Advances in Bioscience and Bioengineering
2019; 7(1): 1-7
http://www.sciencepublishinggroup.com/j/abb
doi: 10.11648/j.abb.20190701.11
ISSN: 2330-4154 (Print); ISSN: 2330-4162 (Online)

Review Article

Genotype by Sequencing Method and Its Application for Crop Improvement (A Review)

Kalkidan Shiferaw Abebe

Department of Plant Sciences, Gambella University, Gambella, Ethiopia

Email address:
kalkidanshif@gmail.com

To cite this article:
Kalkidan Shiferaw Abebe. Genotype by Sequencing Method and Its Application for Crop Improvement (A Review). Advances in Bioscience and Bioengineering. Vol. 7, No. 1, 2019, pp. 1-7. doi: 10.11648/j.abb.20190701.11

Received: April 12, 2019; **Accepted**: June 3, 2019; **Published**: June 19, 2019

Abstract: Genotype by sequencing (GBS) is a next generation sequencing based method that takes advantage of reduced representation to enable genotyping of large numbers of individuals at a large number of SNP markers. It is relatively straightforward, robust, and cost-effective method to reduce problems in crop caused by a large genome size, reduced representation libraries are produced using a restriction enzyme that targets genomic regions while multiplexing with barcodes reduces the cost for individual sample. Several types of molecular markers, such as single nucleotide polymorphism (SNP), have been identified and effectively used in plant breeding. The application of next-generation sequencing (NGS) technologies has led to remarkable advances in whole genome sequencing, which provides ultra- throughput sequences to revolutionize plant genotyping and breeding. The GBS approach includes the digestion of genomic DNA with restriction enzymes followed by the ligation of barcode adapter, PCR amplification and sequencing of the amplified DNA pool on a single lane of flow cells. GBS has been successfully used in implementing genome-wide association study (GWAS), genomic diversity study, QTL mapping, genetic linkage analysis, molecular marker discovery and genomic selection under a large scale of plant breeding programs. GBS will have broad application in genomics-assisted plant breeding programs.

Keywords: Single Nucleotide Polymorphism (SNP), Next Generation Sequencing (NGS), Reduced Representation Library (RRL), Marker Assisted Selection (MAS)

1. Introduction

Conventional plant breeding requires a significant amount of time for the selection and evaluation of desirable traits over many generations. Marker-assisted selection (MAS) provides a more accurate and faster approach to select the desired phenotypes in a breeding population [1].

Plant molecular breeding has advanced so rapidly that several types of molecular markers have been developed and used for decades. The restriction fragment length polymorphism (RFLP) was firstly applied as DNA markers in plant genotyping [29]. RFLP technique is useful in the construction of genetic linkage maps, but it is challenged by the complicated hybridization, radio activity, and time consuming and limited by the number of available probes [3]. With further advance of biotechnology, several types of PCR-based markers were developed and used in plant breeding programs. These PCR-based markers mainly include random amplification of polymorphic DNA (RAPD) [4], sequence characterized amplified region (SCAR) [16], cleaved amplified polymorphic sequences (CAPS) [24], simple sequence repeats (SSRs) [19, 25], amplified fragment length polymorphisms (AFLPs) [22] and direct amplification of length polymorphisms (DALP) [22].

Advances in next-generation sequencing (NGS) technologies have taken the implementation of SNPs for genetic analysis to a new level. Genotype-by-sequencing (GBS) has provided new opportunities for breeders with cost-effective, genome-wide scanning, and multiplexed sequencing platforms. In principle, GBS can simultaneously perform SNP discovery and genotyping, which is particularly advantageous for under studied species that lack reference genome sequences [18].

NGS technologies have been game changers for genomic and genetic studies owing to their reduced cost of nucleotide

sequencing. Whole genome sequence data can be an ideal tool to call SNP variants and was used to genotype SNPs for crop species with a reference genome but low coverage and high sequencing cost per useful data point have been the biggest obstacle for this approach. Because it is not necessary or practical to generate deep whole-genome sequence data solely for genotyping purposes, several approaches have been applied to decrease the cost of NGS to a level where it can be implemented for SNP genotyping. Recently, the reduced representation library (RRL) has been widely adopted as an efficient approach to genotyping using NGS technologies [18].

