High-throughput NGS-based genotyping and phenotyping: Role in genomics-assisted breeding for soybean improvement

Javaid Akhter Bhat | Deyue Yu

Soybean Research Institution, National Center for Soybean Improvement, Key Laboratory of Biology and Genetics and Breeding for Soybean, Ministry of Agriculture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China

Correspondence
Javaid Akhter Bhat and Deyue Yu, Soybean Research Institution, National Center for Soybean Improvement, Key Laboratory of Biology and Genetics and Breeding for Soybean, Ministry of Agriculture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China.

Email: javid.akhter69@gmail.com; dyyu@njau.edu.cn

Funding information
Ministry of Science and Technology of China, Grant/Award Number: 2017YFE0111000

Abstract
Soybean is an important food crop that provides edible protein and oil for human and animal nutrition. Conventional phenotypic-based breeding approaches have made significant contribution in the last century by developing many improved soybean varieties. However, due to the longer time taken to develop a variety, low genetic gain per unit time, and adverse environmental influence of phenotypic-based selection, conventional approaches are not sufficient to maintain pace with growing population and climate change. In this context, the recent method of genotypic selection, that is, genomics-assisted breeding (GAB) is considered a promising approach to address the challenges in soybean breeding. However, to harness the true potential of GAB in soybean improvement, the great coverage and precision in the genotyping and phenotyping are required. Previously, a huge gap was observed between the discovery and practical use of quantitative trait loci (QTLs) in soybean improvement. It has been suggested that low marker density and manually collected phenotypes are the major reasons for this failure. Hence, high-throughput genotyping (HTG) providing higher genome-wide marker density, as well as accurate and precise phenotyping using high-throughput digital phenotyping (HTP) platforms, can significantly increase the success of QTL and candidate gene identification in soybean. These approaches can greatly increase the practical utility of GAB in soybean and also offer a faster characterization of germplasm and breeding materials. This review provides the detailed information on how the recent innovations in genomics and phenomics can assists in improving the efficiency and potential of GAB in soybean improvement.

KEYWORDS
genomics, genomics-assisted selection, improved varieties, phenomics, soybean

INTRODUCTION

Soybean (Glycine max L. Merr.), being a rich source of edible oil and protein, is an economically important crop. It provides significant benefits in health, biofuel, and soil fertility improvement (Kulkarni et al., 2016). In China, soybean production has continuously declined due to lower yield increases in the past 50 years (Liu et al., 2018). China imports >80% of soybean for their total domestic use; hence, it is essential to increase the domestic production of soybean to make the country self-sufficient (Liu et al., 2018). Soybean production can be increased mainly by breeding high-yielding stress-tolerant cultivars (Sharmin et al., 2020). In this regard, the different yield-related traits are targeted by plant breeders to increase soybean production (Karikari et al., 2019). In the last century, conventional breeding
approaches have tremendously contributed by successfully developing many high-yielding and superior quality soybean varieties (Ahmar et al., 2020). However, the progress achieved through the conventional approaches are not sufficient to keep pace with growing population and climate change (Athar & Ashraf, 2009; Bhat et al., 2016). As most of the economically important traits in soybean such as yield, yield-related traits, and stress tolerance are complex quantitative traits (Karikari et al., 2019; Hina et al., 2020), the time-consuming and expensive conventional methods employed for the phenotypic selection of the desired plant are not sufficient to maintain global food security (Bhat et al., 2016; Mir et al., 2019).

The recent progress made in genomics and phenomics provides exciting opportunities to conduct precision breeding at higher efficiency (Mir et al., 2019; van Bezouw et al., 2019). For example, the genomics-assisted breeding (GAB) have allowed more genetic gain for complex traits at a relatively lower cost, but it requires genomic resources and molecular understanding of the trait (Chaudhary et al., 2015; Steiner et al., 2019). GAB involves two main important approaches, namely, marker-assisted selection (MAS) and genomic selection (GS) (Bhat et al., 2016). MAS depends upon the availability of markers associated with the trait of interest which can be identified either through linkage mapping or genome-wide association studies (GWAS). Many previous studies have demonstrated successful utilization of MAS in soybean by incorporating major genes and large-effect quantitative trait loci (QTLs) for different traits (Concibido et al., 1996; Gupta et al., 2017; Saghai Maroof et al., 2008). However, the minor genes govern most of the inheritance in the complex traits, but they have been never considered because of the limitations of MAS (Spindel et al., 2015). Further, the environment influence, epistatic interactions, and effect of genetic background have made the breeding of complex traits very complicated. Therefore, plant breeders have realized that MAS is not a right approach to breed complex plant traits. In this context, Meuwissen and Goddard (2001) have put forward the concept of GS as an alternative strategy.

In contrast to traditional MAS, GS utilizes whole genome-wide marker profile of breeding lines to predict the genomics-estimated breeding value (GEBV) by using different models; thus, it prevents the loss of the substantial portion of variation governed by minor effect QTLs/gens (Spindel et al., 2015). However, the accuracy in the identification of marker-trait associations and estimation of GEBV depends upon precise genotyping and phenotyping analysis, which in turn determine the success of GAB. The manual low-throughput phenotyping and genotyping often lead to false positive or negatives selection (Tuberosa, 2012). In this regard, high-throughput next generation sequencing (NGS)-based genotyping and phenotyping allow effective MAS and GS, and offer increased success of breeding programs (Cobb et al., 2019; Khan et al., 2016). Availability of high-throughput NGS-based genotyping techniques has considerably accelerated the process of gene discovery and GS, that too in crop plants with larger and complex genomes such as soybean (Bhat et al., 2016). To this effect, phenomics and genomics are equally critical for precise gene discovery and deriving GS model to estimate the GEBV of the breeding population (BP). Therefore, combining these approaches with appropriate genetic diversity, soil and weather data, analytical

**FIGURE 1**  Diagram showing the critical role of high-throughput phenotyping (HTP) and high-throughput genotyping (HTG) in the precise and accurate identification of quantitative trait loci (QTLs)/quantitative trait nucleotides (QTNs)/genes as well as genomic selection (GS)-based estimation of genomics-estimated breeding values (GEBVs)
tools, and databases would result in development of new varieties with enhanced yield, quality, and stress tolerance at a fast pace.

In the present review, we provide detailed information on how the advances in the high-throughput phenotyping (HTP) and high-throughput genotyping (HTG) approaches will help in increasing the precision of gene mapping and genomic prediction. We also discuss how these techniques are going to prove fruitful for developing next-generation crop improvement strategies in soybean and other crop species.

2 | GENOMICS-ASSISTED BREEDING

GAB involves two main approaches, namely, MAS and GS (Figure 1; Bhat et al., 2016). The MAS (includes MABC and MARS) utilizes molecular markers known to be associated with a particular trait/phenotype to identify a desirable individual carrying favorable allele for the trait of interest (Jiang, 2013). However, MAS is applicable only for the major-effect genes (Bhat et al., 2016); for example, most of the successful MAS-based studies carried out earlier in soybean involved mainly the major-effect QTLs/genes that govern large portion of phenotypic variation for the trait of interest (see details in Concibido et al., 1996; Gupta et al., 2017; Kim et al., 2020; Saghai Maroof et al., 2008; Viganó et al., 2018). In case of minor genes, which contribute only a small invisible phenotypic variation for the complex trait, MAS has not led to successful results in soybean (Bhat et al., 2016; Zhang et al., 2015). Most of the economically important traits in soybean such as yield, oil and protein content, and the stress tolerance are complex in nature; the maximum portion of phenotypic variation for these traits are controlled by minor effect genes (Hina et al., 2020). In this regard, GS offers promising results for complex quantitative traits (Figure 2; Bhat et al., 2016; Stewart-Brown et al., 2019). The GS combines marker profile and phenotypic data of the training population to develop the prediction model, and this model is used to estimate the GEBV of all individuals of BP (Gao et al., 2018; Yin et al., 2020). However, prior to the use of prediction model for the selection of individuals from BP, the cross-validation on subsets of the training population are used to check the accuracy of this model (Bhat et al., 2016). After successful validation, this model can be used to select individuals from the BP based on the GEBVs calculated only from the marker/genotypic data; hence, only the genotypic information is used to predict the phenotypic performance of individuals of the BP (Stewart-Brown et al., 2019). Recently, some of the GS studies carried out in soybean using high-throughput single nucleotide polymorphism (SNP) genotyping have revealed promising results for different complex traits with moderate to high prediction accuracy (Table 1).

