International Journal of
Plant Biology

MDPI

Review

Single-Nucleotide Polymorphism Chip Platforms in Rice (*Oryza sativa*) Genomics and Their Potential for Precision Breeding

Ha Duc Chu [1,2], Trung Quoc Nguyen [3], Anh Quynh Ho [3], Gioi Huy Dong [3], Hong Viet La [4], Nguyen Thi Phuong Thao [2] and Linh Hong Ta [5,*]

1 Faculty of Agricultural Technology, University of Engineering and Technology, Vietnam National University Hanoi, Hanoi City 100000, Vietnam; cd.ha@vnu.edu.vn
2 VinUni Bigdata Research Institute, VinUniversity, Hanoi City 100000, Vietnam; thao.ntp9@vinuni.edu.vn
3 Faculty of Biotechnology, Vietnam National University of Agriculture, Hanoi City 100000, Vietnam; nqtrung@vnua.edu.vn (T.Q.N.); hqanh.sci@gmail.com (A.Q.H.); dhgioi@vnua.edu.vn (G.H.D.)
4 Institute of Scientific Research and Applications, Hanoi Pedagogical University 2, Phuc Yen City 35000, Vietnam; laviethong@hpu2.edu.vn
5 Department of Science and International Cooperation, Vietnam Academy of Agricultural Sciences, Hanoi City 100000, Vietnam
* Correspondence: thlinh3@mae.gov.vn; Tel.: +84-968226788

Abstract

Single-nucleotide polymorphism (SNP) arrays provide reproducible, accurate, and scalable genotyping for rice research and cultivar development. This review critically evaluates the design and development of major rice SNP platforms and defines their value within a rapidly changing genotyping landscape. In contrast to earlier platform-centered summaries, this review connects array selection with the complete breeding pathway from germplasm assessment and locus discovery to functional validation, marker-assisted transfer, genomic selection, varietal authentication, and pre-breeding. SNP arrays are compared with genotyping-by-sequencing, double-digest restriction site-associated DNA sequencing, low-pass sequencing, whole-genome resequencing, targeted sequencing, and long-read sequencing in terms of cost, throughput, data completeness, computational demand, marker discovery, and population transferability. Candidate-locus validation through gene-expression analysis, haplotype analysis, near-isogenic lines, RNA interference, transgenic complementation, clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas) editing, and field evaluation is also examined. Pangenome resources and the contributions of structural variants, presence/absence variations, and copy number variations to probe design and trait analysis receive particular attention. The review further assesses ascertainment bias, weak detection of rare alleles, genotype-cluster errors, limited portability across distant germplasm, and reduced genomic prediction across populations or environments. Future priorities include multi-omics integration, climate-resilient rice improvement, modular panels, imputation anchor loci, and economical deployment through low-density assays, shared facilities, and kompetitive allele-specific polymerase chain reaction conversion. The value of SNP arrays is greatest when stable genotype calls, large sample numbers, and repeated use across breeding cycles are required. Their practical contribution should be judged by prediction accuracy, selection cost, cycle duration, donor-segment control, realized genetic gain, and cultivar deployment.

Keywords: rice genomics; SNP arrays; genomic selection; marker-assisted breeding; germplasm characterization

Check for updates

Academic Editor: Hui Xia

Received: 25 July 2026
Revised: 22 August 2026
Accepted: 25 August 2026
Published: 31 August 2026

Copyright: © 2026 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.

1. Introduction

Rice (*Oryza sativa*) improvement increasingly depends on the accurate characterization of genome-wide variation within landraces, modern cultivars, breeding lines, and wild *Oryza* accessions [1]. This requirement has gained importance as breeding programs seek higher yield and grain quality under drought, salinity, heat, flooding, new pathogen populations, and variable production environments [2]. Morphological descriptors remain useful for field evaluation, although many traits change across environments and developmental stages. Simple sequence repeat (SSR) markers and other low-density assays often provide insufficient genome coverage for closely related lines [3]. Whole-genome resequencing and long-read sequencing can discover new single-nucleotide changes and complex genomic variants across rice populations [4,5]. Genotyping-by-sequencing (GBS), double-digest restriction site-associated DNA sequencing (ddRAD-seq), skim sequencing, and low-pass sequencing can lower sequencing requirements, whereas targeted sequencing restricts analysis to selected regions [5]. Data completeness and genotype accuracy may vary with read depth, library quality, reference choice, and imputation method. These approaches also require bioinformatic processing, data storage, and quality control for each dataset. Single-nucleotide polymorphism (SNP) arrays offer a distinct balance of features, including predefined markers, standardized genotype clusters, low missing-data rates, high sample throughput, and predictable analysis. Their main advantage occurs in large populations that undergo repeated genotyping across breeding cycles or laboratories [6]. Sequencing-based approaches retain greater value for new variant discovery, rare alleles, structural changes, and germplasm poorly represented during array design [5]. Platform choice should therefore follow sample number, target diversity, required marker resolution, analytical capacity, and the breeding decision supported by the data [7].

Rice SNP-array development has progressed from general diversity platforms to arrays designed for defined populations and breeding tasks [7]. The GeneChip Rice 44K SNP Genotyping Array (Affymetrix 44K) supported early population analyses and genome-wide association studies (GWAS). The Rice SNP 50K Array (RiceSNP50) and 50K Rice SNP Chip (OsSNPnks 50K) increased genic coverage and supplied markers for trait mapping and germplasm analysis [8–10]. The High-Density Rice Array 700K (HDRA700K) provided dense genome coverage for fine-scale association and haplotype analysis. Lower-density platforms, such as whole-genome 6K rice SNP array (RICE6K), Cornell_6K_Array_Infinium_Rice (C6AIR), and Cornell_7K_Array_Infinium_Rice (C7AIR), reduced per-sample cost for fingerprinting, linkage mapping, background recovery, introgression analysis, and genomic selection (GS) [11–14]. Recent platforms, such as the 580K Axiom Rice Genotyping Chip developed at Kongju National University (KNU 580K) and pangenome-derived panels, have added population-specific alleles, functional variants, organellar polymorphisms, and presence/absence variation (PAV) [15,16]. Marker density, however, is only one measure of platform value. A fixed array can genotype only variants selected during its design [7]. Discovery panels dominated by a narrow genetic group can introduce ascertainment bias and reduce polymorphism rates in distant populations. Rare alleles and structural variants (SVs), PAVs, and copy number variations (CNVs) may receive poor coverage, especially in landraces and wild relatives. Probe-binding differences and genotype-cluster definitions may also reduce call accuracy outside the original validation panel. Platform assessment must therefore consider the discovery population, genomic distribution, probe performance, allele frequency, transferability, cost, and intended use.

The purpose of this review is to provide a critical, breeding-oriented assessment of rice SNP arrays and to connect platform choice with measurable genetic and breeding outcomes. The review first examines SNP discovery, probe design, platform density, and the competitive position of arrays relative to GBS, ddRAD-seq, skim sequencing, low-pass

sequencing, whole-genome resequencing, targeted sequencing, and long-read sequencing. It then follows the breeding sequence from germplasm evaluation and parent selection to locus discovery, candidate-gene analysis, functional validation, marker-assisted transfer, GS, variety identification, seed quality control, and pre-breeding. Candidate associations are distinguished from causal evidence through gene-expression analysis, haplotype analysis, near-isogenic lines (NILs), ribonucleic acid (RNA) interference (RNAi), transgenic complementation, clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas) editing, and field tests. The discussion also addresses pangenome-informed marker development, non-SNP variation, ascertainment bias, rare-allele detection, poor portability across germplasm, and reduced prediction across populations or environments. Future priorities cover the integration of SNP-chip data with transcriptomics, epigenomics, and metabolomics, as well as climate-resilient rice breeding and cost-sensitive deployment in developing countries. Platform value is evaluated through genotype quality, prediction accuracy, selection cost, breeding-cycle duration, donor-segment control, genetic gain, and cultivar use in breeding programs.

2. Development and Comparative Position of Rice SNP Arrays

2.1. SNP Discovery, Marker Selection, and Array Construction

The performance of an SNP array originates from the genetic composition of its discovery panel. Early rice arrays obtained candidate variants primarily from comparisons among a limited number of *indica* and *japonica* accessions, which produced markers capable of separating the major varietal groups but provided fewer polymorphic loci within individual breeding populations [8]. Later arrays drew on resequencing datasets that included aus, aromatic, tropical *japonica*, temperate *japonica*, landraces, elite cultivars, breeding lines, and wild *Oryza* accessions [9,17]. Expansion of the discovery panel increased the number and geographical range of detectable variants, although the distribution of those variants remained dependent on the proportions of accessions assigned to each genetic group. A discovery panel dominated by one ancestry group yielded a larger proportion of common alleles from that group and fewer lineage-specific alleles from less represented materials. Reference-genome selection introduced another source of variation. Alignment against the Nipponbare genome preferentially detected polymorphisms in sequences conserved with temperate *japonica*, whereas highly divergent sequences and genomic segments absent from Nipponbare were incompletely represented. Consequently, the candidate marker pool was already a filtered representation of rice diversity before probe design began.

Variant calling and quality filtering reduced millions of initial sequence variants to a smaller set of assayable loci. Candidate SNPs were commonly removed when they showed insufficient read depth, excessive missing data, uncertain alignment, low genotype quality, or abnormal heterozygosity. Rice is predominantly self-pollinated, and recurrent heterozygous calls across inbred accessions frequently indicated alignment of reads from duplicated loci, collapsed paralogs, residual heterogeneity, or sequencing errors. Minor allele frequency (MAF) filtering increased the probability that both alleles would be represented during assay validation and genotype-cluster training [18]. This procedure also removed rare and geographically restricted variants. The KNU 580K array assembled variants from 3475 accessions and applied thresholds of MAF ≥ 0.05 and missing data ≤ 0.10 to its primary sequence-derived marker set before incorporating additional markers from legacy arrays and functional databases [15]. The C6AIR used an observation rate above 80% and an MAF above 20% within specified subpopulations for part of its marker content, while another component was selected according to segregation in crosses represented in the design process [13,14]. These contrasting filters produced different allele-frequency spectra and levels of informativeness within rice subpopulations.

Marker selection also determined the distribution of array content among genomic regions. Genome-wide markers were selected according to chromosome position and the size of gaps between neighboring loci. Population-informative markers were chosen for allele-frequency differences among rice groups, while gene-oriented markers represented coding sequences, introns, untranslated regions, promoters, cloned genes, and previously characterized quantitative trait loci (QTLs). Exonic SNPs included synonymous, nonsynonymous, splice-site, and putative loss-of-function variants. Gene enrichment increased coverage of known functional regions but reduced the proportion of array content available for intergenic and poorly annotated sequences. Physical marker density also differed from effective genetic density. Multiple SNPs within an extended linkage disequilibrium (LD) block frequently represented the same haplotype, whereas genomic regions with rapid LD decay contained fewer correlated loci. Therefore, arrays with similar mean physical spacing could provide different numbers of independent polymorphic markers in *indica*, aus, aromatic, and *japonica* populations.

Probe design imposed an additional technical filter. Successful hybridization depended on the uniqueness, nucleotide composition, and sequence conservation of the regions surrounding each SNP. Repetitive sequences, transposable elements, paralogous genes, segmental duplications, and high-copy genomic regions increased the probability of cross-hybridization. A second polymorphism within the probe-binding region could weaken hybridization of one haplotype and generate allele-specific signal loss. C6AIR excluded candidates with another variant within 10 base pairs of the target SNP, common variants within the surrounding 35-base-pair sequence, nearby insertions or deletions, and repetitive flanking sequences [13]. C7AIR applied related sequence-quality criteria when new probes were added to the platform [14,19]. In the Axiom design used for KNU 580K, predicted probe-conversion scores incorporated sequence uniqueness, nucleotide composition, nearby polymorphisms, and nonspecific binding. Candidate probes were also removed when their flanking sequences contained duplications, long homopolymer tracts, or several neighboring variants [15]. The two DNA strands often differed in predicted binding performance, which allowed selection of the strand with the more distinctive flanking sequence.

Array manufacture was followed by empirical classification of samples and loci. Sample-level measurements included DNA concentration and purity, total call rate, observed heterozygosity, agreement between technical replicates, and correspondence between declared and molecular identities. Marker-level measurements included call frequency, fluorescence-cluster separation, reproducibility, Mendelian consistency, and concordance with sequence-derived genotypes. Inbred accessions usually formed two homozygous clusters, while F_1 plants and segregating progeny supplied heterozygous genotypes for the definition of the intermediate cluster. Markers without a minor homozygous class could still produce reliable calls, although their cluster boundaries were less well defined than those of loci represented by all three genotype classes. A marker with a high overall call rate could remain monomorphic within a particular rice subgroup. Manufacturing conversion rate, call rate, polymorphic rate, and population-specific informativeness consequently represented distinct components of technical performance.

2.2. Design Progression of Major Rice SNP Arrays

Rice SNP arrays have progressed from moderate-density platforms based on a limited number of reference genomes to high-density and pangenome-informed platforms based on thousands of accessions [20]. Their discovery resources, marker-selection criteria, marker composition, array density, and validation materials are summarized in Table 1. The Affymetrix 44K array contained 44,100 SNPs across the 12 rice chromosomes and provided

one of the first standardized genome-wide marker systems for cultivated rice [17]. Its application to 413 diverse accessions generated a common genotype dataset for the major *O. sativa* groups [8]. The discovery resources remained concentrated on variants accessible through the reference genomes and diversity panels available at the time of development.

Table 1. Design characteristics, marker composition, genomic coverage, and validation performance of major single-nucleotide polymorphism arrays developed for rice genotyping. C6AIR, Cornell_6K_Array_Infinium_Rice; C7AIR, Cornell_7K_Array_Infinium_Rice; HDRA, High-Density Rice Array; KNU, Kongju National University; MAF, minor allele frequency; PAV, presence/absence variation; RPGA, Rice Pangenome Genotyping Array; SNP, single-nucleotide polymorphism.

