



## Association mapping in rice: basic concepts and perspectives for molecular breeding

Holden Verdeprado, Tobias Kretzschmar, Hasina Begum, Chitra Raghavan, Priya Joyce, Prakash Lakshmanan, Joshua N. Cobb & Bertrand C.Y. Collard

To cite this article: Holden Verdeprado, Tobias Kretzschmar, Hasina Begum, Chitra Raghavan, Priya Joyce, Prakash Lakshmanan, Joshua N. Cobb & Bertrand C.Y. Collard (2018) Association mapping in rice: basic concepts and perspectives for molecular breeding, Plant Production Science, 21:3, 159-176, DOI: [10.1080/1343943X.2018.1483205](https://doi.org/10.1080/1343943X.2018.1483205)

To link to this article: <https://doi.org/10.1080/1343943X.2018.1483205>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 02 Jul 2018.



Submit your article to this journal [↗](#)



Article views: 4065



View related articles [↗](#)



Citing articles: 12 View citing articles [↗](#)

## Association mapping in rice: basic concepts and perspectives for molecular breeding

Holden Verdeprado<sup>a</sup>, Tobias Kretzschmar<sup>a,b</sup>, Hasina Begum<sup>a,c,d</sup>, Chitra Raghavan<sup>a</sup>, Priya Joyce<sup>e</sup>, Prakash Lakshmanan<sup>e</sup>, Joshua N. Cobb<sup>a</sup> and Bertrand C.Y. Collard<sup>a,e,f</sup>

<sup>a</sup>Rice Breeding Platform, International Rice Research Institute (IRRI), Los Baños, Philippines; <sup>b</sup>Southern Cross Plant Science, Southern Cross University, Lismore, Australia; <sup>c</sup>Genetic Resources and Seed Division, Bangladesh Rice Research Institute (BRRI), Gazipur, Bangladesh; <sup>d</sup>Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany; <sup>e</sup>Sugar Research Australia (SRA), Indooroopilly, Australia; <sup>f</sup>NSW Department of Primary Industries, Yanco, Australia

### ABSTRACT

In the last decade, association mapping (AM) has become a well-established method to detect genes and quantitative trait loci (QTLs) associated with agronomically important traits. The identification of a large number of single nucleotide polymorphisms (SNPs) from genome sequencing and concurrent development of high-throughput genotyping platforms has led to AM being widely used for a range of crops. These technologies have been used in rice (*Oryza sativa*) to explore the abundant diversity and there is enormous potential to identify novel QTLs for traits of interest. Due to the availability of cost-effective high-throughput SNP genotyping methods and rapid developments in rice genomics, it is inevitable that these AM approaches will become more popular in the future, especially in the context of genome-wide association studies (GWAS). In this paper, we review the fundamental concepts, critical considerations and limitations of AM focusing on rice, and reiterate the importance of accurate phenotypic data. We also include a section about connecting GWAS to molecular breeding, covering practical consideration for breeders, which is required to use GWAS results in actual rice molecular breeding programs and which has not received adequate attention in the scientific literature.

### ARTICLE HISTORY

Received 23 August 2017  
Revised 30 April 2018  
Accepted 19 May 2018

### KEYWORDS

Biparental QTL mapping (BPQM); association mapping (AM); genome-wide association study (GWAS); linkage disequilibrium (LD); precision phenotyping; marker-assisted selection (MAS)

## Introduction

The development of quantitative trait loci (QTL) mapping methods to identify genes or QTLs controlling quantitative traits was a landmark achievement in plant genetics research in the late 1980s (Doerge, 2002; Mohan et al., 1997; Semagn, Bjornstad, & Xu, 2010). Since then literally thousands of research papers have reported the identification of genes or QTLs for important traits across a diverse range of plant species. This wealth of genetic information was built on the foundation of decades of research in crop molecular genetics and genomics.

The selection and enrichment of QTLs within breeding material was the inevitable next step from QTL discovery to application, and there have been numerous reports of successful tracking of genes and QTLs within breeding programs. DNA (or molecular) markers have enabled selection of major genes or QTLs for critical or important traits in a process called marker-assisted selection (MAS), which has revolutionized plant breeding. DNA markers are used as tools by breeders to improve the accuracy or efficiency of selection (Dwivedi et al., 2007; Xu & Crouch, 2008) as well cost and time efficiency. Apart from MAS, DNA markers

have numerous applications including DNA fingerprinting, genetic diversity analysis and parental characterization (Collard & Mackill, 2008). Despite the availability of information, the extent of actual product development using MAS in rice breeding is limited to a few major QTLs for biotic and abiotic stress tolerance (Gregorio, Islam, Vergara, & Thirumeni, 2013; Singh et al. 2015b).

Recent developments in genotyping platforms and systems to implement molecular breeding schemes offer new tools for modern rice breeders (Thomson, 2014). Single nucleotide polymorphism (SNP) markers are biallelic, co-dominant markers which are abundant in the genome (Mammadov, Aggarwal, Buyyarapu, & Kumpatla, 2012). Insertion-deletion (Indel) markers are also prevalent and can be screened using high-throughput genotyping platforms (Misra et al., 2017; Yonemaru et al., 2015). These platforms are cost-efficient, can handle large sample sizes and provide fast data turn-around time. The availability of genome sequences and genomic resources continues to provide a wealth of SNP and Indel markers which will undoubtedly be the marker type of choice for decades to come (McCouch et al., 2010).

Rice was the first crop genome to be sequenced (Goff et al., 2002; Yu et al., 2002; (International Rice Genome Sequencing Project, 2005), and is uniquely poised as a major food crop and a model crop species, with abundant genetic diversity and publicly available genomic and phenotypic resources. Genomics resources include several *de novo* genome builds, thousands of re-sequenced genomes, numerous SNP resources and validated high-throughput SNP genotyping platforms. Developments in next-generation sequencing technologies have led to a wealth of genome sequence data that was not conceivable only a few years ago (Edwards, Batley, & Snowdon, 2013; Van, Rastogi, Kim, & Lee, 2013; Varshney & Dubey, 2009). While the *japonica* accession Nipponbare is still the gold standard reference genome, several other representative rice varieties from all major subspecies were subsequently assembled *de novo*, including IR64, DJ123 (Schatz et al., 2014) and Shuhui498 (Du et al., 2017). Recently, 3000 rice genomes were re-sequenced and a resulting set of over 29 million SNPs has been characterized and made accessible (3,000 rice genomes project, 2014; Alexandrov et al., 2015).

The availability of high-throughput SNP platforms in conjunction with advances in computational methods has also led to the molecular dissection of traits using genome-wide association studies (GWASs) across diversity panels and the development of genomics-based molecular breeding strategies (Lipka et al., 2015; Morrell, Buckler, & Ross-Ibarra, 2011; Varshney, Graner, & Sorrells, 2005). Association mapping (AM) has now emerged as a standard method for gene and QTL discovery for major crops (Gupta, Kulwal, & Jaiswal, 2014; Rafalski, 2002, 2010; Zhu, Gore, Buckler, & Yu, 2008).

A number of excellent general review articles have been written about AM in crops (Buckler, Ersoz, Yu, & Buckler, 2009; Ersoz, Yu, & Buckler, 2009; Gupta et al., 2014; Hamblin, Buckler, & Jannink, 2011; Ingvarsson & Street, 2011; Mackay & Powell, 2007; Myles et al., 2009; Rafalski, 2002, 2010; Zhang, Zhong, Shahid, & Tong, 2016; Zhu et al., 2008). In this article, we review the fundamental concepts and use of AM specifically focusing on rice, a globally important food crop. We also discuss applied research activities and relevant topics that are required to move from QTL detection to actual molecular breeding, which has rarely been discussed in the scientific literature to date. In our view, consideration of the latter topic is critical to ensure that AM is fully integrated with rice molecular breeding programs in the future.

## Approaches to QTL mapping

Traditional QTL mapping utilizes biparental population (BP) mapping populations from controlled crosses (Zhu

et al., 2008) which we refer to as biparental population QTL mapping (BPQM) in this paper. The selected parents usually differ for traits of interest, and the resulting mapping population should segregate for these traits, although transgressive segregation can be detected even when the parents do not differ for the trait (Collard, Jahufer, Brouwer, & Pang, 2005; Mackay, Stone, & Ayroles, 2009). BP mapping populations include  $F_2$ , backcross (BC), recombinant inbred lines (RILs) and near-isogenic lines (NILs).  $F_2$  and  $BC_1$  populations only require two generations to develop. Although they are the simplest population types, these are not homozygous (i.e. 'fixed') and cannot be repeatedly phenotyped. On the other hand, more complex populations, such as NILs and RILs, usually require seven or eight generations to develop, but can be repeatedly phenotyped over multiple years and across environments. Chromosome segment substitution lines (CSSLs) contain chromosome segments from the donor parent in the recurrent parent background. These secondary mapping populations are required to facilitate more comprehensive analysis of target QTLs (Yano, 2001). More recently in rice, multi-parent populations called multi-parent advanced generation intercross (MAGIC; Bandillo et al., 2013), CSSLs (Bessho-Uehara et al., 2017; Ogawa et al., 2016) and nested-association mapping populations (Fragoso et al., 2017) have been used for QTL mapping.

Constructing linkage maps is necessary to identify chromosomal locations and effects of genes and QTLs associated with traits of interest. A linkage map (or genetic map) shows the relative position and genetic distances between markers or genes along chromosomes (Collard et al., 2005). QTL mapping is based on the co-segregation of QTLs and DNA markers. Conceptually, this is based on the principle of chromosome recombination which occurs during meiosis. The principle of QTL detection is based on the association between phenotype and the genotype of markers. The mapping population is partitioned into groups based on the marker each individual carries. A significant difference between the phenotypic means of the groups indicates that the specific marker used is linked to a QTL controlling the trait (Collard et al., 2005).

In contrast to BPQM, association analysis uses diverse accessions from germplasm collections of varieties, landraces or breeding material referred to as a 'panel' (also 'diversity panel' or 'association panel'). Identifying novel QTLs from these panels is the most important use of GWAS for breeding. QTL identification is performed by examining the associations of the markers with the trait that can be explained by the 'linkage disequilibrium' (LD) between markers and polymorphisms across a set of diverse germplasm (Zhu et al., 2008). QTLs are identified based on historical recombination events between SNPs

and QTL at the population level (Nordborg & Tavaré, 2002). It is based on the principle that over multiple generations of recombinations, markers that are tightly linked to genes for the trait of interest will generally remain to be associated with the trait.