The major advantage of GAB is the genotypic data collected from an early plant growth stage (such as seedling) can be used to predict the phenotypic performance of mature individuals. Thus, it can considerably save time, money, and labor needed for the extensive phenotypic evaluation in multiple environments and years (Bohra et al., 2020; Kole et al., 2015). In addition, GAB allows a greater number of breeding selection cycles and genetic gain per unit time (Bhat et al., 2016). In the past few decades, the complete reliance on phenotypic selection has steadily changed to more increased use of genotypic-based methods for plant selection, especially made possible by NGS-based genotyping platforms (Jarquín et al., 2014; Matei et al., 2018; Qin et al., 2019; Ravelombola et al., 2020). The NGS technologies have enabled increased throughput, cost-effectiveness, and speed of genome-wide genotyping (Getachew, 2019). Before the emergence of NGS-based marker genotyping, the development of

![FIGURE 2](image_url) The role of NGS-based marker genotyping and high-throughput phenotyping (HTP) on the success of genomic selection (GS) in crop/soybean breeding. The NGS and HTP offers promising results of GS in the development of improved soybean varieties in less period of time span.
markers was expensive and laborious, and GAB, especially the GS, was constrained by the number of markers that could efficiently be assayed (Bhat et al., 2016). As a result, only markers in critical genomic regions were utilized to predict the presence or absence of agriculturally valuable traits (Varshney et al., 2014). Both GAB approaches, namely, MAS and GS, have different limitations and strengths, and use of either approaches in soybean breeding depends upon the trait architecture and heritability as discussed above (Mir et al., 2019; Varshney et al., 2014).

3 | WHOLE GENOME RESEQUENCING

Whole genome resequencing (WGRS) is the ultimate approach to detect all available genetic variations/polymorphism present in a BP of a particular crop species. However, the use of WGRS in marker genotyping requires the availability of high-quality reference genome sequence for a particular crop species. In this context, since the publication of the first high-quality soybean reference genome (Schmutz et al., 2010), the follow-up WGRS studies have allowed the whole genome investigations in soybean (Zhou et al., 2015). The comparison of the resequencing information generated from the BP with the high-quality reference genome has led to the identification of all kinds of genetic variation present in genetic populations. This approach has the advantage of providing potential access to whole genome level variations, such as SNPs, insertions, deletions, transversions, and copy number variants (Zhou et al., 2015). WGRS has increased the resolution of gene mapping and candidate gene identification. This has ultimately increased the applicability and success of GAB in soybean improvement (Figure 1; Jaganathan et al., 2020).

Among the various NGS-based marker genotyping approaches, the WGRS approach is the only approach to detect all genome-wide polymorphisms for their ultimate use in GAB (Bhat et al., 2016). The WGRS will provide highest marker density and resolution of gene mapping, as well as precise estimation of GEBVs; hence, among all the NGS genotyping methods, the WGRS will provide the maximum benefits from genomic revolution in crop improvement (Zhou et al., 2015).

However, it has some important limitation especially in the large and complex genomes such as soybean that prevents its routine commercial use in the soybean breeding programs. These limitations include high sequencing costs (long reads and chromosome scale scaffolding techniques are needed), requirement of complex data analysis, and demand for high-performance computing platforms. The continuous decline in sequencing cost will make use of WGRS feasible and cost effective in different applications of soybean breeding in the near future. Until then, the reduced representation sequencing especially

### Table 1: Genomic selection (GS) studies for different quantitative traits in soybean using high-throughput SNP genotyping platforms

| Trait                        | High-throughput SNP genotyping platform | Number of SNPs | Population size | Training population | Prediction accuracy | Model               | Reference                                      |
|-----------------------------|----------------------------------------|----------------|-----------------|---------------------|---------------------|---------------------|-----------------------------------------------|
| Yield and related traits    | SNP Chip                               | 2647           | 483             | 483                 | 0.26–0.81           | RR-BLUP             | Stewart-Brown et al. (2019)                   |
| Amino acid contents         | GBS                                    | 23,279         | 249             | 199                 | 0.25–0.61           | RR-BLUP             | Qin et al. (2019)                             |
| Chlorophyll content         | SNP Chip                               | 4089           | 172             | 100                 | 0.31–0.76           | RR-BLUP             | Ravelombola et al. (2019)                     |
| Grain yield                 | GBS                                    | 2395           | 139             | 55                  | 0.64                | G-BLUP              | Jarquin et al. (2014)                         |
| Yield-related traits        | SNP Chip                               | 4947           | 324             | 324                 | 0.56                | eBLUP               | Matei et al. (2018)                           |
| Yield-related traits        | SNP Chip                               | 5361           | 235             | –                   | 0.47–0.86           | RR-BLUP             | Ma et al. (2016)                              |
| Yield and protein content   | SNP Chip                               | 4141           | 1248            | 252–492             | 0.55–0.62           | G-BLUP              | Duhnen et al. (2017)                          |
| Yield                       | GBS                                    | 3000           | 227             | 227                 | 0.6                 | RR-BLUP and Bayesian models                   | Dordević et al. (2019)                       |
| Seed weight                 | SNP Chip                               | 31,045         | 309             | 97–197              | 0.75–0.87           | RR-BLUP             | Zhang et al. (2016)                           |
| Yield, plant height, and maturity | GBS                                  | 8140           | 301             | –                   | 0.71                | G-BLUP              | Jarquin et al. (2019)                         |
| Nematode resistance         | SNP Chip                               | 3782           | 234             | 117–201             | 0.43–0.48           | gBLUP               | Ravelombola et al. (2020)                     |
| Disease resistance          | GoldenGate assay                       | 1536           | 282             | 248                 | 0.30–0.64           | RR-BLUP             | Bao et al. (2015)                             |
| Trait                                | NGS genotyping (No. SNP/bin markers) | Genetic approach | Population type | Population size | Marker-trait associations/GS prediction accuracy | $R^2$   | Candidate genes                                                                 | Reference                  |
|--------------------------------------|--------------------------------------|-------------------|-----------------|-----------------|-----------------------------------------------|---------|---------------------------------------------------------------------------------|---------------------------|
| Agronomic traits                     | GBS (47,000)                         | GWAS              | NDP             | 304             | 25 QTNs                                       |         | Glyma19g41210                                                                    | Sonah et al. (2015)       |
| Seed protein and oil                 | GBS (12,072)                         | GWAS              | NDP             | 185             | 31 QTN                                        |         | —                                                                               | Li et al. (2019b)         |
| Disease resistance                   | GBS (11,811)                         | GWAS              | NDP             | 420             | 7 QTNs                                        |         | Glyma.01G106000, Glyma.11G084000, Glyma.11G084200, and Glyma.11G086600      | Wei et al. (2017)         |
| Disease resistance                   | GBS (32,836)                         | GWAS              | NDP             | 295             | 1 QTN                                         |         | Glyma.14G024300, Glyma.14G026300, Glyma.14G026500, and Glyma.14G026700      | dos Santos et al. (2019)  |
| Salt tolerance                       | WGRS (37,573)                        | GWAS              | NDP             | 234             | 6 QTNs                                        |         | Glyma.08g146100, Glyma.08g224400, Glyma.02g204300, Glyma.14g211300, and Glyma.08g157400 | Do et al. (2019)          |
| Salt tolerance                       | WGRS (37,400)                        | GWAS              | NDP             | 106             | 3 QTNs                                        |         | —                                                                               | Patil et al. (2016)       |
| Mycorrhizal colonization             | WGRS (47,000)                        | GWAS              | NDP             | 350             | 6 QTNs                                        |         | —                                                                               | Pawlowski et al. (2020)   |
| Disease resistance                   | GBS (7893)                           | GWAS              | NDP             | 130             | 4 QTNs                                        |         | —                                                                               | Bastien, et al. (2014)    |
| Disease resistance                   | GBS (8397)                           | GWAS              | NDP             | 101             | 4 QTNs                                        |         | Iquira et al. (2015)                                                           |                          |
| Disease resistance                   | WGRS (41,000)                        | GWAS              | NDP             | 127             | 1 QTN                                         |         | Glyma.01g048000                                                                 | Boudhrioua et al. (2020)  |
| Flooding tolerance                   | GBS (60,109)                         | GWAS              | NDP             | 25 QTNs         |                                               |         | Glyma.13g248000                                                                 | Yu et al. (2019)          |
| Seed quality traits                  | WGRS (16,469)                        | Linkage mapping   | RIL (Williams 82 × PI 483460B) | 188             | 18 QTLs                                       | 4.6–29%| —                                                                               | Patil et al. (2018)       |
| Disease resistance                   | WGRS (1536)                          | Linkage mapping   | RIL (Magellan’ × PI 438489B) | 247             | 2 QTLs                                        | 6.4–16.8%| —                                                                               | Klepadlo et al. (2018)    |
| Agronomic and seed quality traits    | GBS (20,084)                         | Linkage mapping   | RIL (CSSL3228 × NN1138–2) | 149             | 35 QTLs                                       | 10.31–14.35%| Glyma.13g249400                                                              | Zhang et al. (2018)       |