Genotyping-by-sequencing (GBS), a method to identify genetic variants and quickly genotype samples, reduces genome complexity by using restriction enzymes to divide the genome into fragments whose ends are sequenced on short read sequencing platforms. While cost-effective, this method produces extensive missing data and requires complex bioinformatics analysis [21].

2. General Feature of GBS Methods

GBS was first coined by Rob Elshire and his colleagues as a simple highly multiplexed system for constructing reduced representation libraries for the Illumina NGS platform.

The GBS method incorporates a multiplex sequencing strategy for constructing reduced representative libraries for the Illumina NGS platform that uses an inexpensive barcoding system for increased efficiency at a lower cost compared to other method [22] GBS greatly reduces complexity by using enzymes to cleave the DNA coupled with DNA barcoded adapters) with small amounts of starting DNA (100-200 ng). Fractionated genomic DNA via restriction digestion reduces representation through size selection or the specific combinations of restriction enzymes (e.g. frequent-rare, rare-rare, or frequent-frequent cutters) for further targeting specific genomic regions of interest. The sequenced portion of the genome is highly consistent within a population because restriction sites are generally conserved across species [1].

GBS is simple, specific, highly reproducible, and rapid due to the simultaneous detection of SNPs and genotyping. Thus, the key components of this system have a lower cost, reduced sample handling, fewer PCR and purification steps, no size fractionation, no reference sequence limits, and efficient barcoding, and the system is easy to scale up [27]. These features make GBS a powerful tool for a number of plant genetic studies.

3. Protocols of GBS Method

3.1. DNA Sample Preparation

High quality genomic DNA (free of contamination either with RNA or with DNA from other species) is crucial to the success of a GBS protocols given that varying efficiency of digestion, ligation and amplification can have significant effects on the final marker set. Most importantly, the quantity of DNA from different samples should be evenly balanced before pooling to avoid losing markers from some individuals owing to lack of coverage. The choice of method may also be influenced by the amount of genomic DNA starting material required [27].

3.2. Choosing Restriction Enzyme

Selection of REs that leave 2 to 3 bp overhangs and do not cut frequently in the major repetitive fraction of the genome is of critical importance. A suitable RE for many crops is *Ape*KI a type II restriction endonuclease that recognizes a degenerate 5 bp sequence creates a 5′ overhang (3 bp), has relatively few recognition sites in the Using an RE that leaves an overhang comprising more than one nucleotide is extremely useful in promoting efficient adapter ligation to insert DNA [22].

3.3. Adapter Design

Two different types of adapters were used in GBS protocol. The "barcode" adapter terminates with a 4 to 8 bp (base pair) barcode on the end of its top strand and a 3 bp overhang on the end of its bottom strand that is complementary to the "sticky" end generated by ApeKI.. Compatible set of 96 barcode sequences that have been used for multiplex sequencing is provided as supporting information. To minimize the possibility of misidentifying samples as a result of sequencing or adapter synthesis error, all pair-wise combinations of barcodes differed by a minimum of three mutational steps. Hence, it should be possible to correctly assign samples with single base barcode sequencing errors, or to identify particular adapters with high rates of synthesis error. To avoid the potential loss of sequence quality due to phasing errors caused by reading through a non-variable restriction site prior to the twelfth base, or through an adapter position with a highly skewed base ratio [22].

