et al*. 2014). In plants, SNP identification is usually facilitated by NGS. SNPs are the single base pair change in plants DNA which are distributed randomly genome-wide (McCouch *et al*.

2010). SNPs gained tremendous importance as a third generation DNA-based molecular markers that were widely practiced in molecular breeding experiments. Currently, extensively being exploited in multiple analysis, which includes association analysis, population polymorphism, marker assisted breeding, QTL analysis, positional cloning, pedigree haplotype and for variety discriminations studies (Rafalski 2002; Elshire *et al*. 2011; You *et al*. 2011). In rice, SNPs have been investigated among indica, Japonica species and of their wild type also (Monna *et al*. 2006; Nagasaki *et al*. 2010; Arai-Kichise *et al*. 2011; Thomson *et al*. 2012; Kim and Tai 2013; Jain *et al*. 2014; Takano *et al*. 2014; Tang *et al*. 2016). Likewise, after complete genome sequence of the rice genome (Project 2005), it initiated a revolutionary sequencing platform for SNPs discovery at the genome-wide scale. Very recently, 2495052 and 660778 of SNPs were detected by whole genome sequencing between the indica and japonica parents of hybrid rice, respectively (Subbaiyan *et al*. 2012; Hu *et al*. 2014). Similarly, 132462 of SNPs were screened out through protocol of whole genome sequencing in a landrace omachi belongs to japonica rice (Arai-Kichise *et al*. 2011).

Hybrid rice research originated in China and successfully attained appreciable production scale, promoting 20% more grain yield compared to indica and japonica inbred lines (Cao and Zhan, 2014). Chinese breeders enabled large-scale cultivation of hybrid rice through adapting the three-line and two-line hybrid rice breeding schemes. The requisites for three line scheme comprised of cytoplasmic-genetic male sterile line (A line), a maintainer line (B line) and a restore line (R line). The two lines hybrid rice breeding system were evaluated in China in 1985 (Shi 1985). The two line hybrid rice breeding scheme was established on a restorer line and a photoperiod genic male sterility (PGMS) having a recessive gene and strongly influenced by temperature or light. During the past 20 years, international rice research institute (IRRI) and China have developed several PGMS and restorer lines of two line breeding system. Since then considerable work has been done to exploit the genetic diversity of mating parents of two line breeding systems. Previous breeding practices reported that parental combinations of two-line hybrid rice provided limited genetic diversity. Therefore, efforts are needed to analyze and expand the genetic diversity of parents of two-line hybrid rice.

The japonica PGMS line (1004S) was developed in Jiangsu Academy of Agricultural Sciences, Nanjing, China. As a pollen receiving parent in two-line hybrid rice breeding system, it showed potential in the development of two-line rice hybrid. Also, it has challengeable high fertility and strong adaptation toward several biotic and abiotic stresses. The F_1 derived from

1004S had high grain yield with resistance against stripe disease (*stv*). The present experiment was carried out to sequence the 1004S using NGS-based GBS method for discovery of genome-wide SNPs. We expect that the results of our experiment will provide an assessment of GBS protocol for exploiting genetic diversities in parents of two-line hybrid rice.

2. Material and Methods

2.1. DNA extraction and library construction

In order to construct a high-quality genomic library for sequencing, young leaves of 1004S were selected and total genomic DNA was isolated with DNA secure plant kit (Tiangen biotech, Beijing, China) under recommended protocol (Lu *et al*. 2016). This kit is specialized, based on silica membrane technology containing a set of unique buffers to eliminate impurities.

The isolated DNA yield was assessed for purity and concentration by three different methods including, agarose gel electrophoresis, spectrophotometer test (Nano Drop 2000) and Qubit 2.0 Fluorometer. After purification, approximately 50µL of DNA was selected to construct DNA library following the standard protocol of genotyping by sequencing method (Poland *et al*. 2012).

