

Review

Utilization of Genotyping-by-Sequencing (GBS) for Rice Pre-Breeding and Improvement: A Review

Vincent Pamugas Reyes^{1,*}, Justine Kipruto Kitony¹, Shunsaku Nishiuchi¹, Daigo Makihara²
and Kazuyuki Doi^{1,*}

¹ Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan

² International Center for Research and Education in Agriculture, Nagoya University, Nagoya 464-8601, Japan

* Correspondence: reyesvp@nagoya-u.jp (V.P.R.); kdoi@agr.nagoya-u.ac.jp (K.D.)

Abstract: Molecular markers play a crucial role in the improvement of rice. To benefit from these markers, genotyping is carried out to identify the differences at a specific position in the genome of individuals. The advances in sequencing technologies have led to the development of different genotyping techniques such as genotyping-by-sequencing. Unlike PCR-fragment-based genotyping, genotyping-by-sequencing has enabled the parallel sequencing and genotyping of hundreds of samples in a single run, making it more cost-effective. Currently, GBS is being used in several pre-breeding programs of rice to identify beneficial genes and QTL from different rice genetic resources. In this review, we present the current advances in the utilization of genotyping-by-sequencing for the development of rice pre-breeding materials and the improvement of existing rice cultivars. The challenges and perspectives of using this approach are also highlighted.

Keywords: molecular breeding; DNA markers; genomic selection; NGS



Citation: Reyes, V.P.; Kitony, J.K.; Nishiuchi, S.; Makihara, D.; Doi, K. Utilization of Genotyping-by-Sequencing (GBS) for Rice Pre-Breeding and Improvement: A Review. *Life* **2022**, *12*, 1752. <https://doi.org/10.3390/life12111752>

Academic Editors: Jitendra Kumar and Krishan Rai

Received: 13 October 2022

Accepted: 28 October 2022

Published: 1 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the completion of the rice genome sequencing in 2005 [1], the identification of agronomically important genes and QTL for rice improvement has been greatly accelerated [2–4]. The information from the sequencing of the rice genome resulted in the success of modern breeding practices in rice. However, the current yield performance of the existing rice cultivars may not be sufficient due to the rapid increase in the human population. In addition, intensive rice breeding practices have narrowed the genetic base of elite germplasm. This has compromised the long-term genetic gain and increased the genetic vulnerability to various stresses. Therefore, an efficient method for the identification and incorporation of new genetic variations is necessary.

Pre-breeding refers to the intermediate step of breeding for rational use of unused landraces, alien germplasm, or wild relatives, which are often not adapted to actual breeding or production sites. The outcome of pre-breeding activities is the development of materials that are suitable for evaluation at actual breeding sites. The materials are often referred to as introgression lines [5] or as chromosome (segment) substitution lines in elite genetic backgrounds. These materials have uniform genetic backgrounds and are well adapted to the breeding sites, thus works as genetic “library” to screen various traits. The traits can be associated with the chromosomal regions defined by DNA markers. In addition to the pre-breeding materials, genetically fixed experimental materials, such as recombinant inbred lines (RILs), nested association mapping (NAM) populations [6,7], multi-parent advanced generation inter-cross (MAGIC) populations [8,9] or genome-wide association (GWA) panels [10–12], are publicly available for gene discovery. Fine genotyping of these genetic resources brings a big advantage to efficiently identify the causal genes (Figure 1).

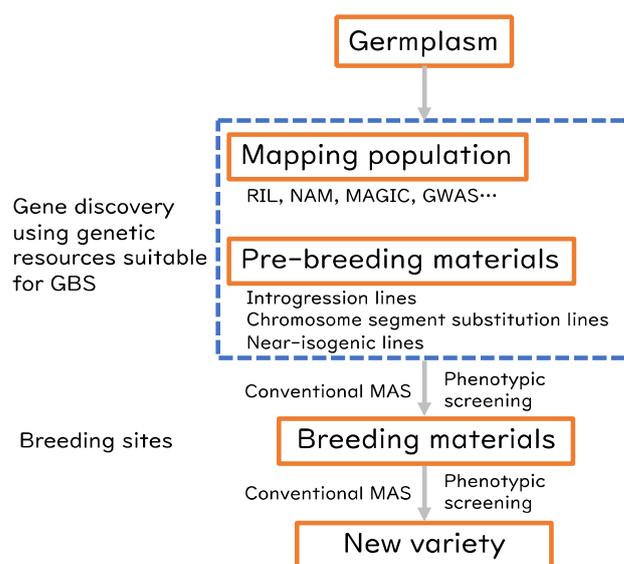


Figure 1. Conventional scheme for gene discovery and breeding. Materials in the hatched blue box are suitable to apply GBS.

Genomic selection (GS), on the other hand, can be utilized to improve the efficiency in breeding of quantitative traits. Although bi-parental mapping and GWAS approaches were successful in dissecting the genetic architecture of a trait, only a limited proportion of the total genetic variance is explained by the markers. GS offers an alternative concept where favorable genetic variations are targeted across the whole genome without defining the threshold to find significant QTL [13]. In brief, this approach develops a prediction model based on the phenotypic and genotypic information of the training population. This information is then used to establish the genomic estimated breeding values (GEBVs) of all individuals of the breeding population from their genotypic information [14,15]. As compared to QTL-based marker-assisted selection (MAS), GS needs more markers than QTL analysis, hence the need for high throughput-sequencing technologies.

Currently, most of the rice breeding projects are designed by combining phenotypic selection and MAS. At the actual breeding sites, MAS is conducted in early to late generations, depending on the breeding targets. It is supposed that most of the breeders are using PCR-fragment-based markers, such as SSR markers, for foreground selection for a limited number of favorable alleles. On the other hand, background or whole-genome selection is becoming common, but its use is still limited. This is probably because most of the actual breeding materials are discarded, and thus the cost for whole genome selection never suits for a program. However, cost-effective genotyping approaches such as genotyping-by-sequencing (GBS) are promising for conducting background selection or GS. In this review, we present the current advances in the utilization of GBS in the pre-breeding and improvement of rice. Challenges and perspectives regarding the utilization of GBS are also presented.

2. Advances in Chemistry of GBS

Over the years, sequencing technologies have become a fundamental approach in the analysis of genetic variation. The rapid advancement in these technologies has led to the development of sequencing platforms that yield millions to billions of DNA bases per run, such as Roche 454, Illumina MiSeq and HiSeq, and Ion torrent [16–19]. This then enabled the identification of single nucleotide polymorphisms (SNPs) within the genome. Additionally, to overcome the drawbacks of hybridization-based and PCR-fragment-based markers, the use of single nucleotide polymorphism (SNP) markers has been adopted in rice genetics research, as advances in chemistry have allowed cost-effective multiplex sequencing.