(Continues)
| Trait                          | NGS genotyping (No. SNP/bin markers) | Genetic approach | Population type                              | Population size | Marker-trait associations/GSprediction accuracy | $R^2$          | Candidate genes                                                                 | Reference          |
|-------------------------------|--------------------------------------|------------------|----------------------------------------------|-----------------|-----------------------------------------------|-----------------|--------------------------------------------------------------------------------|-------------------|
| Chlorophyll content traits    | GBS (2356)                           | Linkage mapping  | RIL (ZhongHuang 24 × Huaxia 3)               | 164             | 78 QTLs                                       | 5.10–16.55%    | Glyma01g15506, Glyma02g08910, Glyma02g11110, Glyma07g15960, Glyma15g19670, and Glyma15g19810 | Wang et al. (2020) |
| Pod dehiscence                | RAD-seq (4593)                       | Linkage mapping  | RIL (Heihe 43 × Heihe 18)                   | 260             | 6 QTLs                                        | 7.22–24.44%    | —                                                                              | Han et al. (2019)  |
| Aluminum tolerance            | RAD-seq (3426)                       | Linkage mapping  | RIL (ZH 24 × HX 3)                          | 160             | 5 QTLs                                        | 0.07–8.98%     | Glyma04g218700                                                                 | Wang et al. (2019) |
| Domestication-related traits  | GBS (35,303)                         | Linkage mapping  | RIL (Williams 82 × PI 468916; Williams 82 × PI 479752 | 151 and 510     | 132 QTLs                                      | 1.5–8.1%       | —                                                                              | Swarm et al. (2019) |
| Long-juvenile trait           | GBS (2958, 2272, 4016, 3329, 2260, and 2702) | Linkage mapping  | F$_2$ (PI 591429 × PI 628930; PI 240664 × BR121; PI 285096 × PI 591429; PI 159925 × PI 285096; H3 × PI 628951; PI 628805 × PI 591429) | 161, 113, 168, 129, 141, and 146 | 7 QTLs                                | 3.58–54.20%    | —                                                                              | Fang et al. (2019) |
| Seed quality and plant height | GBS (2977)                           | Linkage mapping  | RIL (A6 and 194D; 194D and A6)              | 100, 95         | 21 QTLs                                       | 4.1–56.80%     | —                                                                              | Heim and Gillman (2017) |
| Flowering and maturity        | GBS (2168)                           | Linkage mapping  | RIL (PI 290136 × PI 54687)                  | 169, 110        | 4 QTLs                                        | 5.78–40.44%    | —                                                                              | Wang et al. (2020) |
| Nitrogen fixation             | GBS (1450)                           | Linkage mapping  | RIL (Bossier’ × Embrapa 20)                 | 113             | 2 QTLs                                        | 14.93–18.13%   | —                                                                              | Grunvald et al. (2018) |
| Agronomic and stress response | GBS (2757)                           | Linkage mapping  | RIL (W05 × C08)                             | 552             | 15 QTLs                                       | 7.80–78.53%    | —                                                                              | Qi et al. (2014)   |
| Yield-related and quality traits | RAD-seq (2629)                    | Linkage mapping  | RIL (Zhonghuang 24 × Huaxia 3)              | 164             | 60 QTLs                                       | 4.30–32.56%    | —                                                                              | Liu et al. (2017)  |
| Nematode resistance           | WGRS (3509)                          | Linkage mapping  | RIL (Magellan × PI 438489B)                 | 246             | 3 QTLs                                        | 4.5–15.4%      | Glyma10g02150 and Glyma10g02160                                                | Xu et al. (2013)   |

Abbreviations: GBS, genotyping by sequencing; GWAS, genome-wide association studies; QTLs, quantitative trait loci; QTN, quantitative trait nucleotide; RIL, recombinant inbred line; WGRS, whole genome resequencing.
the genotyping by sequencing (GBS) seems to be more cost-effective option for large-scale marker discovery and GS (Tables 1 and 2; Bhat et al., 2016; Nepolean et al., 2018).

Efforts have been made in soybean utilizing the WGRS for genotyping and marker generation in gene mapping and domestication studies (Table 2). But such studies are very limited in soybean especially the use of WGRS for large-sized genetic population. For example, Lam et al. (2010) were the first group to resequence 31 wild and cultivated soybean genotypes to identify the pattern of genetic diversity and selection. More recently, Zhou et al. (2015) also resequenced 302 wild and cultivated soybean accessions, used this data in GWAS, and identified genes related to domestication, oil content, plant height, and pubescence form. Similarly, the resequencing data of the diverse panel of 350 soybean plant introductions were used to identify QTLs associated with mycorrhizal colonization in soybean (Pawlowski et al., 2020). Kim et al. (2019) also resequenced 245 soybean accessions to examine their genomic architectures based on domestication and improvement traits, and they classified accessions into three wild-type, two landrace, and two improvement subgroups based on various population analyses. However, such studies are very few. In the near future, the cost of WGRS is predicted to decrease to a level where it becomes feasible to use WGRS in commercial soybean breeding programs.

4 | GBS: SIGNIFICANCE IN SOYBEAN BREEDING

In the past decades, advances in the genome sequencing has provided multiple NGS-based reduced representation genotyping platforms. These include reduced representation libraries (RRLs), restriction-site associated DNA sequencing (RAD-seq), GBS, complexity reduction of polymorphic sequence (CRoPS), genome reduction on restriction-site conservation (GR-RSC), double digest RAD-Seq (ddRADSeq), and its modified forms (Andrews et al., 2016; Campbell et al., 2018). Among them, GBS is the most powerful, simple, high-throughput, and cost-effective technique for obtaining SNP markers (Bhat et al., 2016; Hap et al., 2019). GBS is a modified form of RAD-seq approach based on NGS and is a highly multiplexed and simple system. The GBS offers extraordinary features such as ease of handling, a smaller number of PCR and purification steps, no need for a reference sequence, low cost, no size fractionation, and efficient barcoding technique (Davey et al., 2011; Getachew et al., 2019). It enables the detection of thousands of millions of SNPs in large collections of lines that can be used for genetic diversity analysis, linkage mapping, GWAS, GS, and evolutionary studies (Figure 1; Beissinger et al., 2013; Wickland et al., 2017). This approach combines marker discovery and genotyping of large populations, making it an excellent marker platform for breeding applications even in the absence of a reference genome sequence or previous polymorphism discovery (Bhat et al., 2016). The GBS method offers a greatly simplified library production procedure more amenable to use on large numbers of individuals/lines (Elshire et al., 2011; Poland et al., 2012).

Recently, reduced cost of GBS has made it increasingly more feasible in soybean compared with WGRS as a method of genotyping (Tables 1 and 2; Alipour et al., 2019; Fang et al., 2019; Gutierrez-Gonzalez et al., 2019; Jiang et al., 2020). GBS has been routinely used in the gene mapping and other studies in soybean (Table 2). GBS has been demonstrated to be an efficient technique for high-resolution genetic mapping and precision breeding in soybean (Getachew, 2019). Recently, GBS has been used to sequence collections of different mapping population (such as RIL and F2) and diversity panels, to analyze and map various traits in soybean (Table 2). GBS has considerably facilitated the high-resolution gene mapping and candidate gene identification in soybean (Fang et al., 2019; Sonah et al., 2015; Wang et al., 2020). GBS have also been used for de novo genotyping of soybean breeding panels and to develop accurate GS models (Table 2; Stewart-Brown et al., 2019). The GEBV prediction accuracies were observed to be moderate to high ranging from 0.26 to 0.81 for yield and yield-related traits by using this genotyping method (Stewart-Brown et al., 2019). By using the GBS, an improvement of 0.1 to 0.2 in the prediction accuracy for GEBV has been compared with array-based and SSR marker systems (Heslot et al., 2013). Through intensive multiplexing, this approach can reduce the cost to below 10 dollars per sample (Bhat et al., 2016). The GBS cost per individual is much lower compared with array-based and other NGS-based markers in soybean (Bassi et al., 2016; Getachew, 2019). Moreover, GBS covers much higher fraction of the genome compared with even the densest SNP arrays currently available in soybean plant (Bhat et al., 2016; Gorjanc et al., 2015). The SNP arrays are typically developed from a limited sample of individuals, whereas GBS can capture genetic variation that is specific to a population or family of interest. GBS has the advantage that markers are discovered using the population to be genotyped, thus minimizing ascertainment bias. The flexibility, low cost, and GEBV prediction accuracy of GBS make this an ideal approach for GAB applications in soybean in the current era (Bhat et al., 2016).