The genomic library of 200 bp was constructed by digesting DNA with two restriction enzymes, followed by a ligation reaction. A set of the developed adapter was used for the ligation reaction. At last, a polymerase chain reaction (PCR) was performed to build the GBS library. The experiment of NGS was performed with Ion Torrent machine using Ion Torrent kits (200 Kit v2) under the manufacturer specifications (Life Technologies, Carlsbad, CA, U.S.A). After sequencing The FASTQ raw DNA sequence data file was submitted to NCBI short read archive (SRA) under the study accession number: SRR2758809.

2.2. Read mapping and SNP calling

To minimize the sequencing errors, all the short reads were trimmed by applying screening steps. All the obtained reads of 1004S were aligned against reference Nippon bare genome (http://rice.plantbiology.msu.edu/) using Bowtie 2 software with default parameters (Langmead 2010). The mapped short reads were further classified into categories of uniqueness, multi-location and unmapped reads. Only certain reads mapped uniquely onto reference were considered for genome-wide SNP calling. The Tassel application of GBS pipeline was used to call genome-wide SNPs (Glaubitz *et al*. 2014). SNP distribution across and within chromosomes was also analyzed.

2.3. In-depth annotations

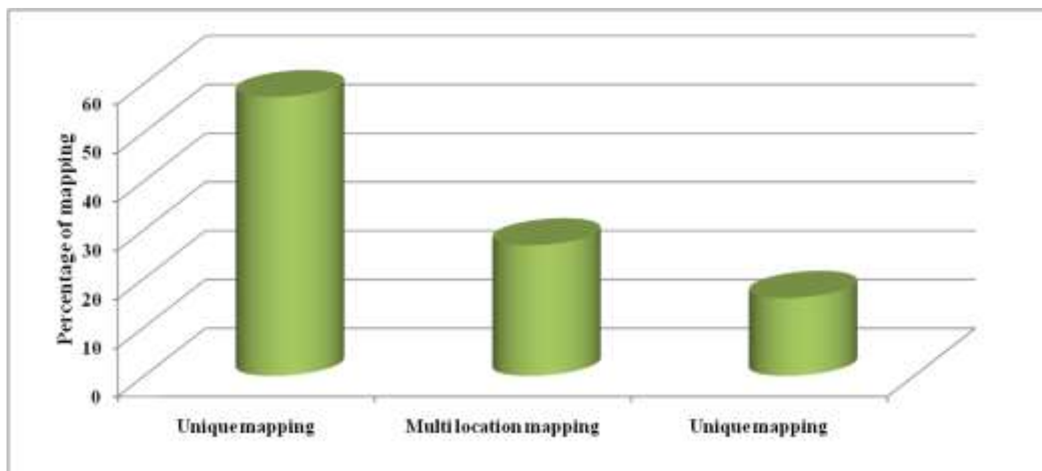
The annotation of SNPs provides the necessary information about SNP location in the sequenced genome. To identify the SNPs within the coding or non-coding region, the information of reference genome was tracked and annotated accordingly. SNPs in genic regions were divided into coding sequence, un-translated regions, and intron variants. The SNPEff (4.2) software was used to annotate and further classify SNPs into synonymous and non-synonymous amino acid substitutions (Cingolani *et al*. 2012).

3. Results and Discussion

3.1. Short read mapping

After de-multiplexing of attached barcodes and removal of undesired sequences, the sequenced file was then evaluated through FastQC software. The short reads of 1004S genome were mapped via the Bowtie algorithm. The 200-bp long single-end reads were mapped against the Nippon bare genome. The mapping results retrieved a total of 496610 high-quality short reads. Among them, 57.2% aligned exactly once and 26.8% aligned more than one time onto the 12 chromosomes of Nipponbare genome (Figure 1). The successfully mapped short repeats against the reference genome contributed to the genome wide SNP discovery (Nielsen *et al*. 2011). Nearly 84% of alignment rate was recorded. About, 16% of the reads remained unmapped against the reference genome. The genomic deletions/duplications in the process of sequencing experiment are the most probable factors for the un-mapped short repeats (Arai-Kichise *et al*. 2011).

Figure 1. Mapping of short reads onto Nipponbare reference genome.



