BioThink E – Magazine, Vol. -3, Issue -2, April 2026
ISSN No. – 3048- 7943
www.biothink.in

Genotyping-by-Sequencing: A Transformative Approach for High-Density Genotyping in Agricultural Genomics

Thamaraikannan Sivakumar ✉, Matta Revanth Kumar Reddy [b], Sathishkumar S

Department of Agronomy, Horticulture and Plant Science, South Dakota State University, Brookings, United States of America – 57007.

[b] Department of Genetics and plant breeding, VNMKV, Parbhani-431202, Maharashtra, India

[c] Division of Genetics, ICAR – Indian Agricultural Research Institute, New Delhi – 110012.

*All authors contributed equally.

✉ sathishkumar09.edu@gmail.com

Abstract

Advances in next-generation sequencing (NGS) technologies have revolutionized plant genomics by enabling high-throughput genotyping and comprehensive genome-wide variant discovery. Among emerging genomic tools, genotyping-by-sequencing (GBS) has become a popular method for high-density single-nucleotide polymorphism (SNP) discovery and genotyping in plant species. By combining genome complexity reduction with multiplexed sequencing, GBS permits the simultaneous analysis of large populations at a relatively low cost.

This technology has greatly accelerated research in plant genetics, crop improvement, and evolutionary genomics by enabling applications such as genetic diversity analysis, genome-wide association studies (GWAS), quantitative trait locus (QTL) mapping, genomic selection (GS), and phylogenetic reconstruction. In agricultural research, GBS has played a crucial role in identifying genetic variants linked to yield, stress tolerance, and disease resistance in major crops like wheat, rice, maize, soybean, and barley.

Recent advances in bioinformatics, the integration of long-read sequencing, and machine learning-based SNP detection are overcoming these limitations and broadening GBS's applications in plant research.

Keywords:

Genotype by Sequencing, Plant Genomics, Variant Calling, Crop Improvement

Introduction

The rapid increase in global population, along with environmental challenges, has driven greater demand for sustainable agricultural production. Contemporary plant breeding relies strongly on genomic technologies to detect and harness genetic variation for key agronomic traits. While molecular markers have opened the studying of plant genetic makeup and implementing it in the breeding programs; previous molecular systems like restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphisms (AFLP), and simple sequence repeats (SSR), for example, were limited due to their time consuming, more involvement of manpower, lack of precisions due to experimental errors, low throughput, limited genome coverage, and relatively high costs (Deschamps et al., 2012). These limitations hindered the ability to conduct large-scale genomic investigations necessary for modern crop improvement.

The introduction of next-generation sequencing (NGS) technologies has revolutionized plant genomics, enabling the detection of genetic polymorphisms across entire genomes in a short time and at low cost. Among NGS approaches, genotyping-by-sequencing (GBS), a highly efficient genotyping platform that uses high-throughput SNP discovery and genotyping, has been developed. Initially introduced as a simple and scalable approach for the simultaneous identification and genotyping of single-nucleotide polymorphisms (SNPs) in taxa with or without a reference genome (Poland & Rife, 2012), GBS is a method to reduce genome complexity by the use of restriction enzymes and multiplexing large numbers of samples in a single sequencing run, thus providing a robust approach to genotyping all kinds of breeding populations at a comparatively low cost.

Over the past decade, GBS has established itself as a major tool in plant genetics and breeding. The technology has enabled genome-wide marker discovery across a wide range of crops, including maize, wheat, barley, rice, sorghum, and soybean. For instance, GBS-based SNP datasets from large genetic populations have been used to develop high-resolution genetic maps and identify genomic regions associated with yield and stress tolerance traits in wheat and maize (He et al., 2014). Likewise, recent research has used SNP markers from GBS to identify candidate genes controlling drought tolerance and flower size traits in crops, mung beans, and watermelons (Ahmed et al., 2024; Nyirahabimana and Solmaz, 2025).