Array	Discovery Resources	Principal Marker-Selection Criteria	Marker Composition	Marker Number and Density or Coverage	Validation Panel and Reported Performance	References
Affymetrix 44K rice array	Early rice reference genomes and diversity panels	Genome-wide distribution, assay suitability, and representation of polymorphisms among major cultivated rice groups	Predominantly biallelic SNPs	44,100 SNPs distributed across all 12 chromosomes	Applied to 413 accessions from the principal *Oryza sativa* groups	[8,17]
RiceSNP50	More than 10 million candidate variants identified from 801 resequenced rice varieties	Genome-wide distribution, probe suitability, representation of gene-rich regions, and inclusion of agronomic loci	51,478 SNPs; approximately 68% located within annotated genes	51,478 loci with genome-wide coverage	Parental accessions, first filial generation (F_1) plants, and diverse rice lines; a rice-specific genotype-cluster file was established	[9]
RICE6K	More than four million SNPs identified from over 500 rice landraces	Genome-wide distribution, technical assay performance, and inclusion of validated functional loci	4473 high-quality SNPs and 45 functional markers within 28 characterized genes	5636 submitted assays; 5102 passed initial detection and 4473 met high-quality criteria	Diverse rice germplasm was used to assess assay conversion, call quality, and polymorphism	[11]
C6AIR	Earlier 384-SNP panels and resequencing data from major rice subpopulations and breeding parents	Observation rate above 80%; MAF above 20% within selected subpopulations; segregation in parental combinations; exclusion of nearby polymorphisms, insertions or deletions, and repetitive flanking sequences	1571 legacy markers and 4429 newly selected markers; the new set included 2000 population-informative and 2429 cross-informative loci	6000 attempted bead types; 5274 passed manufacturing quality control; 4606–4940 markers were retained in evaluated datasets	International Rice Research Institute and Cornell University datasets; cultivated accessions and breeding populations were used for genotype-cluster assessment	[13]

Table 1. *Cont.*

Array	Discovery Resources	Principal Marker-Selection Criteria	Marker Composition	Marker Number and Density or Coverage	Validation Panel and Reported Performance	References
C7AIR	C6AIR, High-Density Rice Array, earlier GoldenGate panels, Affymetrix 44K array, and trait-associated loci	Removal of poorly resolved or consistently uninformative C6AIR probes; selection for sequence quality, genome coverage, population informativeness, and continuity with legacy platforms	7098 unique loci, including 4007 markers retained from C6AIR	7098 loci distributed across the rice genome	Asian cultivated rice, African cultivated rice, selected wild relatives, and segregating populations; informativeness differed among *indica*, tropical *japonica*, aus, temperate *japonica*, and aromatic rice	[14]
OsSNPnks	Rice gene annotations, rice-wheat conserved genes, cloned agronomic genes, and selected multi-copy genes	Preference for single-copy genes, functional relevance, genic position, sequence uniqueness, and probe suitability	50,051 SNPs in 18,980 genes; 38% exonic, 41% intronic, 21% in untranslated regions, and approximately 96% in single-copy genes	50,051 loci concentrated within genic regions	192 cultivated and wild rice accessions; reported call and reproducibility rates were approximately 99.9%	[10]
HDRA	Genome-wide rice diversity resources and annotated gene regions	Genome-wide physical coverage, high marker density, and enrichment near annotated genes	Approximately 700,000 SNPs	Mean density of approximately one SNP per 0.54 kilobase	A purified panel of 1568 rice accessions that represented the principal subpopulations	[12]
KNU 580K	Korean World Rice Collection, 3000 Rice Genomes Project, HDRA, Affymetrix 44K array, functional gene databases, chloroplast and mitochondrial sequences, and genetically modified event markers	MAF $\geq$ 0.05 and missing data $\leq$ 0.10 for the primary sequence-derived set; predicted probe conversion, sequence uniqueness, genomic distribution, functional annotation, and population specificity	Genome-wide nuclear SNPs, functional markers, legacy-array markers, organellar variants, and genetically modified event-associated markers	7,682,442 candidates; approximately 3.20 million technically suitable loci; 620,852 prioritized assays; 581,006 markers in the final platform	4720 samples from cultivated accessions, landraces, wild rice, breeding lines, mutants, and segregating populations	[15]
Rice3K56	Resequencing data from 3024 rice accessions	Genome-wide distribution, allele frequency, physical spacing, and probe-design suitability	56,606 genome-wide SNPs	Mean physical spacing of approximately 6.7 kilobases	192 representative rice accessions; estimated genotyping reliability was approximately 99.6%	[21]

Table 1. *Cont.*

Array	Discovery Resources	Principal Marker-Selection Criteria	Marker Composition	Marker Number and Density or Coverage	Validation Panel and Reported Performance	References
RPGA	Core and dispensable genomic regions from the 3000-rice pangenome	Representation of pangenome diversity, probe suitability, genome-wide distribution, and direct coverage of sequences absent or poorly represented in Nipponbare	Approximately 80,000 SNP and PAV assays	Approximately 80,000 assays across core and dispensable pangenome regions	271 rice accessions; evaluation included repeatability of SNP calls and PAV signal classes	[16]
WDR6K	Genome-wide rice variants; upland-lowland differentiated loci; known genes and QTLs for drought resistance, water and nutrient use, and direct seeding; stress-responsive expression data	Even genomic distribution; missense variation; upland-lowland differentiation; drought-related functional evidence; capture-probe suitability	5895 SNP/InDel loci in 5001 genes; >90% missense SNPs; approximately 41% upland-lowland differentiated loci; 922 drought-resistance-contributing loci	5895 loci across 12 chromosomes; updatable liquid-phase capture panel	Good repeatability across technical and biological replicates; high genotyping efficiency; functional enrichment and stress-expression support for drought-related content	[22]

RiceSNP50 expanded marker discovery through analysis of more than ten million candidate variants from 801 resequenced varieties [9]. The final Infinium array contained 51,478 loci, and approximately 68% occurred within annotated genes. The design included markers across all chromosomes, gene-rich regions, and previously reported agronomic loci. Validation materials comprised parental accessions, F_1 plants, and diverse rice lines. These materials supported construction of a rice-specific genotype-cluster file. The cluster file reduced classification errors that arose when generic parameters were applied to the characteristic homozygous and heterozygous signal distributions of rice. RiceSNP50 combined broad discovery-panel representation with population-specific genotype classification [9].

The RICE6K and C6AIR represented lower-density strategies that emphasized marker informativeness. RICE6K derived candidate loci from more than four million SNPs detected in over 500 landraces [11]. Among 5636 submitted assays, 5102 passed initial detection, and 4473 were classified as high-quality markers. The array also contained 45 validated functional markers within 28 characterized genes. C6AIR began with 6000 attempted bead types, of which 5274 passed manufacturing quality control [13]. Standardized analytical pipelines retained 4606 markers in the dataset evaluated at the International Rice Research Institute and 4940 markers in the Cornell University dataset. Its design combined 1571 markers from earlier 384-SNP panels with 4429 markers selected from resequencing data. Of the new markers, 2000 represented loci with intermediate allele frequencies within five major *O. sativa* subpopulations. The remaining 2429 loci represented polymorphism in parental combinations within the design program. The platform contained fewer loci than high-density arrays, but much of its content was polymorphic in the intended germplasm.

C7AIR was developed through empirical reassessment of C6AIR [14]. Probes with poor resolution or consistently low informativeness were removed. New markers came

from the HDRA, earlier GoldenGate panels, the Affymetrix 44K array, and trait-associated loci. The final platform contained 7098 unique loci, of which 4007 originated directly from C6AIR [14]. Its cluster file used cultivated Asian rice, African cultivated rice, selected wild relatives, and segregating populations. Markers shared with earlier platforms connected genotype data across array generations through a common set of anchor loci. Differences in polymorphism among *indica*, tropical *japonica*, aus, temperate *japonica*, and aromatic rice also demonstrated the effect of population ancestry on effective marker content.

The OsSNPnks and HDRA used contrasting high-density strategies. Specifically, OsSNPnks concentrated 50,051 SNPs within 18,980 genes [10]. Approximately 96% of its markers occurred in single-copy genes. This feature reduced ambiguous hybridization from closely related paralogs. The represented genes comprised 3710 genes conserved between rice and wheat, 14,959 genes classified as single-copy in rice, 194 cloned agronomic genes, and 117 selected multi-copy genes. Approximately 38% of the SNPs were exonic, 41% were intronic, and 21% occurred in untranslated regions. Validation across 192 cultivated and wild accessions produced reported call and reproducibility rates of approximately 99.9%. The concentration of markers within genes generated detailed genic haplotypes but provided less direct coverage of intergenic regulatory regions. Meanwhile, HDRA contained approximately 700,000 SNPs with an average physical density close to one SNP per 0.54 kilobase [12]. The platform combined genome-wide coverage with additional enrichment near annotated genes. Application to a purified panel of 1568 accessions generated a high-density reference dataset across the principal rice subpopulations. Despite its uniform assay manifest, the number of variable loci differed markedly among genetic groups. Closely positioned markers also showed extensive correlation within genomic regions that contained long LD blocks. The number of physically assayed SNPs exceeded the number of statistically independent haplotypes in many intervals.

KNU 580K combined population-specific variants, international rice diversity data, legacy-array markers, functional annotations, and organellar polymorphisms [15]. Its eight marker sources comprised 475 accessions from the Korean World Rice Collection [23], variants from the 3000 Rice Genomes Project [24], markers from HDRA and the Affymetrix 44K array, functional gene databases, chloroplast and mitochondrial sequences, and markers associated with genetically modified events. Integration of these sources produced 7,682,442 candidate loci. Probe-quality prediction reduced this pool to approximately 3.20 million technically suitable candidates. A subsequent selection step prioritized 620,852 loci for manufacture. The final platform contained 581,006 markers. Its evaluation included 4720 samples from cultivated accessions, landraces, wild rice, advanced lines, mutants, and progeny populations. This validation set permitted comparison of probe behavior across materials with marked differences in ancestry and sequence divergence.

Rice3K56 was developed from resequencing data for 3024 rice accessions [21]. The array contained 56,606 SNPs with an average physical spacing of approximately 6.7 kilobases. Marker selection considered genome-wide distribution, allele frequency, and probe suitability. Evaluation in 192 representative accessions produced an estimated genotyping reliability of approximately 99.6%. Its density occupied an intermediate position between the 6K-7K platforms and arrays with several hundred thousand markers. The combination of a moderate marker number and broad discovery resources distinguished Rice3K56 from earlier low-density panels based on smaller population samples. The Rice Pangenome Genotyping Array (RPGA) extended marker discovery beyond variants represented adequately by the Nipponbare reference [16]. The array contained approximately 80,000 SNP and PAV assays from core and dispensable regions of the 3000-rice pangenome. The platform contained probes within sequences that were absent or poorly represented in conventional Nipponbare-based pipelines. The RPGA differed from previous arrays in

both marker type and reference structure. Its manifest represented variation across multiple constituent genomes instead of variation detected exclusively through alignment to one linear reference.

More recently, the Water-saving and drought-resistance rice 6K SNP array (WDR6K) introduced a trait-directed genotyping design for water-saving and drought-resistance breeding in rice. The liquid-phase panel contains 5895 SNP/InDel loci distributed across the 12 rice chromosomes and associated with 5001 genes. More than 90% of the SNPs are missense variants, approximately 41% differentiate upland and lowland rice, and 922 loci were classified by the developers as contributors to drought resistance. Gene Ontology analysis, pathway enrichment, and transcriptional responses to drought, osmotic stress, oxidative stress, and abscisic acid provided additional support for the drought-related marker composition. The platform applies pinpoint sequencing to captured targets, which permits the addition of new markers in subsequent panel versions. WDR6K complements RPGA, as RPGA represents core and dispensable pangenome variation, whereas WDR6K prioritizes functional and ecotype-differentiated loci for the selection of drought-resistance and water-saving traits [22].

2.3. Pangenome-Informed Markers and Non-SNP Variation

A linear reference genome represents only part of the sequence diversity in cultivated and wild rice. Analysis of 3010 cultivated rice genomes identified approximately 29 million SNPs, 2.4 million small insertions and deletions, more than 90,000 SVs, and over 10,000 full-length protein-coding genes absent from the Nipponbare annotation [24]. Subsequent pangenome studies confirmed extensive non-reference sequence diversity across cultivated and wild rice [20,25,26]. Analysis of 33 genetically diverse rice accessions detected 171,072 SVs and 25,549 gene CNVs [20]. A genus-level graph pangenome from 251 Asian and African cultivated and wild accessions contained 1.52 gigabases of non-redundant sequence. This total included 1.15 gigabases absent from Nipponbare [26]. Among 51,359 non-redundant genes, 21,888 were classified as core and 29,471 as dispensable. These datasets showed that reference-based SNP discovery omitted many variable genes and genomic segments before candidate loci entered an array-design pipeline.

SVs comprise deletions, insertions, inversions, duplications, translocations, and complex rearrangements that exceed the usual size range of SNPs and small insertions or deletions. PAVs describe genes or genomic segments that occur in some accessions but not in others. CNVs represent differences in the number of copies of a sequence. These variants can alter coding regions, promoters, gene dosage, chromosomal organization, and recombination patterns. An array SNP near an SV may act as an indirect marker through LD. This association depends on the haplotype that carries the structural allele and may differ among rice populations. Direct SV assays instead interrogate sequences inside the variable region, sequences that span a breakpoint, or combinations of internal and flanking signals. PAV genotyping differs from conventional biallelic SNP classification. Weak hybridization may indicate biological absence of the target sequence. It may also arise from poor DNA quality, divergence within the probe-binding region, or technical assay failure. PAV assays use distributions of normalized signal intensity and may include several probes within the same variable segment. Successful signals from flanking reference probes and concordant loss of internal signals provide stronger evidence of sequence absence than failure of one probe. RPGA incorporated predefined PAV assays from core and dispensable genomic regions. These assays assigned presence and absence classes from repeatable signal patterns [16].

CNV classification presents a separate analytical issue. Standard diploid SNP clustering uses two homozygous classes and one heterozygous class. This model does not directly

represent several copy-number states. Array-based CNV detection instead uses normalized fluorescence intensity across multiple probes within the same genomic region. Materials with known copy numbers provide reference distributions for dosage classification. Signal intensity can also vary with nucleotide composition, probe affinity, DNA concentration, and experimental batch. CNV inference thus depends on intensity normalization and agreement across neighboring probes. Large inversions and complex insertions are difficult to define from array signals alone. Their direct characterization depends on prior knowledge of the altered sequence or breakpoint. Long-read sequencing of 111 rice genomes identified 604 megabases of non-reference sequence among 105 *O. sativa* accessions. The total reached approximately 879 megabases after six wild accessions were added [27]. These assemblies identified insertions, rearrangements, and highly variable gene regions that were inaccessible or fragmented in short-read alignments. After characterization of an SV, its internal sequence or breakpoint junction could enter a fixed genotyping assay. Pangenome arrays consequently assay a predefined subset of known structural variation and do not independently discover novel rearrangements in newly genotyped samples.

Pangenome-derived markers also require coordinate relationships among the source assembly, graph genome, and conventional linear reference. A locus within a dispensable segment may lack a direct Nipponbare coordinate. Related sequences may also occur at different positions across several genomes. RPGA associated its markers with the pangenome resources used during design [16]. Nevertheless, its genotypes remained limited to alleles encoded in the array manifest. Multi-allelic insertions, variable tandem repeats, nested rearrangements, and previously uncharacterized structural alleles could not always be represented by one biallelic assay. Pangenome-informed arrays expanded the sequence space covered by fixed genotyping but retained the predefined character of array content.

2.4. Technical Comparison with Sequencing-Based Genotyping

SNP arrays and sequencing-based methods differ principally in whether marker loci are defined before or after sample analysis. Arrays interrogate the same predefined sites in every sample and convert fluorescence intensities into genotype clusters. GBS, ddRAD-seq, skim sequencing, low-pass whole-genome sequencing, whole-genome resequencing, targeted sequencing, and long-read sequencing obtain reads from each sample [5,28]. Genotypes are then derived through alignment, variant calling, genotype-likelihood estimation, or imputation. Fixed markers give arrays high consistency across batches. Sequence data retain information outside a predefined marker catalog and permit reanalysis after changes in reference assemblies or variant databases.

GBS reduces genome complexity through restriction-enzyme digestion and sequencing of fragments adjacent to restriction sites [28], while ddRAD-seq uses two restriction enzymes and a specified fragment-size interval. This design gives tighter control over the genomic fraction represented in each library [29]. Both methods can detect variants absent from an existing array manifest. Their recovered loci depend on restriction-site presence, digestion efficiency, fragment-size selection, library balance, read depth, and bioinformatic filters. A mutation within a restriction site prevents the corresponding fragment from entering the library and creates allele dropout. Population differences in restriction-site sequences can thus generate non-random missing data. Variation in read depth may also cause a locus to pass genotype-quality thresholds in one dataset but fail in another. Arrays lack restriction-site dropout and usually return a more consistent locus set, but they produce no genotype information outside the manufactured probes. Skim sequencing and low-pass whole-genome sequencing distribute reads across the genome without deliberate restriction-site reduction. Both terms describe coverage below the level required

for confident independent genotype calls at every genomic position, although their exact depth ranges vary among studies. Rice provided an early example through resequencing of a recombinant inbred population at approximately 0.02-fold coverage [30]. Deeply sequenced parental genomes and long inherited haplotype blocks supported genotype inference across genomic windows despite sparse reads from each progeny line. In diverse rice collections, accuracy depends more strongly on reference-panel composition and local haplotype representation. At low coverage, a heterozygous site may appear homozygous if only one allele is sampled. Rare alleles and divergent introgressed segments also show lower imputation accuracy when related haplotypes are absent from the reference panel.