GBS is a reduced representation sequencing approach that generates thousands of markers at a low cost. As compared to other genotyping techniques, it is more flexible as it is independent of prior genomic information in most cases. To enable different GBS approaches, different chemistries for library construction have been developed. As summarized in Table 1, approaches for GBS can be categorized into three different types: (1) restriction enzyme-based, (2) PCR-based, and (3) target capture. Among these three, restriction-enzyme based GBS has been widely used. However, the digest of sample DNA and ligation to adaptors depends on the quality of DNA, whereas the PCR-based methods can accept small amounts of low-quality DNA, and labor is reduced in library development. Thus, it is more feasible as compared to other methods and will be more popular in the future. On the other hand, the target capture needs preliminary information about the target sites and the design of the probes, making it less cost-effective for breeding.

Table 1. Genotyping-by-sequencing methods based on the type of chemistry used.

Approach	Type	Advantages	Disadvantages	Examples	References
Restriction enzyme-based	Single enzyme		<ul style="list-style-type: none"> Decreased sequencing quality caused by a high proportion of short fragments 	RRL, RAD-seq, Elshire's GBS,	[20–22]
	Double enzyme	Provides a greater degree of complexity reduction	<ul style="list-style-type: none"> Repeatability maybe dependent on size-selection step 	ddRAD, Poland's GBS, SGB, eZRAD	[23–25]
PCR-based	Random	<ul style="list-style-type: none"> Only a small amount of DNA is required High-quality and high-quantity DNA is not necessary Target SNPs can be added flexibly 	<ul style="list-style-type: none"> Optimization is necessary when applied to novel genome 	Genotyping by random amplicon sequencing, direct (GRAS-Di); MIG-Seq	[26–28]
	Targeted	<ul style="list-style-type: none"> High reproducibility 		GT-seq; MTA-Seq;	[29–31]
Target capture		<ul style="list-style-type: none"> Efficiently enrich the target sites 	<ul style="list-style-type: none"> Cost for designing the probes Library development is costly 	Exon capture; Capture of known polymorphic sites	[32]

2.1. Restriction Enzyme-Based (RE-Based) GBS

The RE-based GBS employs the use of a restriction enzyme to digest the genomic DNA. This GBS approach is generally categorized into two types: single-enzyme and double-enzyme. The use of the single-enzyme approach for reduced genome complexity was first described by Baird et al. [21] in restriction association DNA sequencing (RAD-seq). In this technique, the restriction enzyme is used to generate genomic fragments (digestion), which are then ligated to a set of unique adapters that enables the pooling of multiple samples (multiplexing). Pooled samples are then size selected (~300–600 bp for Illumina sequencing), amplified by PCR, and then sequenced. In 2011, Elshire et al. [20] simplified the RAD-seq, herein referred to as Elshire's GBS, by removing random shearing and size selection steps. In Elshire's GBS, barcoded adapters and common adapters have an overhanging site that matches the restriction sites. These sites are then ligated onto digested fragments in a single sticky end-ligation. As compared to RAD-seq, Elshire's GBS is less complicated as the generation of restriction fragments is more straightforward.

The double-enzyme approach was introduced by Peterson et al. [23] using the double-digest RAD (ddRAD). Unlike the single-enzyme GBS, this approach capitalizes on the use

of two kinds of REs. This provides a greater degree of complexity reduction. The ddRAD shares almost the same library construction steps as the RAD-seq with minor modifications. In ddRAD, the barcoded adapter is ligated to one end and the common adapter is ligated to the other. The random shearing step was also eliminated in ddRAD, but size selection was retained to recover regions that are randomly distributed in the genome. Another double-enzyme approach, the ezRAD, was introduced by Toonen et al. [33]. In this approach, the flexibility to use any restriction enzyme (or combination) that frequently cuts to generate a desired fragment size has been introduced. Unlike ddRAD, the adapters of ezRAD are not custom designed for the preferred enzyme. This allows researchers to try different enzymes without a costly investment for each enzyme of choice. Elshire's GBS technique was modified by Poland et al. [24] by adopting the two-enzyme approach. In Poland's GBS, a combination of rare- and common-cutting REs is used to digest the DNA sample. The digested DNA fragments will contain alternate ends which are fitted to the barcode adaptors and the reverse (Y) adaptor. This approach has been demonstrated to capture fragments that are associated with rare-cutting enzymes. In addition, the use of a Y adaptor on the common restriction avoids the amplification of more common fragment; this is a preferential situation for larger and more complex genomes.

To date, the RE-based GBS approaches have been widely used due to their advantage in terms of scalability, which is dependent on the choice and combinations of restriction enzymes used. However, the enzyme of choice can be a major limitation of these approaches, as the genomic distribution of SNPs is dependent on the specific choice and combination of REs.

2.2. PCR-Based GBS

Although RE-based GBS approaches have been utilized in marker-assisted studies, the need for high quality and quantities of DNA has become their limiting factor to being widely adopted. To address this issue, Suyama and Matsuki [29] proposed changing the RE-based steps to PCR-based steps. This approach was based on conventional RE-based markers such as restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (AFLP), which were later replaced by the PCR-based marker, simple sequence repeat (SSR).

Generally, PCR-based GBS can be categorized as either random or targeted. The random PCR-based GBS was first introduced by Suyama and Matsuki [29] using multiplexed ISSR genotyping by sequencing (MIG-seq). In brief, a two-step PCR is carried out in the development of the library. The first PCR step is conducted to amplify the ISSR regions in the genomic DNA using 12 bp SSR sequences with 2 bp anchor oligos at the 3' tail. The products from the first PCR are then used as a template for the second PCR step. In the second PCR step (tailed PCR), complimentary sequences for the Illumina sequencing flow cell and indices are added, and PCR products are pooled into a single sequencing library. Like RAD, this method has a size selection (300–800 bp) step, and size selected fragments are then sequenced. In a recent study by Nishimura et al. [34], MIG-seq has been successfully applied in different crops for genetic analysis. It was highlighted that the number of bases sequenced using this method is associated with genome size. Hence, this method is more suited for crops with large genomes such as wheat (~17 Gb). Given this, few loci could be sequenced in plant species with small genomes. However, it is worthy of note that polymorphism is not only dependent on the number of loci but also on the genetic distance between accessions. Therefore, MIG-seq also be used in plant species with small genome as long as the samples have a high nucleotide diversity ($\pi > 0.01$). Another type of random PCR-based GBS is the genotyping by random amplicon sequencing-direct (GRAS-Di). Similar to MIG-seq, the development of GRAS-Di sequencing libraries consists of two sequential PCR steps and a final purification step. However, in GRAS-Di, amplification at the first PCR step uses a 10 bp Illumina Nextera adaptor plus 3 bp random oligomers at the 3'-end, which allows amplification of loci higher than the MIG-seq [26–28]. To date, this method has been only applied in several crop genetic research [27,35–37]. For

example, Enoki and Takeuchi [28] developed this method and applied it to a population of rice backcross inbred lines (BIL). Over 10,000 SNPs that were uniformly distributed in the genome, were detected with a very low missing rate (~1.5%) in the data. Similarly, Kumawat and Xu [36] used this approach and generated ~4000 markers that were used for QTL mapping of seed size and shape in a soybean RIL population.