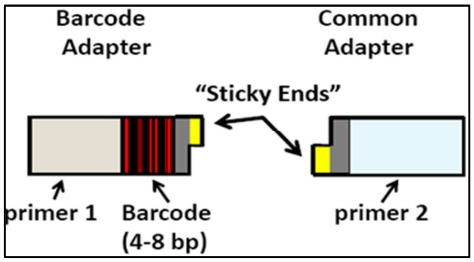

Figure 1. *Adapter design.*

3.4. Ligation

The ligation is completed in the same tube/plate as the digestion. The ligation reaction is conducted in NEB Buffer4 with the addition of ATP (NEB T4 DNA Ligase #M0202). A very High concentration of T4 is used to ensure adequate ligation of all the fragments. The concentration of Adapter1 needs to be adjusted depending On the species. For wheat and Barley 0.1 pmol is close to the optimal amount for 200ng of genomic DNA. The Adapter2 is A Y-adapter and can be added in excess, as it will not amplify unless the PCR

reaction has first proceeded from Adapter1 on the other end of the same fragment. The ligase should be inactivated prior to multiplexing the samples by holding at 65C for 20min after the ligation is complete [16].

3.5. GBS Library Construction

Up to 96 DNA samples can be processed simultaneously. (1) DNA samples, barcode, and common adapter pairs are plated and dried; (2-3) samples are then digested with *Ape*KI and adapters are ligated to the ends of genomic DNA fragments; (4) T4 ligase is inactivated by heating and an aliquot of each sample is pooled and applied to a size exclusion column to remove unreacted adapters; (5) appropriate primers with binding sites on the ligated adapters are added and PCR is performed to increase the fragment pool; (6-7) PCR products are cleaned up and fragment sizes of the resulting library are checked on a DNA analyzer. Libraries without adapter dimers are retained for DNA sequencing [22].

3.6. Multiplexing and PCR Amplification

The ligated samples are multiplex and PCR amplified in a single tube. This produces a single library From 96 sample, which is sequenced on a single lane of Illumina HiSeq [16]. The multiplex library is PCR amplified using a short extension time. This will enrich for fragments that are in the 200---500bp range and suitable for bridge amplification. Only fragments that have a *PstI* cut site and an *MspI* (or second enzyme) cut site will amplify. The *MspI-MspI fragments* will be common but will not amplify due to the Y adapter.

GBS 96-Plex Protocol

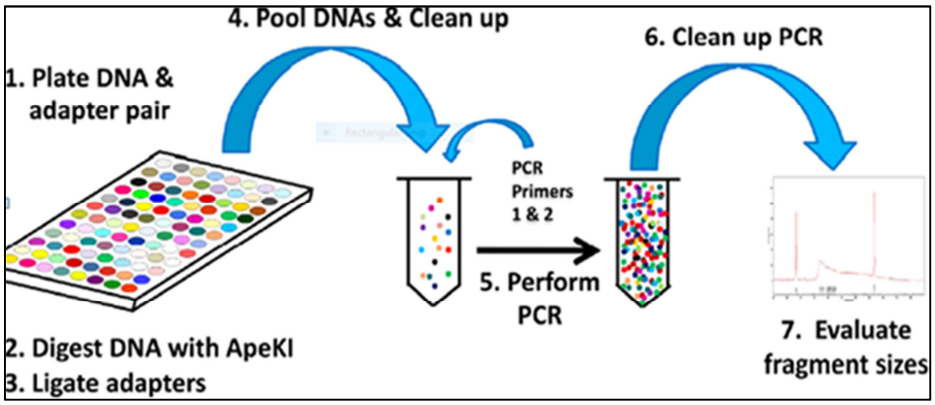

Figure 2. *Steps in GBS library construction.*

3.7. Sequencing on Miseq

The GBS protocol uses the Miseq "Generate FASTQ" workflow, the "FASTQ Only" application and "TruSeq HT" assay to generate a de-multiplexed set of FASTQ files with the adapter sequences removed upon completion of the sequencing run. The freshly denatured and diluted library containing PhiX is loaded onto a Miseq Reagent Kit v3 600-cycle cartridge. The run is initiated and monitored according to the protocol outlined by Illumina for the Miseq instrument. A Miseq run typically lasts up to 48 h, and the run data, including the FASTQ files, are downloaded. Each sample has two FASTQ files representing the forward and reverse sequencing reads labelled with the respective terms "R1" and "R2*"* [17].