GBS has also contributed greatly to the development of genomic-aided breeding approaches, including genomic selection and marker-assisted breeding. By enabling the genotyping of thousands of individuals quickly, GBS can help accurately understand marker-trait associations and fast-track the development of superior crop varieties (Tyagi et al., 2024). This technology is particularly beneficial for crops with large or complex genomes, such as plant species for which whole-genome sequencing is economically unfeasible. This chapter provides an overview of the GBS methodology and its variations, discusses the bioinformatics pipelines used for SNP discovery and genotype analysis, and highlights the many applications of GBS in plant sciences and agriculture.

2. Genotyping-by-Sequencing Methodology in Plants

The success of genotyping-by-sequencing (GBS) in plant genomics is largely due to its simple library preparation protocol and its ability to produce genome-wide markers at high efficiency. The methodology combines restriction enzyme-mediated genome complexity reduction with next-generation sequencing, thereby enabling the simultaneous discovery and genotyping of single-nucleotide polymorphism (SNP) markers in plant populations, and the methodology is illustrated in **Figure 1.**

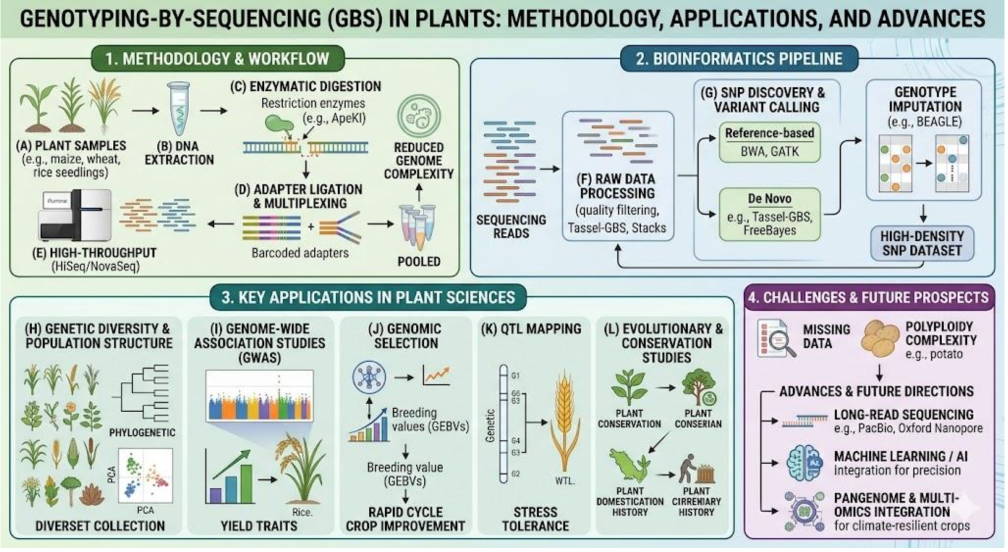

Figure 1. Overview of GBS techniques in Plants

The first step in GBS involves enzymatic digestion of genomic DNA with restriction endonucleases. These enzymes cut at specific recognition motifs and produce a predictable subset of genome fragments, thereby reducing the full genome sequence. This reduction strategy is especially beneficial when a taxon's genome is large, for which thorough genome and whole-genome sequencing would be prohibitively expensive. The methylation-sensitive endonuclease *ApeKI*, which prefers gene-rich sequences, is a commonly used enzyme in early GBS protocols. By exploiting the preferred methylation of repetitive sequences, *ApeKI* digestion enriches for low-copy fragments, which are more likely to contain informative SNPs. The selection of restriction enzymes is an important factor in marker density and genome coverage. While single-enzyme digests were initially used for GBS, subsequent refinements introduced double-enzyme digestion protocols that enhance reproducibility and equal distribution of sequenced fragments throughout the genome. When the double-digest strategy is overused, a pair of enzymes with varying cut frequencies produces a reliable set of fitted genomic fragments. This method forms the basis for the widely used double-digest restriction site-associated DNA sequencing (ddRAD-seq) method, which has been successfully applied to many plant species.