5 | NGS-BASED GENOTYPING: GENE DISCOVERY AND MAS IN SOYBEAN

Prior to the use of MAS approach in crop breeding, it is a prerequisite to have initial knowledge of the major-effect gene/QTL serving as a target in the selection process. In this regard, the NGS-based genotyping technologies have greatly benefited in increasing the resolution of gene mapping and tagged the gene/QTL very close to the adjacent maker (Figure 1). For example, in the GWAS analysis, the use of NGS has made it possible to genotype large populations of plants with a higher density of markers that was previously impossible, and this contributes directly to increased mapping resolution (Jing et al., 2018; Li et al., 2019a). Use of diverse and large population in GWAS analysis allows the detection of more recombination break points that in turn help in the identification of candidate genes at higher precision.
Many studies have utilized NGS-based genotyping for GWAS analysis in soybean for the different traits; and these studies have revealed considerable success in the candidate gene identification for particular traits of interest (see details in Table 2). For example, Yu et al. (2019) utilized the RAD-seq approach and identified candidate gene underlying the major QTL regulating flooding tolerance in soybean. Similarly, many other studies have demonstrated NGS-based genotyping mediated candidate gene identification such as nitrogen fixation (Torkamaneh et al., 2020), plant height and primary branches in soybean (Borah et al., 2018), agronomic traits (Sonah et al., 2015), disease resistance (Zhao et al., 2015), and protein content (Sui et al., 2020). The NGS-based WGRS has considerably increased the power of bulk segregant analysis (BSA) and its modified approaches and is being widely used in many plant species including soybean. For example, Song et al. (2017) used WGRS to resequence contrasting DNA pools in the BSA analysis, and they have simultaneously identified two major genes regulating the cotyledon color in soybean. In addition, many studies have used NGS-based techniques in the BSA approach to identify candidate genes for various traits in soybean such as soybean mosaic virus (Yang et al., 2020), charcoal rot resistance (da Silva et al., 2020), flowering time (Watanabe et al., 2017a), phytophthora resistance (Cheng et al., 2017), and powdery mildew resistance (Jiang et al., 2019). The modified approach of BSA such as MutMap based on WGRS of pooled DNA samples collected from the phenotypic extremes of a segregating population (derived from mutant x wild-type cross) was successfully demonstrated in soybean leading to the identification of mutant gene regulating the mutant phenotype (Al Amin et al., 2019). Similarly, the gene identification via QTL-Seq based on WGRS of bulked DNA samples collected from the phenotypic extremes of a segregating population (derived from mutant x wild-type cross) was successfully demonstrated in soybean leading to the identification of mutant gene regulating the mutant phenotype (Al Amin et al., 2019). Similarly, Zhong et al. (2019) combined the QTL-Seq approach with genetic mapping to identify RpsX gene in soybean cultivar Xiu94-11, which provides broad-spectrum resistance against the Phytophthora sojae isolates. Compared with previous marker systems, NGS is very efficient for map-based gene cloning, as it can perform SNP discovery, SNP validation, and SNP genotyping simultaneously in a mapping or mutant population (Table 2). It has been demonstrated that NGS also possesses great capability to resolve the genome duplication issues in a complex and paleopolyploid soybean crop (Xu et al., 2013).

MAS is an earliest form of GAB and has been extensively used for the breeding of improved crop varieties (Cobb et al., 2019). Application of MAS in crop breeding needs a marker tagged to a particular gene/QTL regulating the phenotypic variation for the trait of interest (Collard & Mackill, 2008). Trait-linked marker is used in MAS, which involves the introgression of tagged gene/QTL into elite background of superior crop variety. As discussed in detail above, high-throughput SNP genotyping has greatly increased the resolution of gene tagging, thereby significantly increasing the success of MAS (Figure 1). HTG has been successfully used to introgress target loci into elite varieties to improve their performance for specific trait (Jaganathan et al., 2020; Varshney et al., 2014), Patil et al. (2016) developed a high-throughput SNP-marker assay for MAS of salinity tolerance in soybean. Similarly, Shi et al. (2015) have generated a SNP-marker assay for high-throughput selection of soybean cyst nematode (SCN), and they demonstrated that the use of one or two of these markers is sufficient for high-throughput MAS of soybean plants showing enhanced SCN resistance. In addition, the NGS-based genotyping has considerably reduced the linkage drag involved in the MAB (Varshney et al., 2014). Breeding program involved in the transfer of favorable allele/gene via MAB from the wild unadapted parent into the elite cultivated recipient parent often leads to linkage drag (Mir et al., 2014). Breeders do not want this linkage drag, that is, transfer of undesirable genes linked to desirable gene into the elite recipient parent, because they often make the new developed cultivar unacceptable to the farmers and consumers (Mir et al., 2014). In this regard, NGS has played a critical role by rapidly identifying the plant lines in which the linkage between undesirable and desirable genes has broken due to recombination process. In one example, NGS was used to identify the recombinants in which the linkage between a favorable allele conferring SCN resistance and a deleterious gene affecting seed quality and 100-seed weight was broken (St-Amour et al., 2020). The wild unadapted genotype (PI 494182) used as donor parent possesses the favorable and the unfavorable alleles in coupling phase; hence, the deep NGS sequencing carried on a bigger sized segregation population within the target region identified a recombinant individual in which the linkage had been broken (St-Amour et al., 2020). In such cases, if the breeders know the functional polymorphism(s) for the desirable and undesirable genes, it would be very easy to identify lines/genotypes in which linkage drag has been broken by recombination event. Therefore, following the identification of recombinants, they can immediately serve as a donor to introduce new genetic variation into a breeding pipeline. Hence, the NGS technology has a great potential to identify recombinants with no linkage drag and might be the quickest technology to provide new sources of genetic variation for their ultimate use in soybean breeding.

**NGS-BASED GENOTYPING VERSUS SUCCESS OF GS IN SOYBEAN**

Sanger sequencing was the initial technique used for the whole genome sequencing (WGS) of model crop plants (Edwards & Batley, 2010). Availability of WGS has increased the scope of sequence-based markers such as SSR and SNP and has greatly revolutionized the marker technology in plants (Zargar et al., 2015). But many factors such as high cost, more time, and information restricted to target individual has limited the use of Sanger technology for specific gene discovery (Ray & Satya, 2014). Hence, this sequencing technology is not feasible for commercial use either in small or large breeding programs (Bhat et al., 2016). The recent advances in the NGS technologies and powerful computational pipelines has reduced the cost of WGS/WGRS to great extent allowing the discovery, sequencing, and genotyping of hundreds of thousands of markers in a
single step (Stapley et al., 2010). As discussed in the previous section, NGS-based WGRS is not currently commercially feasible for large breeding programs due to its higher cost, but in the near future, the cost for resequencing a genome will decrease to only a few hundred US dollars. Until then, the NGS-marker technologies based on reduced representation sequencing is the best option for large-scale marker discovery, especially for the large and complex soybean genome. These NGS-based marker techniques represent the partial genome of the soybean, and they even can be performed in the absence of a reference sequence (Ray & Satya, 2014; Toonen et al., 2013). Among these NGS technologies, RAD-seq (or its variants) and GBS have already proven to be efficient and effective techniques for GAB and were frequently used for GS studies in different crop plants (Tables 1 and 2; Glaubitz et al., 2014; Yang et al., 2012). In addition, the NGS has allowed the development of high-density SNP chips for conducting HTG in soybean (Table 1).