4. Application of GBS in Plant Breeding

Genotyping by sequencing is an ideal platform for studies ranging from single gene markers to whole genome profiling. GBS is one of the most powerful applications in the field of plant breeding. GBS provides a rapid and low-cost tool to genotype breeding populations, allowing plant breeders to implement GWAS, genomic diversity study, genetic linkage analysis, molecular marker discovery and genomic selection (GS) under a large scale of plant breeding programs. There is no requirement for a prior knowledge of the species genomes as the GBS method has been shown to be robust across a range of species and SNP discovery and genotyping are completed together [18].

Identification of high density SNP markers through GBS to construct genetic linkage maps has a great value for numerous applications in plant breeding [18]. In Arabidopsis sequenced the whole genome shotgun sequencing on the Illumina platform [19]. Mapping population consisting of 176 rice recombinant inbred lines and mapped the recombined hot and cold spots and quantitative trait loci (QTLs) for leaf width and aluminum tolerance. After the efficiency of multiplexed SNP genotyping for diversity, mapping and breeding applications were evaluated, [28] demonstrated that 384 plex SNP genotyping on the Bead Xpress platform is a robust and efficient method for marker genotyping and mapping in rice [12, 13, 15]. GBS was applied to bread wheat, resulting in the incorporation of 1000s of markers in the bread wheat map [25]. The high resolution of SNP markers were identified in barley and the GBS mapping data were used to confirm that the semi-dwarfing gene (ari-e) is located on barley chromosome 5H [28]. Construction of a GBS linkage map using the sequence-based markers leads to the RAD technique [29], which has been used in barley QTL analysis [24].

Orphan plant species without a known genomic sequence represent the vast majority of crops over the world. The GBS

protocol for wheat and barley and subsequent genetic analyses [25] were carried out when a draft genomic sequence was not available yet. An available reference genome can simplify the data analyses, but it is not essential in GBS, indicating a great advantage of the GBS technique in accelerating plant breeding and crop improvement. This reality has been confirmed with the recent GBS applications on different oat accessions. The depth of genomic sequencing is important to identify stable and representative SNPs which can be generated to improve crop genotypes [22].

GBS approaches helps to rapid discovery of sequence-based molecular markers are used to construct a genetic map without a reference genome [18]. High density genetic maps integrated with an extremely large number of markers discovered through GBS can serve as a reference genome for anchoring and ordering physical maps and refining or correcting unordered sequence contigs [25]. If a reference genome is available, markers can be ordered along the physical map without calculating recombination frequencies for linkage maps. Even at low coverage, sparsely placed markers can be used to narrow points of recombination from 100 to 200 kb intervals [8, 13].

By using GBS genetic linkage maps show the relative distances between markers along the chromosomes as determined by their recombination frequency. Such maps are important in breeding programmes as they facilitate QTL and association analysis. These analyses are powerful tools to identify genetic loci governing traits of interest using the principle of genetic.

Two genotyped BC4 F1 populations by GBS to identify introgression from donor parent in each line. The two populations respectively involved inbred strains of cabbage and cauliflower (Orange) as donor parents and an inbred rapid cycling line, to 1434 as recurrent parent. In the cauliflower BC4 F1 population, 89 lines were genotyped by 693 SNP markers, identifying a total of 164 introgressed segments, with zero to seven introgressed segments (averaging 2.4) per line. In cabbage BC4 F1 population75 lines were genotyped by 641 SNP markers, finding 162 introgressed segments, with zero to seven (averaging 2.6) per line. The respective average of the two populations, 2.4 and 2.6 segments, closely approximate the a priori expectation of 2.5 introgressed segments per line in the BC4F1 generation [18].

A genetic map of pearl millet using GBS generated high-quality SNPs to construct a genetic map with an average interval of 2.1 (± 0.6) cM between the SNP markers. Their study demonstrated that GBS can quickly produce a denser and more uniform genetic map than previously published maps. This type of map will be useful for the identification of genomic regions associated with important agronomic traits by using GBS 254 lines from the Cycle 29 SAWSN [1].