Another type of PCR-based GBS is the targeted approach, which was firstly introduced by Campbell et al. [31] in their method, genotyping-in-thousands by sequencing (GT-seq). Later on, a method that has the same concept as GT-seq was introduced by Onda et al. [30], called multiplex PCR targeted amplicon sequencing (MTA-seq). These methods also share the same concept as the two-step PCR of MT-Seq and GRAS-Di. However, preliminary information on the identified SNPs is necessary, which makes it less flexible for species that lacks genomic resources.

Among all the PCR-based GBS, although still in its infancy, GRAS-Di has great potential to be routinely used in breeding for the following reasons: (1) it can be applied to thousands of samples by using primer sets at a relatively low cost; (2) the amplified regions are reproducible, hence suppressing the missing data will not be a problem; and (3) it can be utilized even with a small quantity of DNA (<100 ng).

2.3. Target Capture

Target capture has been proposed as an alternative for complete sequencing of a large and complex genome. To date, methods such as exome capture have been developed [32,38]. This method is focused on the coding and regulatory regions of the genome, which are primary interests for functional genomics research [39]. In this method, genomic DNA is hybridized to microarrays that contain oligonucleotides that are complementary to the target sequences, followed by elution and sequencing of hybridized DNA. Over the years, this method has been modified where biotin-labeled oligonucleotides are used and retrieved using streptavidin beads, hence the solution-based hybridization [39,40].

The target capture GBS has been described to be advantageous over PCR-based GBS as oligonucleotides in target capture have less specificity compared to PCR primers. However, several drawbacks were also documented. For example, the efficiency of this method can be reduced for probes that overlap multiple exons as intron sequences prevent “probe:exon” hybridization [39]. This is a major issue for samples that do not have a reference genome. Although this has been proposed in the past few years, a low adoption rate of this approach in breeding programs is seen due to the costliness of the design of probes and library development.

3. GBS for Rice Pre-Breeding

Over the years, GBS methods have been used in different rice pre-breeding studies. In fact, these GBS methods were already used in rice complex crosses such as MAGIC and NAM populations [6,8,9]. Bandillo et al. [8] developed four multi-parent populations (Indica MAGIC, MAGIC plus, Japonica MAGIC, and Global MAGIC) and were subjected to GBS. The sequence data that were obtained were used for GWAS which led to the detection of known (*Sub1*, *Xa4*, and *xa5*) and novel QTL [8]. Ogawa et al. [9] developed an improved allele mining approach using the Japan-MAGIC (JAM) population. Using GBS, a total of 16,345 SNPs were identified and used to predict the haplotype blocks. Similarly, Fragozo et al. [7] also used GBS in the genotyping of rice NAM populations. The genotyping data obtained were successfully used for simple QTL and joint QTL mapping. Arbelaez et al. [41] developed two populations of interspecific introgression lines derived from *O. meridionalis* × *O. sativa* cv Curinga and *O. rufipogon* × *O. sativa* cv. Curinga. The utilization of GBS in their study increased the marker density by over 50-fold and led to the identification of small donor introgressions that may have been missed using a lower density of markers. Similarly, Spindel et al. [42] used GBS to generate 30,984 markers for 176 RILs derived from an Indica × Japonica cross. Using these markers, QTL for leaf width and aluminum tolerance were mapped. In 2018, Furuta et al. [43] modified the GBS

method for rice by changing the rare-cutting enzyme *Pst*I to *Kpn*I. As a result, a higher per-site coverage of sequence reads were generated, and more samples were multiplexed in the sequencing library. However, a trade-off in the number of sites per sample was observed, but the read information was still sufficient for certain applications. The modified GBS method of Furuta et al. [43] has been successfully applied for linkage mapping, introgression studies, recurrent genome recovery analysis, and GWAS in rice [6,44–46]. In a study conducted by Liang et al. [47], GBS was efficiently utilized to study cadmium (Cd) accumulation in rice by conducting GWAS on 270 Indica rice varieties. A total of 79,545 genetic markers were identified and were used to identify QTL associated with Cd accumulation. In a study by Goto et al. [48], GBS was utilized and obtained a total of 2221 SNPs. These SNPs were used to map QTL regions that are associated with sodium removal ability in rice leaf sheaths. Similarly, Waheed et al. [49] utilized this approach and identified *qWU7* and *qWU1* which are associated to drought response. Other QTL/genes that were identified using GBS approach are presented in Table 2.

Table 2. Detected QTL using the genotyping-by-sequencing technique.

Trait	QTL	Type of Population	Reference
Hybrid weakness	<i>hwj1</i> and <i>hwj2</i>	F ₂	[45]
Shoot Na ⁺ concentration	<i>qSNC1-1</i> , <i>qSNC1-2</i> , and <i>qSNC11</i>		
Leaf sheath Na ⁺ concentration	<i>qSHNC1</i> and <i>qSHNC11</i>	F ₂	[48]
Leaf blade K ⁺ -Na ⁺ ratio	<i>qBKNR11</i>		
Water uptake	<i>qWU7</i> and <i>qWU11</i>	RIL	[49]
Zinc content in polished rice	<i>qZPR1.1</i>	BRIL	[50]
Salinity Tolerance	<i>qSIS5.1b</i> and <i>qSIS6.30</i>	RIL	[51]
Grain quality	<i>qGS5.2</i> , <i>qGS7.1</i> , and <i>qPGWC8</i>	RIL	[52]

The success of the GBS approaches in genetic mapping of complex traits and its application in complex populations such as NAM and MAGIC shows that it has great potential to be implemented across various pre-breeding programs.

4. Implementation of GS and GBS for Rice Improvement

GS can be used to predict the performance of progeny and thus enrich the starting pedigrees. Previous simulation and empirical studies have shown that GS selection can speed up crop improvement [53,54]. In rice, several simulations and much empirical research on GS have been conducted [55–63]. One of the benefits of integrating GBS to GS programs is that it can effectively use genome-wide molecular markers from GBS. For example, Spindel et al. [59] conducted GS and GWAS using 363 elite breeding lines from the International Rice Research Institute (IRRI). Using GBS, the population was genotyped with 73,147 markers, and used for predicting grain yield, flowering time, and plant height. The authors demonstrated that subsetting the SNP markers from 73,147 to 7142 (approximately 1 SNP for every 0.2 cM) and 73,147 to 13,101 SNPs (approximately 1 SNP for every 0.1 cM) does not have a significant difference in GS models for a given trait or validation season. However, when the markers were lower than 7142, prediction accuracies began to decrease in most traits and models. Collectively, their results showed that using a SNP every 0.2 cM (~10,000 markers) is a sufficient marker density for GS in inbred rice populations.

A simplified adoption of GS combined with population breeding (bulk method) is proposed in this paper (Figure 2). The breeding population (BP) is generated by single seed descent or by bulk method without selection until F₄ or F₅ generation. All plants in BP are genotyped, and a part of BP is subjected to phenotyping to construct the GS model. The GEBV of all plants in BP is calculated and the plants with the best GEBV are selected. The selected progeny are handled in a similar manner to pedigree breeding. If necessary, the GS model can be refined by repeating the whole or a part of the process.