Multiple features of NGS-marker technologies, namely, low cost, genome-wide marker coverage, higher speed and throughput, and higher marker density have allowed the geneticist to study the inheritance of many traits at the nucleotide level precision (Chung et al., 2017). In contrast, the previous low-throughput marker technologies (such as RFLP, AFLP, and SSR) have restricted the use of markers only in critical genomic regions such as MAS, to predict the presence or absence of agriculturally important traits (Bhat et al., 2016; Varshney et al., 2014). These marker types were not efficient and cost effective to get the better results from the GS, and thus, the GAB was limited to only MAS. However, the availability of NGS genotyping platforms has widened the scope of GAB in crop improvement leading to the great success of GS, leading to substantial shift from the phenotyping-based selection to increased reliance on genotyping-based selection (Houston et al., 2020). Compared with array-based marker systems, the NGS genotyping has offered many advantages such as lack of ascertainment bias, low genotyping cost (per sample cost < $20 USD), and higher marker density that makes them marker of choice for GS (Bhat et al., 2016; Poland et al., 2012). In case of GS, the marker density and size of reference population are the most important factors affecting the efficiency and accuracy of GS in the prediction of superior genotypes, and both of these factors are limited by high genotyping cost (Qin et al., 2019; Varshney et al., 2014) that are now resolved by NGS genotyping technologies (Bhat et al., 2016; Jarquin et al., 2014). Moreover, the prediction accuracy of GS is also affected by population structure that often leads to biased accuracies of genomic predictions (Stewart-Brown et al., 2019; Wray et al., 2013). NGS-based marker genotyping compared with earlier marker systems (SSR and array-based markers) provides higher genome-wide marker density to precisely determine the population structure of TP; thus, it enables accurate estimation of the GEBV of BP (Isidro et al., 2015; Qin et al., 2019).

The beneficial features of NGS technology have made the NGS-based genotyping as a cost-effective and efficient agrigenomics tool for performing GS in soybean (Kirst et al., 2011; Metzker, 2010; Poland et al., 2012; Toonen et al., 2013). For example, Poland et al. (2012) used the NGS-based genotyping in GS, and they demonstrated an increase of 0.1 to 0.2 GEBV prediction accuracies compared with other established marker platform in crop species. The GS has been attempted in soybean for grain yield traits using GBS, and genome-wide prediction accuracies were investigated to be high (average of 0.64). GS was suggested to have good potential for grain yield in soybean (Jarquin et al., 2014). Zhang et al. (2016) also demonstrated higher prediction accuracy of GS (0.75–0.87) compared with MAS (0.62–0.75) for grain weight in soybean using high-throughput SNP genotyping. Thus, GS outperforms MAS for grain weight in soybean. In addition, the GS was also demonstrated to be superior compared with phenotypic selection for multiple traits such as grain yield, plant height, insertion of first pod, days to maturity and 1000-grain weight in soybean, and selection time is reduced in 50%, that is, from 6 to 3 years (Matei et al., 2018). Besides, many studies have demonstrated that GS can reduce the time required to complete a selection cycle in crop plants, which can lead to increased production of this commercially important crop (Bhat et al., 2016; Qin et al., 2019). List of studies involving the use of high-throughput SNP- and NGS-marker genotyping for conducting GS in soybean for the different traits showing moderate to high predication accuracy are presented in Table 1. All these studies have demonstrated the great potential of GS in accelerating the soybean production and productivity. Presently, the cost of NGS and WGRS genotyping is still high to commercially use these platforms for conducting GS in a soybean (Bhat et al., 2016; Varshney et al., 2019). The high cost has limited the use of these genotyping platforms to the smaller plant population. The reduction of sequencing cost in the future will allow use of genotyping at commercial scale in the large collection of soybean germplasm stored in the genebanks to predict the phenotype at higher accuracy (Bhat et al., 2016, 2020). The NGS-marker genotyping have been documented to give higher GS accuracy than array-based markers on the same lines of crop plants, even in the case of 43.9% missing data (Heslot et al., 2013). All of these features have made the sequence-based genotyping an ideal approach for GS and its successful application in soybean breeding (Figure 2).

### 7 | HIGH-THROUGHPUT PHENOMICS: CURRENT SNAG IN SOYBEAN BREEDING

In the past decade, the advances made in the NGS and HTG technologies have greatly improved our understanding of crop genomes. These advances have considerably increased the scope of crop breeding and genetics (Zargar et al., 2015). Sequencing of plant and crop genomes has become a routine process (Chen et al., 2019). Due to the advances in NGS technology, draft genome sequences of multiple legume crops including soybean have now become available (Legume Information System, https://legumeinfo.org). The allelic variation has also been identified by using WGRS in soybean (Bhat et al., 2016; Patil et al., 2019; Zhou et al., 2015). Genotyping cost has also considerably reduced due to the availability of HTG platforms in soybean (Pandey et al., 2016). However, to harness the true potential of the genomic revolution in soybean improvement, high-quality genomic data are
needed to be integrated with precise and accurate phenotypic data in a real-world environment (Figures 1 and 2; Furbank & Tester, 2011). The traditional manual phenotyping used in soybean was laborious, time consuming, destructive, and biased that often led to inaccurate results and at times resulted in failure of breeding programs (Bhat et al., 2016). The use of HTP technologies can considerably reduce the cost and labor for the phenotyping of different traits in soybean and increase the accuracy and precision in the crop breeding compared with traditional manual phenotyping (Moreira et al., 2019, 2020; Singh et al., 2016; Trevisan et al., 2020). So far, very limited efforts have been made to develop high-throughput advanced phenotypic approaches for crop breeding (Mir et al., 2019). Thus, in the current era of soybean breeding research, this has become a major snag in utilizing the abundant high-quality available genomic data (www.soybase.org) that in turn affects the progress of breeding research in soybean (Edwards et al., 2013; Stamp & Visser, 2012).

Soybean breeders often face a huge genotype–phenotype gap (GP Gap). Hence, an immediate need has been felt to revolutionize crop phenomics in order to reap the true fruits of genomic revolution.

Recently, researchers and scientist from different scientific disciplines are collaborating to alleviate this phenomic snag by generating high-throughput phenotypic facilities that can have potential to scan and record the data from hundreds of thousands of plants in a day in a sophisticated manner (Cobb et al., 2013; Mir et al., 2019). By considering the need of phenomics, recently, an International Plant Phenomics Initiative was launched (http://www.plantphenomics.org/). The high-throughput phenomics facilities make use of sophisticated noninvasive imaging, spectroscopy, image analysis, robotics, and high-performance computing facilities, thus saving time, labor, and cost. The phenomics revolution is suggested to assist in the collection of the high-quality, accurate phenotyping data, which is necessary and useful for meaningful genetic dissection and GAB, such as GWAS, linkage mapping, gene cloning, QTL meta-analysis, MAS, GS, and TILLING (Bhat et al., 2016). Currently, the high cost of HTP limits its scope to some small experimental programs. There are limited reports in soybean demonstrating the combined use of HTP and HTG for gene discovery or GAB. However, the combined use of HTP and HTG has been reported in other crop plants such as maize (Wang et al., 2019; Zhang et al., 2017), rice (Chattopadhyay et al., 2019; Tanger et al., 2017), wheat (Condorelli et al., 2018; Lyra et al., 2020; Rutkoski et al., 2016), sorghum (Watanabe et al., 2017b), and cotton (Pauli et al., 2016). These studies have revealed promising results and have suggested a great potential of HTP in crop breeding. Presently, there are a number of factors that restrict the use of HTP in commercial soybean breeding such as higher construction and maintenance cost of HTP facility, high cost of HTP, statistical complexity for the analysis of HTP data, and lack of skilled person handling the HTP data. Hence, for the promotion of HTP in soybean breeding in the future, it is required to develop cost-effective, easy to handle, and well-sophisticated data analysis infrastructure such as HTPheno (Hartmann et al., 2011) and IAP (Klukas et al., 2014): incorporating the open-source software ImageJ needs to be developed and popularized.

8 | PERSPECTIVES AND CONCLUSION

In the past century, conventional breeding has contributed significantly to soybean improvement and development of soybean varieties with improved yield, quality, and tolerance to multiple stresses. However, conventional methods of crop improvement are not able to maintain pace with the accelerated growing population and climate change (Gosal et al., 2009). GAB is suggested to have great potential in overcoming the future challenges in soybean improvement (Collard & Mackill, 2008). But the true potential of GAB can be harnessed only by precisely and accurately identifying the marker–trait associations (via linkage mapping and GWAS) and estimation of GEBV (via GS). The gene identification process in both QTL and GWAS mapping needs precise genotyping and phenotyping data. Previously, crop breeders have observed a large gap between the QTL/gene identification and their subsequent deployment in soybean improvement. Till now, only few successful stories of MAS are available in soybean (www.soybase.org). The main reason behind the failure of MAS in soybean improvement is the use of manual phenotyping and low-throughput marker system that has created a GP gap. Recently, the advances in HTP and HTG have narrowed this GP gap in crop selection and gene identification (Cobb et al., 2013; Mir et al., 2019).