Arrays can also provide marker scaffolds for imputation. The rice HDRA dataset was expanded from approximately 700,000 directly assayed loci to more than 4.8 million variants after optimization against sequence-based reference genotypes [31]. Directly genotyped markers supplied stable haplotype anchors. The reference panel supplied unobserved sequence variants. Accuracy of the inferred genotypes varied with allele frequency, genetic relatedness, local LD, and reference-panel representation. Low-pass sequencing and array-based imputation both depend on the density and population compatibility of observed markers and reference haplotypes.

Whole-genome resequencing provides direct short-read access to genome-wide SNPs, small insertions and deletions, and population-specific alleles. The 3000 Rice Genomes Project generated a variant catalog that contributed to KNU 580K, Rice3K56, RPGA, and rice imputation resources [15,16,21,32]. Sequence-derived marker sets remain sensitive to read depth, alignment quality, reference choice, variant-detection software, and quality-filter parameters. Short reads also align poorly within repeated regions, duplicated genes, resistance-gene clusters, and structurally rearranged segments. Long-read sequencing resolves a larger proportion of these regions and detects insertions, inversions, tandem duplications, and complex breakpoints [27]. Its greater sequence span has distinct requirements for DNA quality, assembly, storage, and structural-variant analysis.

Targeted sequencing limits data generation to predefined genes or genomic intervals. Amplicon panels use multiplexed primers, whereas capture panels use hybridization probes to enrich selected sequences. Unlike SNP arrays, targeted sequencing can detect several alleles within a sequenced interval. These alleles may include variants that were not selected explicitly when the panel was constructed. Missing data can arise from primer-binding polymorphisms, uneven amplification, capture efficiency, competition among multiplexed targets, and read-depth variation. The 1K-Rice Custom Amplicon assay (1k-RiCA) contained 995 SNPs with strong representation of elite *indica* breeding materials [33]. Its marker content provided high information density within closely related target germplasm but lower representation of variation specific to other rice groups.

These methods occupy distinct technical positions with respect to marker stability, discovery capacity, the structure of missing data, and computational demand (Table 2). Arrays generate standardized genotypes at known loci with little variation in marker identity among batches. GBS and ddRAD-seq recover population-specific variants, but their marker sets depend on restriction sites and library construction. Skim and low-pass sequencing produce sparse genome-wide observations that become dense after haplotype inference. Whole-genome resequencing provides broader discovery of short variants. Long-read sequencing extends analysis to complex genomic regions and structural alleles. Targeted sequencing provides deeper information within a limited genomic fraction. Nominal marker number alone does not establish a universal ranking among these platforms.

Table 2. Technical comparison of single-nucleotide polymorphism arrays and sequencing-based platforms used for rice genotyping and variant discovery. ddRAD-seq, double-digest restriction site-associated DNA sequencing; GBS, genotyping-by-sequencing; HDRA, High-Density Rice Array; PAV, presence/absence variation; SNP, single-nucleotide polymorphism.

Genotyping Platform	Marker Definition and Primary Data	Genome Coverage and Locus Stability	Principal Sources of Missing Data or Error	Imputation Dependence	Variant-Discovery Capacity	Computational Requirements	References
SNP array	Fluorescence intensities are measured at predefined SNP or PAV assays. The same physical loci are tested in every sample.	Genome-wide coverage depends on the manufactured probe set. Marker identity remains highly consistent across samples and experimental batches.	Poor DNA quality, polymorphisms within probe-binding sequences, cross-hybridization, target-sequence absence, weak cluster separation, and experimental batch effects	Optional. Directly assayed markers can serve as haplotype anchors for inference of unassayed variants. The HDRA dataset increased from approximately 700,000 assayed SNPs to more than 4.8 million variants through reference-panel imputation.	No discovery outside the array manifest. Pangenome-informed arrays can genotype previously characterized PAVs and structural alleles if suitable probes are present.	Low to moderate. Principal steps include fluorescence normalization, genotype-cluster assignment, sample quality control, marker quality control, and optional imputation.	[8,9,17]
GBS	Sequence reads are obtained from genomic fragments adjacent to restriction-enzyme recognition sites. Marker loci are defined after sequence analysis.	Reduced and non-uniform genome representation. The recovered locus set can differ among samples, populations, libraries, and experiments.	Restriction-site polymorphisms, incomplete digestion, uneven library representation, low read depth, read-alignment errors, and genotype-quality filters	Commonly used to reduce missing genotypes and increase marker density. Accuracy depends on local haplotypes, population structure, marker density, and reference-panel compatibility.	Detects SNPs and small insertions or deletions within recovered fragments. Variants outside the represented fragments remain undetected.	Moderate to high. Requirements include sequence de-multiplexing, read-quality control, alignment, variant detection, genotype filtering, and often imputation.	[28]
ddRAD-seq	Two restriction enzymes and a selected fragment-size interval define the sequenced genomic fraction. Variants are identified after read analysis.	Reduced genome representation with greater control over fragment selection than conventional GBS. Locus recovery still varies with restriction-site conservation and library quality.	Mutations at either restriction site, variation in size selection, incomplete digestion, unequal amplification, low read depth, and read-alignment errors	Often used when read depth is insufficient for complete genotype recovery. Its accuracy depends on the observed marker distribution and reference-panel representation.	Detects SNPs and small insertions or deletions within fragments that contain both required restriction sites and fall within the selected size interval.	Moderate to high. Library-specific quality control and fragment-based variant analysis increase processing requirements relative to fixed arrays.	[29]

Table 2. *Cont.*

Genotyping Platform	Marker Definition and Primary Data	Genome Coverage and Locus Stability	Principal Sources of Missing Data or Error	Imputation Dependence	Variant-Discovery Capacity	Computational Requirements	References
Skim sequencing and low-pass whole-genome sequencing	Sparse sequence reads are distributed across the genome without deliberate restriction-site reduction. Genotypes are estimated from read evidence, genotype likelihoods, haplotypes, or reference-panel imputation.	Genome-wide but shallow coverage. Directly observed sites differ among samples due to stochastic read sampling.	Low read depth, uneven genomic coverage, failure to sample both alleles at heterozygous sites, reference bias, read-alignment errors, and absence of suitable reference haplotypes	High for construction of a dense genotype dataset. Rice recombinant inbred lines were genotyped at approximately 0.02-fold coverage through parental haplotypes and long inherited chromosomal segments.	Direct discovery is limited at very low coverage. Common variants represented in a reference panel can be inferred, but rare alleles and divergent introgressions show lower accuracy.	Moderate to high. Major tasks include read alignment, genotype-likelihood estimation, haplotype inference, and large-scale imputation. Per-sample storage remains lower than that of deep resequencing.	[30,31]
Short-read whole-genome resequencing	Genome-wide short reads are aligned to a reference genome. Variants are identified from sample-specific sequence evidence.	Broad genome-wide coverage. Coverage declines within repetitive sequences, duplicated genes, highly divergent segments, and complex rearrangements.	Insufficient read depth, uneven coverage, read-alignment ambiguity, reference-genome bias, sequencing errors, variant-detection parameters, and quality-filter thresholds	Optional at moderate or high coverage. Imputation can improve low-confidence genotypes and harmonize datasets with different marker densities.	High capacity for SNPs and small insertions or deletions. Detection of large insertions, inversions, tandem duplications, and complex structural variants remains limited by short read length.	High. Requirements include substantial data storage, read-quality control, reference alignment, variant detection, joint genotyping, and multi-sample quality control.	[24]
Targeted sequencing	Multiplexed primers or hybridization probes enrich predefined genes or genomic intervals. Sequence variants are detected within each captured region.	Restricted to selected targets. Coverage within those targets may be deep, but no direct information is obtained from unselected genomic regions.	Primer-binding polymorphisms, uneven amplification, variable capture efficiency, competition among multiplexed targets, low read depth, and off-target reads	Usually optional within well-covered targets. Its value increases when coverage is uneven or when a suitable haplotype reference is available.	Detects known and previously uncharacterized SNPs, small insertions or deletions, and multiple alleles within the targeted intervals. The 1k-RiCA assay contained 995 SNP targets, selected primarily from elite *indica* materials.	Moderate. Data volumes are smaller than those of whole-genome resequencing, but primer-specific or target-specific quality control remains necessary.	[33]

Table 2. *Cont.*

Genotyping Platform	Marker Definition and Primary Data	Genome Coverage and Locus Stability	Principal Sources of Missing Data or Error	Imputation Dependence	Variant-Discovery Capacity	Computational Requirements	References
Long-read sequencing	Individual long DNA molecules are sequenced and analyzed through reference alignment, de novo assembly, or both approaches.	Broad genome representation with improved access to repetitive regions, duplicated genes, non-reference sequences, and complex genomic rearrangements.	Degraded or fragmented DNA, insufficient molecule length, uneven depth, platform-specific sequence errors, assembly errors, and uncertainty at complex structural breakpoints	Not required for structural-variant discovery. Reference panels may support genotype refinement and comparison of structural alleles across populations.	Highest capacity for large insertions, deletions, inversions, tandem duplications, translocations, complex rearrangements, and non-reference sequence discovery. Long-read analysis of 111 rice genomes identified extensive sequence absent from conventional reference assemblies.	High to very high. Requirements include high-capacity storage, long-read quality control, genome assembly or long-read alignment, structural-variant detection, and cross-genome coordinate analysis.	[27]

3. Applications of Rice SNP Chips from Germplasm Evaluation to Breeding

3.1. Germplasm Evaluation and Selection of Genetic Resources

The first function of SNP-chip analysis is to define the genetic composition of materials that will enter subsequent research or breeding (Table 3). The Affymetrix 44K dataset for 413 *O. sativa* accessions from 82 countries resolved *indica*, *aus*, temperate *japonica*, tropical *japonica*, and aromatic groups. Pairwise Fixation Index values of 0.23–0.53 confirmed strong differentiation among the major cultivated rice populations [8]. The analysis identified admixed accessions with genomic contributions from divergent ancestry groups, which provided material for crosses that combine favorable alleles. HDRA700K expanded this analysis to 1568 accessions and assigned them to five major subpopulations and three admixed classes under a 70% ancestry threshold [12]. In wild and weedy germplasm, OsSNPnks 50K separated 272 Indian accessions of *O. nivara*, *O. rufipogon*, and *O. sativa* f. *spontanea* into three ancestry groups, five phylogenetic clusters, and nine subclusters [34]. These classifications define the population structure that must be accounted for during association analysis and help breeders select panels with sufficient variation for target traits. Sampling imbalance, marker ascertainment, and poor coverage of rare wild alleles can alter inferred relationships, so chip-based group assignment should be examined with passport information and phenotypic data.

Table 3. Representative applications of rice single-nucleotide polymorphism arrays to germplasm characterization, population structure analysis, accession authentication, and diversity assessment. AMOVA, analysis of molecular variance; C6AIR, Cornell_6K_Array_Infinium_Rice; DAPC, discriminant analysis of principal components; F_{ST}, fixation index; IBS, identity by state; LD, linkage disequilibrium; NJ, neighbor joining; PCA, principal component analysis; SNP, single-nucleotide polymorphism; UPGMA, unweighted pair-group method with arithmetic mean.

#	Germplasm Panel	Platform and Markers Used	Main Analyses	Principal Findings	References
1	413 *O. sativa* accessions from 82 countries	Affymetrix 44K, 44,100 SNPs	PCA, F_{ST}, IBS, LD	Five groups comprising *indica*, aus, temperate japonica, tropical japonica, and aromatic rice; pairwise F_{ST} ranged from 0.23 to 0.53	[8]
2	128 accessions selected from 192 cultivated, landrace, and wild samples	OsSNPnks 50K, 50,051 SNPs	Haplotype analysis, neighbor-joining tree	Four groups separated wild rice, wild-aus materials, *indica* cultivars, and *japonica* cultivars	[10]
3	1568 accessions from global diversity panels	HDRA700K, approximately 700,000 SNPs	fastSTRUCTURE, PCA, LD	Five principal subpopulations and three admixed classes; the first three PCs explained 40% of genomic variation	[12]
4	More than 40,000 cumulative breeding and germplasm samples	C6AIR, 4606 standard high-quality SNPs	Diversity, fingerprinting, pedigree, and introgression analyses	Demonstrated routine large-scale characterization of cultivated rice, breeding lines, and wild-relative introgression materials	[13]
5	1282 farmer samples and 24 breeder references from Bangladesh; 1171 farmer samples retained	C6AIR, 4503 polymorphic SNPs after filtering	IBS, UPGMA, reference matching	Classified 247 samples as modern varieties, 216 as Swarna-related, and 708 as traditional varieties; more than half of genetically identified modern varieties had incorrect farmer names	[35]
6	4591 accessions obtained through integration of the 3K Rice Genomes and HDRA panels	HDRA genotypes imputed against sequence data	fastSTRUCTURE, ancestry assignment, imputation	Integrated array and sequence resources and assigned accessions to rice subpopulations using a 75% ancestry threshold	[24]
7	44 weedy rice, 28 cultivated *japonica*, 16 traditional red rice, 4 *aus*, and 4 *indica* accessions	Combined Affymetrix 44K and RICE6K panel	PCA, phylogeny, F_{ST}, selective-sweep analysis	Nagano weedy rice formed a distinct group with low diversity and was genetically separated from local traditional red rice	[36]
8	272 wild accessions of *O. nivara*, *O. rufipogon*, and *O. sativa* f. *spontanea*	OsSNPnks 50K, 50,051 SNPs	Population structure, PCA, diversity, kinship	The wild rice panel was divided into three major subpopulations and supplied diverse donors for resistance breeding	[34]
9	154 Yunnan local varieties, three wild accessions, and Nipponbare	High-density whole-genome SNP array	Cluster analysis, genetic-background and functional-allele assessment	The local varieties formed three groups; favorable resistance alleles were uncommon, whereas favorable *Wx* and *ALK* alleles were frequent and the favorable *BADH2* allele occurred in about one fifth of the collection	[37]

Table 3. *Cont.*

#	Germplasm Panel	Platform and Markers Used	Main Analyses	Principal Findings	References
10	4720 genotypes, including 1123 breeding lines, 252 landraces, 488 weedy lines, 96 *O. rufipogon* accessions, and 245 wild materials	KNU 580K, 581,006 variants	Genotype quality assessment and polymorphism classification	All accepted samples had call rates above 95%; the panel demonstrated broad coverage of cultivated, weedy, wild, mutant, and breeding materials	[15]
11	298 landraces from six regions of India	OsSNPnks 50K; 32,782 SNPs retained	fastSTRUCTURE, DAPC, NJ tree, AMOVA	SNP data identified three populations and five DAPC clusters; Uttar Pradesh landraces had the highest SNP-based gene diversity of 0.49	[38]
12	192 representative rice accessions	Rice3K56, 56,606 SNPs	Variety differentiation, diversity and population analyses	The array achieved 99.6% genotyping reliability and separated genetically distinct materials with uniform genome coverage at a mean interval of 6.7 kb	[21]

Accurate population definition also requires correction of mislabeled, duplicated, or heterogeneous accessions before phenotypic records are linked to genotypes. Farm-level genotyping in Bangladesh found extensive disagreement between farmer-assigned names and molecular identity [35]. Such errors can generate false pedigree relationships, bias allele-frequency estimates, and weaken association signals. Genome-wide similarity can identify candidate duplicates and unexpected seed mixtures, but genebank curation also requires passport records, collection sites, seed history, phenotypes, and authenticated references. Closely related materials may carry different alleles at a few loci that control disease resistance, maturity, grain quality, or adaptation. Consequently, SNP chips provide an initial molecular audit that improves the composition of diversity panels and the accuracy of downstream locus discovery.