Population sizes used in preliminary genetic mapping studies generally range from 100 to 250 individuals (Collard et al., 2005). A larger population size (approximately >500) is required for the analysis of QTLs having small effects on the target trait. Population sizes used for AM are usually larger than BPQM and, depending on the population structure and diversity, generally require several hundred individuals to identify QTLs.

### Comparison between BPQM and AM

Both BPQM and AM are based on the co-segregation of DNA markers with traits of interest (Zhu et al., 2008). While the development of a mapping population is required for linkage analysis, AM usually uses diverse populations, or individuals with contrasting geographical origin (Lipka et al., 2015). Therefore, AM requires less time and resources because phenotypic data are sometimes available for the populations (i.e. panels) that are used for analysis. There is no need to perform controlled crosses to develop mapping populations, although a large number of markers (in the order of thousands or tens of thousands) are required compared to only a few hundred markers for BPQM.

The mapping populations used in BPQM include limited recombination events resulting from fewer generations to establish these populations (i.e. two generations for  $F_2$  population and six to ten generations for RILs). For this reason, QTLs are usually span 10 to 20 cM intervals. In contrast, AM utilizes populations which have undergone many generations of recombination since domestication and therefore in general, only markers that are physically located close to the QTL will be detected as significant. This also explains why linkage analysis has a lower resolution of QTL detection compared to AM (Flint-Garcia, Thornsberry, & Buckler, 2003). As the size of the interval for localizing the QTL decreases, the number of individuals required to detect at least one recombinant in the region of interest increases, as does the number of molecular markers necessary to detect recombination events.

BPQM uncovers only a small portion of the genetic architecture for a trait because only alleles that differ between the two parental lines will segregate while AM provides an alternative route to identifying QTL that has effects across a broader range of germplasm. On the other hand, BPQM can lead to the discovery of very rare

alleles provided the donor accession carries it, while in AM, alleles below a certain minor allele frequency (usually a MAF of < 5%) will be filtered out and hence rare QTLs are usually not detectable. By highlighting the differences between traditional BPQM and AM, it becomes clear that these methods are complementary to each other (Figure 1).

### Key concepts in AM

The main purpose of AM is to dissect complex traits and identify QTLs (Zhu et al., 2008). QTLs detected using AM (also called 'signals', 'peaks' or 'hits') are usually represented using 'Manhattan plots' which show the association of markers with the trait along a chromosome. The y-axis indicates  $-\log_{10}(P \text{ value})$  for the association plotted against the SNPs along each chromosome on the x-axis, so the map positions of all markers used must be known (Figure 2).

### Linkage disequilibrium (LD)

The fundamental basis of detecting QTLs using AM is due to LD. LD is the 'non-random association of alleles at two or more different loci' in a population (Flint-Garcia et al., 2003; Slatkin, 2008). It measures the strength of correlation between markers caused by their shared genetic history. It is a characteristic of pairs of SNP that describes the degree to which an allele of one SNP is inherited or correlated with an allele of another SNP within a population (Bush & Moore, 2012). The terms LD mapping and AM are often used interchangeably.

AM is dependent upon the extent of LD across the genome. Thus, the extent of LD should be known before AM can be performed. Two markers that are in LD show non-random association between alleles, but do not necessarily correlate with a particular phenotype. An association is defined when there is significant correlation between the covariance of a marker polymorphism and a trait of interest. This is the basis of identifying QTLs associated with markers by AM (Soto-Cerda & Cloutier, 2012). It is important to note that LD is not the same as physical linkage; many allelic variants that are very close to each other may have low LD either due to recombination or because the variants are not at equal frequencies. It is also worth noting that SNPs in LD can be located on different chromosomes. If LD decays within a short distance, mapping resolution is expected to be high, but a larger number of markers are required (Rafalski, 2002). If LD extends over a long distance, mapping resolution will be low, but a relatively small number of markers are required.

RESOURCE or STAGE	TRADITIONAL QTL MAPPING	ASSOCIATION MAPPING
EXPERIMENTAL PLANT MATERIAL	<b>Bi-parental population</b> e.g. F <sub>2</sub> , backcross or RIL derived from A x B	<b>Panel</b> e.g. diverse accessions A, B, C, D, E, F, G, H, I, J, ... Z....
GENOTYPING	<b>Linkage map construction</b> Number of markers ~ 100–400 for framework map	<b>High marker density</b> Number of markers generally >10,000 (GWAS) aligned to reference map position
PHENOTYPING	<b>Field, greenhouse or lab-based trials</b> Requires experimental design and appropriate replication May need to account for differences in growth and morphology	
ANALYSIS	<b>Interval mapping</b> e.g. composite interval mapping (CIM) or multiple interval mapping (MIM)	<b>Association analysis</b> Generally includes population structure analysis and/or marker-based kinship analysis
SIGNIFICANCE TESTING	e.g. Permutation tests	e.g. Bonferroni correction or False discovery rate (FDR)

Figure 1. Comparison of QTL mapping methods.

In practice, combinations of closely linked, adjacent SNPs or ‘haplotypes’ are used to characterize a QTL region or allele of a gene rather than a single SNP (Buntjer, Sørensen, & Peleman, 2005; Rafalski, 2002). Generally, about 5–15 SNP markers per locus are usually sufficient to characterize haplotypes in crops (Famoso et al., 2011). Haplotypes based on multiple SNPs are generally multi-allelic, in contrast to a single SNP which is biallelic. Variations in SNP haplotypes are of interest to breeders in order to identify genomic regions under selection. Recently in rice, efforts have been initiated to compile data and investigate haplotype blocks in breeding material (Yamamoto et al., 2010), or SNP haplotypes for genes or QTLs (Yonemaru, Ebana, & Yano, 2014).

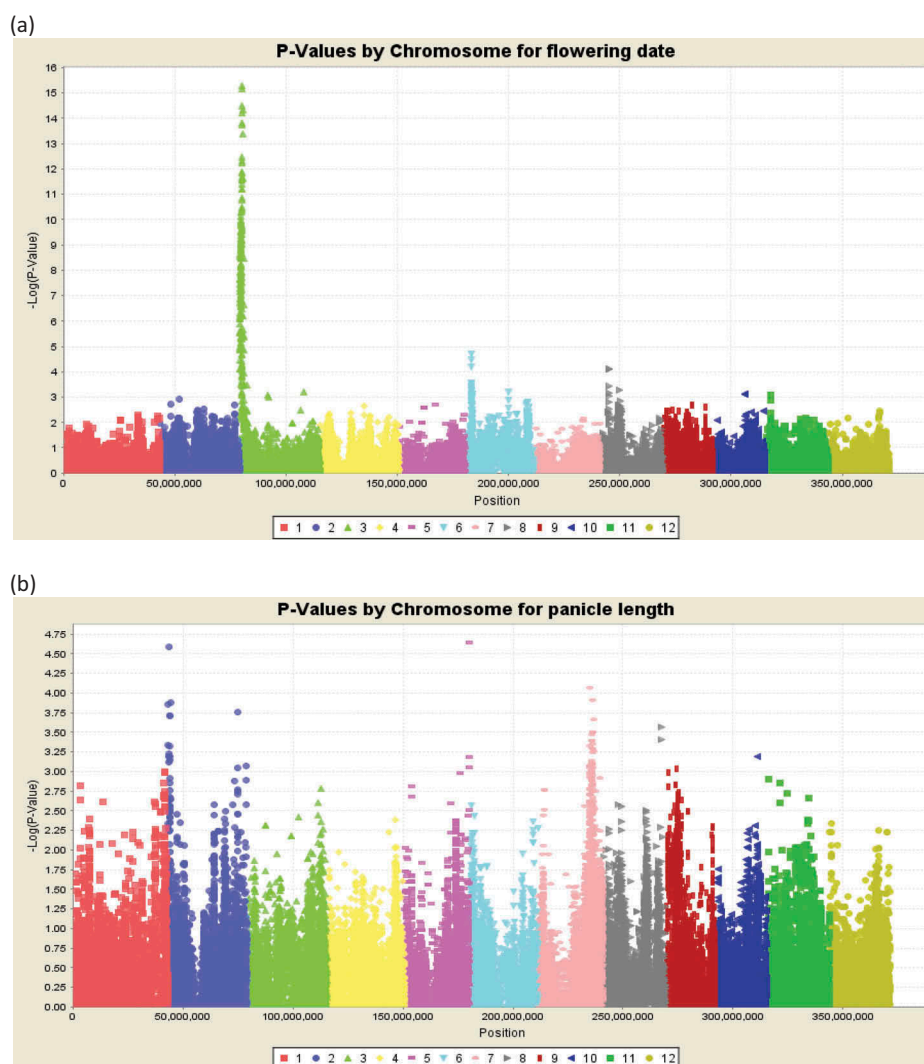
One of the first estimates of LD in *O. sativa* was ~100 kb based on the region around *xa5*, a recessive gene conferring resistance to bacterial leaf blight (Garris, McCouch, & Kresovich, 2003). Targeting six genomic regions on chromosomes 1 and 4 and unlinked background SNPs, LD was estimated and ranged from 75 kb in *indica*, ~150 kb in tropical *japonica* and >500 kb in temperate *japonica* (Mather et al., 2007). Huang et al. (2010) estimated that genome-wide LD rates were ~123 kb for *indica* and ~167 kb for *japonica* subspecies. These estimates (< 1 cM) represent a significant improvement in comparison with the confidence interval of QTLs detected by BPQM.

It is worth noting that these average rates of LD reflect ~10,000 years of historical recombination. In a well-designed breeding program, these rates will be higher due to the crossing structure and use of elite x elite crosses. Based on the estimated LD in rice, a minimum of 5000 markers should theoretically be sufficient to cover the ~400 Mb rice genome using the estimate of 75 kb (i.e.  $400,000/75 = 5333$  markers) as proposed by Courtois et al. (2013). However, filtering for monomorphic markers and low allele frequencies may lead to sub-optimal densities in practice, and therefore genotyping at higher densities is recommended.