As mentioned above, the combination of phenotypic and allelic data will greatly facilitate the identification of genetic loci that are linked to key agronomic traits in soybean (Bohra et al., 2020; Torkamaneh et al., 2020; Yu et al., 2019). Recent advancements in crop phenomics and genomics providing many high-throughput platforms, along with statistical methods and computational tools, have allowed next-generation GAB in crop plants. The combined use of these advanced technologies can precisely and accurately identify genes/QTLs, as well as their effective utilization in soybean breeding (Figures 1 and 2; Patil et al., 2016; Moreira et al., 2020). Although high-throughput SNP genotyping technologies have entirely changed marker application in soybean breeding, they have enabled research communities to use GWAS and GS routinely for soybean improvement. However, to harness real benefits from genomic studies, these marker technologies must be combined with HTP to achieve valuable genetic gain from complex traits. In soybean, very few studies involving the combined use of both HTP and HTG have been published so far. The reason for this is the higher cost involved in the large-scale field-based HTP. It has been predicted that in the near future, new innovations in crop phenotyping technology will allow HTP affordable for its commercial use in the soybean breeding programs. This will definitely expand the scale of germplasm evaluation and will allow the rapid development of improved soybean varieties. The declining cost of DNA sequencing will also make WGRS-based genotyping more feasible and cost-effective. At present, for breeding-based applications such as GS, sequencing-based genotyping using genome-reduction techniques such as GBS and RAD-sequencing seems to be more cost-effective.

ACKNOWLEDGMENT

This research was funded by the Ministry of Science and Technology of China (2017YFE0111000).
CONFLICT OF INTEREST
The authors declare no competing financial interests.

ETHICS STATEMENT
Not applicable.

AUTHOR CONTRIBUTIONS
Javaid Akhter Bhat and Deyue Yu have contributed equally in the development of this manuscript.

DATA AVAILABILITY STATEMENT
The data sets used during the current study are available from the corresponding author on reasonable request.

ORCID
Javaid Akhter Bhat https://orcid.org/0000-0001-9136-6275

REFERENCES
Ahmar, S., Gill, R. A., Jung, K. H., Faheem, A., Qasim, M. U., Mubeen, M., & Zhou, W. (2020). Conventional and molecular techniques from simple breeding to speed breeding in crop plants: Recent advances and future outlook. International Journal of Molecular Sciences, 21(7), 2590. https://doi.org/10.3390/ijms21072590

Al Amin, G. M., Kong, K., Sharmin, R. A., Kong, J., Bhat, J. A., & Zhao, T. (2019). Characterization and rapid gene-mapping of leaf lesion mimic phenotype of spl-1 mutant in soybean (Glycine max (L.) Merr.). International Journal of Molecular Sciences, 20(9), 2193. https://doi.org/10.3390/ijms20092193

Allpouir, H., Bai, G., Zhang, G., Bihamta, M. R., Mohammadi, V., & Peyghambari, S. A. (2019). Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. PLoS One, 14(1), e0208614. https://doi.org/10.1371/journal.pone.0208614

Alqudah, A. M., Sallam, A., Baenziger, P. S., & Börner, A. (2020). GWAS: A l i p o u r, H., Bai, G., Zhang, G., Bihamta, M. R., Mohammadi, V., & Peyghambari, S. A. (2019). Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. PLoS One, 14(1), e0208614. https://doi.org/10.1371/journal.pone.0208614

Aldquah, A. M., Sallam, A., Baenziger, P. S., & Börner, A. (2020). GWAS: Fast-forwarding gene identification and characterization in temperate cereals: Lessons from barley—a review. Journal of Advanced Research, 22, 119-135. https://doi.org/10.1016/j.jare.2019.10.013

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Athar, H. R., & Ashraf, M. (2009). Strategies for crop improvement against salinity and drought stress: An overview. In Salinity and water stress (pp. 1–16). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-9065-3_1

Bao, Y., Kurle, J. E., Anderson, G., & Young, N. D. (2015). Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Molecular Breeding, 35(6), 128. https://doi.org/10.1007/s11032-015-0324-3

Bass, M. F., Bentley, A. R., Charnet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Science, 242, 23-36. https://doi.org/10.1016/j.plantsci.2015.08.021

Bastien, M., Sonah, H., & Belzile, F. (2014). Genome wide association mapping of Sclerotinia sclerotiorum resistance in soybean with a genotyping-by-sequencing approach. The Plant Genome, 7(1), plantgenome2013-10.

Beissinger, T. M., Hirsch, C. N., Sekhon, R. S., Foerster, J. M., Johnson, J. M., Muttoni, G., Vaillancourt, B., Buell, C. R., Kaeppler, S. M., & de Leon, N. (2013). Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics, 193(4), 1073–1081. https://doi.org/10.1534/genetics.112.147710

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., Tyagi, A., Mushtaq, M., Jain, N., Singh, P. K., & Singh, G. P. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Frontiers in Genetics, 7, 221.

Bhat, J. A., Deshmukh, R., Zhao, T., Patil, G., Deokar, A., Shinde, S., & Chaudhary, J. (2020). Harnessing high-throughput phenotyping and genotyping for enhanced drought tolerance in crop plants. Journal of Biotechnology, 324, 248–260.

Bohra, A., Saxena, K. B., Varshney, R. K., & Saxena, R. K. (2020). Genomics-assisted breeding for pigeonpea improvement. Theoretical and Applied Genetics, 133(5), 1721–1737.

Borah, J., Singode, A., Talukdar, A., Yadav, R. V., & Sarma, R. N. (2018). Genome-wide association studies (GWAS) reveal candidate genes for plant height and number of primary branches in soybean [Glycine max (L) Merrill]. Indian Journal Genetics and Plant Breeding, 78, 460–469.

Boudhrioua, C., Bastien, M., Torkamanef, D., & Belzile, F. (2020). Genome-wide association mapping of Sclerotinia sclerotiorum resistance in soybean using whole-genome resequencing data. BMC Plant Biology, 20, 1–9.

Campbell, E. O., Brunet, B. M., Dupuis, J. R., & Sperling, F. A. (2018). Would an RRS by any other name sound as RAD? Methods in Ecology and Evolution, 9(9), 1920–1927. https://doi.org/10.1111/2041-210X.13038

 Chattapatdhyay, K., Behera, L., Bagchi, T. B., Sardar, S. S., Moharana, N., Patra, N. R., Chakraborti, M., Das, A., Mamdi, B. C., Sarkar, A., & Ngankham, U. (2019). Detection of stable QTLs for grain protein content in rice (Oryza sativa L.) employing high throughput phenotyping and genotyping platforms. Scientific Reports, 9(1), 1–16.

Chaudhary, J., Patil, G. B., Sonah, H., Deshmukh, R. K., Vuong, T. D., Vailiyodan, B., & Nguyen, H. T. (2015). Expanding omics resources for improvement of soybean seed composition traits. Frontiers in Plant Science, 6, 1021.

Chen, F., Song, Y., Li, X., Chen, J., Mo, L., Zhang, X., Lin, Z., & Zhang, L. (2019). Genome sequences of horticultural plants: Past, present, and future. Horticulture Research, 4(1), 1–23.

Cheng, Y., Ma, Q., Ren, H., Xia, Q., Song, E., Tan, Z., Li, S., Zhang, G., & Nian, H. (2017). Fine mapping of a Phytophthora-resistance gene RpsWY in soybean (Glycine max L.) by high-throughput genome-wide sequencing. Theoretical and Applied Genetics, 130(5), 1041–1051. https://doi.org/10.1007/s00122-017-2869-5

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. https://doi.org/10.1007/s13580-017-0297-8

Collard, B. C., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics, 126(4), 867–887. https://doi.org/10.1007/s00122-013-2066-0

Collard, B. C., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., & Ng, E. H. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: Lessons from the breeder’s equation. Theoretical and Applied Genetics, 132(3), 627–645. https://doi.org/10.1007/s00122-019-03317-0

Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, 363 (1491), 557–572. https://doi.org/10.1098/rstb.2007.2170

Concibido, V. C., Denny, R. L., Lange, D. A., Orf, J. H., & Young, N. D. (1996). RFLP mapping and marker-assisted selection of soybean cyst nematode resistance in PI 209332. Crop Science, 36(6), 1643–1650. https://doi.org/10.2135/cropsci1996.00111833x003600600638x

Condorelli, G. E., Maccaferri, M., Newcomb, M., Andrade-Sanchez, P., White, J. W., French, A. N., Sciarra, G., Ward, R., & Tuberosa, R. (2018). Comparative aerial and ground based high throughput phenotyping
Klukas, C., Chen, D., & Pape, J. M. (2014). Integrated analysis platform: an LAM, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., Li, M. W., Kirst, M., Resende, M., Munoz, P., & Neves, L. (2011, December). Capturing and genotyping the genome-wide genetic diversity of trees for association mapping and genomic selection. In BCM Proceedings (Vol. 5, No. 7, pp. 1-2). BioMed Central.