In wheat breeding panel they identified a set of 41,371 SNPs that were at an allele frequency greater than 1% and had more than 20% data present. Removing multiple SNPs in the same tag reduced the marker number to 34,749 SNPs that were used for subsequent analysis. As is typical of sequence based genotyping at low coverage, many markers had a large proportion of missing data. There was limited power to confirm low-frequency alleles in the presence of sequencing errors. This was evidenced by a decrease in the number of identified SNPs with minor allele frequency below 5% [25].

Genotyping by sequencing (GBS) was used to construct the high-density linkage map. The map contained 3,641 markers distributed on 21 chromosomes and spanned 1,959 cM with an average distance of 1.8 cM between markers. The constructed linkage map revealed strong collinearity in marker order across 21 chromosomes which were based on a high-density linkage map. The reliability of the linkage map for QTL mapping was demonstrated by co-localizing the genes to previously mapped genomic regions for two highly heritable traits, chaff color, and leaf cuticular wax [27].

By using GBS mapping data and phenotypic measurements they show that ari-e. GP maps to a small genetic interval on chromosome 5H and that alternative alleles at a region encompassing Vrs1 on 2H along with a region on chromosome 3H also influence plant height. The location of Ari-e is supported by analysis of near-isogenic lines containing different ari-e alleles. They explored use of the GBS to populate the region with sequence contigs from the recently released physically and genetically integrated barley genome sequence assembly as a step toward Ari-e gene identification [28].

4.1. Genome Wide Association Studies

Genome wide association studies (GWAS) use ancestral recombination events to identify the genetic loci underlying traits at high resolution. By employing association panels consisting of diverse genotypes, GWAS is able to pinpoint candidate genes precisely when linkage disequilibrium is relatively low, overcoming the limitations of less exact methods such as QTL mapping. Although commercial SNP arrays have been widely used for GWAS in crops such as rice, maize and soybean [36, 37, 44], GBS methods are increasingly contributing data for GWAS. This is advantageous because GBS produces raw sequence reads, which can be reused more easily by other researchers.

In the potential energy crop *Miscanthus sinensis*, more than 100,000 SNPs were identified using RRS, which were used for a GWAS to detect associations between genetic variants and phenotypic traits such as cell wall composition, biomass and plant height [42]. Using Elshire GBS, 14 loci were identified in sorghum for the inflorescence branch length trait [40], and in soybean, loci associated with resistance to fungal stem rot and oil and protein content could be detected with similar methods [34, 43].

4.2. Linkage and QTL Mapping

Genetic linkage maps show the relative distances between markers along the chromosomes as determined by their recombination frequency. Such maps are important in breeding programmes as they facilitate QTL and association analysis. These analyses are powerful tools to identify genetic

loci governing traits of interest using the principle of genetic linkage [35, 39].

4.3. Genomic Selection

Genomic selection (GS) has emerged as a robust approach that is directly benefitting the plant breeders. GS comprises the method that uses dense, genome wide molecular markers to predict the GEBV (Genomic Estimated Breeding Value) of individuals and perform selection on individuals based on GEBV without taking them out in field [38]. GS provides the ability to select complex quantitative traits based on marker data alone and combines the benefit of high throughput technologies like GBS and developments in the statistical methods needed for data analysis. GS can greatly accelerate the breeding cycle while also using marker information to maintain genetic diversity and potentially prolong gain beyond what is possible with phenotypic selection [41]. The accuracy of genomic prediction using GBS is currently undergoing investigation in several important crops including maize and wheat.

5. Potential Limitation of GBS Method

Genotype by sequencing method is complexity reduction method that has been based on restriction enzyme digestion or transcriptome sequencing. One of the most popular and widely used methods is the GBS-RAD (restriction site-associated DNA) method of Elshire *et al*. [45], which uses a restriction endonuclease to digest DNA samples before library preparation. Variations of the GBS-RAD method have been extensively reviewed by Davey *et al*. (2011) [46]. Notably, this method suffers from large amounts of missing data and, due to generation of a significant number of dominant markers, is unsuitable for outbreeding species as it is unable to unambiguously identify heterozygous sites.

