Advances in Biotechnology & Microbiology
ISSN 2474-7637

Juniper Publishers
key to the Researchers

Review Article
Volume 14 Issue 4 - August 2019
DOI: 10.19080/AIBM.2019.14.555891

Adv Biotechnol Microbiol

Genotyping by Sequencing for Plant Breeding- A Review

Essubalew Getachew Seyum[1]*, Ngalle Hermine Bille[1], Joseph Martin Bell[2] and Wosene Gebreselassie[2]

[1,2]*Department of Plant Biology and Physiology, University of Yaoundé I, Cameroon*

[2]*Department of Horticulture and Plant Sciences, Jimma University College of Agriculture and Veterinary Medicine, Ethiopia*

Submission: June 08, 2019; **Published:** August 30, 2019

***Corresponding author:** Seyum Essubalew Getachew, Department of Plant Biology and Physiology, Faculty of Sciences, University of Yaoundé I, Yaoundé, Cameroon

Department of Horticulture and Plant Sciences, Jimma University College of Agriculture and Veterinary Medicine, Jimma, Ethiopia

Abstract

Molecular plant breeding using DNA marker, or Marker-Assisted Selection (MAS), plays a pivotal role in a breeding program of crops to release a new variety within a short period of time compared to conventional method of plant breeding. Different types of marker have been used for breeding of plants and currently SNPs (single nucleotide polymorphisms) have become a reference type of DNA markers for plant breeding. Food production decline due to climate change, population growth, polypoid level and others result in a different level of DNA sequencing. Discovery of SNP without getting previous information about sequencing genome is called Genotyping by Sequencing (GBS). It is rapid, cost-effective and high throughput approach in next-generation sequencing. It is a new approach for implementing molecular tools in plant breeding. This paper briefly reviews the current status of Genotyping by Sequencing (GBS) and its application in plant breeding. It is efficiently applied in a wide range of plant breeding programs such as Genomic selection (GS), Genomic Diversity (GD), Genome-Wide Association (GWA), Linkage Analysis (LA), Marker discovery. It combines discovery of molecular marker and genome genotyping. This method has been developed and applied sequencing of multiplexed genomic samples. World economically most important crops, GBS used as tool for plant breeder to select them for better yield and quality. In spite of the above facts this method also have some limitation in its application.

Keywords: Molecular plant breeding; Single nucleotide polymorphism

Abbreviations: GBS: Genotyping-by-Sequencing; MAS: Marker-Assisted Selection; LA: Linkage Analysis; GWA: Genome-Wide Association; GD: Genomic Diversity; GS: Genomic selection; MAS: Marker-Assisted Selection

Introduction

Increasing production and productivity of crops for food and feed with the changing climate is one of the key slogans in our world in the 21st century [1]. Nowadays, agricultural productivity is becoming lower down due to biotic and abiotic stresses [2]. In this century world population will grow form 7 billion to 12.3 billion [3]. Reduction in crop production and productivity due to water scarcity, decreasing area and land degradation due to environmental change, pollution, occurrence of new pathogens and pests, and change in climate have major impact in food security of the world [2]. Improving production and productivity of major food, feed, and industrial crops in parallel alleviating food security problem plant breeding remains the main driving force [4]To increase food production plant breeding will play a key role and breeders face an endless task in order to developing new crop varieties [5]. For this purpose, predicting population with the increasing climate change and considering both quantitative and qualitative traits, yield stability should be a major focus of plant breeding.

Breeding of crops can be accomplished through two major approaches i.e., conventional and molecular. Variety development through the former approach requires continuous hybridization between distinct parents and selection over several generations. Long time (5-12 years) to develop crop variety, genotype by environmental interaction, low efficiency for complex and low heritable traits are the major limitations of this approach [4,6,7]. Applications of molecular biology tools that used to improve (develop) new cultivar is known as molecular plant breeding [8]. Unlike conventional method, this method used in DNA marker for selection of a given trait. This method helps to increase the efficiency, speed and precision of plant breeding in which it reduced cost and time [7,9].

Marker-Assisted Selection (MAS) selection process based on DNA marker for a given trait. It is a new discipline in the area of molecular breeding [10]. It is the method applied without phenotypic information in some individuals. It was started to solve the gaps in crop improvement program through conventional method [9]. It is tremendously useful in plant breeding and genetics. It is precondition for various biological applications such as mapping and tagging genes, segregation analysis, genetic diagnosis study, phylogenetic study etc. [11,12]. Selection of a trait and to know its association with a trait of interest in a target plants this method use DNA marker. It is more efficient for a character controlled by few Quantitative Trait Loci (QTLs) having major effect on trait expression. In contrary, this method is inferior over conventional breeding method in which a character controlled by a complex quantitative character [13-15]. A newly introduced approach in marker-assisted selection is known as genomic selection. This method uses high density genetic markers covering the whole genome in all Quantitative Trait Loci (QTL) and a genome linked with at least one marker. [16]. GS is used to estimate the genetic makeup an individual based on large set of markers distributed across the whole genome and selection was undertaken based on the relationship between training and validation sets, unlike the former it is not based on few markers [17-19] Genotyping-by-Sequencing (GBS) is newly introduced method and widely used range of crop improvement program in which it is used for detecting SNPs using high-throughput sequencing [20]. It is a modified RAD-seq based library preparation protocol for NGS [21]. The most important feature of this methods are reduced sample handling and fewer PCR purification steps, low cost, no reference sequence limits, no size fractionation and efficient barcoding technique [4] GBS was developed as a tool for genomic association studies and marker-assisted breeding. It is mainly works for species with large complex genomes and inimitable tool for genomics-assisted breeding in a wide range of plant species [22]. Presently, this technology has been used for whole genome sequencing and re-sequencing schemes in which the genomes of several specimens are sequenced to discover large numbers of Single Nucleotide Polymorphisms (SNPs) to discovering within-species diversity, constructing haplotype maps (blocks) and performing genome-wide association studies. Based on the above listed major problems and feeding the fastly growing population along with the problems it is essential to study modern breeding techniques. In the other hand around 7.4 million accessions of the world most economically important crops have no any non-model species it needs genotype sequencing [23].Therefore, the objective of this paper is to review the role of genotyping by sequencing (GBS) in plant breeding and its application.