3.2. SNP detection and distributions

The recent discovery of molecular markers such as SNPs, insertions, and deletions through NGS-based GBS method proved to be exclusive and time saving (Elshire *et al*. 2011; Poland *et al*. 2012; İpek *et al*. 2016). The GBS protocol was applied to many crop species for genome-wide variants discovery. The main features of GBS assay include its low cost, few quality and standard checks, efficient bar-coding system, reliable sequencing of the large-scale genome and multiplexed genotyping (Davey *et al*. 2011). Within the mapped sequenced reads of 1004S japonica PGMS line, a total of 1894 SNPs were identified across 12 chromosomes of rice (Table 1). The SNPs in the sequenced line were identified by individual comparison with the Nippon bare reference genome. The mapped file obtained from bowtie software was submitted in Tassal-GBS application for genome-wide SNPs discovery. The discovered SNPs in the PGMS line were distributed non-randomly over the 12 chromosomes of rice. The maximum number of 547 SNPs was found on chromosome 2, while minimum number (37) of SNPs was located on chromosome 4. The 12 chromosome of rice revealed an average of 157 SNPs. Our findings of low polymorphisms in japonica based PGMS line clearly exhibited the limited diversity and narrow genetic background. (Yang *et al*. 1994; Garris *et al*. 2005).

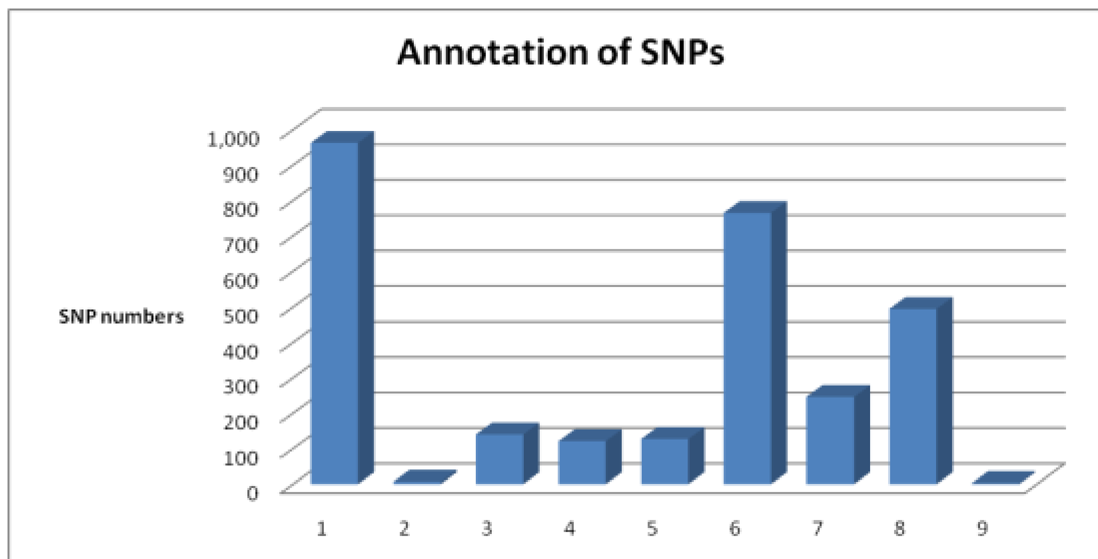
Table 1. SNPs observed in 1004S in comparison with Nipponbare reference genome.

Chromosome	Length (MB)	SNPs	SNP rate
1	43,270,923	211	152,362
2	35,937,250	547	192,177
3	36,413,819	119	316,641
4	35,502,694	37	163,606
5	29,958,434	55	475,530
6	31,248,787	137	201,605
7	29,697,621	70	204,811
8	28,443,022	64	160,695
9	23,012,720	73	511,393
10	23,207,287	87	164,590
11	29,021,106	387	90,975
12	27,531,856	107	199,506
Total	**373,245,519**	**1,894**	**187,938**