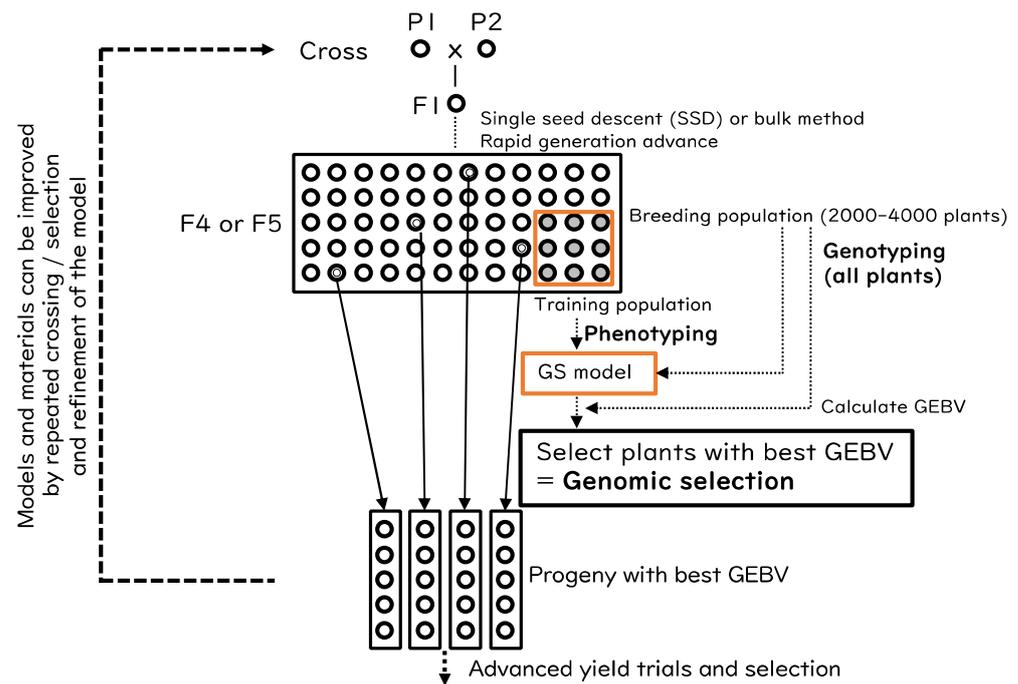


Figure 2. Schematic diagram of a simplified implementation of genomic selection in rice.

A study by Bassi et al. [64] demonstrated that in wheat, GS in F₃ or F₄ generation is the most cost-effective approach with respect to genetic gain and breeding cycles. The same may hold true for rice. We propose to wait until F₄ or F₅ to obtain sufficient selection accuracy because GS is most beneficial for polygenic traits. Another important factor is the size of BP and training populations, a larger population provides more possibility to obtain appropriate plants, but this is always limited by external factors such as spaces.

Another viewpoint on the use of GS is to use it for a trait that is not the primary target. In our study using a rice NAM population, more than 90% of the phenotypic variance of heading time can be explained by a model based on genotypes (Kitony et al. unpublished). This means that breeders can enrich the materials by discarding those that do not have the expected heading time. The primary target trait can be selected from the enriched population that have the desired heading time.

5. Challenges in Informatics

Although GBS has become a popular genotyping approach, just like any other tool, this approach has some drawbacks. The major drawbacks associated with GBS are: (1) a large amount of missing data, (2) errors, and (3) undercalled heterozygous genotypes [65] (Figure 3). Missing data in GBS can happen by chance and is primarily due to low coverage sequencing (Figure 3A), whereas sequencing errors are inevitable (Figure 3B). The undercalled heterozygous genotypes result when true genotypes are heterozygous, yet the call is homozygous (Figure 3C). These problems result in wrong calls of genotypes and may affect GS prediction models as they need good molecular datasets [65]. To address these issues, filtering by parents, filtering by minor allele frequency (MAF), and imputation and error correction are necessary.

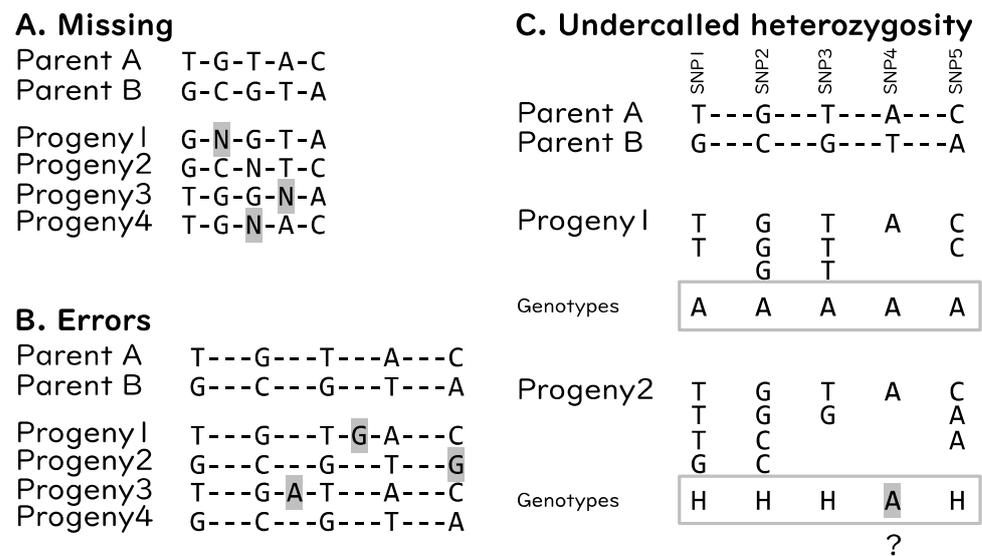


Figure 3. General drawbacks of GBS. (A) missing reads resulting in missing genotypes. (B) sequencing errors resulting in incorrect genotypes. (C) undercalled heterozygosity, SNP4 of progeny 2 is likely to be “H” (heterozygous) but called as “A” (same as parent A). Hyphens (“-”) indicate monomorphic sites. Letters highlighted in gray represent the type of drawback.

Sequencing errors can be filtered out by several approaches. Filtering based on minor allele frequency (MAF) < (proportion of 2 different samples against all samples) can be carried out to remove sequencing errors, because if an allele is detected in 2 or more samples, it is likely that the allele is true. Filtering by MAF is not suitable for certain breeding populations such as ILs/CSSLs, but these populations can still be utilized for GBS by using two or more indices for each sample, especially parents. Similarly, filtering by parental genotypes is also an efficient way. This is carried out by selecting SNPs that are (1) not variable in a set of replicated parental samples, (2) non-heterozygous in both parents, and (3) polymorphic between parents.