### Analysis methods for AM

In early reports of AM in crops, relative simple statistical association tests (e.g. general linear models for normally distributed traits or non-parametric tests) were performed (Li & Zhu, 2013), analogous to the initial use of single-marker analysis in BPQM. More advanced methods for AM were subsequently developed, which have become routinely used in crops (Lipka et al., 2015).

The effect of population structure (referred to as ‘Q’) must be accounted for when performing AM. The most commonly used approach to assess the level of population structure is to use marker information to detect sub-groups within the experimental population. This is



**Figure 2.** AM results for agronomic traits on all 12 chromosomes (2012 dry season data), analyzed from Begum et al. (2015). Manhattan plots for flowering date (a) and panicle length (b). Analysis was performed using TASSEL software Version 4.1.34 (Bradbury et al., 2007). Results indicate a major locus for flowering time on chromosome 3 and multiple QTLs for panicle length, typical of quantitative traits.

sometimes called 'structured association' and pieces of information on population structure are considered as fixed effects and used as cofactors in the analysis. Using this method, a set of random markers is used to infer the structure of the population as well as the ancestry of the panel. Common methods used to calculate population structure include: (1) using a computer program called STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) or (2) principal component analysis (PCA) (Price et al., 2006). One of the main advantages of PCA is that the computational analysis is considerably simpler.

It was subsequently determined that taking into account the level of genetic relatedness (called 'kinship'; referred to as '*K*') improved the accuracy of AM (Yu et al.,

2006). In this landmark paper, mixed linear models (MLM) were used to include information on population structure and kinship (i.e. '*Q + K*' mixed-model) that was superior in terms of reducing the false positive rate while maintaining statistical power (Zhao et al., 2007). This has now become commonly used in crops and algorithms have been streamlined to improve the efficiency (i.e. speed) of the 'data crunching'. The most commonly used approaches in rice include efficient mixed model association (EMMA; Kang et al., 2008), EMMA eXpediated (EMMAX, Kang, Sul, & Service et al., 2010), compressed MLM and population parameters previously determined (P3D) (Zhang et al., 2010). Recently, even more advanced methods have been developed (e.g. Settlement of MLM Under

Progressively Exclusive Relationship (SUPER, Wang, Tian, Pan, Buckler, & Zhang, 2014) and multi-locus MLM (Segura et al., 2012)).

False positive signals are an inherent problem in GWAS, which is often referred to as the problem of ‘multiple testing’ because a large number of markers are tested and a  $P$  value is generated for each individual test (reviewed by Balding (2006) and Gupta et al. (2014)). Therefore, the cumulative possibilities of false positive results are large when all of the multiple tests are considered. Among the methods to define significance levels (i.e. that will reduce the chance of false positives but increase the chance of false negatives and vice versa), the Bonferroni correction and the false discovery rate (FDR; Benjamini & Hochberg, 1995) are most commonly used.

### Practical considerations – scale of AM

There are two broad categories of AM: (1) candidate gene (CG) analysis and (2) GWAS. While CG analysis is a hypothesis-driven approach based on prior studies about genes involved in the trait of interest, GWAS is a more comprehensive approach which does not require any initial information about the genetic control of a trait of interest (Zhu et al., 2008). Due to the increasing use of GWAS over CG, the term is now used interchangeably with AM.

Genotyping is a critical component of GWAS experiments, and two methods are usually employed. Genotyping by sequencing (GBS) is a cheaper way to generate medium to high marker densities (i.e. 10,000–100,000 filtered data points per sample) (Elshire et al., 2011; He et al., 2014). Although GBS has been optimized and successfully applied in rice (Table 1), it requires considerable wet-lab and computational expertise for library preparation and bioinformatics data analysis, respectively. Fixed rice arrays (i.e. ‘SNP chips’) are easier to use and several options for different marker densities are available (Chen et al., 2014a; McCouch et al., 2010; Singh et al. 2015a). At present, the cost of rice is generally about US\$30–US\$50/sample. User-friendly, open-access software to perform AM, such as TASSEL (Bradbury et al., 2007) and GAPIT (Lipka et al., 2012; Tang et al., 2016), have been commonly used in rice and allow data formatting, data filtering, data visualization for quality control and advanced analyses.

### Limitations of AM

There are several inherent limitations of AM which researchers must consider (reviewed by Korte & Farlow, 2013) when interpreting results, especially for downstream applications. Some of these limitations are

based on the fundamental architecture of complex quantitative traits and are briefly discussed below.

### Rare variants

Previous research has indicated that combinations of rare alleles are usually involved in controlling complex traits. The power to detect rare alleles, however, is low especially since markers used in GWAS are often excluded based on the MAF (i.e.  $< 0.05$ ), since the statistical methods used for AM are not reliable for very low MAF. Furthermore, these variants may also be in strong/complete LD with non-causative rare variants (e.g. specific SNPs within an individual which have nothing to do with the target trait, but appear to be associated). This is referred to as a ‘synthetic association’ (Korte & Farlow, 2013). Importantly, indirect associations may be detected due to LD between multiple factors affecting a single trait, especially caused by adaptation (Platt, Vilhjálmsson, & Nordborg, 2010). Thus, conventional QTL mapping approaches still are and will likely remain more effective in discovering rare alleles. The deliberate inclusion of lines derived from BPs with parents containing rare alleles (if known) to supplement the panel and the use of ‘joint-linkage association mapping’ are two approaches to improve the detection for rare variants.

### Small-effect QTLs

Many complex traits are controlled by multiple loci, many of which are small effect QTLs (Holland, 2007). Therefore, their presence may simply be too small to be reliably detected, which is a problem when using any QTL mapping method. Detection of these small-effect QTLs can be increased by using a larger population size and accurate phenotypic measurements (i.e. to increase heritability).

### Ascertainment bias

This refers to the fact that SNPs are pre-selected during the SNP development stage prior to their use for association analysis. This is relevant for array-based SNP genotyping platforms, which are usually derived by sequencing a small number of accessions for SNP discovery (McCouch et al., 2010; Myles et al., 2009). Therefore, SNPs that are associated with a trait might not be included in SNP chips, and therefore cannot be detected. Furthermore, fixed arrays might be designed to be maximally informative across certain subgroups (e.g. *indica*), but might detect many monomorphic markers when deployed across panels of a different subgroup (e.g. *japonica*) and hence may not be useful.

**Table 1. Representative AM experiments in rice (2013 to 2018), sorted by traits.**

Population	Sample size <sup>a</sup>	Marker type	Platform	No. of markers <sup>b</sup>	Traits	Highlights	Reference
Elite <i>indica</i> breeding lines	369	SNP	GBS	71,710	Agronomic traits	Emphasis on elite breeding material. Also mapped plot yield.	Begum et al. (2015)
Diverse <i>indica</i> accessions from 25 countries	469 (523)	SNP	SNP chip (5 k)	4,136	Agronomic traits	Focus on plant architecture traits.	Lu et al. (2015)
Vietnamese rice accessions ( <i>indica</i> and <i>japonica</i> )	270	DAIT and SNP	GBS	241 and 25,971	Agronomic traits	Validation of results using flowering time.	Phung et al. (2014)
Landrace and elite accessions	529 (533)	SNP	GBS	4,358,600	Agronomic traits	High-throughput rice phenotyping methods used.	Yang et al. (2014)
Japanese rice varieties ( <i>japonica</i> )	176	SNP and indels	Whole genome sequencing	426,337 SNPs and 67,544 indels	Agronomic traits	Rapid identification of new genes.	Yano et al. (2016)
Diverse accessions	248 (198 <i>indica</i> subset (380))	SNP	SNP chip and GBS	384 and 40,396	Bacterial blight resistance	Nine isolates used. Some novel QTLs detected. Others overlapped with known <i>Xa</i> genes.	Dilla-Ermita et al. (2017)
<i>Indica</i> accessions	172	SNP	SNP chip (HDRA)	700,000	Bacterial blight resistance	Novel significant associations on seven chromosomes.	Zhang et al. (2017a)
<i>Japonica</i> accessions (adapted to Italian conditions)	138	SNP	GBS	31,752	Bakanae disease resistance	New QTLs discovered. Candidate genes identified within largest effect QTLs.	Volante et al. (2017)
Rice diversity panel 1	>336		SNP chip (44 k)	44,000	Blast resistance	Field sites containing diverse <i>M. oryzae</i> populations in three representative regions in China. Detected 16 loci associated with field blast resistance. Twenty cultivars were highly resistant to rice blast disease under both field and growth chamber conditions.	Zhu et al. (2016)
<i>Indica</i> and <i>japonica</i> panels	190 and 151	SNP	GBS (DAIT-Seq)	14,147 and 12,438	Blast resistance (field)	Phenotyping conducted at two upland locations with naturally occurring blast epidemics (i.e. blast hot spots). Higher level of quantitative resistance in the <i>japonica</i> panel than the <i>indica</i> panel.	Raboin et al. (2016)
Diverse accessions	529	SNP	Re-sequencing	1 to 3 million SNPs, depending on the population group	Chlorophyll content	Grain number, plant height and heading date7 (Ghd7) associated with variation of chlorophyll content at the heading stage. Three F2 mapping populations with parents selected from the panel were used for validation of GWAS signals.	Wang et al. (2015)
Traditional landraces, modern elite and old cultivar	200	SNP	SNP chip	1,672	Cold tolerance (germination stage)	Markers associated with low temperature (LTG) and coleoptile elongation were determined.	Sales, Vinuel, Domingo and Marqués (2017)
Rice mini-core collection	202	SSR, InDel and SNP	Capillary system (SSR markers)	148, 3 and 6	Cold tolerance	Novel QTLs identified, including subspecies- specific alleles.	Schläppi et al. (2017)
Rice diversity panel 1	421	SNP	SNP chip (HDRA)	700,000	Cold tolerance	Subspecies-specific and common QTLs identified. Candidate genes identified for QTLs.	Shakiba et al. (2017)
Chinese <i>indica</i> varieties	249	SNP	SNP chip (5 k)	5,000	Cold tolerance (germination and reproductive stage)	Detected 47 SNPs loci associated with severity of damage and seed survival.	Zhang et al. (2018)
Diverse landrace accessions ( <i>indica</i> and <i>japonica</i> )	324 and 267 (591)	SNPs and Indels	Whole genome sequencing	2,933,037	Cooked grain dimensions	Novel genetic region associated with grain dimensions was identified.	Misra et al. (2017)
Diverse landraces and cultivars	270	SNP	Re-sequencing	393,429	Drought tolerance (including agronomic traits)	Compared GWAS between irrigated and drought conditions. Many known genes located with QTLs detected. Novel QTLs also detected.	Ma et al. (2016)
Malaysian rice accessions	75	SSR	Gel electrophoresis	119	Drought tolerance (yield and yield related traits under drought stress)	Several new large effect QTLs and candidate genes identified.	Swamy et al. (2017)
PRAY ( <i>indica</i> )	301 and 151 (452)	SNP	GBS	42,500	Environment responsiveness	ER candidate alleles associated with relative response of panicle weight to low density were validated.	Kikuchi et al. (2017)
MAGIC ( <i>indica</i> ) <sup>c</sup>	432	SNP	SNP chip	5,291	Flood tolerance (during germination)	Integrated gene expression information to identify promising candidate gene.	Zhang et al. (2017c)
Diverse <i>indica</i> accessions from 19 countries	203	SNP	GBS	16,232	Flowering time	Eighteen significant QTLs were identified.	Dingkuhn et al. (2017a)
<i>Indica</i> accessions from ORYTAG diversity panel	222	SNP	SNP chip (HDRA)	700,000	Iron and zinc toxicity	Five toxicity tolerance QTLs were detected	Zhang et al. (2017b)
<i>Indica</i> accessions (from IRR)	95 (58 and 37)	SSR	Gel electrophoresis	263	Grain filling rate	Time course phenotyping of grain filling. Predicted best parental combinations based on allelic combinations.	Liu et al. (2015)
Landrace and elite varieties	1568	SNP	SNP chip (HDRA)	700,000	Grain length	High GWAS resolution due to large population size and SNP density. Grain length QTL used as a specific example.	McCouch et al. (2016)