After accession identity and ancestry have been established, marker data can reduce large collections to manageable panels without losing major components of genetic variation. Roy Choudhury et al. (2021) [39] examined 2242 accessions from eastern coastal India with a 36-SNP assay and derived a core set of 247 accessions plus a mini-core of 30 accessions. Gene diversity reached 0.34 and 0.40 in the two reduced panels compared with 0.29 in the complete collection, and the mini-core retained 99% of the detected distinctiveness [39]. Kumar et al. (2020) [32] used SNPs, 18 agronomic traits, and geographic origin to select 520 accessions from 3004 materials. The resulting panel represented *indica*, *japonica*, *aus* or boro, and aromatic or basmati groups and reproduced major marker-trait associations found in the original population [32]. Analysis of 298 Indian landraces with 32,782 SNPs also revealed three genetic populations and regional differences in diversity, which supported the selection of contrasting parents [38]. Parent choice should consider genetic distance, target alleles, flowering compatibility, adaptation, grain-market requirements, and expected linkage drag. This step converts germplasm evaluation into a defined population for trait mapping or cross development.

3.2. Genetic Locus Discovery, Candidate-Gene Analysis, and Functional Validation

SNP arrays support genetic locus discovery through GWASs, linkage mapping, and combined analyses of diversity panels and experimental populations. Mixed models

can reduce confounding by kinship and population structure, while analyses within rice subpopulations can detect alleles whose frequencies or phenotypic effects differ among *indica*, aus, aromatic, and *japonica*. However, a significant SNP identifies a phenotype-associated haplotype and does not necessarily represent the causal variant. A soybean (*Glycine max*) GWAS similarly reported marker-trait associations and positional candidates, illustrating an evidence level that should remain distinct from functional validation [40]. The biological resolution of an association depends on marker density, local LD, historical recombination, allele frequency, reference-panel representation, population composition, and phenotype quality. Candidate genes can subsequently be prioritized through haplotype analysis, overlap with linkage QTLs, sequence variation, gene annotation, expression profiles, and prior biological information. For example, genome-wide analysis of the rice GRAS transcription factor family combined phylogenetic relationships, promoter *cis*-acting elements, protein features, and expression patterns to prioritize genes potentially involved in abiotic-stress responses [41]. This evidence strengthens candidate selection but does not demonstrate that a particular natural allele causes the associated phenotype. Representative rice studies that progressed from locus discovery to different levels of candidate-gene or functional evidence are summarized in Table 4.

Table 4. Representative rice studies illustrating the progression from genetic locus discovery to candidate-gene prioritization and functional validation using single-nucleotide polymorphism arrays and other dense genotyping resources. CRISPR/Cas9, clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9; GBS, genotyping-by-sequencing; GWAS, genome-wide association study; HDRA, High-Density Rice Array; LD, linkage disequilibrium; QTL, quantitative trait locus; RNC4, Root Na^+ Content 4; SNP, single-nucleotide polymorphism.

Genotyping Resource and Study Design	Population and Phenotype	Locus-Discovery and Candidate-Gene Evidence	Functional-Validation Evidence	Evidence Achieved and Unresolved Limitation	References
Affymetrix 44K array GWAS	44,100 SNPs in 413 accessions evaluated for 34 morphological, developmental, grain-quality, and resistance traits	Associations occurred near *GS3* and *qSW5* for grain dimensions, *SSIIa* and *Wx* for starch properties, *Rc* for pericarp color, *SD1* for plant height, *OsMADS13* for flowering time, and *Pi-ta* for blast resistance. Panicle-length loci differed among rice subpopulations.	No new functional experiments were conducted for the detected loci.	The recovery of characterized genes demonstrated the capacity of the array to detect major common-effect loci. Novel causal genes or variants were not validated.	[8]
HDRA GWAS	Approximately 700,000 SNPs in 1568 diverse accessions; grain-length analyses were conducted across the complete panel and individual subpopulations	Major associations were detected on chromosomes 3 and 5. The chromosome 3 lead SNP corresponded to the characterized functional polymorphism in *GS3*, whereas the chromosome 5 signal tagged the *GW5* deletion. Additional minor and subpopulation-specific associations were identified.	Functional information for *GS3* and *GW5* was obtained from prior studies.	The analysis directly recovered a known functional SNP and indirectly detected a structural allele. Newly detected minor loci remained at the association level.	[42]

Table 4. *Cont.*

Genotyping Resource and Study Design	Population and Phenotype	Locus-Discovery and Candidate-Gene Evidence	Functional-Validation Evidence	Evidence Achieved and Unresolved Limitation	References
HDRA imputed with the 3000 Rice Genomes reference panel	HDRA genotypes were expanded to approximately 4.8 million SNPs; amylose content was examined as a representative phenotype	The strongest association at *Wx* shifted from an intergenic array SNP to a characterized splice-site functional nucleotide polymorphism within *Wx*. The surrounding LD interval remained approximately 207 kb and contained 62 gene models.	The identified nucleotide had previously established effects on *Wx* transcript processing and amylose content.	Imputation recovered a known causal nucleotide but did not eliminate the broad associated interval. The result demonstrated that local LD can limit resolution after marker density has increased.	[31]
KNU Axiom Oryza 580K array GWAS	1288 rice lines evaluated for culm length, heading date, and panicle length	Significant loci were detected for all three traits. Prominent marker signals occurred within or near *OsPRR37*, *DEP1*, and *OsLP1*.	Candidate genes were assigned through physical position and comparison with characterized genes and QTLs. No targeted functional experiments were performed.	The analysis identified associated loci and positional candidates but did not establish causal genes or natural alleles.	[43]
HDRA panicle-trait GWAS combined with linkage mapping	Forty-nine panicle traits were measured in 242 tropical rice accessions, with complementary QTL mapping in a recombinant inbred population	Ten candidate genes occurred within 30 kb of significant GWAS SNPs, and five candidates were also located within QTLs detected by linkage mapping. Some supported regions contained several linked genes with potential roles in panicle development.	Candidate selection incorporated GWAS-QTL overlap, gene annotation, LD, trait correlations, and co-expression information.	Independent mapping increased confidence in the genomic regions, but linked candidates within complex loci could not be distinguished without direct functional tests.	[42]
Affymetrix 44K array GWAS combined with biparental QTL analysis	Aluminum tolerance was evaluated in 383 diverse accessions and complementary mapping populations	Most associated loci were subpopulation-specific. An aus-specific susceptible haplotype at *Nrat1* explained 40% of aluminum-tolerance variation within aus and contained three nonsynonymous substitutions associated with aluminum sensitivity.	Haplotype and sequence evidence was interpreted with previously reported information on *Nrat1* transporter function and mutant phenotypes.	*Nrat1* received strong candidate-gene and haplotype support, but the contributions of the three linked substitutions were not experimentally separated.	[44]

Table 4. *Cont.*

Genotyping Resource and Study Design	Population and Phenotype	Locus-Discovery and Candidate-Gene Evidence	Functional-Validation Evidence	Evidence Achieved and Unresolved Limitation	References
HDRA GWAS and haplotype analysis of the Root Na^+ Content 4 locus	GWAS used 397,812 SNPs and eight salinity-related traits in 365 accessions	The major RNC4 association covered approximately 575 kb. Haplotype analysis identified a 9.7 kb block containing *HKT1;1* and *HKT1;4*. Three nonsynonymous substitutions differentiated the principal *HKT1;1* allele classes.	Independent RNA-interference lines distinguished *HKT1;1* from *HKT1;4*. Electrophysiological analysis in *Xenopus laevis* oocytes demonstrated different Na^+ transport activities, and contrasting natural *HKT1;1* isoforms produced different ion-accumulation phenotypes in a common *japonica* background.	*HKT1;1* was validated as the gene underlying RNC4, and the principal natural allele classes were functionally differentiated. The individual effects of the three linked substitutions remained unresolved.	[45]
Dense-genotype GWAS of rice root architecture	Twenty-one root traits were evaluated in 529 accessions under normal and drought conditions	The analysis identified 264 loci, of which 225 overlapped previously reported root-related QTLs. Association and haplotype analyses prioritized *NAL1* and *OsJAZ1*.	Genetic complementation, overexpression, and RNA interference supported the function of *NAL1*. *OsJAZ1* overexpression increased root length, crown-root number, and root biomass.	Reciprocal genetic evidence validated *NAL1* at the gene level, whereas evidence for *OsJAZ1* was based mainly on overexpression. The associated natural polymorphisms were not directly validated.	[43]
GBS followed by imputation and GWAS	Eight salt-tolerance traits were evaluated in 268 rice accessions using approximately 2.54 million imputed variants	Ten candidate genes were prioritized from 83 associated loci. Haplotype and annotation analyses identified *OsWRKY53* and *OsMKK10.2* as major candidates and connected them with *OsHKT1;5*.	CRISPR/Cas9 knockouts, overexpression lines, ion-flux measurements, double mutants, chromatin immunoprecipitation, yeast one-hybrid assays, electrophoretic mobility shift assays, surface plasmon resonance, and reporter assays established regulatory relationships among *OsWRKY53*, *OsMKK10.2*, and *OsHKT1;5*.	Gene functions and an OsWRKY53-OsMKK10.2/OsHKT1;5 regulatory mechanism were validated. The phenotypic effects of individual natural GWAS variants were not isolated.	[46]

The array-based studies distinguish increased mapping resolution from functional confirmation. Particularly, the 44K array recovered numerous characterized genes and exposed subpopulation-dependent genetic architectures, whereas the HDRA and KNU 580K arrays generated denser regional signals and additional positional candidates [8,42,43]. Imputation provided a particularly informative result at *Wx*, where the strongest association moved to the characterized functional nucleotide polymorphism. Marker analysis at the *Wx* locus has also supported selection for rice amylose content and grain quality [47]. However, the width of the associated region changed little, which showed that additional markers cannot resolve candidate genes when extensive local LD preserves correlations among many variants [31] (Table 4).

Candidate-gene prioritization becomes more persuasive when independent evidence supports the same genomic region. The panicle-architecture study combined association mapping with linkage QTLs, gene annotation, and co-expression information, thereby reducing the number of candidates that depended only on proximity to a lead SNP [42]. The aluminum-tolerance analysis illustrated a complementary strategy in which subpopulation-specific GWAS, haplotype effects, coding-sequence variation, and previous functional information converged on *Nrat1* [44]. Nevertheless, neither approach independently identified a causal nucleotide. Several linked genes remained within some panicle QTLs, while the three nonsynonymous substitutions in the susceptible *Nrat1* haplotype were not separated experimentally (Table 4). Campbell et al. (2017) [45] represented the most complete transition from array-based association to natural-allele validation among the examples considered here. Haplotype analysis reduced the RNC4 locus to two transporter genes, gene-specific silencing distinguished their effects, and transport and transformation assays demonstrated functional differences between the principal *HKT1;1* allele classes [45]. The root-architecture study shows why the type of genetic perturbation also matters. Complementation, overexpression, and RNA interference provided reciprocal evidence for *NAL1*, whereas overexpression of *OsJAZ1* supported gene function without testing whether loss of the gene produced the opposite phenotype [43]. Yu et al. (2023) extended validation from individual genes to a regulatory mechanism through reciprocal mutant phenotypes, downstream-gene analysis, ion measurements, and DNA-binding assays [46]. However, these experiments still did not establish which natural polymorphism generated the original population-level association. Evidence from SNP-based locus discovery should be reported at distinct levels. A significant marker defines an associated locus. A gene supported by position, haplotype, annotation, expression, or QTL overlap remains a candidate gene. Reproducible knockout, silencing, overexpression, complementation, or rescue phenotypes support a functionally validated gene. A causal allele requires functional comparison of alternative natural alleles under comparable genetic backgrounds. A causal nucleotide requires direct editing or genetic separation of that nucleotide from other linked variants. Maintaining these distinctions prevents marker significance, proximity to an annotated gene, differential expression, gene function, and natural-variant causality from being treated as equivalent forms of evidence [43–46].

3.3. Marker-Assisted Transfer of Validated Loci

Marker-assisted transfer uses different marker systems for foreground and background selection. Functional, gene-based, or tightly linked markers are generally more suitable for foreground selection because they directly identify the favorable allele or its associated genomic region. In contrast, SNP chips contribute primarily to genome-wide background selection by estimating recurrent-parent genome recovery, constructing graphical genotypes, detecting residual donor segments, identifying heterozygous regions, and assessing linkage drag. Their principal value is not to replace diagnostic markers for validated loci,

but to determine whether the target allele has been incorporated into an otherwise restored elite genome. This distinction was demonstrated by the 50K genic SNP-chip analysis of Swarna-*Sub1*, CR1009-*Sub1*, IR64-*Sub1*, and Samba Mahsuri-*Sub1*, which had previously been developed using SSR-based background selection and phenotypic evaluation [10]. The chip estimated recurrent-parent genome recovery at 97.7%, 91.9%, 91.7%, and 78.7%, respectively, and detected donor-derived segments across chromosomes other than chromosome 9, which carries *Sub1A*. Recovery of recurrent-parent haplotypes at agronomically important genes was also lower than the corresponding genome-wide estimates in several lines. Thus, low-density markers correctly supported development of submergence-tolerant varieties but did not fully describe the amount and distribution of residual donor DNA. A comparable division of marker functions was used during the incorporation of *Pi2* and *Pi54* for blast resistance and *xa13* and *Xa21* for bacterial blight resistance into Pusa Basmati 1121 and Pusa Basmati 6 [48]. Target genes were selected with locus-specific markers, whereas a 50K SNP array was applied to advanced pyramided lines for genome-wide background analysis. Although SSR- and SNP-based estimates of recurrent-parent recovery were broadly similar, the SNP panel provided denser genome coverage and included more genic positions [3]. The selected lines retained Basmati grain and cooking quality and expressed resistance across pathogen isolates and test locations, but these outcomes depended on the integration of gene-specific selection, SNP-based genome assessment, disease screening, and quality evaluation. In an eating-quality backcrossing program, target-linked and background kompetitive allele-specific PCR (KASP) markers produced recurrent-parent genome recovery estimates of 91.22–98.65% [49]. Trait-focused arrays can connect foreground selection with genome-wide background assessment in a single genotyping assay. WDR6K combines genome-wide markers with 922 loci identified by its developers as contributors to drought resistance and includes numerous missense variants that differentiate upland and lowland rice [22]. This marker composition allows breeding programs to screen drought-related alleles and assess the recipient genome within the same platform. However, functional annotation, ecotype differentiation, pathway enrichment, and stress-responsive expression do not establish a direct causal relationship between every selected locus and drought resistance. Application in foreground selection requires confirmation that the relevant allele is polymorphic between the selected parents, associated with the target phenotype in the chosen genetic background, and predictive under drought conditions representative of the target production environment.