For missing data and undercalled heterozygosity, imputation and error correction can be implemented. Imputation refers to a statistical procedure that replaces the missing values in a dataset based on probability. In genotyping, imputation is carried out to predict the untyped loci from the sequencing call and is necessary in any genomic study for a more reliable result [66]. Based on a similar concept, error correction can be conducted. To date, several imputation pipelines have been implemented and evaluated in rice GBS datasets [67–70]. A comparative analysis by Nazzicari et al. [71] showed that the performance of four general imputation methods (K-nearest neighbors, Random Forest, singular value decomposition, and mean value) and two genotype-specific methods (“Beagle” and “FILLIN”) on rice GBS datasets with up to a 67% missing rate. For general imputation methods, random forest showed the highest accuracy, 90%, whereas Beagle with ordered markers performed well in genotype-specific methods. As a result, the comparison of all methods showed that Beagle with ordered markers outperformed all other imputation methods. An R package called “ABHgenotypeR” for imputation and error-correction on F₂ populations was previously developed by Furuta et al. [43]. One of the main features of this package is the easy visualization of graphical genotypes for direct comparison (Figure 4). In this analytical tool, imputation is carried out based on the flanking alleles

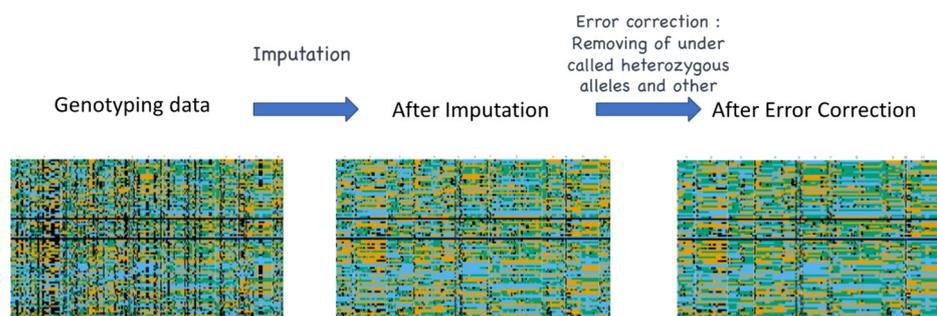


Figure 4. Visual representation of imputation and error correction in GBS data. Black = missing data, Blue = parent 1, Orange = parent 2, Green = heterozygous allele.

In this tool, if the genotypes on the left and right of the missing data are identical, then the genotypes are filled in. The package also offers a two-way error correction for GBS datasets: (i) correction of undercalled heterozygous genotypes, and (ii) correction of other genotyping errors. However, in a report presented by Lorieux et al. [72], ABHgenotypeR is not suited for noisy low-coverage sequence datasets. Although several studies have demonstrated the importance of informatics pipelines, error-free GBS dataset is still impossible. However, as demonstrated by Furuta et al. [43] the structural differences between the genomes of the parental lines are more likely to be the major source of the erroneous markers. Therefore, checking the GBS dataset even after the imputation and error correction steps is necessary. For example, manual curation (removal of suspicious SNP calls or markers) of the dataset should be implemented for the outputs of the informatic pipeline.

The high volume of genetic data provided by parallel short-read sequencing also brings serious challenges in the analysis [73]. To fully utilize these GBS datasets, development of software and informatics pipelines that can effectively assemble reads, identify alleles and genotypes, and monitor those genotypes in hundreds of individuals across several populations using a statistically rigorous framework is necessary [74]. In a GBS informatics pipeline, factors such as SNP calling strategies and ease of use must be considered. As summarized in Table 3, Wickland et al. [75] conducted a comparative study on five GBS pipelines.

Table 3. Comparison of available GBS pipelines.

GBS-Pipeline	SNP Calling Strategy	Ease of Use ¹	Reference
Trait Analysis by aSSociation, Evolution and Linkage (TASSEL)	Binomial likelihood ratio	Needs extra steps to improve SNP call accuracy. Not built into the pipeline	[68]
IBIS genotyping by sequencing tools (IGST)	Bayesian	-	[76]
Fast-GBS	Haplotype-based	Needs extra steps to improve SNP call accuracy. Not built into the pipeline	[77]
Stacks	Multinomial-based likelihood	-	[78]
GB-eaSy	Bayesian	Additional steps for SNP call accuracy is not needed or built into the pipeline itself.	[75]

¹ In terms of carrying out all the steps needed to produce accurate SNPs.

Among these tools, the GB-eaSy, TASSEL-GBS, and IGST were found to have the highest accuracy (~99%) in terms of SNP calling in comparison with the whole-genome sequence (WGS). Interestingly, a low percentage of common SNPs were detected in these SNP calling tools. Looking at the common SNPs detected using these tools, Wickland et al. [75] only found 12.08% common SNPs between TASSEL, GB-eaSy, and IGST. The difference in the SNPs detected using these tools could be attributed to the SNP calling strategy and read aligners. For example, Hwang et al. [79] conducted a comparison of three read aligners and four variant callers and identified that BWA-MEM together with SAMtools has the greatest accuracy for SNP identification. Similarly, the combination of GB-eaSy

with BCFtools/SAMtools showed a great allelic concordance in reference to the WGS data in Soybean lines.

6. Future Perspective

The cost reduction in sequencing is allowing a wider application of GBS. In terms of cost-efficiency, multiplexing a greater number of samples for a single GBS run is beneficial for breeders. Several approaches, such as the modification of sequencing adapters, can be carried out to achieve a higher multiplexing capacity. Reyes et al. [80] demonstrated that the addition of indexing reads to barcode adapters of Poland et al. [14] enabled the multiplexing of 2304 samples from independent populations in a single sequencing run. This method improved the convenience of genotyping different populations from multiple breeders.

GBS has enabled the genotyping of thousands of individuals. However, a lack of phenotype information prevents the understanding of polygenic traits because the population is a limiting factor in statistical genetics. Simultaneous utilization of high-throughput phenotyping technologies and GBS will allow a new way of dissecting complex traits. Currently, these new phenotyping technologies are first used for GWAS panels. However, the construction of new genetic resources will greatly accelerate rice breeding.