Genotyping-by-sequencing (GBS) potentially offers a cost-effective alternative for SNP discovery and genotyping. Both *Ape*KI and *Pst*I/*Msp*I enzymes are used for library preparation. *Ape*KI yielded more markers than *Pst*I/*Msp*I but provided a lower read coverage per marker, resulting in more missing data and limiting effective genotyping to the tetraploid mode.

6. Conclusion

Genotype by sequencing (GBS) is a next generation sequencing based method that takes advantage of reduced representation to enable genotyping of large numbers of individuals at a large number of SNP markers. The low cost of GBS makes it an attractive approach to saturate the mapping and breeding populations with a high density of SNP markers. GBS has become a cost effective alternative to other whole genome genotyping platforms. It can be anticipated that high density of SNP markers from NGS will be extensively applied to GWAS and MAS. It is one of the most powerful methods for crop improvement. GBS has been successfully used in implementing genome-wide association study (GWAS), genomic diversity study, QTL mapping, genetic linkage analysis, molecular marker discovery and genomic selection under a large scale of plant breeding programs. Future applications of GBS to crop improvement may allow plant breeders to conduct MAS on a novel germplasm or species without first having to develop any prior molecular tools. GBS will stand to be one of the major components of breeding for crop improvement.

References

[1] Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., & Jiang, R. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature, 465 (7298), 627.

[2] Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoSONE 3: e3376. doi: 10.1371/journal.pone.0003376.

[3] Bernatsky, R., and Tanksley, S. D. (1986). Toward a saturated linkage map in tomato based on iso zymes and random cDNA sequences. Genetics 112, 887-898.

[4] Botstein, D., White, R. L. Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314-331.

[5] Brown, P. (2011). Genotyping-by sequencing Protocol PstI-MspI Chung, Y. S., Choi, S. C., Jun, T. H., & Kim, C. (2017). Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. Horticulture, Environment, and Biotechnology, 58 (5), 425-431.

[6] Chutimanitsakun, Y., Nipper, R. W., Cuesta-Marcos, A., Cistue, L., Corey, A., Filichkina, T., et al. (2011). Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. BMC Genomics 12: 4. doi: 10.1186/1471-2164-12-4.

[7] Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12 (7), 499.

[8] Deschamps, S., L laca, V. and May G. D. (2012). Genotyping-by-sequencing in plants. Biology 1, 460-483. doi: 10.3390/biology1030460

[9] Desmarais, D., Zhong, Y., Chakraborty, R., Perreault, C. and Busque, L., 1998. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). Journal of Forensic Science, 43 (5), pp. 1046-1049.

[10] Elshire, R. J., Glaubitz, J. C., Sun, Q., Pol, J. A., Kawamoto, K., & Buckler, E. S. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6: 19379.

[11] He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in plant science, 5, 484.

[12] Heffner, E. L., Sorrells, M. E., andJannink, J. L. (2009). Genomic selection for crop improvement. Crop Sci. 49, 1-12. doi: 10.2135/cropsci2008.08.0512.

[13] Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009) High through put genotyping by whole genome resequencing. Genome Res. 19, 1068- 1076. doi: 10.1101/gr.089516.108.

[14] Hussain, W., Baenziger, P. S., Belamkar, V., Guttieri, M. J., Venegas, J. P., Easterly, A.,... & Poland, J. (2017). Genotyping-by-Sequencing Derived High-Density Linkage Map and its Application to QTL Mapping of Flag Leaf Traits in Bread Wheat. Scientific reports, 7 (1), 16394.

[15] Jannink, J. L., Lorenz, A. J., andIwata, H. (2010). Genomicselectioninplant breeding from theory to practice. Brief. Funct. Genomics 9, 166-177. doi: 10.1093/bfgp/elq001.

[16] Jiang, Z., Wang, H., Michal, J. J., Zhou, X., Liu, B., Woods, L. C. S., & Fuchs, R. A. (2016). Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development. International journal of biological sciences, 12 (1), 100.

[17] Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L. S., & Paterson, A. H. (2016). Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Science, 242, 14-22.

[18] Konieczny, A., and Ausubel, F. M. (1993 A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4, 403-410. doi: 10.1046/j.1365-313X.1993.04020403.