Molecular markers

Currently, molecular plant breeding has reached an advanced stage. For the last few decades different types of molecular markers have been used and develop [24]. The first DNA markers applied for plant genotyping were Restriction Fragment Length Polymorphism (RFLP) [25]. It is more suitable method in the construction of genetic linkage maps. Despite its numerous advantages this approach becoming less applicable due to complicated hybridization, radioactivity, time consuming and limited number of available probes. Molecular plant breeding development resulted in the establishment numerous types of PCR-based markers mainly used in different crop improvement and research programs [24]

These PCR- based markers include Random Amplification of Polymorphic DNA (RAPD), Cleaved Amplified Polymorphic Sequences (CAPS) [26], Amplified Fragment Length Polymorphisms (AFLPs) [27], Simple Sequence Repeats (SSRs) [28], Sequence Characterized Amplified region (SCAR) [29] and Direct Amplification of Length Polymorphisms (DALP) [30]. Unlike RFLP, all these methods are relatively inexpensive and requires short period of time to undertake amplification and genome sequencing of a given populations [31]. Among all PCR based markers, the most applicable ones were Simple Sequence Repeats (SSR) and it was relatively inexpensive, abundant in plant genomes and more informative than bi-allelic markers [32]. In the year 1990s new techniques were developed by [33], for a given model plant species this method combines genome and Expressed Sequence Tags (ESTs). Identification of variations at the single base pair the development of Sanger sequencing highly accelerated the identification process [32]. The most recent DNA markers developed is Single Nucleotide Polymorphisms (SNPs) [34]. Plant genotyping through this technique has increased the potential to score variation in specific DNA targets. In addition, compared others it has small missing marker and also increases information on potentially millions of genome wide marker and their surrounding sequences sets in which it is the foundation of high-throughput genotyping [7,31,32]. Over the past 10 years, as compared to the earlier genotyping approaches, SNP-based marker techniques increased marker density, reducing cost of genotyping and requires less time for SNP discoveries [31]. The most common system in fluorescent detection of SNP-specific hybridization probes on PCR products are Taqman, Molecular Beacons and Invader [35]. In line with this, SNP-specific PCR primer extension products uses in homogeneous Mass-Extend (hME) assay. However, its output are read on a MALDI-TOF mass spectrophotometer [36]. Application all this method results around 100–1000s of SNPs per day. The current interest results an increasing demand for higher throughput, end-point fluorescent assays such as Taqman and Invader have been significantly enhanced by the use of array tape technology in place of 96, 384 or 1,536-well microtiter plates. This method reduced cost per assay and increasing throughput in a format [32].

Currently, there is enormously parallel array system enabled parallel scoring of up to hundreds and thousands of markers in plants genome. Depending on the application, assay simplicity, cost, throughput and accuracy, these ultra-high throughput technologies are used in wide range of researches. All these systems follow a similar pattern for DNA template preparation. The two most widely used array-based systems in plants genomic

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891

are Golden Gate and Infinium assays and these arrays consist on multistep protocols based on Illumina's Bead Array/Bead Chip technology [37]. The former assay is allowing screening of many samples using a single multiplexed assay that can include as many as 3,072 SNPs. While, the latter assay provides considerably higher throughput, of up to four million SNPs from a single sample, or up to several hundred thousand on multiple samples in the same array. In Infinium, samples are incubated on bead chips where they anneal to locus-specific 50-mers covalently linked to beads. After hybridization, oligos are subject to allele-specific single-base extension; followed by fluorescent staining, signal amplification, scanning in a dual-color channel reader, and analysis. The use of pre-made arrays reduces cost considerably although the actual number of markers derived from this array will be considerably lower, depending on the relationship to the reference and gene representation in the interrogated plants. Beckman Coulter's Genome Lab SNP stream is another method which allows the processing of up to three million genotypes in 384 samples/day/instrument (Table 1). Affymetrix Gene Chip system is most widely used method and it is not only detect hundreds of thousands of SNPs in a single array but, it can also be used for SNP discovery by Sequencing by Hybridization (SbH) [7,15,32,].

Table 1. Evaluation of representative NGS technologies.