Author Contributions: Conceptualization, V.P.R. and K.D.; writing—draft, V.P.R. and K.D.; writing—review and editing, V.P.R., K.D., S.N., J.K.K. and D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: K.D. is grateful for the research support from Cross-ministerial Strategic Innovation Promotion Program (SIP), National Bioresource Project (NBRP), RIKEN Cluster for Science, Technology and Innovation Hub (RCSTI), and JSPS KAKENHI Grant Number 22H02310/21K05522/20KK0138.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. International Rice Genome Sequencing Project; Sasaki, T. The map-based sequence of the rice genome. *Nature* **2005**, *436*, 793–800. [[CrossRef](#)] [[PubMed](#)]
2. Ashikari, M.; Matsuoka, M. Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends Plant Sci.* **2006**, *11*, 344–350. [[CrossRef](#)] [[PubMed](#)]
3. Fan, C.; Xing, Y.; Mao, H.; Lu, T.; Han, B.; Xu, C.; Li, X.; Zhang, Q. *GS3*, a Major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **2006**, *112*, 1164–1171. [[CrossRef](#)]
4. Doi, K.; Izawa, T.; Fuse, T.; Yamanouchi, U.; Kubo, T.; Shimatani, Z.; Yano, M.; Yoshimura, A. *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. *Genes Dev.* **2004**, *18*, 926–936. [[CrossRef](#)] [[PubMed](#)]
5. Brar, D.S.; Dalmacio, R.; Elloran, R.; Aggarwal, R.; Angeles, R.; Khush, G.S. Gene transfer and molecular characterization of introgression from wild *Oryza* species into rice. In *Rice Genetics III (In 2 Parts), Proceedings of the Third International Rice Genetics Symposium, Manila, Philippines, 16–20 October 1995*; Khush, G.S., Hettel, G., Rola, T., Eds.; World Scientific: Singapore, 2008; pp. 477–486.
6. Kitony, J.K.; Sunohara, H.; Tasaki, M.; Mori, J.-I.; Shimazu, A.; Reyes, V.P.; Yasui, H.; Yamagata, Y.; Yoshimura, A.; Yamasaki, M.; et al. Development of an *aus*-derived nested association mapping (*aus*-NAM) population in rice. *Plants* **2021**, *10*, 1255. [[CrossRef](#)] [[PubMed](#)]
7. Frago, C.A.; Moreno, M.; Wang, Z.; Heffelfinger, C.; Arbelaez, L.J.; Aguirre, J.A.; Franco, N.; Romero, L.E.; Labadie, K.; Zhao, H.; et al. Genetic architecture of a rice nested association mapping population. *G3 Genes Genomes Genet.* **2017**, *7*, 1913–1926. [[CrossRef](#)]
8. Bandillo, N.; Raghavan, C.; Muyco, P.; Sevilla, M.A.L.; Lobina, I.T.; Dilla-Ermita, C.; Tung, C.-W.; McCouch, S.; Thomson, M.; Mauleon, R.; et al. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice* **2013**, *6*, 11. [[CrossRef](#)]

9. Ogawa, D.; Yamamoto, E.; Ohtani, T.; Kanno, N.; Tsunematsu, H.; Nonoue, Y.; Yano, M.; Yamamoto, T.; Yonemaru, J. Haplotype-based allele mining in the Japan-MAGIC rice population. *Sci. Rep.* **2018**, *8*, 4379. [[CrossRef](#)]
10. Norton, G.J.; Travis, A.J.; Douglas, A.; Fairley, S.; Alves, E.D.P.; Ruang-areerate, P.; Naredo, M.E.B.; McNally, K.L.; Hossain, M.; Islam, M.R.; et al. Genome wide association mapping of grain and straw biomass traits in the rice Bengal and Assam aus panel (BAAP) grown under alternate wetting and drying and permanently flooded irrigation. *Front. Plant Sci.* **2018**, *9*, 1223. [[CrossRef](#)]
11. Eizenga, G.C.; Ali, M.L.; Bryant, R.J.; Yeater, K.M.; McClung, A.M.; McCouch, S.R. Registration of the rice diversity panel 1 for genomewide association studies. *J. Plant Regist.* **2014**, *8*, 109–116. [[CrossRef](#)]
12. Hoang, G.T.; Van Dinh, L.; Nguyen, T.T.; Ta, N.K.; Gathignol, F.; Mai, C.D.; Jouannic, S.; Tran, K.D.; Khuat, T.H.; Do, V.N.; et al. Genome-wide association study of a panel of Vietnamese rice landraces reveals new QTLs for tolerance to water deficit during the vegetative phase. *Rice* **2019**, *12*, 4. [[CrossRef](#)] [[PubMed](#)]
13. Gorjanc, G.; Jenko, J.; Hearne, S.J.; Hickey, J.M. Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. *BMC Genom.* **2016**, *17*, 30. [[CrossRef](#)] [[PubMed](#)]
14. Poland, J.; Endelman, J.; Dawson, J.; Rutkoski, J.; Wu, S.; Manes, Y.; Dreisigacker, S.; Crossa, J.; Sánchez-Villeda, H.; Sorrells, M.; et al. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* **2012**, *5*, 103–113. [[CrossRef](#)]
15. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)] [[PubMed](#)]
16. Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M.; et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **2011**, *475*, 348–352. [[CrossRef](#)]
17. Thudi, M.; Li, Y.; Jackson, S.A.; May, G.D.; Varshney, R.K. Current state-of-art of sequencing technologies for plant genomics research. *Brief. Funct. Genom.* **2012**, *11*, 3–11. [[CrossRef](#)]
18. Bentley, D.R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **2006**, *16*, 545–552. [[CrossRef](#)]
19. Quail, M.; Smith, M.E.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A Tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* **2012**, *13*, 341. [[CrossRef](#)]
20. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)]
21. Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid snp discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **2008**, *3*, e3376. [[CrossRef](#)]
22. Van Tassel, C.P.; Smith, T.P.L.; Matukumalli, L.K.; Taylor, J.F.; Schnabel, R.D.; Lawley, C.T.; Haudenschild, C.D.; Moore, S.S.; Warren, W.C.; Sonstegard, T.S. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* **2008**, *5*, 247–252. [[CrossRef](#)] [[PubMed](#)]
23. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double digest RADseq: An inexpensive Method for De Novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **2012**, *7*, e37135. [[CrossRef](#)] [[PubMed](#)]
24. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **2012**, *7*, e32253. [[CrossRef](#)] [[PubMed](#)]
25. Truong, H.T.; Ramos, A.M.; Yalcin, F.; de Ruiter, M.; van der Poel, H.J.A.; Huvenaars, K.H.J.; Hogers, R.C.J.; van Enckevort, L.J.G.; Janssen, A.; van Orsouw, N.J.; et al. Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. *PLoS ONE* **2012**, *7*, e37565. [[CrossRef](#)] [[PubMed](#)]
26. Hosoya, S.; Hirase, S.; Kikuchi, K.; Nanjo, K.; Nakamura, Y.; Kohno, H.; Sano, M. Random PCR-based genotyping by sequencing technology GRAS-Di (genotyping by random amplicon sequencing, direct) reveals genetic structure of mangrove fishes. *Mol. Ecol. Resour.* **2019**, *19*, 1153–1163. [[CrossRef](#)]
27. Enoki, H. The construction of pseudomolecules of a commercial strawberry by DeNovoMAGIC and new genotyping technology, GRAS-Di. In Proceedings of the Plant and Animal Genome Conference XXVII, San Diego, CA, USA, 12–16 January 2019.
28. Enoki, H.; Takeuchi, Y. New genotyping technology, GRAS-Di, using next generation sequencer. In Proceedings of the Plant and Animal Genome Conference XXVI, San Diego, CA, USA, 13–17 January 2018.
29. Suyama, Y.; Matsuki, Y. MIG-Seq: An effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing Platform. *Sci. Rep.* **2015**, *5*, 16963. [[CrossRef](#)]
30. Onda, Y.; Takahagi, K.; Shimizu, M.; Inoue, K.; Mochida, K. Multiplex PCR targeted amplicon sequencing (MTA-Seq): Simple, flexible, and versatile SNP genotyping by highly multiplexed PCR amplicon sequencing. *Front. Plant Sci.* **2018**, *9*, 201. [[CrossRef](#)]
31. Campbell, N.R.; Harmon, S.A.; Narum, S.R. Genotyping-in-thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Mol. Ecol. Resour.* **2015**, *15*, 855–867. [[CrossRef](#)]
32. Telfer, E.; Graham, N.; Macdonald, L.; Li, Y.; Klápště, J.; Resende, M., Jr.; Neves, L.G.; Dungey, H.; Wilcox, P. A high-density exome capture genotype-by-sequencing panel for forestry breeding in *Pinus radiata*. *PLoS ONE* **2019**, *14*, e0222640. [[CrossRef](#)]
33. Toonen, R.J.; Puritz, J.B.; Forsman, Z.H.; Whitney, J.L.; Fernandez-Silva, I.; Andrews, K.R.; Bird, C.E. EzRAD: A simplified method for genomic genotyping in non-model organisms. *PeerJ* **2013**, *1*, e203. [[CrossRef](#)]
34. Nishimura, K.; Motoki, K.; Yamazaki, A.; Takisawa, R.; Yasui, Y.; Kawai, T.; Ushijima, K.; Nakano, R.; Nakazaki, T. MIG-Seq is an effective method for high-throughput genotyping in wheat (*Triticum* spp.). *DNA Res.* **2022**, *29*, dsac011. [[CrossRef](#)]