Klepadlo, M., Meinhardt, C. G., Vuong, T. D., Patil, G., Bachleda, N., Ye, H., Robbins, R. T., Li, Z., Shannon, J. G., Chen, P., & Meiksem, K. (2018). Evaluation of soybean germplasm for resistance to multiple nematode species: Heterodera glycines, Meloidogyne incognita, and Rotylenchulus reniformis. Crop Science, 58(6), 2511–2522. https://doi.org/10.2135/cropsci2018.05.0327

Kulkas, C., Chen, D., & Pape, J. M. (2014). Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. Plant Physiology, 165(2), 506–518.

Kole, C., Muthamilarasan, M., Henry, R., Edwards, D., Sharma, R., Abberton, M., Batley, J., Bentley, A., Blakeney, M., Bryant, J., & Cai, H. (2015). Application of genomics-assisted breeding for generation of climate resilient crops: Progress and prospects. Frontiers in Plant Science, 6, 563.

Kulkarni, K. P., Kim, M., Shannon, J. G., & Lee, J. D. (2016). Identification of quantitative trait loci controlling soybean seed weight in recombinant inbred lines derived from PI 483463 (Glycine soja) × ‘Hutcheson’ (G. max). Plant Breeding, 135(5), 614–620. https://doi.org/10.1111/plb.12407

Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., Li, M. W., He, W., Qin, N., Wang, B., & Li, J. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nature Genetics, 42(12), 1053–1059. https://doi.org/10.1038/ng.715

Li, D., Zhao, X., Han, Y., Li, W., & Xie, F. (2019a). Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. Genomics, 111(1), 90–95. https://doi.org/10.1016/j.jgen.2018.01.004

Li, X., Kong, C., Yu, H., Liu, X., Fang, Z., Liu, Y., Yang, L., Zhuang, M., Wang, Y., Lv, H., & Zhang, Y. (2019b). Identification of a major QTL for seed number per silique in cabbage (Brassica oleracea L. var. capitata) using genotyping by sequencing. Euphytica, 215(7), 133. https://doi.org/10.1007/s10681-019-2409-2

Liu, D., Yan, Y., Fujita, Y., & Xu, D. (2018). Identification and validation of QTLs for 100-seed weight using chromosome segment substitution lines in soybean. Breeding Science, 68, 442–448. https://doi.org/10.1070/bbs.17127

Liu, N., Li, M., Hu, X., Ma, Q., Mu, Y., Tan, Z., Xia, Q., Zhang, G., & Nian, H. (2017). Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean RILs population by RAD-sequencing. BMC Genomics, 18(1), 466. https://doi.org/10.1186/s12864-017-3854-8

Lyra, D. H., Virlet, N., Sadeghi-Tehran, P., Hassall, K. L., Wingen, L. U., Orford, S., Griffiths, S., Hawkesford, M. J., & Slavov, G. T. (2020). Functional QTL mapping and genomic prediction of canopy height in wheat measured using a robotic field phenotyping platform. Journal of Experimental Botany, 71(6), 1885–1898. https://doi.org/10.1093/jxb/errz545

Ma, Y., Reif, J. C., Jiang, Y., Wen, Z., Wang, D., Liu, Z., Guo, Y., Wei, S., Wang, S., Yang, C., & Wang, H. (2016). Potential of marker selection to increase prediction accuracy of genomic selection in soybean (Glycine max L.). Molecular Breeding, 38(8), 113. https://doi.org/10.1007/s11032-016-0504-9

Matei, G., Woyan, L. G., Milioni, A. S., de Bem Oliveira, I., Zdzierski, A. D., Zanella, R., Coelho, A. S. G., Finatto, T., & Benin, G. (2018). Genomic selection in soybean: Accuracy and time gain in relation to phenotypic selection. Molecular Breeding, 38(9), 117. https://doi.org/10.1007/s11032-018-0872-4

Metzker, M. L. (2010). Sequencing technologies—The next generation. Nature Reviews Genetics, 11(1), 31–46. https://doi.org/10.1038/nrg2626

Meuwissen, T. H., & Goddard, M. E. (2001). Prediction of identity by descent probabilities from marker-haplotype types. Genetics Selection Evolution, 33(6), 605–634. https://doi.org/10.1186/1297-9686-33-6-605

Mir, R. R., Bhat, J. A., Jan, N., Singh, B., Razdan, A. K., Bhat, M. A., Kumar, A., Srivastava, E., & Malviya, N. (2014). Role of molecular markers. In Alien Gene Transfer in Crop Plants (Vol. 1) (pp. 165–185). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-8585-8_7

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., & Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. Plant Science, 282, 60–72. https://doi.org/10.1016/j.plantsci.2019.01.007

Moreira, F. F., Hearst, A. A., Cherkauer, K. A., & Rainey, K. M. (2019). Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. Plant Methods, 15(1), 1–9.

Moreira, F. F., Oliveira, H. R., Volencen, J. J., Rainey, K. M., & Brito, L. F. (2020). Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. Frontiers in Plant Science, 11, 681. https://doi.org/10.3389/fpls.2020.00681

Nepolean, T., Kaul, J., Mukri, G., & Mittal, S. (2018). Genomics-enabled next-generation breeding approaches for developing system-specific drought tolerant hybrids in maize. Frontiers in Plant Science, 9, 361. https://doi.org/10.3389/fpls.2018.00361

Pandey, M. K., Roorkiwal, M., Singh, V. K., Ramalingam, A., Kudapa, H., Thudi, M., Chitikineni, A., Rathore, A., & Varshney, R. K. (2016). Emerging genomic tools for legume breeding: Current status and future prospects. Frontiers in Plant Science, 7, 455.
Sharmin, R. A., Bhuivyan, M. R., Lv, W., Yu, Z., Chang, F., Kong, J., Bhat, J. A. & Zhao, T. (2020). RNA-Seq based transcriptomic analysis revealed genes associated with seed-flooding tolerance in wild soybean (Glycine soja Sieb. & Zucc.). *Environmental and Experimental Botany*, 171, 103906.

Shi, Z., Liu, S., Noe, J., Arelli, P., Meksem, K. & Li, Z. (2015). SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance. *BMGC Genomes*, 16(1), 314. https://doi.org/10.1186/s12864-015-1531-3

Singh, A., Ganapathysubramanian, B., Singh, A. K. & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, 21(2), 110–124.

Sonah, H., O’Donohue, L., Cober, E., Rajcan, I., & Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soybean. *Plant Biotechnology Journal*, 13(2), 211–221. https://doi.org/10.1111/pbi.12249

Song, J., Li, Z., Liu, Z., Guo, Y. & Qiu, L. J. (2017). Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Frontiers in Plant Science*, 8, 919. https://doi.org/10.3389/fpls.2017.00919

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., Atlínn, G., Jannink, J. L. & McCouch, S. R. (2015). Genomic selection and association mapping in rice (Oryza sativa): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genetics*, 11(2), e1004982. https://doi.org/10.1371/journal.pgen.1004982

St-Amour, V. T. B., Mimeo, B., Torkamaneh, D., Jean, M., Belzile, F. & O’Donohue, L. S. (2020). Characterizing resistance to soybean cyst nematode in PI 494182, an early maturing soybean accession. *Crop Science*, 60(4), 2053–2069.

Stampley, J., Reger, J., Feulner, P. G., Smadja, C., Galindo, J., Ekblom, R., Benninson, C., Ball, A. D., Beckham, A. P. & Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology & Evolution*, 25(12), 705–712. https://doi.org/10.1016/j.tree.2010.09.002

Steiner, B., Michel, S., Maccferri, M., Lemmens, M., Tuberosa, R. & Buerstmayr, H. (2019). Exploring and exploiting the genetic variation of Fusarium head blight resistance for genomic-assisted breeding in the elite durum wheat gene pool. *Theoretical and Applied Genetics*, 138(4), 969–988. https://doi.org/10.1007/s00122-019-02533-9

Stewart-Brown, B. B., Song, Q., Vaughn, J. N. & Li, Z. (2019). Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3: Genes, Genomes, Genetics*, 9(7), 2253–2265. https://doi.org/10.1534/g3.118.200917

Sui, M., Jing, Y., Li, H., Zhan, Y., Luo, J., Teng, W., Qiu, L., Zheng, H., Li, W., Zhao, X. & Han, Y. (2020). Identification of loci and candidate genes analyses for tocopherol concentration of soybean seed. *Frontiers in Plant Science*, 11, 1366.