The same complementary strategy has been applied to insect- and blast-resistance breeding [50]. RICE6K analysis of three Huahui938-derived NILs carrying Bph14 and Bph15 estimated recurrent-parent genome recovery between 86.62% and 87.88%, confirmed introgression of the two target regions, and detected no residual heterozygous loci at the surveyed SNP positions [51]. Brown planthopper bioassays nevertheless revealed that resistance depended on the allelic state of the transferred genes: the NILs and hybrids homozygous for *Bph14* and *Bph15* were resistant, whereas hybrids heterozygous for both genes remained susceptible. SNP-chip confirmation of target-region inheritance and genome recovery therefore did not establish the expression of resistance in every hybrid combination. During the transfer of BPH31 into Jaya, foreground selection was conducted with flanking markers PA26 and RM2334, while the Infinium 6K SNP chip was used for background analysis [52]. Only 575 array SNPs were polymorphic between the donor and Jaya, but these markers permitted construction of a graphical genotype for the selected BC_2F_1 plant and confirmed the intended chromosome 3 introgression. The derived lines subsequently expressed resistance to brown planthopper populations from India and the Philippines. A related application transferred *Pi40* into the Turkish cultivars Osmancık-97 and Halilbey [53]. Foreground selection relied on a gene-specific marker, whereas the 6K

SNP chip was used to evaluate 11 advanced backcross lines. Of 4606 high-quality array markers, only 676 and 795 were polymorphic in the two parental combinations. Maximum recurrent-parent recovery reached 90.07% in the Osmancık-97 derivatives but only 70.78% in the Halilbey derivatives, and several regions on chromosomes 1, 3, 4, and 12 remained poorly characterized because the array markers were monomorphic. The effective resolution of an SNP chip is determined by the number and genomic distribution of informative markers between the donor and recipient, irrespective of the array's total probe number.

High-density SNP chips can also refine or revise background-recovery estimates obtained with sparse markers. Three dominant blast-resistance genes, *Pi54*, *Pi1*, and *Pita*, were introduced into the aromatic cultivar Mushk Budji using gene-based or closely linked foreground markers and genome-wide SSR and sequence-tagged site markers [54]. Subsequent analysis with the 50K OsSNPnks array resolved additional recombination events, delineated the donor segments surrounding the target genes, and produced lower recurrent-parent genome recovery estimates than the original SSR analysis. The array restricted the *Pi54* introgression to an approximately 2 Mb region and distinguished recurrent-parent genome recovery from overall genomic similarity between the donor and recipient. This distinction is important because parental genomes may already share alleles at numerous loci; consequently, high genomic similarity does not necessarily indicate that most informative donor alleles have been eliminated. SNP-chip analysis has also supported the characterization of lines carrying multiple quantitative stress-tolerance regions. In Swarna-derived lines containing combinations of *qDTY1.1*, *qDTY2.1*, *qDTY3.1*, and *Sub1*, linked markers were used to select the target regions, whereas the 6K SNP Infinium array estimated recurrent-parent genome recovery of 93–98% in the three principal NILs [55]. Their breeding value was subsequently demonstrated in 60 multi-location trials, national trials across India and Nepal, and release as CR Dhan 801, Bahuguni Dhan-1, and Bahuguni Dhan-2. The array established their genome-wide similarity to Swarna, while multi-environment phenotyping demonstrated improved yield under drought or submergence without a consistent penalty under non-stress conditions. This application provides stronger evidence of breeding utility than studies that ended with marker confirmation or preliminary phenotyping because the transferred QTL combinations were evaluated under the environmental conditions for which the resulting varieties were intended.

More recent breeding programs have extended this application to denser and more representative SNP content. During the incorporation of *xa13-Xa21*, *Xa38*, *Pi9-Pib*, and *Pita* into Pusa Basmati 1509, gene-based or linked markers were used for foreground selection, and SSR markers supported background selection during early backcross generations [56]. Advanced NILs were subsequently analyzed with an 80K Rice Pan Genome Array, which estimated genomic similarity to Pusa Basmati 1509 at 97.38–98.92% across lines carrying different resistance-gene combinations. Multi-location agronomic evaluation, grain and cooking-quality analysis, and pathogen assays identified lines that retained the principal characteristics of the recurrent cultivar while expressing the intended resistance. However, because the SNP array was applied only to advanced generations, it validated the products of the breeding program rather than directly guiding selection during each backcross cycle [49]. Trait-focused arrays can connect foreground selection with genome-wide background assessment in a single genotyping assay. WDR6K combines genome-wide markers with 922 loci identified by its developers as contributors to drought resistance and includes numerous missense variants that differentiate upland and lowland rice [22]. This marker composition allows breeding programs to screen drought-related alleles and assess the recipient genome within the same platform. However, functional annotation, ecotype differentiation, pathway enrichment, and stress-responsive expression do not establish a direct causal relationship between every selected locus and drought resistance.

Consequently, application in foreground selection requires confirmation that the relevant allele is polymorphic between the selected parents, associated with the target phenotype in the chosen genetic background, and predictive under drought conditions representative of the target production environment.

3.4. Genomic Selection for Polygenic Traits

Marker-assisted selection is most effective when phenotypic variation can be traced to a small number of validated loci with stable effects. Grain yield, drought adaptation, nutrient-use efficiency, milling quality, and other polygenic traits cannot generally be improved through the independent transfer of a few favorable alleles. GS addresses this limitation by estimating genomic breeding values from genome-wide markers fitted simultaneously in a training population with phenotypic and genotypic records. The fitted model can then rank genotyped selection candidates without complete phenotypic evaluation. SNP arrays and other standardized SNP panels provide reproducible genotypes across populations and breeding cycles, while imputation can increase effective marker coverage from lower-density assays. However, prediction performance depends on the training-population size, relatedness between training and candidate populations, trait architecture, linkage disequilibrium, phenotype quality, environmental correspondence, and validation design. Representative applications in rice are summarized in Table 5 [57–61]. Cross-crop evidence also indicates that correlated, less costly aerial traits can support genomic prediction of difficult-to-measure root traits [62].

Table 5. Representative applications of genome-wide single-nucleotide polymorphism data to genomic prediction and selection for polygenic traits in rice. GBLUP, genomic best linear unbiased prediction; GBS, genotyping-by-sequencing; GEBV, genomic estimated breeding value; LASSO, least absolute shrinkage and selection operator; LD, linkage disequilibrium; NCII, North Carolina Design II; RKHS, reproducing kernel Hilbert space; RR-BLUP, ridge-regression best linear unbiased prediction; SNP, single-nucleotide polymorphism.

Genotyping Resource and Study Design	Population and Target Traits	Prediction Strategy	Principal Result	Evidence Achieved and Remaining Limitation	Reference
GBS; 73,147 SNPs and historical trials	363 elite tropical lines; yield, flowering, and height across wet and dry seasons	Six genomic models versus pedigree prediction; training season, marker number, and population varied	Maximum accuracy: yield 0.31, height 0.34, flowering 0.63; plateau at ~1 marker per 0.2 cM; RR-BLUP performed best for yield	Trait architecture affected performance; cross-validation only, with no realized selection response	[57]
Genome-wide SNPs in recurrent-selection populations	343 S_2:$_4$ lines from four upland-rice populations; four agronomic traits	Repeated cross-validation across marker filters, training size and composition, kinship, and models	Accuracy 0.12–0.54 (mean 0.30); trait and relatedness effects exceeded marker-filtering effects	Supported prediction within related populations; no completed cycle or realized gain	[58]
GBS of a diversity panel and progenies	284 *japonica* accessions and 97 F_5–F_7 lines from 36 crosses; three traits	Panel-trained models predicted progenies using subsets differing in size and relatedness	Progeny accuracy 0.41–0.54; related small training sets were effective; little gain beyond ~1 SNP per 22 kb	True progeny prediction, but parents were in the panel; unrelated-cross transfer was untested	[59]

Table 5. *Cont.*

Genotyping Resource and Study Design	Population and Target Traits	Prediction Strategy	Principal Result	Evidence Achieved and Remaining Limitation	Reference
Parental resequencing and NCII hybrids	575 hybrids from 115 varieties and five male-sterile lines; eight traits	Six methods, marker subsets, training sizes, and additive or dominance models; 362,760 hybrids ranked	GBLUP and LASSO performed best; accuracy plateaued at ~5000 SNPs, whereas reducing training size caused a larger loss	Could reduce field testing, but top predicted hybrids were not produced or evaluated	[60]
Genome-wide SNPs under two water regimes	284 accessions and 97 progenies; flowering, panicle weight, and nitrogen balance	Single- and multi-environment models, relative-performance index, and joint-regression slope	Mean predictive ability was 0.31 for the index and 0.64 for slope; multi-environment models improved ~30% with partial data	Supported cross-regime prediction; no selection followed by progeny field comparison	[63]
~215,000 SNPs in a rainfed-lowland panel	280 accessions in favorable and two drought environments; yield, flowering, and height	GBLUP and RKHS with genome-wide, LD-pruned, or trait-specific SNPs and multiple validation scenarios	Trait-specific markers improved prediction up to 22%; multi-environment models improved up to 32% only with correlated-environment records	Benefits were scenario-dependent; no independent candidate population	[61]
GBS; 9928 SNPs in early recurrent selection	334 S_0 genotypes with $S_{0:2}/S_{0:3}$ progenies at irrigated and rainfed sites; four traits	Single- and two-site GBLUP and RKHS with alternative phenotyping allocations	Two-site prediction increased Zn accuracy from 0.31 to 0.51 and yield from 0.35 to 0.42	Relevant to early selection; no realized response after recombination	[64]
International elite panel; 882 SNPs	111 elite *indica* lines; flowering, height, and yield across 15 environments	Seven GBLUP models with genomic, environmental, covariate, and genotype-by-environment effects	Accuracy ranged 0.06–0.79 for flowering, 0.25–0.88 for height, and −0.29–0.62 for yield; interaction terms were inconsistent	Low density worked within related material; limited diversity and weak cross-environment yield correlation restricted transferability	[65]
1K-RiCA; 814–889 SNPs retained	Thousands of irrigated-rice lines in Bangladesh multi-location trials, 2018–2022	Genomic BLUP integrated with sparse testing, rapid generation advance, and marker-assisted selection	Annual accuracy 0.454–0.701; yield gain ~117 kg ha^{-1} $year^{-1}$; breeding cycle shortened by ~1.5 years	Operational evidence; concurrent breeding changes obscured the independent genomic-selection effect, and no randomized control was used	[66]

The early studies established that training-population design is generally more influential than very high marker density once the genome has been covered at a resolution compatible with population LD. In the elite IRRI population, prediction accuracy approached a plateau after the marker interval reached approximately 0.2 cM [57]. A similar plateau was observed in the japonica reference panel, while prediction of biparental progenies depended strongly on including accessions closely related to their parents [58]. Hybrid prediction produced the same general pattern: increasing the training population had a much greater effect than increasing the number of markers beyond approximately

5000 SNPs [59]. These results indicate that replacing a sufficiently informative panel with a much denser array may add little predictive value if the number of training individuals, phenotype quality, or genetic relevance of the training population remains unchanged.

Trait architecture determines both the attainable accuracy and the most suitable prediction model. Grain yield generally produced lower accuracy than flowering time or plant height owing to its lower heritability, dispersed small-effect genetic control, and stronger environmental sensitivity [57,58]. Genome-wide shrinkage models such as RR-BLUP and GBLUP were effective for yield when no dominant QTL was detected, whereas marker selection or nonlinear models sometimes performed better for traits influenced by one or several larger-effect loci [57,61]. This does not imply that a single complex model should replace simpler approaches for all traits. A model that performs well for flowering time can rank yield lines poorly, and a marker subset selected for one environment may not capture the effects expressed under another environment.

Environmental correspondence is especially important for drought adaptation, water-management response, and grain yield. Multi-environment models improved prediction under alternate wetting and drying and managed drought when phenotypic information from correlated environments was available for the training population or part of the candidate population [61,63]. Early recurrent-selection data similarly showed that an irrigated secondary site improved prediction at a rainfed target site for traits with positive inter-site correlations but provided a smaller gain for yield [64]. Nguyen et al. (2023) presented an important counterexample, as models incorporating genotype-by-environment interactions did not consistently improve predictions, and yield prediction remained weak in several environments despite the availability of multi-location data [65]. Multi-environment models therefore cannot compensate for weak genetic correlations, inadequate environmental representation, limited training size, or poor yield precision merely by adding additional interaction parameters.

Most rice studies have evaluated genomic prediction through random or structured cross-validation, which measures the capacity of a fitted model to reproduce withheld observations from an existing dataset. Prediction of progenies absent from the training set provides stronger evidence, while prediction across later breeding cycles or new environments more closely represents the decisions faced by breeders [63]. The use of genomic values to select parents, followed by recombination, progeny evaluation, and comparison with a conventional-selection scheme, is required to demonstrate a realized genomic-selection response. The recurrent-selection study tested early-generation prediction but did not report the response after a completed genomic-selection cycle [64]. The Bangladesh program provided the clearest operational evidence through repeated annual predictions, sparse testing, early parent recycling, and an observed increase in yield gain [66]. Nevertheless, the simultaneous introduction of several breeding changes prevents the separate contribution of genomic selection from being quantified.

3.5. Variety Identification and Seed Quality Control

SNP-based genotyping provides a stable molecular framework for rice variety identification, pedigree confirmation, and seed-lot quality assessment, particularly when morphological descriptors are influenced by environment or when closely related cultivars share similar phenotypes (Table 6). High-density platforms can establish reference fingerprints and detect residual heterogeneity, while reduced SNP panels are generally more economical for routine seed testing. Firstly, Yu et al. (2014) used RICE6K to characterize rice accessions and individual plants from an elite cultivar and identified duplicate entries as well as several genetically distinct plant types within the same seed source [11]. Large segregating regions on chromosomes 2 and 9 were detected by the SNP array but were not identified

with the standard 24-SSR panel, which showed that low-density conventional markers may miss genomic heterogeneity outside their target loci. RiceSNP50 extended this capacity through 51,478 SNPs and a rice-specific genotype cluster file developed from parents and F_1 progenies, which improved allele assignment and supported accurate confirmation of varietal identity [9]. These platforms are useful when a reference fingerprint must be established for a breeder line, parental stock, or registered cultivar, although routine seed testing with tens of thousands of loci is rarely economical. Their greater value lies in the discovery of informative diagnostic markers that can later be converted into smaller assays.

Table 6. Representative applications of single-nucleotide polymorphism genotyping to rice varietal authentication, genetic-purity assessment, pedigree verification, and seed quality control. C6AIR, Cornell_6K_Array_Infinium_Rice; KASP, kompetitive allele-specific polymerase chain reaction; Mb, megabase; RICE6K, whole-genome 6K rice SNP array; SNP, single-nucleotide polymorphism; SSR, simple sequence repeat.