(Continued)



Table 1. (Continued).

Population	Sample size <sup>a</sup>	Marker type	Platform	No. of markers <sup>b</sup>	Traits	Highlights	Reference
Diverse rice accessions	258	SNP	WGS (from 3K RGP)	22,488	Grain quality (eating and cooking characteristics and protein content)	Nineteen QTLs affecting four traits were identified.	Wang et al. (2017b)
Diverse accessions	529	SNP	GBS	6,428,770	Metabolites	Profiling of 840 metabolites. Candidate genes identified.	Chen, Gao and Xie et al. (2014b)
Rice diversity panel	300	SNP	SNP chip (44 k)	36,900	Mineral concentration	Phenotyping grain concentrations of arsenic, copper, molybdenum and zinc in brown rice in five locations across 2 years. Candidate genes identified for uptake/transport of the minerals.	Norton et al. (2014)
Accessions and elite breeding lines (CMS-WA lines)	217	SNP	SNP chip (44 k)	43,394	Outcrossing in CMS lines	Traits related to hybrid seed production.	Guo et al. (2017)
Diverse accessions	328	SNP	SNP chip	32,175	Ozone tolerance	Physiological and agronomic traits measured in greenhouse trials. Candidate genes identified.	Ueda et al. (2015)
Accessions from East and Southeast Asia	540	SSR	Gel electrophoresis	262 (CG approach)	Panicle length	Led to map-based cloning of <i>LONG PANICLE 1 (LP1)</i> using both linkage and AM.	Liu et al. (2016)
Diverse accessions	315 (of 421 accessions in RDPT)	SNP	SNP chip (44 k)	44,100	Panicle traits	Phenotyping performed in two environments. Investigated QTL alleles in population.	Zhang, Ma and Chen et al. (2015)
Diverse accessions, majority temperate rice	153 and 93 (391)	SNP	GBS	57,179 and 37,827	Plant and grain morphology traits; root traits	Many QTLs were located in regions with previously detected QTLs. Candidate genes were identified.	Biscarini et al. (2016)
<i>Indica</i> and <i>Japonica</i> accessions	162 and 169 (331)	SNP	GBS	16,000	Root cone angle	QTLs were detected; 15 for <i>Indica</i> , 40 for <i>Japonica</i> .	Bettembourg et al. (2017)
<i>Japonica</i> accessions	167	SNP	SNP chip (HDRA) GBS (DART-Seq)	300,000 16,664	Root system architecture and biomass traits	Precision phenotyping using 'Rhizoscope phenotyping platform'.	Courtois et al. (2013)
Accessions from China and world core collection	529	SNP	Whole genome sequencing	4,358,600	Root traits	Many (225) loci identified to be associated with root trait and drought tolerance.	Li et al. (2017a)
Diverse accessions from 46 countries	478	SNP	Whole genome sequencing	6,361,920	Salinity tolerance	Many QTLs were located in regions with previously detected QTLs. Haplotypes and candidate genes identified.	Shi et al. (2017)
<i>Indica</i> accessions from ORYTAGE diversity panel	203	SNP	GBS	16,232	Spikelet sterility	One to five plausible candidate genes per QTL were identified.	Dingkuhn et al. (2017b)
Elite and landraces accessions: <i>Indica</i>	529	SNP	Whole genome sequencing	2,767,191 1,857,866 1,925,362 3,916,415	Nonphotochemical quenching (NPQ) capacity	Identified 33 significant association regions for NPQ capacity.	Wang et al. (2017a)
<i>Japonica</i> whole panel	220	SNP	SNP chip (6 k)	4,191	Salinity tolerance (reproductive) traits	SNPs selected based on abiotic and biotic stress response genes. Validation of <i>Saltol</i> QTL and many novel QTLs identified.	Kumar et al. (2015)
Accessions (India and Philippines)	380	SSR	Capillary system (SSR markers)	72	Straight head (physiological disorder)	Five markers were identified to be associated with straight head.	Li, Hu, Pan, Zhang and Wu (2017)
USDA rice core collection	453	SNP	SNP chip	5,291	Seed dormancy	Identified nine significant SNPs.	Lu et al. (2018)

<sup>a</sup>Actual sample size used. Initial panel size indicated in parentheses.<sup>b</sup>Actual number of SNPs used in analysis<sup>c</sup>PRAY: Phenomics of Rice Adaptation and Yield; MAGC: Multi-parent Advanced Generation Inter-Cross

## Genetic heterogeneity

This can occur when different genes controlling a single trait under investigation are included in the panel (i.e. the mechanism of genetic control differs within the population and causes confounding results because the ability to detect associations is lower). This can be understood by considering biology – plants originating from different areas may have evolved different mechanisms of adaptation. In rice, genetic heterogeneity is present (i.e. in *O. sativa*). *Indica* and *japonica* varietal groups have evolved through independent domestication process (Ikehashi, 2009). The consequence of genetic heterogeneity reduces the strength of the association between markers and traits.

## Will AM always ‘land’ on genes?

One interesting finding from several GWAS experiments in animal and plant species is that the SNPs at the QTL peak are not always within the functional gene. For example, in *Arabidopsis*, the highest associated SNPs were not the causal SNPs for the vernalization-response gene (*FRI*); this was attributed to complexities due to LD with specific alleles and population structure (Atwell et al., 2010). Similar observations have been reported in independent GWASs in rice (Huang et al., 2010; Yano et al., 2016) and may be confusing for researchers who are familiar with BPQM. In the pioneering work done by Yano et al. (2016), this issue of undetectable causal SNPs was addressed by using GWAS based on whole genome sequencing, followed by the screening of candidate genes based on the estimated effect of nucleotide polymorphisms.

## AM in rice

A range of AM studies have been reported for rice in the last decade (reviewed by Zhang et al., 2016). Associations with important agronomic and morphological traits have been the focus of most studies, while others analyzed yield-related, abiotic and biotic stress tolerance, reproductive and metabolic traits (Table 2). SSRs were initially used in some of the earliest reports of AM in rice (Agrama, Eizenga, & Yan, 2007; Borba et al., 2010). SNPs have now clearly become the marker of choice for GWAS in rice (McCouch et al., 2010). In this section, we briefly review some of the seminal GWAS papers in rice.

A landmark GWAS was reported by Huang et al. (2010), who investigated 14 agronomic traits (morphological, yield components, grain quality and agronomic or physiological) in a panel consisting of 517 landraces and detected many novel QTLs with relatively small effects. The panel was re-sequenced to identify a large number of SNPs, which permitted a high resolution for AM. Of the 3.6 million SNPs, 167,514 were located in coding regions of >25,000 annotated genes, which was important to investigate potential function effects of the detected SNP signals. Due to the larger sample size and genetic diversity, Huang and co-workers focused on the *indica* subset of the panel. Putative QTLs explained about 36% of the phenotypic variation and QTLs for six traits were located close to known genes (i.e. *OsC1*, *Rc*, *ALK*, *Waxy*, *qSW5*, *GS3*) that had been previously characterized.

Zhao et al. (2011) conducted a large-scale GWAS, exploring 34 traits (including morphological, agronomic, yield component, stress tolerance, seed/grain

**Table 2.** Comparison of BPQM and AM.

Criteria	Traditional QTL mapping (BPQM)	Population-based AM
Number of markers	Relatively few (100–200)	Many (100s–1000s)
Populations	Few parents or grandparents with many offspring (100–500)	Many individuals with unknown or mixed relationships. If pedigreed, family sizes are typically small relative to sampled population (>500)
QTL analysis	Easy. Less sophisticated tools minimize ghost QTL and increase mapping precision	Complex. Sophisticated tools reduce risk of false positives
Detection depends on	QTL segregation in offspring, and marker-trait linkage within-family(s)	QTL segregation in population, and marker-trait LD in mapping population
Mapping precision	Low (5–15 cM) QTL regions may contain many positional candidate genes	Can be high (10–1000 kb). Depends on population LD
Variation detected	Subset (only the portion segregating in sampled pedigrees)	Larger subset (theoretically all variation segregating in targeted regions of genome)
Extrapolation to other families or populations	Poor (other families not segregating QTL, changes in marker phase, etc.)	Good to excellent (although not all QTL will segregate in all population/pedigree subsamples)
Time spent	More research time in establishing a mapping population	Less research time in establishing an association
Meiosis cycles	Six to seven	Many
Detection power	Low resolution but high statistical power in detecting rare allele	High resolution but low statistical power in detecting rare allele
Population structure	Absent	May be present

morphology and quality) using a 44k SNP array platform across a panel of 413 diverse accessions including *indica*, *aus*, tropical and temperate *japonica*, and aromatic accessions. A large number of QTLs were detected for all traits, including many signals located near the locations of known major genes for the relevant trait. A major finding was that significant genetic heterogeneity was associated with subpopulation structure. In other words, different QTLs were observed when the subpopulations were analyzed separately compared to when the entire diversity panel was used. Another important finding was the influence of environmental effects. Flowering time was investigated in three locations and revealed season specificity, even for well-known major genes such as *Hd1*. Furthermore, some of the strongest signals were relatively far away from known candidate genes, which were attributed to ascertainment bias.