No.	Sequencing Platform	Sequencing Chemistry	Detection Chemistry	RunTimea	Read Length (bp)	Reads per Run (million)	Throughput per Run (Gbp)
1	Roche 454 FLX Titanium	Sequencing by Synthesis	Light	23 hours	~800	~1	~0.7
2	Illumina MiSeq	Sequencing by Synthesis	Fluorescence	39 hours	2 × 250 b	~1	~8
3	Illumina HiSeq2500	Sequencing by Synthesis	Fluorescence	11 days (high output)/27 hours (rapid run)	2 × 100 b	~3,000	~600 (highoutput)/~120 (rapid run)
4	Life Technologies 5500xl	Sequencing by Ligation	Fluorescence	8 days	75 + 35 b	~5,000	~310
5	Ion Torrent PGM	Sequencing by Synthesis	pH	4 hours	100	1	~0.1

a Not including library construction; b Paired end read sequencing.

Genotyping-by-sequencing (GBS)

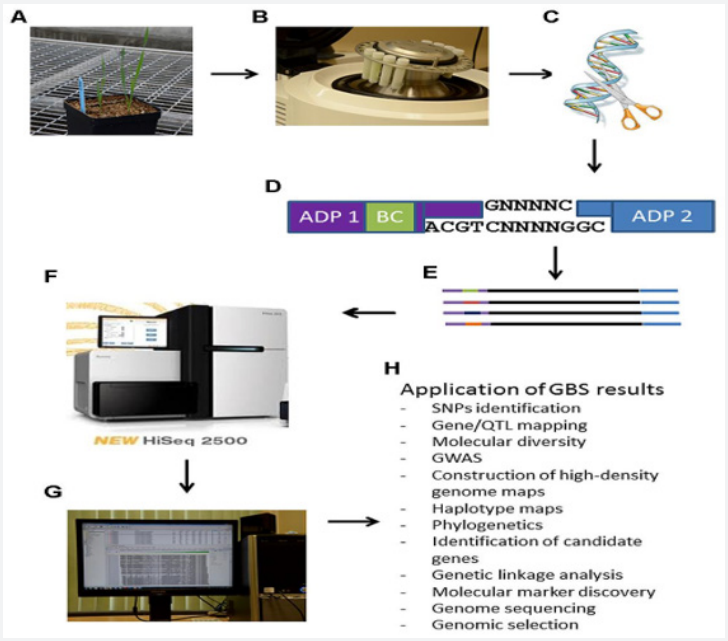

Figure 1: Major steps of genotyping-by-sequencing (GBS) protocol used in plant breeding. Step **(a).** Tissue is obtained from any plant species; Step **(b).** Ground leaf tissues for DNA isolation, quantification and normalization. N.B: Take care of any cross-contamination among samples at this step; Step **(c).** DNA digestion with restriction enzymes; Step **(d).** Ligations of adaptors **(e).** including a bar coding (BC) region in adapter 1 in random PstI-MseI restricted DNA fragments; Step Representation of different amplified **(f).** DNA fragments with different bar codes from different biological samples/lines. N.B: These fragments represent the GSB library; **(g).** Analysis of sequences from library on a NGS sequencer; Step Bioinformatic analysis of NGS sequencing data; Step (H): Application of GBS results in breeding.

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891

Genotyping by Sequencing (GBS) is the discovery of SNPs without prior knowledge about the genome sequence [20]. Nowadays, the advancement of NGS issue is the cost of DNA sequencing reduction to this end GBS is now feasible for large and complex genome species [21]. A thousand millions of SNPs can be detected in the large size lines that can be used for GWAS, GS, gd-study, linkage mapping, evolutionary studies and conservation and ecological genomics study [4,20,38]. It combines both discovery and genotyping of large populations genome applied in plant breeding even in the absence of a reference genome sequence. Its importance dramatically increases due to it's cost-effective and unique tool for genomics-assisted breeding in a wide range of plant species [38, 39]. It is amenable to use on large numbers of individuals/lines due to library production procedure [4, 32]. Application of GBS technology in any plant species are summarized in Figure 1.

Genetic linkage map construction in a given test lines/individuals GBS is more efficient and simpler in line with it combines with genome-independent imputation [21,40]. Originally the system used Ape KI protocol. Currently, modified to a two-enzyme namely PstI and MspI protocol, which reduced genome complexity and uniform library for sequencing than the original protocol [39]. Now a days, GBS is applicable for different world most important economical food crops [41]. It is increase both SNPs call number and depth, allow an important reduction in per sample cost [4, 32].

Presently, it is an efficient approach for plant genotyping in NGS technologies is Reduction of Representation Library (RRL) [20]. The main component through this approach is cutting the entire genome with specific restriction enzyme(s) that reduce genome complexity for the organism of interest. Its results sequence dataset which can provide higher read coverage per locus while allowing higher level of multiplexing with uniquely bar-coded adapters for different samples [39]. The main limitation regrading RRL is that the important genomic regions may not be captured by GBS libraries when restriction sites are not available surrounding those regions. To overcome this problem, it is advisable to use multiple GBS libraries with different combinations of enzyme. Data depicted in Table 2 showed that different methods of GBS with their specific features for technical comparisons [6]. Different researches have been conducted in GBS for species with reference genomes and because of reference genome is available SNP genotyping becomes much easier than the other. Source.

Table 2: Representative GBS protocols published in peer-reviewed journals.