#	Materials Evaluated	SNP Platform or Panel	Quality-Control Purpose	Main Outcome	References
1	112 inbred accessions and individual plants from the elite variety YU	RICE6K; 5102 converted markers and 4473 high-quality SNPs	Duplicate detection, varietal fingerprinting, and within-variety purity analysis	Six duplicate accessions were detected. Four heterogeneous plant types were identified within YU, with major segregating regions at 9.0–11.5 Mb on chromosome 2 and 0–7.0 Mb on chromosome 9 that were not detected by the standard 24-SSR panel.	[11]
2	Elite rice varieties and samples with known parent-F_1 relationships	RiceSNP50; 51,478 SNPs	Variety verification and accurate genotype assignment	A rice-specific cluster file supported reliable allele calling, fine differentiation among elite varieties, and confirmation of varietal identity.	[9]
4	More than 40,000 germplasm accessions and breeding samples	C6AIR; 4606 standard high-quality SNPs	Fingerprinting, pedigree verification, sample identity, and introgression control	The platform provided standardized fingerprints across cultivated rice, wild accessions, and breeding populations and supported routine detection of pedigree or sample inconsistencies.	[13]
5	1171 farmer seed samples and 24 authenticated breeder references from Bangladesh	C6AIR; 4503 polymorphic SNPs	Authentication of farmer-named varieties and verification of varietal adoption	The analysis classified 247 samples as released modern varieties, 216 as Swarna-related materials, and 708 as traditional varieties. More than half of the genetically identified modern varieties were recorded under an incorrect name or described as traditional rice.	[35]
7	530 diverse rice accessions	Core KASP array; 467 polymorphic markers retained from 565 successful assays	Germplasm identity, population assignment, and development of reduced SNP panels	The markers were distributed at an average density of 1.24 SNPs Mb^{-1} and supplied a validated source from which smaller panels could be selected for routine identification.	[67]
8	41 approved Basmati and 19 non-approved rice varieties	364 KASP designs; 327 successful and 255 polymorphic assays	Authentication of premium rice and discrimination of potential adulterants	Thirty-seven approved Basmati varieties were confirmed with three to eight markers selected from a pool of 98 KASP assays, while 24 markers separated the materials into major variety families.	[68]

Table 6. *Cont.*

#	Materials Evaluated	SNP Platform or Panel	Quality-Control Purpose	Main Outcome	References
9	Whole-genome SNP profiles of 3024 rice varieties and two independent validation datasets	Conditional Random Selection panel; 390 core SNPs	Variety identification and identification of essentially derived varieties	The reduced panel distinguished closely related varieties and provided an automated procedure for separating essentially derived varieties from non-derived materials.	[69]
10	518 conventional and hybrid rice varieties	Reduced panel of 48 KASP markers	Identification of conventional varieties, hybrids, and breeding lines	The panel achieved 100% discrimination among 53 conventional *indica* varieties and 193 hybrids and distinguished 89.1% of the conventional japonica varieties.	[70]

The practical importance of molecular identification is well illustrated by farm-level and breeding-pipeline studies. For example, Kretzschmar et al. (2018) genotyped 1171 farmer seed samples from Bangladesh with 4503 polymorphic C6AIR markers and compared them with 24 authenticated breeder references [35]. Only 247 samples matched released modern varieties, 216 were related to Swarna, and 708 were assigned to traditional groups, while more than half of the genetically identified modern varieties were recorded under an incorrect name or described as traditional rice. This disagreement shows that variety names alone cannot provide dependable evidence of genetic identity, particularly when seed is exchanged informally, cultivars are renamed locally, or closely related selections circulate under the same name. C6AIR has also been used across more than 40,000 research and breeding samples for fingerprinting, pedigree verification, genetic-stock confirmation, and detection of sample inconsistencies [13]. In breeding programs, such markers can confirm whether a cross was successful, identify accidental self-pollination, detect seed mixtures, and trace a line through successive generations. However, molecular similarity should be interpreted against authenticated reference samples, pedigree records, and seed-production history. An incomplete reference database may assign an unknown sample to its nearest available relative without establishing its exact cultivar identity.

Reduced SNP and KASP panels offer a more economical route for routine variety authentication and seed quality control. More specifically, Yang et al. (2019) developed a core KASP resource with 467 polymorphic markers from 530 diverse rice accessions, which provided a basis for selecting smaller marker subsets for germplasm identification and population assignment [67]. Then, Steele et al. (2021) showed that 37 approved Basmati varieties could be identified with combinations of three to eight KASP markers selected from a larger pool, while 24 markers separated the materials into major variety families [68]. Tang et al. (2022) developed a 48-marker KASP panel that distinguished all tested conventional *indica* varieties and hybrids and separated 89.1% of conventional japonica varieties [70]. The lower discrimination among *japonica* materials indicates that very small panels can lose power when cultivars share narrow pedigrees or differ by only a few genomic regions. Yuan et al. (2022) addressed this problem through a 390-SNP panel selected from 3024 rice varieties, which improved discrimination among closely related cultivars and supported assessment of essentially derived varieties [69]. Thus, a tiered identification system is preferable to a single fixed panel. A small set of highly informative markers can be used for rapid screening, parental confirmation, and obvious contamination, while a larger standardized panel should be applied when sibling lines, essentially derived varieties, legal disputes, or premium-market authentication are involved. Seed quality control should also include representative sampling from the seed lot, replicated genotyping, predefined

similarity thresholds, and comparison with authenticated breeder seed, since a technically accurate assay cannot compensate for poor sampling or an unreliable reference database.

3.6. Pre-Breeding and Utilization of Landraces, Indigenous Germplasm, and Wild Relatives

SNP-based platforms have strengthened rice pre-breeding through genome-wide characterization of chromosome segment substitution lines, introgression libraries, landraces, and wild *Oryza* accessions that carry alleles absent from elite breeding pools. SNP-chip analysis of chromosome segment substitution lines derived from *O. rufipogon* enabled genome-wide localization of wild donor segments and supported the assignment of phenotypic effects to defined genomic regions [13,71]. These marker numbers are sufficient for genome-wide tracking of donor regions and recovery of recurrent-parent backgrounds, although narrow introgressions can remain unresolved when recombination occurs between widely spaced markers. A coordinated collection of six wild-introgression libraries later introgressed genomic segments from three geographically diverse *O. rufipogon* donors into the elite *indica* cultivar IR64 and the tropical japonica cultivar Cybonnet, with 434 lines carrying average donor contributions of 2.9–6.9% per line [19]. The use of the same donor genomes in two cultivated backgrounds is particularly valuable, since it permits assessment of whether a wild allele retains its effect across *indica* and *japonica* genetic backgrounds or depends on background-specific interactions (Table 7).

Table 7. Representative applications of single-nucleotide polymorphism platforms to rice pre-breeding, donor selection, introgression mapping, and management of linkage drag. C6AIR, Cornell_6K_Array_Infinium_Rice; C7AIR, Cornell_7K_Array_Infinium_Rice; cM, centimorgan; CSSL, chromosome segment substitution line; GWAS, genome-wide association study; QTL, quantitative trait locus; 1k-RiCA, 1K-Rice Custom Amplicon panel; SNP, single-nucleotide polymorphism.

#	Donor Germplasm and Population	SNP Platform	Pre-Breeding Purpose	Main Outcome	References
1	*O. rufipogon* introgressions in *indica* cultivar 9311; 198 CSSLs	Whole-genome SNP chip used to refine selected CSSLs	Construction of wild-rice introgression resources and QTL detection	The population covered 84.9% of the donor genome; mean segment ~6 cM; seven distinctive CSSLs were densely genotyped	[71]
2	Interspecific populations between *O. sativa* and *O. rufipogon* or *O. meridionalis*	C6AIR; 4940 SNPs used for interspecific populations	Introgression tracking, negative selection, and CSSL development	Interspecific and *indica* × *japonica* crosses yielded 1900–2500 polymorphic markers for donor tracking and background recovery	[13]
4	Reciprocal CSSLs between *indica* cultivar Takanari and japonica cultivar Koshihikari	Genome-wide SNP-defined CSSLs	Identification of hidden favorable alleles across divergent rice groups	A Koshihikari chromosome 5 segment increased culm bending stress in Takanari and overlapped cell-wall-density QTLs	[72]
5	272 wild accessions of *O. nivara*, *O. rufipogon*, and *O. sativa* f. *spontanea*	OsSNPnks 50K Affymetrix array	Identification of donors for rice root-knot nematode resistance	Phenotyping, GWAS, and expression analysis identified resistant wild donors and transferable regions	[34]
7	Three geographically diverse *O. rufipogon* complex donors in IR64 and Cybonnet backgrounds	C7AIR; 1496–2962 informative SNPs per library	Development of coordinated wild-introgression libraries	Six libraries contained 434 lines (63–81 each) and 2.9–6.9% mean donor genome, enabling comparison across *indica* and *japonica* backgrounds	[19]

Table 7. *Cont.*

#	Donor Germplasm and Population	SNP Platform	Pre-Breeding Purpose	Main Outcome	References
8	Nipponbare chromosome segments in *indica* cultivar Zhenshan 97	Rice8K; 3610 polymorphic SNPs	Discovery and validation of favorable alleles from divergent cultivated germplasm	The 221-line set covered 96% of Nipponbare; *qGL11* was narrowed to 25 kb, *OsGH3.13* was prioritized, and yield increased ~20% in a NIL	[73]
9	Dongxiang wild rice introgressions in *indica* restorer line R974	Whole-genome SNP chip applied to five salt-tolerant CSSLs	Utilization of wild alleles for salt tolerance	The population covered 93.37% of the donor genome (6.27 homozygous segments per line); five salt-tolerant CSSLs were confirmed	[74]
10	154 Yunnan local varieties and three wild accessions	High-density whole-genome SNP array	Donor selection for resistance and grain-quality improvement	Three groups were identified; resistance alleles were scarce, whereas favorable *Wx* and *ALK* alleles were frequent	[37]
11	Eleven donors crossed into MTU 1010 and IR91648-B-89-B; 48 introgression lines	1k-RiCA with genome-wide and trait-linked SNPs	Development of pre-breeding lines for dry direct-seeded rice	Nineteen target genes/QTLs were combined; 2–49 additional loci per line revealed useful or undesirable nontarget introgressions	[75]

Diverse germplasm has supplied favorable alleles for yield, lodging resistance, salinity tolerance, disease resistance, and adaptation to changing production systems. Reciprocal chromosome segment substitution lines between Takanari and Koshihikari identified a japonica-derived region on chromosome 5 that increased culm bending strength in the *indica* background, with overlapping QTLs for cellulose, hemicellulose, and holocellulose density [72]. This finding illustrates that useful alleles may remain hidden in cultivars with moderate field performance and become valuable only after transfer into a complementary genetic background. Recently, Wang et al. (2021) used Rice8K markers to characterize 221 substitution lines derived from Nipponbare and Zhenshan 97, which covered 96% of the Nipponbare genome and supported localization of *qGL11* to a 25 kb interval [73]. *OsGH3.13* was proposed as the candidate gene, and the favorable allele increased grain yield by approximately 20% in a near-isogenic background. Wild rice introgression panels have supplied candidate alleles for salinity and nematode resistance, although their effects require confirmation in elite genetic backgrounds and target environments [34,74]. These examples confirm the breeding value of diverse germplasm, although favorable phenotypes identified in substitution lines should be validated across environments and genetic backgrounds before a donor allele is treated as broadly useful.

SNP-chip data can also guide parental selection and the assembly of multi-donor pre-breeding populations through the joint assessment of ancestry, functional alleles, donor-fragment size, and undesirable linkage drag. More specifically, Huang et al. (2022) genotyped 154 Yunnan rice varieties and found that favorable *Wx* and *ALK* alleles were common, while useful blast-resistance alleles and the favorable *BADH2* allele occurred at lower frequencies, which allowed breeders to distinguish accessions suited for grain-quality improvement from those needed to broaden resistance [37]. Paul et al. (2025) combined 19 target genes or QTLs for crop establishment, root traits, lodging resistance, yield, and pest resistance in 48 pre-breeding lines for dry direct-seeded rice, and 1k-RiCA genotyping detected an additional 2–49 genes or QTLs in individual lines [75]. These unexpected introgressions may contain favorable alleles, but they can also introduce poor grain quality, delayed flowering, shattering, sterility, or weak adaptation, which shows why target-gene confirmation is not sufficient for pre-breeding decisions. In our assessment,

the most informative use of SNP chips in pre-breeding is the reconstruction of complete donor-segment maps, followed by phenotypic testing of lines that carry small and clearly defined introgressions. Large donor segments should not be removed automatically, since they may contain interacting genes needed for expression of a complex trait, but their agronomic effects should be examined through near-isogenic lines, recombinant progeny, and multi-environment trials. Pre-breeding programs should therefore combine genome-wide marker data with field performance, grain quality, reproductive compatibility, and functional validation so that diverse alleles can be transferred without weakening the adaptation and commercial value of elite rice cultivars.

4. Future Directions

Future rice SNP chips should move beyond marker sets derived from a single reference genome and adopt pangenome-guided designs that represent core genes, dispensable genes, PAV, SV, transposable-element insertions, and haplotypes from cultivated and wild rice [76]. The Rice Pangenome Genotyping Array demonstrated the value of this approach through approximately 80,000 SNP and presence-absence markers distributed across the 3K rice pangenome. Its application detected 42 loci for grain size and weight, including eight dispensable-genome loci that were not recovered through conventional single-reference GWAS [16]. The need for broader marker content has become more evident from the graph-based super-pangenome of 251 cultivated and wild accessions and the wild-cultivated pangenome constructed from 145 chromosome-level assemblies [26,76]. The latter resource identified 3.87 Gb of non-reference sequence, 13,728 wild-rice-specific genes, 855,122 differentiated SNPs, and 13,853 PAVs between *indica* and *japonica*. Future arrays should include markers that tag wild resistance-gene clusters, non-reference genes, large inversions, copy-number changes, and lineage-specific haplotypes. However, the simple addition of pangenome markers will not remove ascertainment bias if probe selection continues to favor common alleles with convenient flanking sequences from elite *O. sativa*. Marker discovery panels should contain balanced numbers of *indica*, *japonica*, *aus*, aromatic, African rice, weedy rice, and wild *Oryza* accessions, while performance should be reported separately for each genetic group. A chip described as universal should not be accepted on the basis of total marker number when large fractions of its loci remain monomorphic or technically unreliable outside the dominant discovery population.

Future genotyping systems should also adopt modular and updatable formats instead of depending on static commercial arrays whose marker content cannot be altered after manufacture. A practical architecture could contain a common set of genome-wide anchor markers, population-specific modules for *indica*, *japonica*, *aus*, aromatic, and African rice, and optional trait modules for grain quality, flowering time, disease resistance, salinity, drought, heat, nutrient-use efficiency, and low accumulation of toxic elements. Target-capture sequencing provides one route toward this architecture, since probes can be added or replaced without redesigning an entire solid array. WDR6K provides a recent rice example of this approach through a liquid-phase capture design that combines genome-wide markers with drought-associated functional loci and permits the incorporation of additional markers in later panel versions [22]. Recently, Lee et al. (2022) developed a capture panel with 2565 SNPs, including 1225 markers selected for temperate japonica and 1339 markers informative for *indica*, of which 2341 produced polymorphic genotypes across 50 rice varieties [77]. A tiered system could use a master panel of 50,000–100,000 markers for discovery and reference-population construction, followed by regional panels of 1000–5000 markers for routine breeding, seed purity testing, and genomic prediction. Common anchor loci would permit genotype integration across panel versions, while imputation could restore denser haplotypes when the reference population represents the

target breeding material. The pursuit of a single permanent chip for all rice programs is unrealistic, since breeding populations change as new parents, wild introgressions, and functional alleles enter the pipeline. Marker panels should be reviewed at fixed intervals and revised when loci become monomorphic, new causal variants are validated, or reference assemblies change. This approach would extend platform life and reduce the current gap between rapidly expanding rice genomic resources and fixed arrays designed from older variant catalogs. Future SNP-chip applications in marker-assisted backcross breeding should progress from post-development characterization of advanced lines to active selection during early backcross generations. Several rice studies used high-density arrays only after foreground selection, backcrossing, and phenotypic recovery had been completed, so the arrays verified recurrent-parent recovery without demonstrating that they accelerated breeding decisions [10,48,51]. An effective system should combine functional or gene-based assays for target alleles, dense flanking markers for defining recombination boundaries, and genome-wide SNPs for removing residual donor segments. Donor and recipient parents should be genotyped before crossing to determine the number and distribution of informative markers and identify genomic regions that cannot be monitored by the selected panel. This step is especially important for closely related parents, for which a nominally dense chip may contain few polymorphic loci. When the target allele is a deletion, insertion, copy-number variant, or non-reference sequence, a complementary locus-specific assay will remain necessary. Studies should report the size and boundary confidence of each target introgression, the number and locations of residual donor segments, genome-wide heterozygosity, recurrent-parent recovery, and the progeny selected or rejected on the basis of chip data. Breeding value should then be assessed through comparison with the recipient under target-stress and non-stress conditions. Evidence that SNP-chip selection reduced the number of backcross generations, restricted linkage drag, or preserved grain quality and adaptation would provide a more meaningful demonstration of utility than post hoc genotyping of the final lines.