35. Umeda, M.; Sakaigaichi, T.; Tanaka, M.; Tarumoto, Y.; Adachi, K.; Hattori, T.; Hayano, M.; Takahashi, H.; Tamura, Y.; Kimura, T.; et al. Detection of a major QTL related to smut disease resistance inherited from a Japanese wild sugarcane using GRAS-Di technology. *Breed. Sci.* **2021**, *71*, 365–374. [[CrossRef](#)]
36. Kumawat, G.; Xu, D. A major and stable quantitative trait locus *qSS2* for seed size and shape traits in a soybean RIL population. *Front. Genet.* **2021**, *12*, 646102. [[CrossRef](#)]
37. Miki, Y.; Yoshida, K.; Enoki, H.; Komura, S.; Suzuki, K.; Inamori, M.; Nishijima, R.; Takumi, S. GRAS-Di system facilitates high-density genetic map construction and qtl identification in recombinant inbred lines of the wheat progenitor *Aegilops tauschii*. *Sci. Rep.* **2020**, *10*, 21455. [[CrossRef](#)]
38. Suren, H.; Hodgins, K.A.; Yeaman, S.; Nurkowski, K.A.; Smets, P.; Rieseberg, L.H.; Aitken, S.N.; Holliday, J.A. Exome capture from the spruce and pine giga-genomes. *Mol. Ecol. Resour.* **2016**, *16*, 1136–1146. [[CrossRef](#)]
39. Neves, L.G.; Davis, J.M.; Barbazuk, W.B.; Kirst, M. Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J.* **2013**, *75*, 146–156. [[CrossRef](#)]
40. Gnirke, A.; Melnikov, A.; Maguire, J.; Rogov, P.; LeProust, E.M.; Brockman, W.; Fennell, T.; Giannoukos, G.; Fisher, S.; Russ, C.; et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **2009**, *27*, 182–189. [[CrossRef](#)]
41. Arbelaez, J.D.; Moreno, L.T.; Singh, N.; Tung, C.-W.; Maron, L.G.; Ospina, Y.; Martinez, C.P.; Grenier, C.; Lorieux, M.; McCouch, S. Development and GBS-genotyping of introgression lines (ILs) using two wild species of rice, *O. meridionalis* and *O. rufipogon*, in a common recurrent parent, *O. sativa* cv. Curinga. *Mol. Breed.* **2015**, *35*, 81. [[CrossRef](#)]
42. Spindel, J.; Wright, M.; Chen, C.; Cobb, J.; Gage, J.; Harrington, S.; Lorieux, M.; Ahmadi, N.; McCouch, S. Bridging the genotyping gap: Using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* **2013**, *126*, 2699–2716. [[CrossRef](#)]
43. Furuta, T.; Ashikari, M.; Jena, K.K.; Doi, K.; Reuscher, S. Adapting genotyping-by-sequencing for rice F2 populations. *G3 Genes Genomes Genet.* **2017**, *7*, 881–893. [[CrossRef](#)]
44. Reyes, V.P.; Angeles-Shim, R.B.; Mendioro, M.S.; Manuel, M.C.C.; Lapis, R.S.; Shim, J.; Sunohara, H.; Nishiuchi, S.; Kikuta, M.; Makihara, D.; et al. Marker-assisted introgression and stacking of major QTLs controlling grain number (*Gn1a*) and number of primary branching (*WFP*) to NERICA Cultivars. *Plants* **2021**, *10*, 844. [[CrossRef](#)]
45. Soe, T.K.; Kunieda, M.; Sunohara, H.; Inukai, Y.; Reyes, V.P.; Nishiuchi, S.; Doi, K. A novel combination of genes causing temperature-sensitive hybrid weakness in rice. *Front. Plant Sci.* **2022**, *13*, 908000. [[CrossRef](#)]
46. Yamada, S.; Kurokawa, Y.; Nagai, K.; Angeles-Shim, R.B.; Yasui, H.; Furuya, N.; Yoshimura, A.; Doi, K.; Ashikari, M.; Sunohara, H. Evaluation of backcrossed pyramiding lines of the yield-related gene and the bacterial leaf blight resistant genes. *J. Intl. Coop. Agric. Dev.* **2020**, *18*, 18–28.
47. Liang, Y.; Dong, X.; Ni, X.; Wang, Q.; Sahu, S.K.; Hou, J.; Liang, M.; Chen, L.; Zhang, G. Genotyping by sequencing of 270 *indica* rice varieties revealed genetic markers probably related to heavy metal accumulation. *Plant Breed.* **2018**, *137*, 691–697. [[CrossRef](#)]
48. Goto, I.; Neang, S.; Kuroki, R.; Reyes, V.P.; Doi, K.; Skoulding, N.S.; Taniguchi, M.; Yamauchi, A.; Mitsuya, S. QTL analysis for sodium removal ability in rice leaf sheaths under salinity using an IR-44595/318 F2 population. *Front. Plant Sci.* **2022**, *13*, 1002605. [[CrossRef](#)] [[PubMed](#)]
49. Waheed, R.; Ignacio, J.C.; Arbelaez, J.D.; Juanillas, V.M.; Asif, M.; Henry, A.; Kretschmar, T.; Arif, M. Drought response QTLs in a super basmati × azucena population by high-density GBS-based SNP linkage mapping. *Plant Breed.* **2021**, *140*, 758–774. [[CrossRef](#)]
50. Suman, K.; Neeraja, C.N.; Madhubabu, P.; Rathod, S.; Bej, S.; Jadhav, K.P.; Kumar, J.A.; Chaitanya, U.; Pawar, S.C.; Rani, S.H.; et al. Identification of promising RILs for high grain zinc through genotype × environment analysis and stable grain zinc QTL using SSRs and SNPs in rice (*Oryza sativa* L.). *Front. Plant Sci.* **2021**, *12*, 587482. [[CrossRef](#)]
51. De Leon, T.B.; Linscombe, S.; Subudhi, P.K. Molecular dissection of seedling salinity tolerance in rice (*Oryza sativa* L.) using a high-density GBS-based SNP linkage map. *Rice* **2016**, *9*, 52. [[CrossRef](#)]
52. Chen, L.; Gao, W.; Chen, S.; Wang, L.; Zou, J.; Liu, Y.; Wang, H.; Chen, Z.; Guo, T. High-resolution QTL mapping for grain appearance traits and co-localization of chalkiness-associated differentially expressed candidate genes in rice. *Rice* **2016**, *9*, 48. [[CrossRef](#)]
53. Bernardo, R. Genomewide selection for rapid introgression of exotic germplasm in maize. *Crop Sci.* **2009**, *49*, 419–425. [[CrossRef](#)]
54. Combs, E.; Bernardo, R. Genomewide selection to introgress semidwarf maize germplasm into U.S. corn belt inbreds. *Crop Sci.* **2013**, *53*, 1427–1436. [[CrossRef](#)]
55. Grenier, C.; Cao, T.-V.; Ospina, Y.; Quintero, C.; Châtel, M.H.; Tohme, J.; Courtois, B.; Ahmadi, N. Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. *PLoS ONE* **2015**, *10*, e0136594. [[CrossRef](#)]
56. Marulanda, J.J.; Mi, X.; Melchinger, A.E.; Xu, J.-L.; Würschum, T.; Longin, C.F.H. Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. *Theor. Appl. Genet.* **2016**, *129*, 1901–1913. [[CrossRef](#)]
57. Onogi, A.; Ideta, O.; Inoshita, Y.; Eban, K.; Yoshioka, T.; Yamasaki, M.; Iwata, H. Exploring the areas of applicability of whole-genome prediction methods for asian rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **2015**, *128*, 41–53. [[CrossRef](#)]
58. Onogi, A.; Watanabe, M.; Mochizuki, T.; Hayashi, T.; Nakagawa, H.; Hasegawa, T.; Iwata, H. Toward integration of genomic selection with crop modelling: The development of an integrated approach to predicting rice heading dates. *Theor. Appl. Genet.* **2016**, *129*, 805–817. [[CrossRef](#)]