3.3. Annotation of identified SNPs

The annotation of Nipponbare rice was used as a reference to confirm the genomic distributions of identified SNPs of 1004S within the various genomic locations. Among the identified 1894 SNPs, the majority 963 (50.8%) of SNPs was detected within the intragenic region of the genome (Figure 2). Previous studies endorsed the similar results regarding the SNP distribution within the intragenic regions (Subbaiyan *et al*. 2012; Hu *et al*. 2014; Srivastava *et al*. 2014). Altogether, 140 (7.3%) and 765 (40.3%) of SNPs were found in the intron and exon regions, respectively. The SNP effect on amino acid substitutions was also analyzed by classification of identified SNPs as synonymous and non-synonymous. The SNPs in the coding region as synonymous nature was counted as 246 (12.9%), while the non-synonymous nature was further divided into missense, 495 (26.1%), and nonsense, 3 (0.1%). The SNPs with non-synonymous nature pertains biological importance that can affect the proteins of different genes (Ng and Henikoff 2006; Kumar *et al*. 2009; Sauna and Kimchi-Sarfaty 2011). Therefore, the presence of non-synonymous SNPs within sequenced genome provides valuable insights for understanding the performance of lines. The 3´-UTR and 5´-UTR regions possessed 121 (6.3%) and 127 (6.7%) of SNPs, whereas, the remaining 7 (0.3%) SNPs were situated inside the splice region of genomes of the 1004S PGMS line.

Figure 2. Annotation of identified SNPs.



1;Intergenic,2;Splice stites,3;intron 4; 3´-UTR,5; 5´-UTR, 6; exon, 7; Synonymous, 8; Missense, 9; Nonsense.

4. Conclusion

The identified SNPs in the present study in addition to already available SNP assets imparts substantial significance in breeding and provide valuable resources for future genomic studies, molecular-assisted breeding lines, QTL mapping, haplotype construction, association studies and pedigree analysis. In conclusion, this study identified that SNPs showing high level of genetic diversity based on genome-wide sequencing in PGMS line. The results deduced that discovered SNPs in this study will helpful for exploiting genetic diversity of parents of hybrid rice.

Author Contributions:

1. Imdad Ullah Zaid performed the experiment and wrote the initial draft
2. Saima Noor, Mudassar Mushtaq, Muhammad Amir Zia, Sabahat Noor, Kainat Bibi, Aitezaz Ali Asad Shahani helped in performing GBS experiment
3. Kainat Bibi reviewed the article, format it according to journal requirement and submitted it to the respective journal for publication.
4. Ghulam Muhammad Ali and Zaheer Abbas supervised the experiment and acquired funding

Funding: National Institute for Genomics and Advanced Biotechnology (NIGAB), National Agricultural Research Centre, Islamabad 45500, Pakistan

5. References

Amaral AJ, Ferretti L, Megens H-J, Crooijmans RP, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MA. 2011. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS One. 6(4):e14782.

Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, Yano M, Wakasa K. 2011. Discovery of genome-wide DNA polymorphisms in a landrace cultivar of japonica rice by whole-genome sequencing. Plant and cell physiology. 52(2):274-282.

Ashkani S, Rafii MY, Shabanimofrad M, Miah G, Sahebi M, Azizi P, Tanweer FA, Akhtar MS, Nasehi A. 2015. Molecular Breeding Strategy and Challenges Towards Improvement of Blast Disease Resistance in Rice Crop. Frontiers in plant science. 6.

Bertini CH, Schuster I, Sediyama T, Barros EGd, Moreira MA. 2006. Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genetics and molecular biology. 29(2):321-329.

Borovkova I, Steffenson B, Jin Ya, Rasmussen J, Kilian A, Kleinhofs A, Rossnagel B, Kao K. 1995. Identification of molecular markers linked to the stem rust resistance gene rpg4 in barley. Phytopathology. 85(2):181-185.

Cao L, Zhan X. 2014. Chinese Experiences in Breeding Three-Line, Two-Line and Super Hybrid Rice. Yan WG, Bao J S Rice: Germplasm, Genetics and Improvement Rijeka, Croatia: InTech.279-308.

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(2):80-92.