Method	Restriction enzyme	Insert size	Barcodes	Sequencing platform	Sequencing mode	Reference
RAD-seq (Restriction association DNA sequencing)	SbfI or EcoRI	Size-selection	~96	Illumina	Paired end	[42]
MSG (Multiplex shotgun genotyping) GBS (Genotype by sequencing)	MseI	Size-selection	~384	Illumina	Single end	[43]
	ApeKI	<350 bp	~384	Illumina	Paired end	[57]
Double-digested RAD-seq	EcoRI and MspI	Size-selection	~48	Illumina	Paired end	[44]
Double-digested GBS	PstI and MspI	<350 bp	~384	Illumina	Paired end	[22]
Ion Torrent GBS	PstI and MspI	<350 bp	~384	Ion Torrent	Paired end	[53]
SBG (Sequence-based genotyping)	EcoRI and MseI PstI and MseI	Size-selection	~32	Illumina	Paired-end	[46]
REST-seq (Restriction fragment sequencing)	TaqI and TruI	Size-selection	~305	Ion Torrent	Paired-end	[55]
Restriction enzyme sequence comparative analysis	MseI or NlaIII	Size-selection	~96	Illumina	Paired-end	[54]

In a GBS there are two different strategies which have been developed with the Ion PGM system for NGS [22]. Restriction enzyme digestion, in which no specific SNPs have been identified and ideal for discovering new markers for MAS programs. Multiplex enrichment PCR, in which a set of SNPs has been defined for a section of the genome. The first strategy works for all complex genome, which reduced its complexity by digesting the DNA with one or two selected restriction enzymes prior to the ligation of the adapters. The second approach designed to amplify the areas of interest by using PCR primers [40,42]. demonstrated that the first restriction site associated DNA sequencing or DNA (RAD) for high density SNP discovery and genotyping. It is a sequence-based marker and used to reduced-representation [32]. This barcoding system increased efficiency and relatively inexpensive. Barcodes included sequences adapter and their locations, just upstream of the RE cut site in genomic DNA, eliminate the need for a second Illumina sequencing read. Unlike, RAD this system has modulation of barcode nucleotide composition and results in fewer length sequence phasing errors [9]. Substantially GBS becoming less complicated; generation of restriction fragments with appropriate adapters is more straight forward, single-well digestion of genomic DNA and adapter ligation results in reduced sample handling, there are fewer DNA purification steps, and fragments are not size selected as compared to the RAD method. Costs can be further reduced via shallow genome sampling tied with imputation of missing internal SNPs in haplotype blocks [20,40].

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891

Libraries construction GBS mainly focuses on the reduction of genome complexity with the help of restriction enzymes [21]. Compared to the other approaches, GBS is simple, quick, extremely specific, highly reproducible, and may reach important regions of the genome that are inaccessible to sequence [40]. To get higher efficiency in GBS with a targeted of two or three-fold it needs the selection of appropriate REs, in order to avoid repetitive regions of genomes and lower copy regions [4,6]. This method tremendously simplifies computationally challenging alignment problems in species with high levels of genetic diversity [21].

Genotyping-by-sequencing (GBS) application in plant breeding

GBS is one of the most powerful tools in genome applications in the area of plant breeding. It is used to study GWAS, GS, gd-study, analysis of genetic linkage and marker discovery of non-model plants [22,40,43]. It is also an ideal platform for studying for a crop ranging from single gene to complex whole genome [4,40]. Generally, it is becoming an excellent tool for many applications and research questions in plant breeding and genetics for different food and industrial crops due to its flexibility and low cost [7,41]. According to it has been shown that this technique becoming valid tool to undertake genomic diversity studies. gd-GBS is new Illumina-based GBS protocol and it is unique from others. Compared to Roche 454 platform, this method yields more SNP genotype data with fewer missing observations. Genotyping a diploid species, it uses of two restriction enzymes that used to reduce genome complexity, application of Illumina multiplexing indexes for barcoding and availability of a custom bioinformatics are the major features of gd-GBS. Like GBS, there are five major steps implemented gd-GBS (Figure 2&3). These are:

a. Overall information about plant genetic diversity analysis;

b. Specific genetic diversity project in mind to pursue;

c. Plant materials prepared and ready to assay; and

d. Access computing resources. The complete gd-GBS protocol, including the bioinformatics pipeline non-model plant genotyping np Geno, is provided in the online supporting materials [44].

As illustrated in Figure 2 & 3, GBS application in genetic diversity study (gd-GBS) involves five major steps:

a. Sample preparation,

b. Library assembly,

c. Sequencing,

d. SNP calling and

e. Diversity analysis [44, 45].