Famoso et al. (2011) combined GWAS with BPQM for investigating aluminum tolerance, using the same panel as Zhao et al. (2011). They discovered that a large component of tolerance was due to subpopulation structure, and several subpopulation-specific QTLs for this trait were detected. Importantly, a detailed haplotype and sequence analysis were performed around the candidate gene *Nrat1*, which indicated a large-effect QTL (explaining 40% of the phenotypic variation within the *aus* subpopulation) and three non-synonymous mutations within *Nrat1* that were predictive of aluminum sensitivity (Famoso et al., 2011).

Huang, Zhao and Wei et al. (2012) extended their previous GWAS results and focused on flowering time and yield-related traits using >1.3 million SNPs (>700,000 for the *indica* subset and >490,000 for the *japonica* subset). In this study, a larger panel of accession ( $n = 950$ , including 508 *indica* and 383 *japonica*) was used. Furthermore, detailed gene annotation, expression data and genetic variation were integrated to refine identification of candidate gene and potential causal polymorphisms for the target traits. Novel QTLs were detected for the traits investigated and the authors identified unknown loci associated with well-characterized traits such as flowering time, which were not identified in their previous report.

Recently, a high-resolution open-access resource for GWAS in rice has been made available to rice researchers (McCouch et al., 2016). The AM panel consists of 1568 diverse accessions including *indica*, *aus*, tropical and temperate *japonica*, and aromatic accessions, comprising two separate rice diversity panels (RDPs). The genotypic data were generated using a high-density rice array (HDRA) with 700,000 SNPs (i.e. approx. 1 SNP every 540 bp) and a suite of bioinformatics tools

including a GWAS viewer, allele finder and Genome Browser were developed to assist in data interpretation.

### From GWAS to molecular breeding

The large number of published QTLs or GWAS signals would imply that there are thousands of trait markers (i.e. highly predictive of target traits) available for breeders to use in selection programs. Most rice breeders, however, would argue that this is certainly not the case. Several authors have addressed the issue why only few reports of QTL discovery from GWAS are resulting in the actual use in MAS. Gupta et al. (2014) and Zhang et al. (2016) suggested that the high FDR was the most likely reason for a lack of application of GWAS signals in breeding.

We believe that the lack of validation of QTL-marker trait associations is a major factor explaining the lack of applied outcomes from GWASs. It is generally accepted that BPQM research needs to be validated in a range of genetic backgrounds and different environments (i.e. field testing) prior to deployment (Nicholas, 2006; Xu & Crouch, 2008). Due to high false discovery rates in AM approaches, there is an even greater need to verify marker-trait associations arising from AM experiments. In other words, the utility of a marker to accurately predict trait phenotype needs to be verified (referred to as 'marker-trait validation'). For example, Breseghello and Sorrells (2006) suggested that this could simply be performed by developing BPs and confirming this at the F2 or F3 stage. In our experience, readily available breeding populations (e.g. F2, BC, RILs or elite material) should be used to validate GWAS results in rice. Other methods involving gene expression of candidate genes can be used if the objective is to identify the causal gene for the trait, but this step is not required for breeding.

For large-scale routine screening in breeding populations, GWAS-based markers need to be converted to low-plex high-throughput marker assays (e.g. KASP, Taqman, Fluidigm) and considerable work is required to design and verify these assays. KASP assays currently range from about US\$0.10–US\$0.36/marker data point (Semagn, Babu, Hearne, & Olsen, 2014). Verification is typically performed for the markers using 48 to 96 samples, which contain known donors and recipients as well as a representative collection of the breeding germplasm. This tests whether the assay is robust in a variety of genetic backgrounds and whether it reliably scores favorable and unfavorable alleles in known backgrounds. Breeders should also be informed by molecular geneticists about details regarding the markers that have been developed (i.e. reliability in terms of selection accuracy for a trait, as well as sensitivity, specificity

and effectiveness in different genetic backgrounds). A versatile analysis tool called 'SNP Seek' (<http://snp-seek.irri.org>) was developed in conjunction with the 3000 rice genomes (Alexandrov et al., 2015) and provides a valuable resource for the validation of markers located in QTL peaks. Single highly significant SNPs or linked SNP clusters/haplotypes of significant association can be queried, analyzed for allele frequencies on a subpopulation specific basis and put into context with global variation within the region of interest, all of which can greatly aid in the development of low-plex assay.

### Phenotyping: critical considerations and new high-throughput phenotyping methods

The importance of accurate trait data for GWAS has been previously emphasized by several researchers (Myles et al., 2009; Rafalski, 2010; Zhu et al., 2008) and we reiterate the importance of phenotypic data for the success of any QTL mapping experiment. Proper experimental design and generation of high-quality phenotypic data are absolutely critical. In practice, accurate phenotyping of panels may be more complicated than molecular geneticists may realize. In rice, for example, ordinal scales are widely used for trait characterization because they are commonly used by rice breeders (for example IRRI standard evaluation system; IRRI, 2014). Ideally, scoring scales should not be based on subjective rating scales (Poland & Nelson, 2011). In order to properly characterize quantitative traits, reliable phenotyping methods based on quantitative measurements are required to accurately dissect genetic variation (Cobb, DeClerck, Greenberg, Clark, & McCouch, 2013). Broad- or narrow-sense heritability should be calculated for each trait in order to understand the proportion of genetic variance that has been explained by the detected QTLs.

Furthermore, GWAS experiments often do not consider genotype by environment (G x E) interactions. Breeders are well aware that G x E interactions may complicate trait phenotyping. Variation usually exists even within and between controlled environment greenhouse trials and certainly occurs between years, seasons and environments in field trials (Atlin, Kleinknecht, Singh, & Piepho, 2011). Further improvement of phenotyping can be achieved under controlled environment conditions, which is useful for reliable and precise phenotyping of plant responses to abiotic stresses (Negin & Moshelion, 2017). However, transferability of controlled environment observations to actual field conditions is difficult for some traits such as drought (Passioura, 2012). In these cases, completely new phenotyping methods are required.

Plant phenomics for accurate high-throughput phenotyping has undergone rapid development in the last

decade. Both ground-based proximal sensing and aerial remote sensing systems are now routinely used for field phenomics. Rapid, GPS-guided, high-throughput semi- or fully automated phenomics systems have been developed for quantitative measurement of above-ground biomass, stem and canopy attributes, photosynthesis and pigment content (Simko, Hayes, & Furbank, 2016), flowering (Guo, Fukatsu, & Ninomiya, 2015), abiotic stress responses (Cobb et al., 2013), pathogenesis (Mahlein, 2016), leaf traits (Yang et al., 2015) and agronomic traits (Duan, Chapman, Guo, & Zheng, 2017). Advancements in aerial vehicle engineering, automation, sensor-based imaging, software capability, data storage and analytical capacity underpin this phenomics revolution. Ground-based platforms have the advantage of generating high-resolution data, but they cannot screen large populations simultaneously, which is critical for many applications.

Remote sensing platforms such as unmanned aerial vehicles (UAVs) and satellites generate data with different spatial resolution and spectral coverage. Manned aircrafts are also used for aerial remote sensing. Most of the commercially available phenomics systems can be deployed directly for many crops though crop-specific features may demand considerable system optimization. UAV-based aerial phenomics platforms are the most commonly used systems for germplasm development and crop research (Watanabe et al., 2017). A range of RGB (red, green, blue), multi-spectral, hyperspectral, thermal-fluorescence, 3D, LIDAR and SONAR sensors are available for plant phenotyping, and many of them can be integrated with UAVs depending on their payload capacity. Rapid progress in automated sensor-based imaging and image-processing software solutions became the cornerstones of field phenomics (Cobb et al., 2013).

The success of high-throughput phenomics depends on the accuracy and precision of data collected (Cobb et al., 2013). Accuracy can be improved with the use of known standards, while the stability of the object sensing system determines the precision, which is important to reduce error variance. Unfortunately increasing the throughput may compromise accuracy and precision and this necessitates crop-specific system optimization. Despite the best efforts to maximize accuracy and precision of primary sensory data, which are collected as images over a relatively long period of time, they carry a significant level of genetic, environmental and experimental noise. Developing algorithms for converting the primary sensory data to useful phenotypic data for performance prediction of breeding lines thus can be quite demanding.

Measuring phenotype during the entire crop cycle requires the collection of a very large volume of data.

This makes storage, processing and management of phenomics data challenging. Successful high-throughput field phenotyping is contingent on several factors (Negin and Moshelion, 2017). These include the type of the phenomics platform used, sensor systems with the required spectral coverage, computational and analytical capacity to convert sensory data to useful phenotypic data and database management capability. Identifying the appropriate phenotypic parameters that can be quantified accurately, rapidly and cost-effectively using a carefully designed experiment established with an appropriate genetic population is also vital.

### Closer integration with breeding is needed

Over the last decade, GWAS results have been poorly utilized in breeding programs which have created an 'application gap' (Collard, Raghavan, & Islam, 2017). Molecular geneticists, who typically perform most of the gene discovery work and who are main users of GWAS, are often not well aligned with breeding programs. The objectives of molecular geneticists are to elucidate and begin to unravel the physiological basis of phenotypic variation at the gene or molecular level. While these are valid and noble endeavors, most of the knowledge obtained from these efforts is not needed in actual breeding programs.