Dao A, Sanou J, Mitchell SE, Gracen V, Danquah EY. 2014. Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. BMC genetics. 15(1):1.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics. 12(7):499-510.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 6(5):e19379.

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH. 2004. An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome research. 14(9):1812-1819.

Fu Q, Zhang P, Tan L, Zhu Z, Ma D, Fu Y, Zhan X, Cai H, Sun C. 2010. Analysis of QTLs for yield-related traits in Yuanjiang common wild rice (Oryza rufipogon Griff.). Journal of Genetics and Genomics. 37(2):147-157.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in Oryza sativa L. Genetics. 169(3):1631-1638.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One. 9(2):e90346.

Hu Y, Mao B, Peng Y, Sun Y, Pan Y, Xia Y, Sheng X, Li Y, Tang L, Yuan L. 2014. Deep re-sequencing of a widely used maintainer line of hybrid rice for discovery of DNA polymorphisms and evaluation of genetic diversity. Molecular Genetics and Genomics. 289(3):303-315.

Huang N, Angeles E, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush G. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theoretical and Applied Genetics. 95(3):313-320.

Huang S, Deng L, Guan M, Li J, Lu K, Wang H, Fu D, Mason AS, Liu S, Hua W. 2013. Identification of genome-wide single nucleotide polymorphisms in allopolyploid crop Brassica napus. BMC genomics. 14(1):717.

İpek A, Yılmaz K, Sıkıcı P, Tangu NA, Öz AT, Bayraktar M, İpek M, Gülen H. 2016. SNP Discovery by GBS in Olive and the Construction of a High-Density Genetic Linkage Map. Biochemical genetics.1-13.

Jain M, Moharana KC, Shankar R, Kumari R, Garg R. 2014. Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. Plant biotechnology journal. 12(2):253-264.

Khaled A, Motawea M, Said A. 2015. Identification of ISSR and RAPD markers linked to yield traits in bread wheat under normal and drought conditions. Journal of Genetic Engineering and Biotechnology. 13(2):243-252.

Kim J-E, Oh S-K, Lee J-H, Lee B-M, Jo S-H. 2014. Genome-wide SNP calling using next generation sequencing data in tomato. Molecules and cells. 37(1):36.

Kim S-I, Tai TH. 2013. Identification of SNPs in closely related Temperate Japonica rice cultivars using restriction enzyme-phased sequencing. PLoS One. 8(3):e60176.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 4(7):1073-1081.

Kumar S, Banks TW, Cloutier S. 2012. SNP discovery through next-generation sequencing and its applications. International journal of plant genomics. 2012.

Kumar S, You FM, Cloutier S. 2012. Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC genomics. 13(1):1.

Kumbhar SD, Kulwal PL, Patil JV, Sarawate CD, Gaikwad AP, Jadhav AS. 2015. Genetic Diversity and Population Structure in Landraces and Improved Rice Varieties from India. Rice Science. 22(3):99-107.

Langmead B. 2010. Aligning short sequencing reads with Bowtie. Current protocols in bioinformatics.11.17. 11-11.17. 14.

Lu Q, Zhu J, Yu D, Xu X. 2016. Genetic and geographical structure of boreal plants in their southern range: phylogeography of Hippuris vulgaris in China. BMC evolutionary biology. 16(1):1.

McCouch SR, Zhao K, Wright M, Tung C-W, Ebana K, Thomson M, Reynolds A, Wang D, DeClerck G, Ali ML. 2010. Development of genome-wide SNP assays for rice. Breeding Science. 60(5):524-535.

Michelmore R. 1995. Molecular approaches to manipulation of disease resistance genes. Annual review of phytopathology. 33(1):393-427.

Michelmore RW, Paran I, Kesseli R. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences. 88(21):9828-9832.

Mondini L, Noorani A, Pagnotta MA. 2009. Assessing plant genetic diversity by molecular tools. Diversity. 1(1):19-35.

Monna L, Ohta R, Masuda H, Koike A, Minobe Y. 2006. Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (Oryza sativa L.) and a wild accession (Oryza rufipogon Griff.). DNA research. 13(2):43-51.