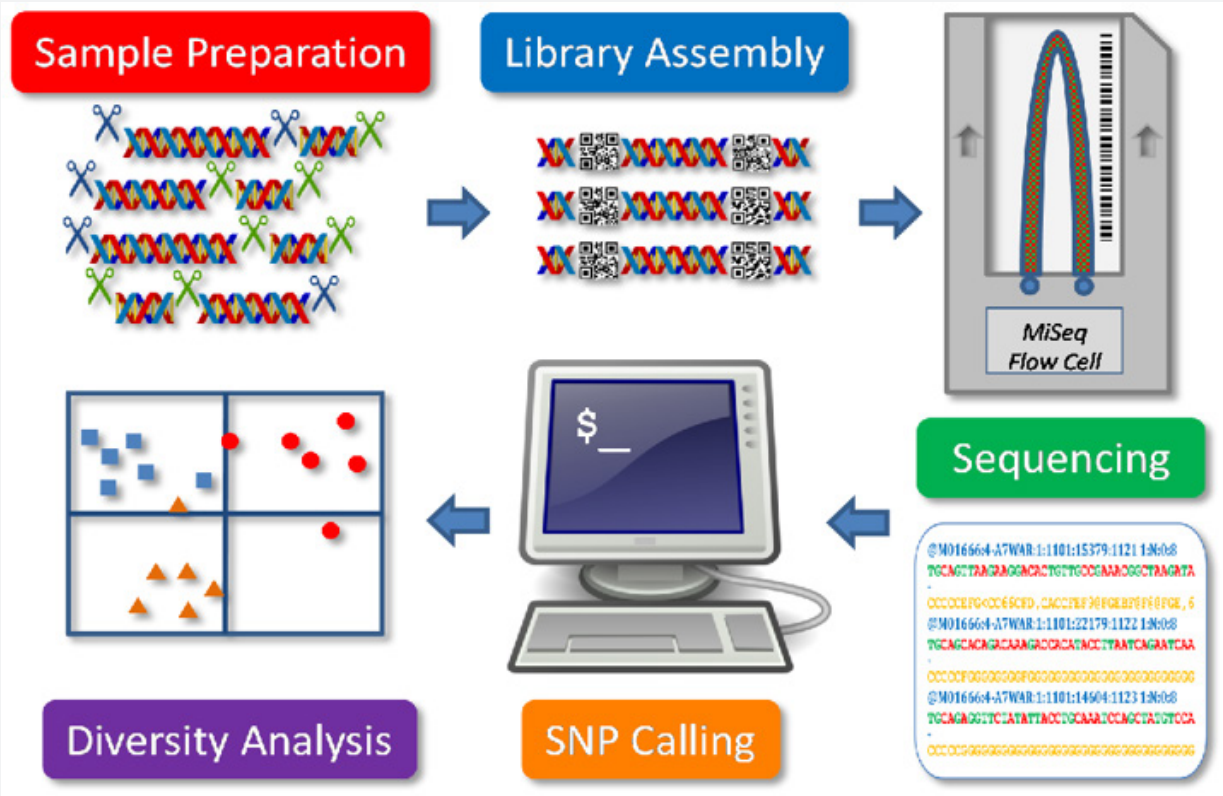

Figure 2: Steps in genotype-by-sequencing application for plant genetic diversity analysis.

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891

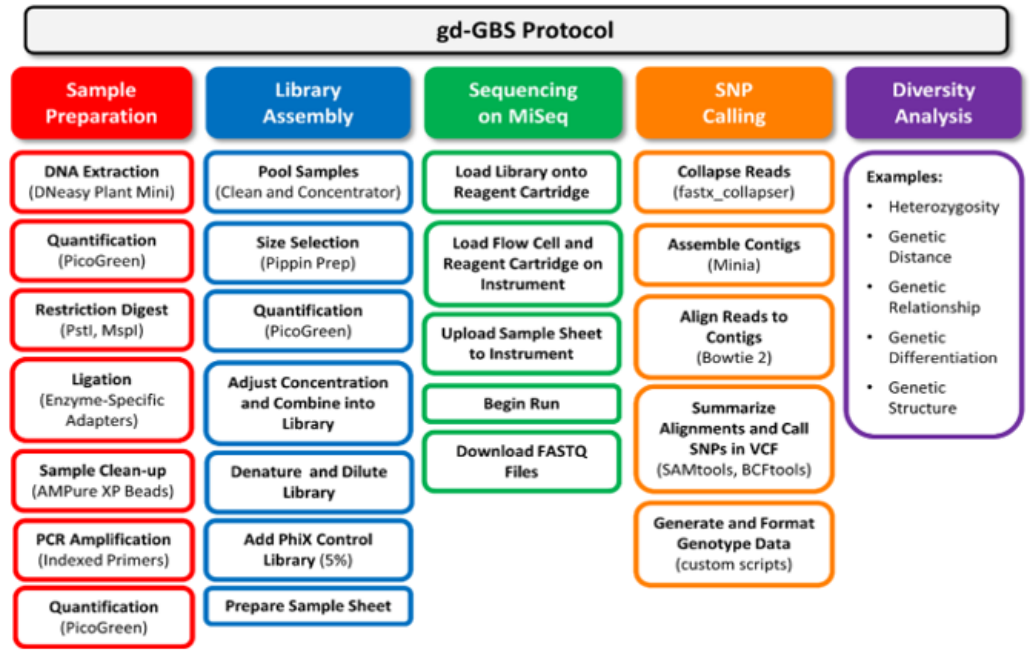

Figure 3: Flow chart of the genetic diversity-focused genotyping-by-sequencing (gd-GBS) protocol.

The above-mentioned steps may vary from one another depending on Restriction Endonuclease (RE) use and NGS platform and bioinformatics analysis for SNP genotyping to explore different objectives. To undertake genetic diversity analysis study the approach focuses on genome-wide sampling of many samples. Whereas, for genome-wide association studies, it emphasizes the accuracy of SNPs call rate with read depth to reveal genetic signals. Specifically, an informative genetic diversity analysis requires SNP data with large genome coverage, high genotyping accuracy, more balanced observation and less bias, which may be technically introduced from sequence mapping, heterozygote detection and data filtering [44]. In plant genetic diversity study analysis GBS approach have several major features. First, it combines the processes of marker discovery and genotyping, provides a rapid, high throughput and cost-effective tool for a genome-wide analysis of genetic diversity. Second, it requires no prior sequencing of the plant genome and provides direct genotyping of plants with complex genomes without prior SNP discovery. Third, and most importantly, it generates many genome-wide SNP data, allowing for better genome sampling. In general, this approach becoming more accessible for crop without model species [4, 6, 44].