Following restriction digestion, DNA adapters containing a sequencing primer and a unique barcode are ligated to the resulting DNA fragments. These sequences of barcodes allow several samples to be pooled into one sequencing library - a process that is called multiplexing. Multiplexing greatly reduces sequencing costs and enables the analysis of hundreds of plant genotypes in parallel. After library construction, the collected fragments are subjected to high-throughput sequencing.

Illumina sequencing platforms continue to dominate GBS technology because they are very accurate and can provide millions of short reads. Instruments such as HiSeq, NovaSeq, and the NextSeq offer the required throughput for studies with large populations. In recent years, other third-generation sequencing methods have been investigated for GBS applications, such as PacBio and Oxford Nanopore. Although these platforms provide higher-quality read lengths that support haplotype reconstruction and structural variant detection, their high error rates and costs currently limit their general use in large-scale GBS experiments.

Post-sequencing, the resulting reads are run through bioinformatics pipelines developed for SNP finding, SNP variant calling, and genotyping. Initial stages include quality filtering, adapter trimming, and sample-specific barcoding demultiplexing. Filtered reads are then either compared to a reference genome or *de novo* assembled. Reference-based alignment is done routinely using tools such as BWA or Bowtie2, while identification of SNPs is done using variant calling software such as GATK, FreeBayes or Stacks. For species without reference genomes, there are methods that allow de novo SNP discovery directly from sequencing reads (TASSEL-GBS). GBS datasets are often incomplete due to missing genotype data resulting from poor sequencing coverage at individual loci. To address this problem, the statistical imputation technique is used to predict missing genotypes from linkage disequilibrium between markers. Software such as BEAGLE has been widely used for genotype imputation in GBS datasets, thereby increasing marker density and accuracy in subsequent analyses.

3. Variants of GBS in Plants

Since its introduction, the original GBS protocol has been modified multiple times to enhance marker density, improve efficient sequencing, and improve data quality. These modified approaches have widened the applicability of GBS across different plant species and research goals. One widely used variant is double-digest RAD sequencing (ddRAD-seq), which uses two restriction enzymes and size selection to produce consistent fragments of genomic DNA. This approach improves reproducibility and reduces sequencing bias, making it very useful for population genetics studies. ddRAD-seq has been used with many crops, including barley and sorghum, to investigate their genetic diversity and population structure.

Another modification aims to enrich genomic regions using PCR amplification or capture-based methods. Techniques such as SeqCap and AgSeq use probes (hybridization) to capture the region of interest in genomes before being sequenced. Such targeted approaches of GBS allow the researcher to focus on gene-rich areas or candidate loci for agronomic traits. Transcriptome-based approaches have also been combined with GBS strategies. In the case of GBS-T, RNA sequencing data is used to identify SNPs within expressed genes, enabling the study of functional genic variation associated with gene expression. This strategy is therefore particularly useful in functional genomics research, where knowledge of gene regulation is important. While whole-genome resequencing covers all genomic variation in its combinations, it remains relatively expensive in large populations. In contrast, GBS provides a cost-efficient alternative for genotyping thousands of individuals and for obtaining sufficient or representative collections of genome-spanning SNP markers. Consequently, GBS has become the method of choice for large-scale plant breeding and population genetics studies.

4. Applications of GBS to Plant Sciences

The power of GBS to generate thousands of SNP markers has completely revolutionized research in plant genetics and crop improvement. The most important use of GBS for the exploration of genetic diversity and population structure of crop germplasm collections. By apportioning genome-wide SNP positions, scientists can investigate the underlying genetic relationships within cultivars and landraces, as well as with their wild progenitor. These relationships play an important role in the conservation of plant genetic resources and in breeding programs that select parental lines. For instance, GBS-based markers (SNP markers) were used to explore the underlying genetic diversity of the germplasm population of sorghum and maize and patterns of population structure associated with geographic origins and domestication history.