In order to align gene discovery programs utilizing GWAS and breeding programs, discovery efforts need to be more closely integrated with actual breeding activities. Traits that gene discovery groups work on should be the same as those used to set breeding priorities. Reviewing the literature, this is seldom the case (examples in Table 1) and gene discovery work and breeding are rarely connected in the public sector. Often morphological or physiological traits are selected for gene discovery because they are easily measured in controlled environments with high accuracy, rather than their relevance for breeding. These secondary traits may be utilized, provided they are more heritable and highly correlated with the trait of interest like yield. Some notable exceptions were studies to specifically identify loci associated with field blast resistance (Raboin et al., 2016; Zhu et al., 2016). The challenge is that high priority complex traits for breeders (e.g. yield) often defy reductionist descriptions using component traits due to complex interactions involving genetics, physiology and the physical environment. When attempts are made to measure complex traits, often the phenotyping methods employed by molecular geneticists are limited to controlled environments like growth chambers and greenhouses, where the correlation with field conditions is either low or unknown. Loci

discovered in this way may be highly reliable for a specific assay (e.g. greenhouse-based disease or abiotic stress tolerance test) but there would be little value using markers developed for these loci if the assay was not an effective predictor of field performance. In private sector breeding organizations, gene discovery and breeding teams are structured in a unified product development pipeline, to efficiently evaluate germplasm and identify new alleles of large effect that can be easily deployed in a breeding program (Eathington, Crosbie, Edwards, Reiter, & Bull, 2007).

In rice and other cereals, however, there are traits (notably disease resistance, abiotic stress tolerance and quality traits) for which single genes of large effect suitable to MAS are well known (Collard & Mackill, 2008; Das, Patra, & Baek, 2017). Any breeding program interested in doing MAS for these traits will need to prioritize the loci of interest. Particularly for traits where well-developed trait markers still need to be developed, GWAS using breeding germplasm is suited to inform marker development priorities by quickly identifying which loci are both polymorphic and at high frequency in the breeding germplasm and thus would have high utility rates for MAS. In this context, GWAS results can provide useful information by identifying loci (i.e. due to the lack of a QTL peak where one might be expected) that are already fixed in the breeding program. If fixed for the favorable allele due to selection, the expenditure of marker development resources can be avoided for a locus that has no utility within the breeding program. For example, one important finding by Begum et al. (2015), who used a panel comprised of elite breeding lines, was that major genes (e.g. for flowering time) were often not segregating in elite *indica* rice breeding germplasm, presumably because fixation had already occurred. This information sheds light on the genetic basis of flowering time among the breeding lines they are using, and can thus help direct future strategy for further modifications of flowering time within the breeding program.

Likewise if a new disease or specific trait suddenly becomes important to a breeding program, GWAS cannot only help to determine if the existing breeding lines contain loci easily amenable to MAS, but also to screen panels of exotic germplasm in hope of finding large effect genes/QTLs (provided they are present at reasonable allele frequencies). However, the utility of GWAS to a pre-breeding program is of limited use when the genetic variants of large effect for a trait are already well known. In these cases, gene discovery efforts are focused on searching for rare alleles of large effect which GWAS is unable to detect. In these situations, a more effective strategy is to screen a large number of

exotic lines to identify donors and create BPs with each donor. Such populations can then be targeted for fine mapping and marker development such that NILs in elite backgrounds can be deployed in the breeding program using marker-assisted forward breeding.

There has been some research using GWAS results as covariates for inclusion in genomic selection (GS) models in a breeding program and potentially increase prediction accuracies (Spindel et al., 2016). GS is a new molecular breeding method that was originally developed in animals (Meuwissen, Hayes, & Goddard, 2001) and has recently been implemented in crops (Heffner, Sorrells, & Jannink, 2009). The method is based on using high-density marker data and advanced computational methods to calculate predict trait performance (in contrast to MAS, which is based on using a single or small number of target alleles). This area of research urgently needs more attention and could represent a critical application of GWAS to breeding programs.

## Conclusion and future perspective

AM has become an increasingly popular approach in crop genetics to understand the architecture of quantitative traits and to identify QTLs controlling important traits. This method complements traditional linkage-based mapping methods based on BPs. By combining GWAS with functional genomics, it is inevitable that the growing use of AM in rice will lead to the identification of new QTLs and candidate genes in the future (Yano et al., 2016). It is likely that future progress using AM approaches will be based on two major factors: (1) greater integration with functional analysis or gene annotation data (i.e. 'post-GWAS research', Zhang, Bailey, & Lupien, 2014); and (2) improvements in statistical and computational methods (e.g. methods using Bayesian methods, haplotypes and SNP imputation). Further advances in rice genome re-sequencing (3,000 rice genomes project, 2014) and developments in mammalian and other model species will surely have a major impact on GWAS (Tak & Farnham, 2015; Visscher et al., 2017). Current research in AM includes methods to combine GWAS results from multiple studies (e.g. meta-analysis), strategies to account for G x E and tests for epistatic interactions. User-friendly analytical tools and genomics resources will always need to be improved and further developed.

In the meantime, researchers engaged in applied AM activities need to focus on the fundamental factors such as population size, marker density and accounting for population structure. Accurate and relevant phenotypic data will always remain to be one of the critical factors for success in

terms of integrating results with breeding. Ultimately, a holistic and integrated approach (i.e. 'systems genetics') for GWAS would be the ideal configuration. Currently, integration and relevance to actual rice breeding programs remain a great challenge. Further application of GWAS research specifically using breeding panels comprising elite breeding germplasm needs further research. Collectively this could facilitate the outstanding recent achievements in rice genomics to be utilized in the development of new and improved rice varieties using molecular breeding.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

Holden Verdeprado  <http://orcid.org/0000-0003-4232-5172>  
 Tobias Kretzschmar  <http://orcid.org/0000-0002-8227-0746>  
 Joshua N. Cobb  <http://orcid.org/0000-0002-1732-2378>  
 Bertrand C.Y. Collard  <http://orcid.org/0000-0003-2566-4148>

## References

- 3,000 rice genomes project. (2014). The 3,000 rice genomes project. *GigaScience*, 3(1), 7.
- Agrama, H. A., Eizenga, G. C., & Yan, W. (2007). Association mapping of yield and its components in rice cultivars. *Molecular Breeding*, 19(4), 341–356.
- Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R. R., ... McNally, K. L. (2015). SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Research*, 43(D1), D1023–D1027.
- Atlin, G. N., Kleinknecht, K., Singh, G. P., & Piepho, H. P. (2011). Managing genotype x environment interaction in plant breeding programs: A selection theory approach. *Journal of the Indian Society of Agricultural Statistics*, 65, 237–247.
- Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., ... Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, 465(7298), 627–631.
- Balding, D. J. (2006, 2006 Oct). A tutorial on statistical methods for population association studies. *Nat Rev Genet*, 7(10), 781–791.
- Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., ... Leung, H. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice*, 6(1), 1–15.
- Begum, H., Spindel, J. E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., ... McCouch, S. R. (2015). Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS ONE*, 10(3), 1–19.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to

- multiple testing. *Journal of the Royal Statistical Society*, 57 (1), 289–300.
- Bessho-Uehara, K., Furuta, T., Masuda, K., Yamada, S., Angeles-Shim, R. B., Ashikari, M., & Takashi, T. (2017). Construction of rice chromosome segment substitution lines harboring *Oryza barthii* genome and evaluation of yield-related traits. *Breeding Science*, 67(4), 408–415.
- Bettembourg, M., Dardou, A., Audebert, A., Thomas, E., Frouin, J., Guiderdoni, E., ... Courtois, B. (2017). Genome-wide association mapping for root cone angle in rice. *Rice*, 10 (1). doi:10.1186/s12284-017-0184-z
- Biscarini, F., Cozzi, P., Casella, L., Riccardi, P., Vattari, A., Orasen, G., ... Greco, R. (2016). Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. *PLoS ONE*, 11(5), 1–28.
- Borba, T. C. D. O., Brondani, R. P. V., Breseghello, F., Coelho, A. S. G., Mendonça, J. A., Rangel, P. H. N., & Brondani, C. (2010). Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). *Genetics and Molecular Biology*, 33(3), 515–524.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635.
- Breseghello, F., & Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, 172(2), 1165–1177.
- Buntjer, J. B., Sørensen, A. P., & Peleman, J. D. (2005). Haplotype diversity: The link between statistical and biological association. *Trends in Plant Science*, 10(10), 466–471.
- Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS computational. Biology*, 8, 12.
- Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., ... Zhang, Q. (2014a). A high-density SNP genotyping array for rice biology and molecular breeding. *Molecular Plant*, 7(3), 541–553.
- Chen, W., Gao, Y., Xie, W., et al (2014b). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature Genetics*, 46, 714–721.
- Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, 126(4), 867–887.
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(12), 169–196.
- Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1491), 557–572.
- Collard, B. C. Y., Raghavan, C., & Islam, M. R. (2017). Critical considerations for molecular breeding in rice. *Oryza*, 53(4), 346–353.
- Courtois, B., Audebert, A., Dardou, A., Roques, S., Ghneim-Herrera, T., Droc, G., ... Dingkuhn, M. (2013). Genome-wide association mapping of root traits in a japonica rice panel. *PLoS ONE*, 8(11), 1–18.
- Das, G., Patra, J. K., & Baek, K. (2017). Insight into MAS: A molecular tool for development of stress resistant and quality of rice through gene stacking. *Frontiers in Plant Science*, 8(6), 1–9.
- Dilla-Ermita, C. J., Tandayu, E., Juanillas, V. M., Detras, J., Lozada, D. N., Dwiyantri, M. S., ... Kretschmar, T. (2017). Genome-wide association analysis tracks bacterial leaf blight resistance loci in rice diverse germplasm. *Rice*, 10 (1). doi:10.1186/s12284-017-0147-4
- Dingkuhn, M., Pasco, R., Pasuquin, J. M., Damo, J., Soulié, J. C., Raboin, L. M., ... Kretschmar, T. (2017a). Crop-model assisted phenomics and genome-wide association study for climate adaptation of indica rice. 1. Phenology. *Journal of Experimental Botany*, 68(15), 4369–4388.
- Dingkuhn, M., Pasco, R., Pasuquin, J. M., Damo, J., Soulié, J. C., Raboin, L. M., ... Kretschmar, T. (2017b). Crop-model assisted phenomics and genome-wide association study for climate adaptation of indica rice. 2. Thermal stress and spikelet sterility. *Journal of Experimental Botany*, 68(15), 4389–4406.
- Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics*, 3(1), 43–52.
- Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., ... Liang, C. (2017). Sequencing and de novo assembly of a near complete indica rice genome. *Nature Communications*, 8(2), 1–12.
- Duan, T., Chapman, S. C., Guo, Y., & Zheng, B. (2017). Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Research*, 210, 71–80.
- Dwivedi, S. L., Crouch, J. H., Mackill, D. J., Xu, Y., Blair, M. W., Ragot, M., ... Ortiz, R. (2007). The molecularization of public sector crop breeding: Progress, problems, and prospects. *Advances in Agronomy*, 95(7), 163–318.
- Eathington, S. R., Crosbie, T. M., Edwards, M. D., Reiter, R. S., & Bull, J. K. 2007. Molecular markers in a commercial breeding program. In *Crop Science*, 47(S3), S154–S163. doi: 10.2135/cropsci2007.04.0015IPBS
- Edwards, D., Batley, J., & Snowdon, R. J. (2013). Accessing complex crop genomes with next-generation sequencing. *Theoretical and Applied Genetics*, 126(1), 1–11.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), 1–10.
- Ersoz, E. S., Yu, J., & Buckler, E. S. (2009). Applications of linkage disequilibrium and association mapping in maize. In A. L. Kriz & B. A. Larkins (eds), *Molecular genetic approaches to maize improvement. Biotechnology in agriculture and forestry*, 63, 173–195. Berlin: Springer.
- Ersoz, E. S., Yu, J., & Buckler, E. S. (2009). Applications of linkage disequilibrium and association mapping in maize applications of linkage disequilibrium and association mapping in maize. In *Molecular genetic approaches to maize improvement* (pp. 173–195). Heidelberg: Springer-Verlag Berlin. doi: 10.1007/978-3-540-68922-5
- Famoso, A. N., Zhao, K., Clark, R. T., Tung, C., Wright, M. H., Bustamante, C., ... McCouch, S. R. (2011). Genetic

- architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genetics*, 7(8), e1002221.
- Flint-Garcia, S. A., Thornsberry, J. M., & Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, 54(1), 357–374.
- Fragoso, C. A., Moreno, M., Wang, Z., Heffelfinger, C., Arbelaez, L. J., Aguirre, J. A., ... Lorieux, M. (2017). Genetic architecture of a rice nested association mapping population. *G3-Genes Genomes Genetics*, 7(6), 1913–1926.
- Garris, A. J., McCouch, S. R., & Kresovich, S. (2003). Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics*, 165, 759–769.
- Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., ... Briggs, S. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 296(5565), 92–100.
- Gregorio, G. B., Islam, M. R., Vergara, G. V., & Thirumeni, S. (2013). Recent advances in rice science to design salinity and other abiotic stress tolerant rice varieties. *Sabara Journal of Breeding and Genetics*, 45(1), 31–41.
- Guo, L., Qiu, F., Gandhi, H., Kadaru, S., De Asis, E. J., Zhuang, J., & Xie, F. (2017). Genome-wide association study of out-crossing in cytoplasmic male sterile lines of rice. *Scientific Reports*, 7(1), 1–10.
- Guo, W., Fukatsu, T., & Ninomiya, S. (2015). Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images. *Plant Methods*, 11(1). doi:10.1186/s13007-015-0047-9
- Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2014). Association mapping in crop plants: Opportunities and challenges. *Advances in Genetics*, 85, 109–147.
- Hamblin, M. T., Buckler, E. S., & Jannink, J. L. (2011). Population genetics of genomics-based crop improvement methods. *Trends in genetics*, 27, 98–106. doi:10.1016/j.tig.2010.12.003
- He, J., Zhao, X., Laroche, A., Lu, Z., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, 5(9), 1–8.
- Heffner, E. L., Sorrells, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1–12.
- Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology*, 10(2), 156–161.
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., ... Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42(11), 961–967.
- Huang, X., Zhao, Y., Wei, X., et al (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics*, 42(1), 32–39.
- Ikehashi, H. (2009). Why are there indica type and japonica type in rice? History of the studies and a view for origin of two types. *Rice Science*, 16, 1–13.
- Ingvarsson, P. K., & Street, N. R. (2011). Association genetics of complex traits in plants. *New Phytologist*, 189(4), 909–922.
- International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature*, 436(7052), 793–800.
- IRRI. (2014). *Standard evaluation system for rice*. Manila, Philippines: International Rice Research Institute.
- Kang, H. M., Sul, J. H., Service, S. K., et al (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348–354.
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, 178(3), 1709–1723.
- Kikuchi, S., Bheemanahalli, R., Jagadish, K. S. V., Kumagai, E., Masuya, Y., Kuroda, E., ... Shimono, H. (2017). Genome-wide association mapping for phenotypic plasticity in rice. *Plant, Cell & Environment*, 40(8), 1565–1575.
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, 9(1), 29–38.
- Kumar, V., Singh, A., Mithra, S. V. A., Krishnamurthy, S. L., Parida, S. K., Jain, S., ... Mohapatra, T. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Research*, 22(2), 133–145.
- Li, G., & Zhu, H. (2013). Genetic studies : The linear mixed models in genome-wide association studies. *The Open Bioinformatics Journal*, 7(Suppl 1, M2), 27–33.
- Li, X., Guo, Z., Lv, Y., Cen, X., Ding, X., Wu, H., ... Xiong, L. (2017a). Genetic control of the root system in rice under normal and drought stress conditions by genome-wide association study. *PLoS Genetics*, 13(7), 1–24.
- Li, X., Hu, B., Pan, X., Zhang, N., & Wu, D. (2017). Association analysis of arsenic-induced straighthead in rice (*Oryza sativa* L.) based on the selected population with a modified model. *BioMed Research International*, (2017b), 1–6.
- Lipka, A. E., Kandianis, C. B., Hudson, M. E., Yu, J., Drnevich, J., Bradbury, P. J., & Gore, M. A. (2015). From association to prediction: Statistical methods for the dissection and selection of complex traits in plants. *Current Opinion in Plant Biology*, 24, 110–118.
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., ... Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28(18), 2397–2399.
- Liu, E., Liu, X., Zeng, S., Zhao, K., Zhu, C., Liu, Y., ... Hong, D. (2015). Time-course association mapping of the grain-filling rate in rice (*Oryza sativa* L.). *PLoS ONE*, 10(3), 1–17.
- Liu, E., Liu, Y., Wu, G., Zeng, S., Tran Thi, T. G., Liang, L., ... Hong, D. (2016). Identification of a candidate gene for panicle length in rice (*Oryza sativa* L.) via association and linkage analysis. *Frontiers in Plant Science*, 7(5), 1–13.
- Lu, Q., Niu, X., Zhang, M., Wang, C., Xu, Q., Feng, Y., ... Wei, X. (2018). Genome-wide association study of seed dormancy and the genomic consequences of improvement footprints in rice (*Oryza sativa* L.). *Frontiers in Plant Science*, 8, 1–16.
- Lu, Q., Zhang, M., Niu, X., Wang, S., Xu, Q., Feng, Y., ... Wei, X. (2015). Genetic variation and association mapping for 12 agronomic traits in indica rice. *BMC Genomics*, 16(1), 1–17.
- Ma, X., Feng, F., Wei, H., Mei, H., Xu, K., Chen, S., ... Luo, L. (2016). Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. *Frontiers in Plant Science*, 7(11), 1–13.
- Mackay, I., & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, 12(2), 57–63.
- Mackay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: Challenges and prospects. *Nature Reviews Genetics*, 10(8), 565–577.
- Mahlein, A.-K. (2016). Present and future trends in plant disease detection. *Plant Disease*, 100(2), 1–11.



- Mammadov, J., Aggarwal, R., Buyyarapu, R., & Kumpatla, S. (2012). International Journal of Plant Genomics. *SNP Markers and Their Impact on Plant Breeding*, Article ID 728398. doi:10.1155/2012/728398
- Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S., & Purugganan, M. D. (2007). The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, 177(4), 2223–2232.
- McCouch, S. R., Wright, M. H., Tung, C., Maron, L. G., McNally, K. L., Fitzgerald, M., ... Mezey, J. (2016). Open access resources for genome-wide association mapping in rice. *Nature Communication* 2016, 7, 10532.
- McCouch, S. R., Zhao, K., Wright, M., Tung, C., Ebana, K., Thomson, M., ... Bustamante, C. (2010). Development of genome-wide SNP assays for rice. *Breeding Science*, 60(5), 524–535.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829. 11290733.
- Misra, G., Badoni, S., Anacleto, R., Graner, A., Alexandrov, N., & Sreenivasulu, N. (2017). Whole genome sequencing-based association study to unravel genetic architecture of cooked grain width and length traits in rice. *Scientific Reports*, 7(1), 1–16.
- Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R., & Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Genome Research*, 3, 87–103.
- Morrell, P. L., Buckler, E. S., & Ross-Ibarra, J. (2011). Crop genomics: Advances and applications. *Nature Reviews Genetics*, 13:85–96. doi:10.1038/nrg3097
- Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., & Buckler, E. S. (2009). Association mapping: Critical considerations shift from genotyping to experimental design. *The Plant Cell*, 21(8), 2194–2202.
- Negin, B., & Moshelion, M. (2017). The advantages of functional phenotyping in pre-field screening for drought-tolerant crops. *Functional Plant Biology*, 44(1), 107–118.
- Nicholas, F. W. (2006). Discovery, validation and delivery of DNA markers. *Australian Journal of Experimental Agriculture*, 46(2), 155–158.
- Nordborg, M., & Tavaré, S. (2002). Linkage disequilibrium: What history has to tell us. *Trends in Genetics*, 18(2), 83–90.
- Norton, G. J., Douglas, A., Lahner, B., Yakubova, E., Lou, G. M., Pinson, S. R. M., ... Price, A. H. (2014). Genome wide association mapping of grain arsenic, copper, molybdenum and zinc in rice (*Oryza sativa* L.) grown at four international field sites. *PLoS ONE*, 9(2). doi:10.1371/journal.pone.0089685
- Ogawa, S., Valencia, M. O., Lorieux, M., Arbelaez, J. D., McCouch, S., Ishitani, M., & Selvaraj, M. G. (2016). Identification of QTLs associated with agronomic performance under nitrogen-deficient conditions using chromosome segment substitution lines of a wild rice relative, *Oryza rufipogon*. *Acta Physiologiae Plantarum*, 38, 4.
- Passioura, J. B. (2012). Phenotyping for drought tolerance in grain crops: When is it useful to breeders? *Functional Plant Biology*, 39(11), 851–859.
- Phung, N. T. P., Mai, C. D., Mournet, P., Frouin, J., Droc, G., Ta, N. K., ... Courtois, B. (2014). Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. *BMC Plant Biology*, 14(1), 371; article.
- Platt, A., Vilhjálmsson, B. J., & Nordborg, M. (2010). Conditions under which genome-wide association studies will be positively misleading. *Genetics*, 186, 1045–1052.
- Poland, J. A., & Nelson, R. J. (2011). In the eye of the beholder: The effect of rater variability and different rating scales on QTL mapping. *Phytopathology*, 101(2), 290–298.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Raboin, L.-M., Ballini, E., Tharreau, D., Ramanantsoanirina, A., Frouin, J., Courtois, B., & Ahmadi, N. (2016). Association mapping of resistance to rice blast in upland field conditions. *Rice*, 9(1), 59.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol*, 5, 94–100.
- Rafalski, J. A. (2010). Association genetics in crop improvement. *Current Opinion in Plant Biology*, 13(2), 174–180.
- Sales, E., Viruel, J., Domingo, C., & Marqués, L. (2017). Genome wide association analysis of cold tolerance at germination in temperate japonica rice (*Oryza sativa* L.) varieties. *PLoS ONE*, 12(8), 1–18.
- Schatz, M. C., Maron, L. G., Stein, J. C., Wences, A. H., Gurtowski, J., Biggers, E., ... McCombie, W. R. (2014). Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biology*, 15(11), 506.
- Schläppi, M. R., Jackson, A. K., Eizenga, G. C., Wang, A., Chu, C., Shi, Y., ... Boykin, D. L. (2017). assessment of five chilling tolerance traits and GWAS mapping in rice using the USDA mini-core collection. *Frontiers in Plant Science*, 8(6), 1–13.
- Segura, V., Vilhjálmsson, B., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed model approach for genome-wide association studies in structured populations. *Nature Genetics*, 44(7), 825–830.
- Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding*, 33, 1–14.
- Semagn, K., Bjornstad, A., & Xu, Y. (2010). The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology*, 13(5), 1–45.
- Shakiba, E., Edwards, J. D., Jodari, F., Duke, S. E., Baldo, A. M., Korniliev, P., ... Eizenga, G. C. (2017). Genetic architecture of cold tolerance in rice (*Oryza sativa*) determined through high resolution genome-wide analysis. *PLoS ONE*, 12(3), 1–22.
- Shi, Y., Gao, L., Wu, Z., Zhang, X., Wang, M., Zhang, C., ... Li, Z. (2017). Genome-wide association study of salt tolerance at the seed germination stage in rice. *BMC Plant Biology*, 17(1), 1–11.
- Simko, I., Hayes, R. J., & Furbank, R. T. (2016). Non-destructive phenotyping of lettuce plants in early stages of development with optical sensors. *Frontiers in Plant Science*, 7(12), 1–19.
- Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V., Singh, B., ... Singh, N. K. (2015a). Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Scientific Reports*, 5, 1–9.