To generate sufficient information and coverage in a GWAS it needs 100s of 1000s to millions of markers. However, the creation of NGS technologies greatly improved the resolution of marker [40]. Nowadays, GBS through the NGS has been used to sequence collections of recombinant inbred lines (RILs) to analyze and map various traits of interest for a specific breeding programs [32]. Cereals crop like maize, wheat, barley, sorghum, oat, rice, root and tuber like potato, cassava and industrial crop like cotton have been reported to optimized by GBS for the efficient, low-cost and large scales of genome sequencing [32, 40, 46]. In maize a collection of 5,000 RILs have been sequenced using a restriction endonuclease-based approach and Illumina sequencing technology, which generated a total of 1.4 million SNPs and 200,000 indels [32, 40]. In maize an inclusive genotyping of 2,815 inbred accessions showed that 681,257 SNP markers are distributed across the entire genome, in which some SNPs are linked to the known candidate genes for kernel color, sweetness, and time of flowering [32, 47]. In soybean 31 genotypes with a set of 205,614 SNPs have been identified after resequencing giving valuable information for a soybean breeding programs. In potato [4, 40], 12.4 GB of high-quality sequence data and 129,156 sequence variants have been identified in breeding program of potato around 2.1 Mb were mapped to reference genome with a median average read depth of 636 per cultivar [32,40].

[48]reported that gd-GBS used the application of Roche 454 GS FLX Titanium technology with reduced genome representation and advanced bioinformatics tools to analyze 16 diverse barley landraces their genetic diversity and reported 2,578 contigs, and 3,980 SNPs, and confirmed a key geographical division in the cultivated barley gene pool [7]. The report from [49] showed that to access genetic diversity of species like switchgrass and they developed a SNP discovery pipeline based on a network approach called the Universal Network-Enabled Analysis Kit (UNEAK). Accordingly, 540 switchgrass plants sampled from 66 populations revealed informative patterns of genetic relationship with respect to ecotype, ploidy level, and geographic distribution to undertake

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891

the diversity study. In addition, in mustard GBS protocol was used to analyze genetic diversity of 24 diverse yellow mustard accessions. The fining showed that 1.2 million sequence reads were generated, and 512 contigs and 828 SNPs were identified. Consequently, the genetic diversity study showed that yellow mustard SNPs revealed 26.1% of total variation over the landrace, cultivar, and breeding lines and 24.7% between yellow-seeded and black-seeded germplasm [7, 50].

In addition, sequencing of Arabidopsis in the whole genome shotgun sequencing on the Illumina platform a pool of 500 F2 plants generated by crossing a recessive Ethane Methyl Sulfonate (EMS)-induced Col-0 mutant characterized by slow growth and light green leaves, with a wild type Ler (Landsberg erecta) line. The result identifying high density SNP markers through GBS to construct genetic linkage maps which has a great value for numerous applications in plant breeding [7,51]. also reported that using a 384 plex GBS protocol to add 30,984 SNP markers to an Indica × japonica mapping population consisting of 176 rice recombinant inbred lines and mapped the recombined hot and cold spots and Quantitative Trait Loci (QTLs) for aluminum tolerance and leaf width. In bread wheat GBS was also applied resulting in the incorporation of 1000s of markers in the bread wheat map [22]. Identification of high resolution of SNP markers in barley and GBS mapping data were used to confirm that the semi-dwarfing gene (ari-e) is located on barley chromosome 5H [42, 49]. After the efficiency of multiplexed SNP genotyping for diversity, mapping and breeding applications were evaluated, and demonstrated that 384 plex SNP genotyping on the Bead Xpress platform is a robust and efficient method for marker genotyping and mapping in rice [32, 47].

The drawbacks of traditional method of plant breeding can be solved by MAS. With GBS, this is mainly achieved with the combination of molecular markers with phenotypic data to increase selection intensity and/or reduced selection interval on genotypic values [7]. Application of both applied and theoretical studies in genomic selection showed a great promise result to accelerate the rate new crop varieties (hybrid) development. GS through the GBS approach stands to be a major supplement to traditional crop improvement and it is a very important feature to move the genomics-assisted breeding into commercial crops [22]. GBS method on barley and wheat study without a reference genome provides a powerful method of developing high density markers by providing valuable tools for anchoring and ordering physical maps and whole genome shotgun sequence [40,47]. GBS approach also gives a very good promising result in cabbage, cauliflower and cotton without the reference genome identification and genetic diversity study. In Miscanthus the application of GBS is difficult due to ploidy level differences [47]. GBS approach also efficient to developed a catalog SNPs both within mapping population and among diverse African cassava varieties in which it allowing the improvements of MAS programs on disease resistance and nutrition in cassava [7].