59. Spindel, J.; Begum, H.; Akdemir, D.; Virk, P.; Collard, B.; Redoña, E.; Atlin, G.; Jannink, J.-L.; McCouch, S.R. 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 Genet.* **2015**, *11*, e1004982. [[CrossRef](#)]
60. Spindel, J.E.; Begum, H.; Akdemir, D.; Collard, B.; Redoña, E.; Jannink, J.-L.; McCouch, S. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity* **2016**, *116*, 395–408. [[CrossRef](#)]
61. Wimmer, V.; Lehermeier, C.; Albrecht, T.; Auinger, H.-J.; Wang, Y.; Schön, C.-C. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics* **2013**, *195*, 573–587. [[CrossRef](#)]
62. Monteverde, E.; Rosas, J.E.; Blanco, P.; Pérez de Vida, F.; Bonnacarrère, V.; Quero, G.; Gutierrez, L.; McCouch, S. Multi-environment models increase prediction accuracy of complex traits in advanced breeding lines of rice. *Crop Sci.* **2018**, *58*, 1519–1530. [[CrossRef](#)]
63. Bhandari, A.; Bartholomé, J.; Cao-Hamadoun, T.-V.; Kumari, N.; Frouin, J.; Kumar, A.; Ahmadi, N. Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. *PLoS ONE* **2019**, *14*, e0208871. [[CrossRef](#)]
64. Bassi, F.M.; Bentley, A.R.; Charmet, G.; Ortiz, R.; Crossa, J. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* **2016**, *242*, 23–36. [[CrossRef](#)] [[PubMed](#)]
65. Heffner, E.L.; Sorrells, M.E.; Jannink, J.-L. Genomic selection for crop improvement. *Crop Sci.* **2009**, *49*, 1–12. [[CrossRef](#)]
66. Phocas, F. Genotyping, the usefulness of imputation to increase SNP density, and imputation methods and tools. In *Genomic Prediction of Complex Traits*; Ahmadi, N., Bartholomé, J., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2022; Volume 2467, pp. 113–138.
67. Howie, B.N.; Donnelly, P.; Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **2009**, *5*, e1000529. [[CrossRef](#)] [[PubMed](#)]
68. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)] [[PubMed](#)]
69. Browning, B.L.; Tian, X.; Zhou, Y.; Browning, S.R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **2021**, *108*, 1880–1890. [[CrossRef](#)]
70. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
71. Nazzicari, N.; Biscarini, F.; Cozzi, P.; Brummer, E.C.; Annicchiarico, P. Marker imputation efficiency for genotyping-by-sequencing data in rice (*Oryza sativa*) and Alfalfa (*Medicago sativa*). *Mol. Breed.* **2016**, *36*, 69. [[CrossRef](#)]
72. Lorieux, M.; Gkanogiannis, A.; Fragoso, C.; Rami, J.-F. NOISYmputer: Genotype imputation in bi-parental populations for noisy low-coverage next-generation sequencing data. *bioRxiv* **2019**. [[CrossRef](#)]
73. Shendure, J.; Ji, H. Next-Generation DNA Sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145. [[CrossRef](#)]
74. Hohenlohe, P.A.; Bassham, S.; Etter, P.D.; Stiffler, N.; Johnson, E.A.; Cresko, W.A. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **2010**, *6*, e1000862. [[CrossRef](#)]
75. Wickland, D.P.; Battu, G.; Hudson, K.A.; Diers, B.W.; Hudson, M.E. A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-EaSy. *BMC Bioinform.* **2017**, *18*, 586. [[CrossRef](#)]
76. Sonah, H.; Bastien, M.; Iquira, E.; Tardivel, A.; Légaré, G.; Boyle, B.; Normandeau, É.; Laroche, J.; Larose, S.; Jean, M.; et al. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* **2013**, *8*, e54603. [[CrossRef](#)]
77. Torkamaneh, D.; Laroche, J.; Bastien, M.; Abed, A.; Belzile, F. Fast-GBS: A new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinform.* **2017**, *18*, 5. [[CrossRef](#)]
78. Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **2013**, *22*, 3124–3140. [[CrossRef](#)]
79. Hwang, S.; Kim, E.; Lee, I.; Marcotte, E.M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci. Rep.* **2015**, *5*, 17875. [[CrossRef](#)]
80. Reyes, V.P. Application of Next-Generation Sequencing Technology for Genetic Analysis and Pre-Breeding of Rice. Ph.D. Dissertation, Nagoya University, Nagoya, Japan, 2021.