The second important application of GBS is in genome-wide association studies (GWAS), which aim to identify regions of the genome associated with traits. Due to GBS's potential to generate thousands of SNPs, GBS-based marker systems allow scientists to identify marker-trait relationships at higher resolution than traditional marker systems. GBS-based marker systems are used to identify genes controlling flowering time and grain yield in rice through genome-wide association studies. Similarly, some recent GWA studies of mung bean using SNPs identified by GBS identified genes associated with drought tolerance and seed size (Ahmed et al., 2024). These studies have shown that GBS can be used to identify genes controlling a range of plant traits. GBS has also become established as an important tool in genomic selection, a technique that uses genome-wide markers to predict the breeding values of individuals. Within this framework, thousands of GBS-developed markers of SNPs with high yields are used for making genomic breeding values (GEBVs) for complex traits such as yield or tolerance to different forms of stress. The use of this technique has greatly shortened the breeding cycle and the time it takes to develop new crop varieties that can be sold.

The second important use of GBS will be in quantitative trait locus (QTL) mapping, where dense marker panels (SNPs) will be used to identify genes controlling quantitative traits. GBS-based maps have been useful for establishing QTL maps for traits such as disease resistance and tolerance to abiotic stress in crops such as wheat and barley. Potato, genes that control the concomitant action of components of plant membrane units, as the major elements in GWA studies, improve search to result in identification of genes controlling tuber yield and quality in autotetraploid populations (Sharma et al., 2024).

GBS has also found many other applications for evolutionary and phylogenetic studies. In comparing species, genome-wide SNP variation is used to assess species relationships and domestication events within the canine family. For example, bar comparison of GBS data plays an important role in understanding the history of barley and maize domestication. In these cases, patterns of gene flow from wild species to domesticated species are evident. Besides efficiency gains in crop improvement, GBS has critical applications in conservation genomics. For example, high-resolution eukaryotic SNP markers enable scientists to measure the genetic diversity and population structure of endangered plant species. In addition, there is an increasing trend of using GBS data to optimize core collections of gene banks for the efficient conservation of plant genetic resources.

5. Deficiencies and Limitations with GBS

GBS has various benefits as discussed above; however, the technique is not without its challenges. One major limitation of the technique is the availability of missing genotype data, as it involves sequencing only a limited portion of the genome, which may not always be possible, leading to missing data points. Even though some techniques have been developed to address this problem, known as genotype imputation, the problem of missing data is still prevalent in some cases. Another problem noted with the technique was the calling of single-nucleotide polymorphisms in polyploid plant genomes, as the presence of multiple copies of the same chromosome makes genotype calling difficult. Polyploid plants, such as wheat, sugarcane, and potato, require specialized bioinformatics software to distinguish between two types of sequences: homoeologous and homologous. The technical challenges of the technique stem from the restriction enzymes used in the experiment, as different enzymes yield different fragment lengths, thereby altering marker density. In addition, the technique is not effective at detecting large structural variants, copy number variation, or insertions and deletions, unlike whole-genome sequencing.

7. Current Advances and Future Prospects of GBS

Recent technological advances in genomics are likely to increase the efficiency of the Genotyping-by-Sequencing technique in plant genomics. The emerging trend is likely to focus on combining the technique with long-read sequencing, as this combination is expected to help researchers assemble the panel and identify structural variants in the genome, which has not been possible with short-read sequencing alone. The second emerging trend is likely to focus on the application of machine learning algorithms, as well as artificial intelligence, as this is likely to help the researcher improve the precision of the Single Nucleotide Polymorphisms, as well as genotype imputation, as well as genomic prediction, as this is likely to help the researcher handle the large datasets that are produced by the high-throughput sequencing technique.