Swarm, S. A., Sun, L., Wang, X., Wang, W., Brown, P. J., Ma, J. & Nelson, R. L. (2019). Genetic dissection of domestication-related traits in soybean through genotyping-by-sequencing of two interspecific mapping populations. *Theoretical and Applied Genetics*, 132(4), 1195–1209. https://doi.org/10.1007/s00122-018-3272-6

Tanger, P., Klassen, S., Mojica, J. P., Lovell, J. T., Movers, B. T., Baraoidan, M., Naredo, M. E. B., McNally, K. L., Poland, J., Bush, D. R. & Leung, H. (2017). Field-based high throughput phenotyping rapidly identifies genomic regions controlling yield components in rice. *Scientific Reports*, 7, 42839. https://doi.org/10.1038/srep42839

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R. & Bird, C. E. (2013). ezRAD: A simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1, e203. https://doi.org/10.7717/peerj.203
Torkamaneh, D., Chalifour, F. P., Beauchamp, C. J., Agrama, H., Boahen, S., Maaroufi, H., Rajcan, I., & Belzile, F. (2020). Genome-wide association analyses reveal the genetic basis of biomass accumulation under symbiotic nitrogen fixation in African soybean. *Theoretical and Applied Genetics*, 133(2), 665–676. https://doi.org/10.1007/s00122-019-04399-7

Trevisan, R., Pérez, O., Schmitz, N., Diers, B., & Martín, N. (2020). High-throughput phenotyping of soybean maturity using time series UAV imagery and convolutional neural networks. *Remote Sensing*, 12(21), 3617.

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, 3, 347.

van Beuzouw, R. F., Keurentjes, J. J., Harbinson, J., & Aarts, M. G. (2019). Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *The Plant Journal*, 97(1), 112–133. https://doi.org/10.1111/tpj.14190

Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., & Saxena, R. K. (2019). Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theoretical and Applied Genetics*, 132(3), 797–816.

Varshney, R. K., Terauchi, R., & McCouch, S. R. (2014). Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLoS Biology*, 12(6), e1001883. https://doi.org/10.1371/journal.pbio.1001883

Vigand, J., Braccini, A. L., Schuster, I., & Menezes, V. M. P. S. (2018). Micro-satellite molecular marker-assisted gene pyramiding for resistance to Asian soybean rust (ASR). *Acta Scientiarum. Agronomy*, 40.

Wang, L., Fang, C., Liu, J., Zhang, T., Kou, K., Su, T., Li, S., Chen, L., Cheng, Q., Dong, L., & Kong, F. (2020). Identification of major QTLs for flowering and maturity in soybean by genotyping-by-sequencing analysis. *Molecular Breeding*, 40(10), 1–12.

Wang, X., Zhang, R., Song, W., Han, L., Liu, X., Sun, X., Luo, M., Chen, K., Zhang, Y., Yang, H., & Yang, G. (2019). Dynamic plant height QTL revealed in maize through remote sensing phenotyping using a high-throughput unmanned aerial vehicle (UAV). *Scientific Reports*, 9(1), 1–10.

Watanabe, K., Guo, W., Araï, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., Yano, K., Tokunaga, T., Fujiwara, T., Tsutsumi, N., & Iwata, H. (2017b). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Frontiers in Plant Science*, 8, 421.

Watanabe, S., Tsukamoto, C., Oshita, T., Yamada, T., Araï, T., & Kaga, A. (2017a). Identification of quantitative trait loci for flowering time by a combination of restriction site–associated DNA sequencing and bulked segregant analysis in soybean. *Breeding Science*, 67(3), 277–285.

Wei, W., Mesquita, A. C. O., Figueiró, A. D. A., Wu, X., Manjunatha, S., Wickland, D. P., Hudson, M. E., Juliatti, F. C., & Clough, S. J. (2017). Genome-wide association mapping of resistance to a Brazilian isolate of *Sclerotinia sclerotiorum* in soybean genotypes mostly from Brazil. *BMC Genomics*, 18(1), 849. https://doi.org/10.1186/s12864-017-4160-1

Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., & Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinformatics*, 18(1), 586. https://doi.org/10.1186/s12859-017-2000-6

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, 14(7), 507–515. https://doi.org/10.1038/nrg3457

Xu, M., Xu, Z., Liu, B., Kong, F., Tsubokura, Y., Watanabe, S., Xia, Z., Harada, K., Kanazawa, A., Yamada, T., & Abe, J. (2013). Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biology*, 13(1), 91. https://doi.org/10.1186/1471-2229-13-91

Yang, H., Tao, Y., Zheng, Z., Li, C., Sweetingham, M. W., & Howieson, J. G. (2012). Application of next-generation sequencing for rapid marker development in molecular plant breeding: A case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genomics*, 13, 318. https://doi.org/10.1186/1471-2164-13-318

Yang, Q., Jin, H., Yu, X., Fu, X., Zhi, H., & Yuan, F. (2020). Rapid identification of soybean resistance genes to soybean mosaic virus by SLAF-seq bulked segregant analysis. *Plant Molecular Biology Reporter*, 38(4), 666–675.

Yin, L., Zhang, H., Zhou, X., Yuan, X., Zhao, S., Li, X., & Liu, X. (2020). KAML: Improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome Biology*, 21(1), 1–22.

Yu, D., Hu, D., Kan, G., Hu, W., Li, Y., Hao, D., Li, X., Yang, H., Yang, Z., He, X., & Huang, F. (2019). Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments. *Frontiers in Plant Science*, 10, 811.

Zargar, S. M., Raatz, B., Sonah, H., Bhat, J. A., Dar, Z. A., Agrawal, G. K., & Rakwal, R. (2015). Recent advances in molecular marker techniques: insight into QTL mapping, GWAS and genomic selection in plants. *Journal of Crop Science and Biotechnology*, 18(3), 293–308.

Zhang, J., Song, Q., Cregan, P. B., & Jiang, G. L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theoretical and Applied Genetics*, 129(1), 117–130. https://doi.org/10.1007/s00122-015-2614-x

Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., Wang, K., Xiao, Y., Chen, G., Liu, Q., & Xiong, L. (2017). High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiology*, 173(3), 1554–1564. https://doi.org/10.1104/pp.16.01516

Zhang, X., Wang, W., Guo, N., Zhang, Y., Bu, Y., Zhao, J., & Xing, H. (2018). Combining QTL-seq and linkage mapping to fine map a wild soybean allele characteristic of greater plant height. *BMC Genomics*, 19(1), 1–12.

Zhang, Y. H., Liu, M. F., He, J. B., Wang, Y. F., Xing, G. N., Li, Y., Yang, S. P., Zhao, T. J., & Gai, J. Y. (2015). Marker-assisted breeding for transgressive seed protein content in soybean (*Glycine max* [L.] Merr.). *Theoretical and Applied Genetics*, 128(6), 1061–1072. https://doi.org/10.1007/s00122-015-2490-4

Zhao, X., Han, Y., Li, Y., Liu, D., Sun, M., Zhao, Y., Lv, C., Li, D., Yang, Z., Huang, L., & Teng, W. (2015). Loci and candidate gene identification for resistance to Sclerotinia sclerotiorum in soybean (*Glycine max* [L.] Merr.) via association and linkage maps. *The Plant Journal*, 82(2), 245–255. https://doi.org/10.1111/tpj.12810

Zhong, C., Li, Y., Sun, S., Duan, C., & Zhu, Z. (2019). Genetic mapping and molecular characterization of a broad-spectrum *Phytophthora sojae* resistance gene in Chinese soybean. *International Journal of Molecular Sciences*, 20(8). https://doi.org/10.3390/ijms20081809

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, D., Wu, Y., Shu, L., Zhao, Y., Ma, Y., & Fang, C. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, 33(4), 408–414. https://doi.org/10.1038/nbt.3096

---

**How to cite this article:** Bhat JA, Yu D. High-throughput NGS-based genotyping and phenotyping: Role in genomics-assisted breeding for soybean improvement. *Legume Science*. 2021;1–15. https://doi.org/10.1002/leg3.81