- Singh, R., Singh, Y., Xalaxo, S., Verulkar, S., Yadav, N., Singh, S., ... Singh, N. K. (2015b). From QTL to variety-harnessing the benefits of QTLs for drought, flood and salt tolerance in mega rice varieties of India through a multi-institutional network. *Plant Science*, 242, 278–287.
- Slatkin, M. (2008). Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6), 477–485.
- Soto-Cerda, B. J., & Cloutier, S. (2012). Association mapping in plant genomes. In M. Çalişkan (ed), *Genetic Diversity in Plants* (pp. 29–54). Rijeka: InTech.
- Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J. L., & McCouch, S. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity*, 116(4), 395–408.
- Swamy, B. P. M., Shamsudin, N. A. A., Rahman, S. N. A., Mauleon, R., Ratnam, W., Sta., C. M. T., & Kumar, A. (2017). Association mapping of yield and yield-related traits under reproductive stage drought stress in rice (*Oryza sativa* L.). *Rice*, 10(1), 21–33.
- Tak, Y. G., & Farnham, P. J. (2015). Making sense of GWAS: Using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics and Chromatin*, 8(1), 1–18.
- Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., ... Zhang, Z. (2016). GAPIT version 2: An enhanced integrated tool for genomic association and prediction. *Plant Genome*. doi:10.3835/plantgenome2015.11.0120
- Thomson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breeding and Biotechnology*, 2(3), 195–212.
- Ueda, Y., Frimpong, F., Qi, Y., Matthus, E., Wu, L., Höller, S., ... Frei, M. (2015). Genetic dissection of ozone tolerance in rice (*Oryza sativa* L.) by a genome-wide association study. *Journal of Experimental Botany*, 66(1), 293–306.
- Van, K., Rastogi, K., Kim, K. H., & Lee, S. H. (2013). Next-generation sequencing technology for crop improvement. *SABRAO Journal of Breeding and Genetics*, 45(1), 84–99.
- Varshney, R. K., & Dubey, A. (2009). Novel genomic tools and modern genetic and breeding approaches for crop improvement. *Journal of Plant Biochemistry and Biotechnology*, 18(2), 127–138.
- Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Trends in Plant Science*, 10(12), 621–630.
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics*, 101(1), 5–22.
- Volante, A., Desiderio, F., Tondelli, A., Perrini, R., Orasen, G., Biselli, C., ... Valè, G. (2017). Genome-wide analysis of japonica rice performance under limited water and permanent flooding conditions. *Frontiers in Plant Science*, 8(10), 1–18.
- Wang, Q., Tian, F., Pan, Y., Buckler, E. S., & Zhang, Z. (2014). A SUPER powerful method for genome wide association study. *PLoS ONE*, 9(9). doi:10.1371/journal.pone.0107684
- Wang, Q., Xie, W., Xing, H., Yan, J., Meng, X., Li, X., ... Wang, G. (2015). Genetic architecture of natural variation in rice chlorophyll content revealed by a genome-wide association study. *Molecular Plant*, 8(6), 946–957.
- Wang, Q., Zhao, H., Jiang, J., Xu, J., Xie, W., Fu, X., ... Wang, G. (2017a). Genetic architecture of natural variation in rice nonphotochemical quenching capacity revealed by genome-wide association study. *Frontiers in Plant Science*, 8(10). doi:10.3389/fpls.2017.01773
- Wang, X., Pang, Y., Zhang, J., Wu, Z., Chen, K., Ali, J., ... Li, Z. (2017b). Genome-wide and gene-based association mapping for rice eating and cooking characteristics and protein content. *Scientific Reports*, 7(1), 17203.
- Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., ... Iwata, H. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Frontiers in Plant Science*, 8, 421.
- Xu, Y., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Science*, 48(2), 391–407.
- Yamamoto, T., Nagasaki, H., Yonemaru, J. I., Ebana, K., Nakajima, M., Shibaya, T., & Yano, M. (2010). Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics*, 11(1). doi:10.1186/1471-2164-11-267
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., ... Xiong, L. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications*, 5, 5087.
- Yang, W., Guo, Z., Huang, C., Wang, K., Jiang, N., Feng, H., ... Xiong, L. (2015). Genome-wide association study of rice (*Oryza sativa* L.) leaf traits with a high-throughput leaf scorer. *Journal of Experimental Botany*, 66(18), 5605–5615.
- Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., ... Matsuoka, M. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genetics*, 48(8), 927–934.
- Yano, M. (2001). Genetic and molecular dissection of naturally occurring variation. *Current Opinion in Plant Biology*, 4(2), 130–135.
- Yonemaru, J., Choi, S. H., Sakai, H., Ando, T., Shomura, A., Yano, M., ... Fukuoka, S. (2015). Genome-wide indel markers shared by diverse Asian rice cultivars compared to Japanese rice cultivar “Koshihikari.”. *Breeding Science*, 65(3), 249–256.
- Yonemaru, J. I., Ebana, K., & Yano, M. (2014). HapRice, an SNP haplotype database and a web tool for rice. *Plant and Cell Physiology*, 55(1), 1–12.
- Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., ... Wang, J. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 296, 79–92.
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., ... Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2), 203–208.
- Zhang, F., Wu, Z. C., Wang, M. M., Dingkuhn, M., Xu, J. L., Zhou, Y. L., & Li, Z. K. (2017a). Genome-wide association analysis identifies resistance loci for bacterial blight in a diverse collection of indica rice germplasm. *PLoS ONE*, 12(3), 1–17.
- Zhang, J., Chen, K., Pang, Y., Naveed, S. A., Zhao, X., Wang, X., ... Xu, J. (2017b). QTL mapping and candidate gene analysis of ferrous iron and zinc toxicity tolerance at seedling stage in rice by genome-wide association study. *BMC Genomics*, 18(1), 1–15.

- Zhang, M., Lu, Q., Wu, W., Niu, X., Wang, C., Feng, Y., ... Wei, X. (2017c). Association mapping reveals novel genetic loci contributing to flooding tolerance during germination in indica rice. *Front. Plant Sci.*, 8, 678.
- Zhang, M., Ye, J., Xu, Q., Feng, Y., Yuan, X., Yu, H., ... Yang, Y. (2018). Genome-wide association study of cold tolerance of Chinese indica rice varieties at the bud burst stage. *Plant Cell Reports*, 37, 529.
- Zhang, P., Zhong, K., Shahid, M. Q., & Tong, H. (2016). Association analysis in rice: From application to utilization. *Frontiers in Plant Science*, 7, 1202.
- Zhang, X., Bailey, S. D., & Lupien, M. (2014). Laying a solid foundation for Manhattan - "setting the functional basis for the post-GWAS era." *Trends in Genetics*, 30(4), 140-149.
- Zhang, Y., Ma, Y., Chen, Z., et al (2015). Genome-wide association studies reveal new genetic targets for five panicle traits of international rice varieties. *Rice Sci*, 22, 217-226. doi: 10.1016/j.rsci.2015.07.001
- Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., ... Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4), 355-360.
- Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., ... Nordborg, M. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS Genetics*, 3(1), 0071-0082.
- Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., ... McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications*, 2(467). doi:10.1038/ncomms1467
- Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome Journal*, 1(1), 5.
- Zhu, D., Kang, H., Li, Z., Liu, M., Zhu, X., Wang, Y., ... Wang, G. L. (2016). A genome-wide association study of field resistance to *Magnaporthe oryzae* in rice. *Rice*, 9(1). doi:10.1186/s12284-016-0116-3