Nagasaki H, Ebana K, Shibaya T, Yonemaru J-i, Yano M. 2010. Core single-nucleotide polymorphisms-a tool for genetic analysis of the Japanese rice population. Breeding Science. 60(5):648-655.

Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet. 7:61-80.

Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics. 12(6):443-451.

Perseguini JMKC, Chioratto AF, Zucchi MI, Colombo CA, Carbonell SAM, Mondego JMC, Gazaffi R, Garcia AAF, Campos Td, Souza APd. 2011. Genetic diversity in cultivated

carioca common beans based on molecular marker analysis. Genetics and molecular biology. 34(1):88-102.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one. 7(2):e32253.

Project IRGS. 2005. The map-based sequence of the rice genome. Nature. 436(7052):793-800.

Qin H, Chen M, Yi X, Bie S, Zhang C, Zhang Y, Lan J, Meng Y, Yuan Y, Jiao C. 2015. Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections. PLoS One. 10(1):e0118073.

Rafalski A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current opinion in plant biology. 5(2):94-100.

Ragimekula N, Varadarajula NN, Mallapuram SP, Gangimeni G, Reddy RK, Kondreddy HR. 2013. Marker assisted selection in disease resistance breeding. Journal of Plant Breeding and Genetics. 1(2):90-109.

Rahman L, Khanam MS, Koh H-J. 2008. QTL analysis for yield related traits using populations derived from an indica-japonica hybrid in rice (Oryza sativa L.). Czech Journal of Genetics and Plant Breeding-UZPI (Czech Republic).

Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. Nature Reviews Genetics. 12(10):683-691.

Shi M. 1985. The discovery and study of the photosensitive recessive male-sterile rice (Oryza sativa L. subsp. japonica). Scientia Agricultura Sinica (China).

Skuza L, Filip E, Szućko I. 2013. Use of organelle markers to study genetic diversity in soybean.

Srivastava SK, Wolinski P, Pereira A. 2014. A strategy for genome-wide identification of gene based polymorphisms in rice reveals non-synonymous variation and functional genotypic markers. PLoS One. 9(9):e105335.

Subbaiyan GK, Waters DL, Katiyar SK, Sadananda AR, Vaddadi S, Henry RJ. 2012. Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. Plant biotechnology journal. 10(6):623-634.

Takano S, Matsuda S, Kinoshita N, Shimoda N, Sato T, Kato K. 2014. Genome-wide single nucleotide polymorphisms and insertion–deletions of Oryza sativa L. subsp. japonica

cultivars grown near the northern limit of rice cultivation. Molecular Breeding. 34(3):1007-1021.

Tang W, Wu T, Ye J, Sun J, Jiang Y, Yu J, Tang J, Chen G, Wang C, Wan J. 2016. SNP-based analysis of genetic diversity reveals important alleles associated with seed size in rice. BMC Plant Biology. 16(1):1.

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung C-W, Reynolds A, Scheffler B, Eizenga G, McClung A. 2012. High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. Molecular Breeding. 29(4):875-886.

Trick M, Adamski NM, Mugford SG, Jiang C-C, Febrer M, Uauy C. 2012. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biology. 12(1):1.

Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature methods. 5(3):247-252.

Wiedmann RT, Smith TP, Nonneman DJ. 2008. SNP discovery in swine by reduced representation and high throughput pyrosequencing. BMC genetics. 9(1):1.

Wu K-S, Tanksley SD. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. Molecular and General Genetics MGG. 241(1-2):225-235.

Yang GP, Maroof MAS, Xu CG, Zhang Q, Biyashev RM. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Molecular & General Genetics. 245(2):187-194.

You FM, Huo N, Deal KR, Gu YQ, Luo M-C, McGuire PE, Dvorak J, Anderson OD. 2011. Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. BMC genomics. 12(1):1.

Zeid MM. 2003. Analysis of genetic diversity based on molecular markers (AFLP) and of heterosis in faba bean (Vicia faba L.). Cuvillier.