A second priority is the integration of SNP-chip genotypes with environmental phenotyping, transcriptomics, metabolomics, epigenomics, and functional genomics. Marker-trait associations should be evaluated through repeated field experiments across water, temperature, soil, disease, and management regimes, followed by expression analysis within associated intervals and experimental testing of candidate alleles through near-isogenic lines, gene editing, transgenic complementation, or allele replacement [2,78]. Future analytical pipelines should retain structural and presence-absence information instead of treating every low-call-rate locus as a technical failure. Recently, Krusenbaum and Wissuwa (2025) showed that low-call-rate SNPs from the dispensable rice genome often represented underlying PAV and that incorporation of this information increased genomic prediction accuracy for several traits and subpopulations [79]. Moreover, Vourlaki et al. (2024) also found that combined SNP and structural-variant data improved prediction in 87% of the tested trait and validation combinations, while deep-learning models exceeded Bayesian models in 75% of their evaluated cases [80]. These results support the inclusion of structural-variant tags and alternative genotype encodings in future chips, but they do not justify indiscriminate use of artificial intelligence. Complex models can overfit small rice panels, obscure the contribution of individual loci, and produce optimistic results when related lines occur in both training and validation sets. Prediction models should therefore be compared with linear benchmarks and evaluated through family-wise, breeding-cycle, geographic, and environmental validation. The strongest evidence will come from predictions tested in later generations and independent breeding programs, not from random cross-validation within the same diversity panel.

Future systems for varietal identification and seed quality control should connect genome-wide SNP platforms with authenticated reference collections and statistically validated sampling procedures. High-density arrays can establish reference fingerprints and identify informative loci, while compact panels can support routine identity, purity, and hybridity testing at lower cost [11,13]. Reference databases should contain breeder-seed samples, pedigree and release information, technical replicates, and versioned genotype profiles that remain traceable across panel updates. Similarity thresholds must be calibrated within the target breeding pool because sister lines, backcross derivatives, and renamed materials may share most assayed alleles. A universal percentage threshold cannot reliably distinguish every cultivar. Seed-lot studies should report the number of individual seeds or seedlings analyzed, the expected confidence interval for the impurity estimate, the minimum detectable contamination level, marker-call failure, heterozygosity thresholds, and concordance among technical replicates. Validation with blinded mixtures containing known proportions of off-type seed would establish whether a panel can detect contamination at the level required for certification. Hybrid-seed testing should compare observed F_1 genotypes with parental predictions at informative loci and include selfed maternal progeny and unintended pollen donors as controls. Molecular fingerprints should remain linked to phenotypic descriptors and official varietal records because identical SNP profiles do not establish complete varietal equivalence, while isolated genotype mismatches do not confirm biological impurity until technical error, residual segregation, and sample-handling errors have been excluded.

Future work should finally place greater emphasis on platform standardization, accessibility, and measurable breeding outcomes. Each chip release should provide a versioned marker manifest with probe sequences, reference coordinates for major rice assemblies, allele orientation, genotype-cluster files, flanking polymorphisms, assay-conversion rates, subgroup-specific minor allele frequencies, and known links to genes or QTLs. A shared benchmark panel containing authenticated *indica*, *japonica*, *aus*, aromatic, African, weedy, and wild accessions would permit laboratories to compare call rate, concordance, heterozygote detection, and cross-platform compatibility [7]. Public breeding programs would benefit from regional genotyping centers, common quality-control protocols, open imputation reference panels, and small trait-focused assays that can be performed without reliance on a single commercial provider. More importantly, platform evaluation should extend beyond technical measures such as call rate and marker density. Future studies should report genotyping cost per selected line, turnaround time, changes in selection accuracy, reduction in backcross generations, prediction across breeding cycles, and realized genetic gain in field trials. Many published studies demonstrate population separation or high cross-validation accuracy but do not demonstrate that SNP-chip decisions led to superior released varieties. In our assessment, the next generation of rice SNP chips should be judged by whether they help breeders identify useful parents, remove unfavorable donor segments, shorten selection cycles, protect varietal identity, and deliver stable gains under target production environments. Without these breeding outcomes, increasingly dense platforms may expand genomic datasets without providing a comparable improvement in cultivar development.

5. Conclusions

Rice SNP chips have progressed from moderate-density arrays derived from limited reference panels to high-density and pangenome-informed platforms that support germplasm characterization, locus discovery, candidate-gene prioritization, marker-assisted transfer, genomic prediction, varietal identification, and seed quality control. Their scientific and breeding value depends on marker informativeness, population representation,

phenotype quality, environmental correspondence, and validation design, not solely on marker number. Association signals require independent evidence to establish candidate genes, gene functions, causal alleles, and causal nucleotides. During marker-assisted transfer, functional or tightly linked markers provide accurate foreground selection, while genome-wide chips estimate recurrent-parent recovery, define donor segments, and detect linkage drag. Genomic prediction depends strongly on the size and genetic relevance of the training population and must be evaluated across families, breeding cycles, locations, and years before improvements in realized genetic gain can be attributed to genomic selection. Varietal identification and seed-purity testing require authenticated reference fingerprints, informative marker subsets, calibrated similarity thresholds, technical repeatability, and representative seed-lot sampling. Future platforms should incorporate non-reference genes, PAV, SVs, wild-rice haplotypes, and population-specific polymorphisms within modular and updatable designs that retain common anchor markers across versions. Their performance should be evaluated through measurable effects on breeding decisions, selection efficiency, number of backcross generations, genetic gain, varietal integrity, and cultivar performance under target environments. Rice SNP chips will remain valuable when improvements in genomic resolution are accompanied by rigorous phenotyping, functional validation, reliable quality control, accessible genotyping services, and demonstrated contributions to cultivar development.

Author Contributions: Conceptualization, H.D.C., T.Q.N. and L.H.T.; methodology, A.Q.H., H.D.C., T.Q.N. and L.H.T.; literature search and data curation, A.Q.H., T.Q.N., G.H.D., H.V.L. and N.T.P.T.; formal analysis, A.Q.H., H.D.C., T.Q.N. and L.H.T.; writing—original draft preparation, T.Q.N. and G.H.D.; writing—review and editing, A.Q.H., H.D.C., T.Q.N., G.H.D., H.V.L., N.T.P.T. and L.H.T.; visualization, A.Q.H., T.Q.N. and G.H.D.; supervision, H.D.C. and L.H.T.; project administration, H.D.C. and L.H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new datasets were generated or analyzed in this review. All data discussed in this article are available from the cited references.

Acknowledgments: During the preparation of this work, the authors used Grammarly and Quillbot to edit grammar. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

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

References

1. Xie, W.; Wang, G.; Yuan, M.; Yao, W.; Lyu, K.; Zhao, H.; Yang, M.; Li, P.; Zhang, X.; Yuan, J.; et al. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5411–E5419. [CrossRef]
2. Roy, S.; Soni, P. Unraveling the epigenetic landscape for salt tolerance in plants. *Int. J. Plant Biol.* **2022**, *13*, 443–462. [CrossRef]
3. Bidyananda, N.; Jamir, I.; Nowakowska, K.; Varte, V.; Vendrame, W.A.; Devi, R.S.; Nongdam, P. Plant genetic diversity studies: Insights from DNA marker analyses. *Int. J. Plant Biol.* **2024**, *15*, 607–640. [CrossRef]
4. Guo, L.; Gao, Z.; Qian, Q. Application of resequencing to rice genomics, functional genomics and evolutionary analysis. *Rice* **2014**, *7*, 4. [CrossRef]
5. Gu, H.; Liang, S.; Zhao, J. Novel sequencing and genomic technologies revolutionized rice genomic study and breeding. *Agronomy* **2022**, *12*, 218. [CrossRef]
6. Tung, C.-W.; Zhao, K.; Wright, M.H.; Ali, M.L.; Jung, J.; Kimball, J.; Tyagi, W.; Thomson, M.J.; McNally, K.; Leung, H.; et al. Development of a research platform for dissecting phenotype-genotype associations in rice (*Oryza* spp.). *Rice* **2010**, *3*, 205–217. [CrossRef]
7. Yang, X.; Yu, S.; Yan, S.; Wang, H.; Fang, W.; Chen, Y.; Ma, X.; Han, L. Progress in rice breeding based on genomic research. *Genes* **2024**, *15*, 564. [CrossRef]

8. Zhao, K.; Tung, C.-W.; Eizenga, G.C.; Wright, M.H.; Ali, M.L.; Price, A.H.; Norton, G.J.; Islam, M.R.; Reynolds, A.; Mezey, J.; et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2011**, *2*, 467. [CrossRef]
9. Chen, H.; Xie, W.; He, H.; Yu, H.; Chen, W.; Li, J.; Yu, R.; Yao, Y.; Zhang, W.; He, Y.; et al. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol. Plant* **2014**, *7*, 541–553. [CrossRef]
10. Singh, N.; Jayaswal, P.K.; Panda, K.; Mandal, P.; Kumar, V.; Singh, B.; Mishra, S.; Singh, Y.; Singh, R.; Rai, V.; et al. Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Sci. Rep.* **2015**, *5*, 11600. [CrossRef]
11. Yu, H.; Xie, W.; Li, J.; Zhou, F.; Zhang, Q. A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J.* **2014**, *12*, 28–37. [CrossRef]
12. McCouch, S.R.; Wright, M.H.; Tung, C.W.; Maron, L.G.; McNally, K.L.; Fitzgerald, M.; Singh, N.; DeClerck, G.; Agosto-Perez, F.; Korniliev, P.; et al. Open access resources for genome-wide association mapping in rice. *Nat. Commun.* **2016**, *7*, 10532. [CrossRef]
13. Thomson, M.J.; Singh, N.; Dwiyanti, M.S.; Wang, D.R.; Wright, M.H.; Perez, F.A.; DeClerck, G.; Chin, J.H.; Malitic-Layaoen, G.A.; Juanillas, V.M.; et al. Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. *Rice* **2017**, *10*, 40. [CrossRef]
14. Morales, K.Y.; Singh, N.; Perez, F.A.; Ignacio, J.C.; Thapa, R.; Arbelaez, J.D.; Tabien, R.E.; Famoso, A.; Wang, D.R.; Septiningsih, E.M.; et al. An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS ONE* **2020**, *15*, e0232479. [CrossRef]
15. Kim, K.-W.; Nawade, B.; Nam, J.; Chu, S.-H.; Ha, J.; Park, Y.-J. Development of an inclusive 580K SNP array and its application for genomic selection and genome-wide association studies in rice. *Front. Plant Sci.* **2022**, *13*, 1036177. [CrossRef]
16. Daware, A.; Malik, A.; Srivastava, R.; Das, D.; Ellur, R.K.; Singh, A.K.; Tyagi, A.K.; Parida, S.K. Rice Pangenome Genotyping Array: An efficient genotyping solution for pangenome-based accelerated genetic improvement in rice. *Plant J.* **2022**, *113*, 26–46. [CrossRef]
17. McCouch, S.R.; Zhao, K.; Wright, M.; Tung, C.-W.; Ebana, K.; Thomson, M.; Reynolds, A.; Wang, D.; DeClerck, G.; Ali, M.L.; et al. Development of genome-wide SNP assays for rice. *Breed. Sci.* **2010**, *60*, 524–535. [CrossRef]
18. Qu, J.; Yu, D.; Gu, W.; Zhao, Y.; Li, K.; Wang, H.; Sun, P.; Vicente, F.S.; Zhang, X.; Zhang, A.; et al. Development of a medium-density genotyping platform to accelerate genetic gain in fresh edible maize. *Plants* **2026**, *15*, 1288. [CrossRef]
19. Singh, N.; Wang, D.R.; Ali, L.; Kim, H.; Akther, K.M.; Harrington, S.E.; Kang, J.-W.; Shakiba, E.; Shi, Y.; DeClerck, G.; et al. A coordinated suite of wild-introgression lines in *indica* and *japonica* elite backgrounds. *Front. Plant Sci.* **2020**, *11*, 564824. [CrossRef]
20. Qin, P.; Lu, H.; Du, H.; Wang, H.; Chen, W.; Chen, Z.; He, Q.; Ou, S.; Zhang, H.; Li, X.; et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **2021**, *184*, 3542–3558. [CrossRef]
21. Zhang, C.; Li, M.; Liang, L.; Xiang, J.; Zhang, F.; Zhang, C.; Li, Y.; Liang, J.; Zheng, T.; Zhang, F.; et al. Rice3K56 is a high-quality SNP array for genome-based genetic studies and breeding in rice (*Oryza sativa* L.). *Crop J.* **2023**, *11*, 800–807. [CrossRef]
22. Li, Z.; Wang, L.; Liu, Y.; Ma, X.; Zhang, A.; Luo, Z.; Yan, M.; Zhou, L.; Chen, L.; Luo, L.; et al. WDR6K, a designed SNP array for the research and improvement of rice drought-resistance. *Plant Stress* **2025**, *15*, 100800. [CrossRef]
23. Ji, H.; Shin, Y.; Lee, C.; Oh, H.; Yoon, I.S.; Baek, J.; Cha, Y.-S.; Lee, G.-S.; Kim, S.L.; Kim, K.-H. Genomic variation in Korean *japonica* rice varieties. *Genes* **2021**, *12*, 1749. [CrossRef]
24. Wang, W.; Mauleon, R.; Hu, Z.; Chebotarov, D.; Tai, S.; Wu, Z.; Li, M.; Zheng, T.; Fuentes, R.R.; Zhang, F.; et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **2018**, *557*, 43–49. [CrossRef]
25. Zhao, Q.; Feng, Q.; Lu, H.; Li, Y.; Wang, A.; Tian, Q.; Zhan, Q.; Lu, Y.; Zhang, L.; Huang, T.; et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **2018**, *50*, 278–284. [CrossRef]
26. Shang, L.; Li, X.; He, H.; Yuan, Q.; Song, Y.; Wei, Z.; Lin, H.; Hu, M.; Zhao, F.; Zhang, C.; et al. A super pan-genomic landscape of rice. *Cell Res.* **2022**, *32*, 878–896. [CrossRef]
27. Zhang, F.; Xue, H.; Dong, X.; Li, M.; Zheng, X.; Li, Z.; Xu, J.; Wang, W.; Wei, C. Long-read sequencing of 111 rice genomes reveals significantly larger pan-genomes. *Genome Res.* **2022**, *32*, 853–863. [CrossRef]
28. 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]
29. Orlando, L.; 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]
30. Huang, X.; Feng, Q.; Qian, Q.; Zhao, Q.; Wang, L.; Wang, A.; Guan, J.; Fan, D.; Weng, Q.; Huang, T.; et al. High-throughput genotyping by whole-genome resequencing. *Genome Res.* **2009**, *19*, 1068–1076. [CrossRef]
31. Wang, D.R.; Agosto-Pérez, F.J.; Chebotarov, D.; Shi, Y.; Marchini, J.; Fitzgerald, M.; McNally, K.L.; Alexandrov, N.; McCouch, S.R. An imputation platform to enhance integration of rice genetic resources. *Nat. Commun.* **2018**, *9*, 3519. [CrossRef]
32. Kumar, A.; Kumar, S.; Singh, K.B.M.; Prasad, M.; Thakur, J.K. Designing a mini-core collection effectively representing 3004 diverse rice accessions. *Plant Commun.* **2020**, *1*, 100049. [CrossRef]