Limitation of genotyping-by-sequencing

Despite the above listed advantages, the applications GBS have some potential drawbacks. In large, complex, polyploid genomes the difficulty getting aligned alleles in a single locus are the major challenges encountered by this method. Compared to others tools available to resolve the above problem GBS has a great potential. In addition, in hexaploid oat data analysis algorithms represent the main limiting factor to ascertain alleles at each single locus in a large polyploidy genome rather than GBS itself given sufficient depth of sequence is available [52]. reported that main weakness of GBS assay, when conducted at low coverage, is the amount of missing data. However, numerous imputation approaches are currently available, and yet more are presently in development, for a wide range of biological scenarios. Selecting appropriate imputation method and the probability of imputation success depends upon the biology of the study population. In the other hand, GBS genome complexity can be reduced by using restriction enzymes if applicable, in case of any mutation at the restriction site, the genomic DNA of this region is not available to be PCR amplified and consequently SNPs of this region will become unavailable and sometimes heterozygote gene may appear as homozygous. However, this drawback is not a problem only related with GBS rather it is shared by all the different methods involving reduction in genome complexity based on the utilization of restriction sites. GBS with two restriction enzymes have been overlooked to each other that the activity of MspI is inhibited in epigenetic studies. Therefore, developmental responses in plants may affect the SNP identification when using the enzyme MspI cannot be ignored but is likely reduced [7]. In addition, most of world food security crops (orphan crops) are neglected plant species and have not any known genomic sequence. An available reference genome can simplify the data analyses, but it is not essential in GBS for the above listed crops [7, 45].

Conclusion

World food security problem is one of the main agenda in the 21th century. To address this problem plant breeding is a main driving force [4] It can be accomplished by both conventional breeding and molecular breeding. However, the former approach has several limitations such as requiring a extended period of time to release high yielding variety. While, the later i.e., Marker-Assisted Selection (MAS) uses DNA markers and it is a new discipline in in the area of 'molecular breeding' [4, 6, 47]. Currently, in a different crop improvement program a novel application in NGS that used to discovering and genotyping SNPs is known as Genotyping-by-sequencing (GBS). GBS has several advantages, including lower costs per samples, and relatively inexpensive to other whole genome genotyping platforms. Due to its use of high density of SNP markers, it is the most attractive approach to saturate mapping and breeding populations. Therefore, to attain the current problems in the area of plant breeding breeder's abele to sequence and resequencing large crop genomes to this effect

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891

it can establish high density genetic linkage maps from large size breeding populations. Even if it has the above listed advantages, it has also numerous biological and technical drawbacks. Among all the following points are considered as the major drawbacks in the application of GBS,

a. Bias during PCR amplification and library construction,

b. Lack of evenly covered regions of interest and within a given populations not all individuals are not sequenced very well,

c. it requires continuous imputation for a missing data using both pedigree and parental information when available.

Future direction in GBS

Nowadays GBS has been reached an advanced stage but, some point regarding the limitation needs attention in the future. According [38] the following points should need more emphasis in the future regarding GBS New technical variation in GBS requires an advanced analytical tool for genomic data in which it can undergone genotyping large numbers of individuals and complete genotyping to the selected targets crops that are considered biologically, economically and socially relevant. Additionally, combination of GBS and RNA sequencing to find out SNPs in association with gene expression patters have a benefit to create a link between genomics, transcriptomics and proteomics. In general, this approach creates an opportunity to expand knowledge in the area of plant breeding and genetics research [53-57].

Acknowledgement

Special thanks goes for Dr David Cros form CIRAD for his very pertinent comment and suggestion which was very helpful in improving the manuscript.

References

1. Spiertz JHJ, Ewert F (2009) Crop production and resource use to meet the growing demand for food, feed and fuel: opportunities and constraints. NJAS - Wagening J Life Sci 56(4): 281-300.
2. Collard, BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans R Soc B Biol Sci 363(1491): 557-572.
3. Kim C, Guo H, Kong W, Chandnani R, Shuang LS, et al. (2016a). Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Sci 242: 14-22.
4. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, et al. (2016) Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. Front. Genet 7: 221.
5. Evans LT, (1997) Adapting and improving crops: the endless task. Philos. Trans R Soc B Biol Sci 352: 901-906.
6. Kim C, Guo H, Kong W, Chandnani R, Shuang LS, Paterson AH (2016b). Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Sci 242: 14-22.
7. He J, Zhao X, Laroche A, Lu ZX, Liu H, et al. (2014a) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5: 484.
8. Xu Y (2010) Molecular plant breeding. Mol Plant Breed 1-734.
9. Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, et al. (2016) Next generation breeding. Plant Sci 242: 3-13.
10. Collard BCY, Mackill DJ (2008b) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Philos Trans R Soc Lond B Biol Sci 363(1491): 557-572.
11. Dale Young N (1999) A cautiously optimistic vision for marker-assisted breeding. Molecular Breeding 5(6): 505-510.
12. Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, et al. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3(2): 87-103.
13. L Heffner E, J Lorenz A, Jannink JL, Sorrells M, (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. Crop Sci - CROP SCI 50(5): 1681-1690.
14. Varshney RK, Mohan SM, Gaur PM, Gangarao NV, Pandey MK, et al. (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv 31(8): 1120-1134.
15. Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. Genetics 182(1): 355-364.
16. Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet Z 124(6): 323-330.
17. Cros D, Tchounke B, Nkague-Nkamba L (2018) Training genomic selection models across several breeding cycles increases genetic gain in oil palm in silico study Mol Breed 38.
18. Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, et al. (2017) Genomic Selection in Commercial Perennial Crops: Applicability and Improvement in Oil Palm (*Elaeis guineensis Jacq*). Sci Rep 7(1): 2872.
19. Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19(9): 592-601.
20. Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J 15(2): 149-161.
21. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE 6(5): e19379.
22. Poland JA, Brown PJ, Sorrells ME, Jannink JL (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.
23. Wambugu PW, Ndjiondjop MN, Henry RJ (2018) Role of genomics in promoting the utilization of plant genetic resources in genebanks. Brief Funct Genomics 17(3): 198-206.
24. Grover A, Sharma PC (2014) Development and use of molecular markers: Past and present. Crit Rev Biotechnol 36(2): 290-302.
25. Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121(1): 185-199.
26. Jarvis P, Lister C, Szabo V, Dean C (1994) Integration of CAPS markers into the RFLP map generated using recombinant inbred lines of Arabidopsis thaliana. Plant Mol Biol 24(4): 685-687.
27. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23(21): 4407-4414.
28. Salimath SS, de Oliveira AC, Godwin ID, Bennetzen JL (1995) Assessment of genome origins and genetic diversity in the genus Eleusine with DNA markers. Genome 38(4): 757-763.