[19] Litt, M., and Luty, J. A. (1986). A hypervariable microsatellite revealed by invitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44, 397-401.

[20] Liu, H., Bayer, M., Druka, A., Russell, J. R., Hackett, C. A., Poland. J. et al. (2014). An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC Genomics 15: 104. doi: 10.1186/1471-2164- 15-104.

[21] Paran, I., and Michelm ore, R. W. (1993). Development of reliable PCR- based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85, 985-993. doi: 10.1007/BF00215038.

[22] Peterson, G. W., Dong, Y., Horbach, C., & Fu, Y. B. (2014). Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity, 6 (4), 665-680.

[23] Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y.... & Jannink, J. L. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome, 5 (3), 103-113.

[24] Poland, J. A., and Rife T. W (2012), Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5, 92-102. doi: 10.3835/plant genome2012.05.0005

[25] Salimath S. S., de Oliveira, A. C. Bennetzen, J. and Godwin, I. D. (1995). Assessment of genomic origin and genetic diversity in the genus Eleusine with DNA markers. Genome 38, 757-763. doi: 10.1139/g95-096.

[26] Scheben, A., Batley, J. and Edwards, D., 2017. Genotyping - by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant biotechnology journal, 15 (2), pp. 149-161.

[27] Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H. Nielsen, K. L., et al. (2009). SHORE map: simultaneous mapping and mutation identification by deep.

[28] Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B. & Belzile, F. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PloS one, 8 (1), 54603.

[29] Thomson, M., Zhao, K., Wright, M., McNally, K., Rey, J., Tung, C. W. et al. (2012). High through put single nucleotide polymorphism genotyping for breeding applications in rice using the Bead Xpress platform. Mol. Breed. 29, 875-886. doi: 10.1007/s11032-011-9663-x.

[30] Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. V. D., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. and Zabeau, M., 1995. AFLP: a new technique for DNA fingerprinting. Nucleic acids research, 23 (21), pp. 4407-4414.

[31] Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic acids research, 18 (22), pp. 6531-6535.

[32] Wickland, Daniel P., et al (2017), a comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GBS BMC bioinformatics, 2017

[33] Yong, L., Thet, Z. and Zhu, Y., 2017. Genetic editing of the androgen receptor contributes to impaired male courtship behavior in zebrafish. Journal of Experimental Biology, 220 (17), pp. 3017-3021.

[34] Bastien M., Sonah H. and Belzile F. (2014) Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. Plant Genome, 7, 1-13.

[35] Collard B. C. Y., Jahufer M. Z. Z., Brouwer J. B. and Pang E. C. K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica, 142, 169-196.

[36] Hwang E. Y., Song Q., Jia G., Specht J. E., Hyten D. L., Costa J. and Cregan P. B. (2014) A genome-wide association study of seed protein and oil content in soybean. BMC Genom. 15, 1.

[37] Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

[38] Meuwissen THT, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps.Genetics 157: 1819-1829

[39] Mohan M., Nair S., Bhagwat A., Krishna T. G., Yano M., Bhatia C. R. and Sasaki T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed. 3, 87-103.

[40] Morris G. P., Ramu P., Deshpande S. P., Hash C. T., Shah T., Upadhyaya H. D., Riera-Lizarazu O. *et al* (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. USA, 110, 453-458.

[41] Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink JL (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 10: 77-120.

[42] Slavov G. T., Nipper R., Robson P., Farrar K., Allison G. G., Bosch M., Clifton-Brown J. C. *et al* (2014) Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. New Phytol. 201, 1227-1239.

[43] Sonah H., Bastien M., Iquira E., Tardivel A., Legare G., Boyle B., Normandeau E. *et al* (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE, 8, e5460.

[44] Zhao K., Tung C. W., Eizenga G. C., Wright M. H., Ali M. L., Price A. H., Norton G. J. *et al* (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat. Commun. 2, 467.

[45] Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6* (5), p. e19379.

[46] Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. and Blaxter, M. L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12* (7), p. 499.