Recent studies have also demonstrated GBS's capacity to successfully survey complex traits, such as resilience to abiotic stress. Genomic analysis in watermelon and other horticultural crops related to genomic regions associated with drought and salinity tolerance. This is a major step towards developing climate-resilient cultivars. The future direction of GBS in plant genomics might be to integrate it with other emerging genomic technologies, such as pangenome sequencing and multi-omics. Pangenome sequencing encompasses the investigation of the species' extensive genetic diversity, including structural variants and gene presence/absence polymorphisms that are invisible to other genomic technologies, such as reference genome-based methods. Rapid genotyping technologies like GBS are likely to play a pivotal role in crop improvement programs through effective screening of the breeding population to identify the desired phenotype. This can then be combined with genomic selection and high-throughput phenotyping to enable the breeder to slow the evolution of improved cultivars and keep pace with changing environmental conditions. Another direction GBS could take in the future is integrating genomic and microbiome data to identify differences between the two, thereby enabling the development of crops with improved nutrient uptake and disease resistance.

8. Conclusion

This genotyping-by-sequencing technique has had a significant impact on plant genomics by enabling low-cost, high-density genotyping across a range of taxa. This technology has made a significant contribution to plant genetics and genomics by enabling the identification of single-nucleotide polymorphisms and high-density genotypes. The practical applications of GBS in genome-wide association studies, genomic selection, and QTL mapping have led to a substantial increase in the discovery of genes for important agronomic traits. Even though this technology faces many challenges and gaps in data structure, the development of polyploid genome complexity is improving with advances in sequencing technology and bioinformatics tools. GBS technology is also likely to remain at the center of plant genomics and the design of precision breeding approaches to establish sustainable agriculture and climate-resilient varieties.

References

Ahmed, S. R., Asghar, M. J., Hameed, A., & Ghaffar, M. (2024). Advancing crop improvement through GWAS and beyond in mung bean. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2024.1436532

Deschamps, S., Llaca, V., & May, G. D. (2012). Genotyping-by-sequencing in plants. *Biology, 1*(3), 460–483. https://doi.org/10.3390/biology1030460

Kumar, R., Das, S. P., Choudhury, B. U., & Kumar, A. (2024). Advances in genomic tools for plant breeding: Harnessing DNA molecular markers and genomic selection. *Biological Research, 57*, 45. https://doi.org/10.1186/s40659-024-00562-6

Mangal, V., Verma, L. K., Singh, S. K., Saxena, K., & Roy, A. (2024). Triumphs of genomic-assisted breeding in crop improvement. *Heliyon, 10*, e11544. https://doi.org/10.1016/j.heliyon.2024.e11544

Nyirahabimana, F., & Solmaz, I. (2025). Cutting-edge of genotyping-by-sequencing (GBS) for improving drought and salinity stress tolerance in watermelon. *Plant Molecular Biology Reporter*. https://doi.org/10.1007/s11105-024-01465-2

Poland, J. A., & Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome, 5*(3), 92–102. https://doi.org/10.3835/plantgenome2012.05.0005

Sharma, S. K., McLean, K., Hedley, P. E., & Dale, F. (2024). Genotyping-by-sequencing targets genic regions and improves resolution of genome-wide association studies in autotetraploid potato. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-024-04651-8

Tyagi, A., Mir, Z. A., Almalki, M. A., Deshmukh, R., & Ali, S. (2024). Genomics-assisted breeding: A powerful breeding approach for improving plant growth and stress resilience. *Agronomy, 14*, 1128. https://doi.org/10.3390/agronomy14061128

Singh, R., Kumar, K., Bharadwaj, C., & Verma, P. K. (2022). Broadening the horizon of crop research: A decade of advancements in plant molecular genetics to divulge phenotype governing genes. *Planta, 256*, 34. https://doi.org/10.1007/s00425-022-03827-0

Altaf, M. T., Tatar, M., Ali, A., & Liaqat, W. (2024). Advancements in QTL mapping and GWAS application in plant improvement. *Turkish Journal of Botany*. https://doi.org/10.55730/1300-008X.2824