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891

29. Jiang C, Sink KC (1997) RAPD and SCAR markers linked to the sex expression locus M in *asparagus*. Euphytica 94(3): 329-333.

30. Desmarais E, Lanneluc I, Lagnel J (1998) Direct amplification of length polymorphisms (DALP), or how to get and characterize new genetic markers in many species. Nucleic Acids Res 26(6): 1458-1465.

31. Amom T, Nongdam, P (2017) The Use of Molecular Marker Methods in Plants: A Review.

32. Deschamps S, Llaca V, May GD (2012) Genotyping-by-Sequencing in Plants. Biology 1(3): 460-483.

33. Rounsley SD1, Glodek A, Sutton G, Adams MD, Somerville CR, et al. (1996) The Construction of Arabidopsis Expressed Sequence Tag Assemblies A New Resource to Facilitate Gene Identification. Plant Physiol 112(3): 1177-1183.

34. Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. (1998) Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. Science 280(5366): 1077-1082.

35. Takatsu K, Yokomaku T, Kurata S, Kanagawa T (2004) A new approach to SNP genotyping with fluorescently labeled mononucleotides. Nucleic Acids Res. 32(7): e60.

36. Elvidge GP, Glenny L, Appelhoff RJ, Ratcliffe PJ, Ragoussis J, et al. (2006) Concordant Regulation of Gene Expression by Hypoxia and 2-Oxoglutarate-dependent Dioxygenase Inhibition THE ROLE OF HIF-1α, HIF-2α, AND OTHER PATHWAYS. J. Biol. Chem. 281(22): 15215-15226.

37. Gupta PK, Rustgi S (2018) Array-Based High-Throughput DNA Markers and Genotyping Platforms for Cereal Genetics and Genomics.

38. Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. Mol Ecol 22(11): 2841-2847.

39. eHe J, eZhao X, eLaroche A, eLu Z-X, eLiu H, eLi Z (2014) Genotyping by sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5: 484.

40. He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z (2014c). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci 5: 484.

41. Poland JA, Rife TW (2012a) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5(3): 92-102.

42. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS One 3(10): e3376.

43. Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, et al. (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res 21 (4): 610-617.

44. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLOS ONE 7(5): e37135.

45. Poland JA, Rife TW (2012b) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome J 5: 92.

46. van Poecke RM, Maccaferri M, Tang J, Truong HT, Janssen A, et al. (2013) Sequence-based SNP genotyping in durum wheat. Plant Biotechnol. J 11(7): 809-817.

47. He J, Zhao X, Laroche A, Lu Z-X, Liu H (2014b) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5.

48. Fu YB, Peterson GW (2011) Genetic diversity analysis with 454 pyrosequencing and genomic reduction confirmed the eastern and western division in the cultivated barley gene pool. Plant Genome 4: 226-237.

49. Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, et al. (2013) Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLOS Genet 9 e1003215.

50. Fu YB, Cheng B, Peterson GW (2014) Genetic diversity analysis of yellow mustard (*Sinapis alba L.*) germplasm based on genotyping by sequencing. Genet. Resour. Crop Evol 61: 579-594.

51. Spindel J, Wright M, Chen C, Cobb J, Gage J, et al. (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet 126 2699-2716.

52. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, et al. (2014) TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. Plos One 9(2): e90346.

53. Mascher M, Wu S, Amand PS, Stein N, Poland J (2013) Application of Genotyping-by-Sequencing on Semiconductor Sequencing Platforms: A Comparison of Genetic and Reference-Based Marker Ordering in Barley. PLOS ONE 8(10): e76925.

54. Monson-Miller J, Sanchez-Mendez DC, Fass J, Henry IM, Tai TH, et al. (2012) Reference genome-independent assessment of mutation density using restriction enzyme-phased sequencing. BMC Genomics 13: 72.

55. Stolle E, Moritz RFA (2013) RESTseq - Efficient Benchtop Population Genomics with RESTriction Fragment SEQuencing. PLOS ONE 8: e63960.

56. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18(22): 6531-6535.

57. Yu J, Hu S, Wang J, Wong GK, Li S, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296(5565): 79-92.

DOI: 10.19080/AIBM.2019.14.555891

How to cite this article: Essubalew G S, Ngalle H B, Joseph M B, Wosene G. Genotyping by Sequencing for Plant Breeding- A Review. Adv Biotechnol Microbiol. 2019; 14(4): 555891. DOI: 10.19080/AIBM.2019.14.555891