33. Arbelaez, J.D.; Dwiyanti, M.S.; Tandayu, E.; Llantada, K.; Jarana, A.; Ignacio, J.C.; Platten, J.D.; Cobb, J.; Rutkoski, J.E.; Thomson, M.J.; et al. 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. *Rice* **2019**, *12*, 55. [CrossRef]
34. Hada, A.; Dutta, T.K.; Singh, N.; Singh, B.; Rai, V.; Singh, N.K.; Rao, U. A genome-wide association study in Indian wild rice accessions for resistance to the root-knot nematode *Meloidogyne graminicola*. *PLoS ONE* **2020**, *15*, e0239085. [CrossRef]
35. Kretzschmar, T.; Mbanjo, E.G.N.; Magalit, G.A.; Dwiyanti, M.S.; Habib, M.A.; Diaz, M.G.; Hernandez, J.; Huelgas, Z.; Malabayabas, M.L.; Das, S.K.; et al. DNA fingerprinting at farm level maps rice biodiversity across Bangladesh and reveals regional varietal preferences. *Sci. Rep.* **2018**, *8*, 14920. [CrossRef]
36. Bi, W.; Sun, J.; Hosoi, J.; Aoki, M.; Sakai, N.; Itani, T.; Xu, Z.; Tominaga, T. Genetic identity based on whole-genome SNP array data of weedy rice in Nagano, Japan. *Agronomy* **2019**, *9*, 472. [CrossRef]
37. Huang, P.; Gu, Q.; Hu, Y.; Li, H.; Wu, Z.; Liu, W.; Zhu, Z.; Yuan, P.; Duan, L.; Zhou, Y.; et al. Genetic analysis of a collection of rice germplasm (*Oryza sativa* L.) through high-density SNP array provides useful information for further breeding practices. *Genes* **2022**, *13*, 830. [CrossRef]
38. Choudhury, D.R.; Kumar, R.; Maurya, A.; Semwal, D.P.; Rathi, R.S.; Gautam, R.K.; Trivedi, A.K.; Bishnoi, S.K.; Ahlawat, S.P.; Singh, K.; et al. SSR and SNP marker-based investigation of Indian rice landraces in relation to their genetic diversity, population structure, and geographical isolation. *Agriculture* **2023**, *13*, 823. [CrossRef]
39. Choudhury, D.R.; Kumar, R.; S, V.D.; Singh, K.; Singh, N.K.; Singh, R. Identification of a diverse core set panel of rice from the East Coast region of India using SNP markers. *Front. Genet.* **2021**, *12*, 726152. [CrossRef]
40. Detranaltes, C.; Ma, J.; Cai, G. Genome-wide association study reveals marker-trait associations with resistance to *Pythium irregulare* from soybean germplasm. *Int. J. Plant Biol.* **2024**, *15*, 769–782. [CrossRef]
41. Zhan, M.; Liu, D.; Peng, Y.; Zhou, Y. Genome-wide identification and functional prediction of the GRAS transcription factor family in rice under abiotic stress conditions. *Int. J. Plant Biol.* **2025**, *16*, 95. [CrossRef]
42. Crowell, S.; Korniliev, P.; Falcão, A.; Ismail, A.; Gregorio, G.; Mezey, J.; McCouch, S. Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nat. Commun.* **2016**, *7*, 10527. [CrossRef]
43. Lin, H.; Li, X.; Guo, Z.; Lv, Y.; Cen, X.; Ding, X.; Wu, H.; Li, X.; Huang, J.; Xiong, L. Genetic control of the root system in rice under normal and drought stress conditions by genome-wide association study. *PLoS Genet.* **2017**, *13*, e1006889. [CrossRef]
44. Famoso, A.N.; Zhao, K.; Clark, R.T.; Tung, C.W.; Wright, M.H.; Bustamante, C.; Kochian, L.V.; McCouch, S.R. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet.* **2011**, *7*, e1002221. [CrossRef]
45. Schroeder, J.I.; Campbell, M.T.; Bandillo, N.; Al Shiblawi, F.R.A.; Sharma, S.; Liu, K.; Du, Q.; Schmitz, A.J.; Zhang, C.; Véry, A.-A.; et al. Allelic variants of OsHKT1;1 underlie the divergence between indica and japonica subspecies of rice (*Oryza sativa*) for root sodium content. *PLoS Genet.* **2017**, *13*, e1006823. [CrossRef]
46. Yu, J.; Zhu, C.; Xuan, W.; An, H.; Tian, Y.; Wang, B.; Chi, W.; Chen, G.; Ge, Y.; Li, J.; et al. Genome-wide association studies identify OsWRKY53 as a key regulator of salt tolerance in rice. *Nat. Commun.* **2023**, *14*, 3550. [CrossRef]
47. Pérez-Almeida, I.; Navia-Pesantes, O.; Celi-Herán, R. Assessment of rice amylose content and grain quality through marker-assisted selection. *Int. J. Plant Biol.* **2025**, *16*, 52. [CrossRef]
48. Ellur, R.K.; Khanna, A.; Yadav, A.; Pathania, S.; Rajashekara, H.; Singh, V.K.; Gopala Krishnan, S.; Bhowmick, P.K.; Nagarajan, M.; Vinod, K.K.; et al. Improvement of Basmati rice varieties for resistance to blast and bacterial blight diseases using marker assisted backcross breeding. *Plant Sci.* **2016**, *242*, 330–341. [CrossRef]
49. Kim, M.-S.; Yu, J.-K.; Ko, S.-R.; Kim, K.-J.; Ji, H.; Kang, K.-K.; Cho, Y.-G. Marker-assisted backcrossing (MABc) to improve eating quality with thin seed coat and aleurone layer of non-glutinous *japonica* variety in rice. *Genes* **2022**, *13*, 210. [CrossRef]
50. Fan, Y.; Zhang, Z.; Huang, D.; Huang, T.; Wang, H.; Zhuang, J.; Zhu, Y. A novel PCR-based functional marker of rice blast resistance gene *Pi25*. *Agriculture* **2023**, *13*, 1926. [CrossRef]
51. Wang, H.; Ye, S.; Mou, T. Molecular breeding of rice restorer lines and hybrids for brown planthopper (BPH) resistance using the *Bph14* and *Bph15* genes. *Rice* **2016**, *9*, 53. [CrossRef]
52. Prahalada, G.D.; Shivakumar, N.; Lohithaswa, H.C.; Sidde Gowda, D.K.; Ramkumar, G.; Kim, S.-R.; Ramachandra, C.; Hittalmani, S.; Mohapatra, T.; Jena, K.K. Identification and fine mapping of a new gene, BPH31 conferring resistance to brown planthopper biotype 4 of India to improve rice, *Oryza sativa* L. *Rice* **2017**, *10*, 41. [CrossRef]
53. Beşer, N.; Del Valle, M.M.; Kim, S.M.; Vinarao, R.B.; Sürek, H.; Jena, K.K. Marker-assisted Introgression of a broad-spectrum resistance gene, *Pi40* improved blast resistance of two elite rice (*Oryza sativa* L.) cultivars of Turkey. *Mol. Plant Breed.* **2016**, *7*, 1–15.
54. Khan, G.H.; Shikari, A.B.; Vaishnavi, R.; Najeeb, S.; Padder, B.A.; Bhat, Z.A.; Parray, G.A.; Bhat, M.A.; Kumar, R.; Singh, N.K. Marker-assisted introgression of three dominant blast resistance genes into an aromatic rice cultivar Mushk Budji. *Sci. Rep.* **2018**, *8*, 4091. [CrossRef]

55. Sandhu, N.; Dixit, S.; Swamy, B.P.M.; Raman, A.; Kumar, S.; Singh, S.P.; Yadaw, R.B.; Singh, O.N.; Reddy, J.N.; Anandan, A.; et al. Marker assisted breeding to develop multiple stress tolerant varieties for flood and drought prone areas. *Rice* **2019**, *12*, 8. [CrossRef]
56. Singh, G.; Singh, N.; Ellur, R.K.; Balamurugan, A.; Prakash, G.; Rathour, R.; Mondal, K.K.; Bhowmick, P.K.; Gopala Krishnan, S.; Nagarajan, M.; et al. Genetic enhancement for biotic stress resistance in Basmati rice through marker-assisted backcross breeding. *Int. J. Mol. Sci.* **2023**, *24*, 16081. [CrossRef]
57. 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]
58. 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]
59. Ben Hassen, M.; Cao, T.V.; Bartholomé, J.; Orasen, G.; Colombi, C.; Rakotomalala, J.; Razafinimpiasa, L.; Bertone, C.; Biselli, C.; Volante, A.; et al. Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. *Theor. Appl. Genet.* **2018**, *131*, 417–435. [CrossRef]
60. Xu, Y.; Wang, X.; Ding, X.; Zheng, X.; Yang, Z.; Xu, C.; Hu, Z. Genomic selection of agronomic traits in hybrid rice using an NCII population. *Rice* **2018**, *11*, 32. [CrossRef]
61. 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]
62. Sagae, V.S.; Costa, N.M.; Suela, M.M.; Ferreira, D.D.; Nascimento, A.C.; Azevedo, C.F.; Silva, F.L.; Nascimento, M. Genomic prediction of root traits via aerial traits in soybean using canonical variables. *Int. J. Plant Biol.* **2024**, *15*, 242–252. [CrossRef]
63. Ben Hassen, M.; Bartholomé, J.; Valè, G.; Cao, T.-V.; Ahmadi, N. Genomic prediction accounting for genotype by environment interaction offers an effective framework for breeding simultaneously for adaptation to an abiotic stress and performance Under normal cropping conditions in rice. *G3 Genes Genomes Genet.* **2018**, *8*, 2319–2332. [CrossRef]
64. Baertschi, C.; Cao, T.-V.; Bartholomé, J.; Ospina, Y.; Quintero, C.; Frouin, J.; Bouvet, J.-M.; Grenier, C. Impact of early genomic prediction for recurrent selection in an upland rice synthetic population. *G3 Genes Genomes Genet.* **2021**, *11*, jkab320. [CrossRef]
65. Nguyen, V.H.; Morantte, R.I.Z.; Lopena, V.; Verdeprado, H.; Murori, R.; Ndayiragije, A.; Katiyar, S.K.; Islam, M.R.; Juma, R.U.; Flandez-Galvez, H.; et al. Multi-environment genomic selection in rice elite breeding lines. *Rice* **2023**, *16*, 7. [CrossRef]
66. Biswas, P.S.; Ahmed, M.M.E.; Afrin, W.; Rahman, A.; Shalahuddin, A.K.M.; Islam, R.; Akter, F.; Syed, M.A.; Sarker, M.R.A.; Ifterkharuddaula, K.M.; et al. Enhancing genetic gain through the application of genomic selection in developing irrigated rice for the favorable ecosystem in Bangladesh. *Front. Genet.* **2023**, *14*, 10832. [CrossRef]
67. Yang, G.; Chen, S.; Chen, L.; Sun, K.; Huang, C.; Zhou, D.; Huang, Y.; Wang, J.; Liu, Y.; Wang, H.; et al. Development of a core SNP arrays based on the KASP method for molecular breeding of rice. *Rice* **2019**, *12*, 21. [CrossRef]
68. Steele, K.; Tulloch, M.Q.; Burns, M.; Nader, W. Developing KASP markers for identification of basmati rice varieties. *Food Anal. Methods* **2021**, *14*, 663–673. [CrossRef]
69. Yuan, X.; Li, Z.; Xiong, L.; Song, S.; Zheng, X.; Tang, Z.; Yuan, Z.; Li, L. Effective identification of varieties by nucleotide polymorphisms and its application for essentially derived variety identification in rice. *BMC Bioinform.* **2022**, *23*, 30. [CrossRef]
70. Tang, W.; Lin, J.; Wang, Y.; An, H.; Chen, H.; Pan, G.; Zhang, S.; Guo, B.; Yu, K.; Li, H.; et al. Selection and validation of 48 KASP markers for variety identification and breeding guidance in conventional and hybrid rice (*Oryza sativa* L.). *Rice* **2022**, *15*, 48. [CrossRef]
71. Qiao, W.; Qi, L.; Cheng, Z.; Su, L.; Li, J.; Sun, Y.; Ren, J.; Zheng, X.; Yang, Q. Development and characterization of chromosome segment substitution lines derived from *Oryza rufipogon* in the genetic background of *O. sativa* spp. indica cultivar 9311. *BMC Genom.* **2016**, *17*, 580. [CrossRef]
72. Mulsanti, I.W.; Yamamoto, T.; Ueda, T.; Samadi, A.F.; Kamahora, E.; Rumanti, I.A.; Thanh, V.C.; Adachi, S.; Suzuki, S.; Kanekatsu, M.; et al. Finding the superior allele of japonica-type for increasing stem lodging resistance in indica rice varieties using chromosome segment substitution lines. *Rice* **2018**, *11*, 25. [CrossRef]
73. Wang, D.; Sun, W.; Yuan, Z.; Sun, Q.; Fan, K.; Zhang, C.; Yu, S. Identification of a novel QTL and candidate gene associated with grain size using chromosome segment substitution lines in rice. *Sci. Rep.* **2021**, *11*, 189. [CrossRef]
74. Ding, G.; Hu, B.; Zhou, Y.; Yang, W.; Zhao, M.; Xie, J.; Zhang, F. Development and characterization of chromosome segment substitution lines derived from *Oryza rufipogon* in the background of the *Oryza sativa indica* restorer line R974. *Genes* **2022**, *13*, 735. [CrossRef]
75. Paul, P.J.; Patil, S.B.; Vipparla, A.K.; Sagare, D.B.; Jain, A.; Balaraju, E.; Venkateswarlu, C.; Alam, S.; Anandan, A.; Dash, S.; et al. Enhancing crop resilience through elite pre-breeding lines with key traits for dry direct-seeded rice. *Plant Genome* **2025**, *18*, e70127. [CrossRef]

76. Guo, D.; Li, Y.; Lu, H.; Zhao, Y.; Kurata, N.; Wei, X.; Wang, A.; Wang, Y.; Zhan, Q.; Fan, D.; et al. A pangenome reference of wild and cultivated rice. *Nature* **2025**, *642*, 662–671. [CrossRef]
77. Lee, C.; Cheon, K.S.; Shin, Y.; Oh, H.; Jeong, Y.M.; Jang, H.; Park, Y.C.; Kim, K.Y.; Cho, H.C.; Won, Y.J.; et al. Development and application of a target capture sequencing SNP-genotyping platform in rice. *Genes* **2022**, *13*, 794. [CrossRef]
78. Gudeta, T.B.; Keneni, G.; Figlan, S. Progress in the use of combined omics for mungbean breeding improvement and Its potential in promoting resistance against cercospora leaf spot. *Int. J. Plant Biol.* **2024**, *15*, 482–504. [CrossRef]
79. Krusenbaum, L.; Wissuwa, M. Low-call-rate SNPs and presence–absence variation identified in the rice pan-genome can improve genomic prediction of rice gene bank accessions. *Theor. Appl. Genet.* **2025**, *138*, 295. [CrossRef]
80. Vourlaki, I.-T.; Ramos-Onsins, S.E.; Pérez-Enciso, M.; Castanera, R. Evaluation of deep learning for predicting rice traits using structural and single-nucleotide genomic variants. *Plant Methods* **2024**, *20*, 121. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.