Published July 12, 2018

Genome-Wide Association Study Using Historical Breeding Populations Discovers Genomic Regions Involved in High-Quality Rice

Gastón Quero, Lucía Gutiérrez,* Eliana Monteverde, Pedro Blanco, Fernando Pérez de Vida, Juan Rosas, Schubert Fernández, Silvia Garaycochea, Susan McCouch, Natalia Berberian, Sebastián Simondi, and Victoria Bonnecarrère

G. Quero, Dep. of Plant Biology, College of Agriculture, Univ. de la República, Garzón 809, Montevideo, Uruguay; G. Quero, J. Rosas, S. Garaycochea, V. Bonnecarrère, Biotechnology Unit, Experimental Station Wilson Ferreira Aldunate, Instituto Nacional de Investigación Agropecuaria (INIA, National Institute of Agriculture Research), Ruta 48, Km 10, Rincón del Colorado, Canelones 90200, Uruguay; L. Gutiérrez, Dep. of Agronomy, Univ. of Wisconsin-Madison, 1575 Linden Dr., Madison, WI 53706; L. Gutiérrez, N. Berberian, Dep. of Statistics, College of Agriculture, Univ. de la República, Garzón 780, Montevideo, Uruguay; E. Monteverde, S. McCouch, Dep. of Plant Breeding and Genetics, Cornell Univ., Ithaca, NY 14850; P. Blanco, F. Pérez de Vida, J. Rosas, National Rice Research Program, Experimental Station INIA Treinta y Tres, INIA, Ruta 8, Km 281, Treinta y Tres 33000, Uruguay; S. Fernández, Information Technology Unit, INIA, Andes 1365 Piso 12, Montevideo, Uruguay; S. Simondi, Mathematics Area, College of Natural and Exact Sciences, Univ. Nacional de Cuyo, Padre Contreras 1300, Mendoza, Argentina.

ABSTRACT Rice (*Oryza sativa* L.) is one of the most important staple food crops in the world; however, there has recently been a shift in consumer demand for higher grain quality. Therefore, understanding the genetic architecture of grain quality has become a key objective of rice breeding programs. Genome-wide association studies (GWAS) using large diversity panels have successfully identified genomic regions associated with complex traits in diverse crop species. Our main objective was to identify genomic regions associated with grain quality and to identify and characterize favorable haplotypes for selection. We used two locally adapted rice breeding populations and historical phenotypic data for three rice quality traits: yield after milling, percentage of head rice recovery, and percentage of chalky grain. We detected 22 putative quantitative trait loci (QTL) in the same genomic regions as starch synthesis, starch metabolism, and cell wall synthesis-related genes are found. Additionally, we found a genomic region on chromosome 6 in the *tropical japonica* population that was associated with all quality traits and we identified favorable haplotypes. Furthermore, this region is linked to the *OsBEI* gene that codes for a starch branching enzyme I, which is implicated in starch granule formation. In *tropical japonica*, we also found two putative QTL linked to *OsBEII*, *OsDEP1*, and *OsDEP2*. Our study provides an insight into the genetic basis of rice grain chalkiness, yield after milling, and head rice, identifying favorable haplotypes and molecular markers for selection in breeding programs.

Abbreviations: GC, percentage of chalky grain; GWAS, genome-wide association study; INIA, Instituto Nacional de Investigación Agropecuaria (National Institute of Agriculture Research); LD, linkage disequilibrium; PCA, principal component analysis; PHR, percentage of head rice recovery; PVE, proportion of phenotypic value explained; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; SSRG, starch synthesis-related genes; YAM, yield after milling.

CORE IDEAS

- Genome-wide association study (GWAS) for rice quality was performed in two breeding populations.
- Twenty-two putative quantitative trait loci (QTL) were associated to rice quality.
- A genomic region on chromosome 6 was associated with all quality traits in the *tropical japonica* population.
- Markers for favorable haplotypes are ready for immediate use for selection.

ONE OF THE MAIN CONCERNS of agricultural research today is intensifying agricultural production in a sustainable manner to feed the 9 billion people expected by 2050 (Godfray et al., 2010). Rice is a staple crop in Asia and Africa, where 3.5 billion people depend on rice for food energy (Food and Agriculture Organization of the United

Citation: Quero, R., L. Gutiérrez, E. Monteverde, P. Blanco, F. Pérez de Vida, J. Rosas, S. Fernández, S. Garaycochea, S. McCouch, N. Berberian, S. Simondi, and V. Bonnecarrère. 2018. Genome-wide association study using historical breeding populations discovers genomic regions involved in high-quality rice. Plant Genome 11:170076. doi: 10.3835/plantgenome2017.08.0076

Received 25 Aug. 2017. Accepted 9 Apr. 2018. *Corresponding author (gutierrezcha@wisc.edu).

This is an open access article distributed under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Copyright © Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA

Nations, 2009). The demand for rice continues to grow, especially in Asia and Africa, where people now require high-quality rice (Hsiaoping, 2005; Zader, 2011; Mohanty, 2013; Yu et al., 2013) regarding traits such as grain length, color, absence of broken grains, aroma, and flavor. The ability to meet this demand can be achieved by understanding the genetic basis of key production traits and accelerating the rate of genetic gain from cultivar development.

Grain quality is a complex quantitative trait (Fitzgerald et al., 2009). Quantitative trait loci mapping via GWAS (Jannink et al., 2001) has successfully identified genomic regions associated with complex traits in diverse crop species and provided targets for marker assisted selection (Begum et al., 2015). Early GWAS studies used large diversity panels to maximize the range of genetic variation and improve the power of detecting QTL (Kraakman et al., 2004; Gore et al., 2009; Huang et al., 2010, 2012; Famoso et al., 2011; Zhao et al., 2011; Chen et al., 2014; McCouch et al., 2016; Zhu et al., 2016). However, in the same way that traditional QTL studies were challenged by their lack of practical use because some of the favorable alleles of major-effect QTL were already fixed in elite germplasm (Langridge et al., 2001); the use of diverse and unadapted germplasm in GWAS studies may yield irrelevant genomic regions for breeding purposes (Kraakman et al., 2004). Furthermore, the genetic background interaction of QTL effects (Langridge et al., 2001) and QTL × environment interactions (Malosetti et al., 2004, 2016; Mathews et al., 2008; Gutiérrez et al., 2015) have been extensively reported, indicating that the choice of germplasm and environments used for mapping studies is relevant, especially regarding its future deployment. Therefore, studies looking for immediate applications in breeding use locally adapted germplasm to map QTL (Begum et al., 2015; Spindel et al., 2015) or nested association mapping with a locally adapted line serving as the common parent (Yu et al., 2008; Brachi et al., 2011; Kump et al., 2011; Tian et al., 2011; Mace et al., 2013; Maurer et al., 2015). The use of breeding populations has been successful in identifying QTL and favorable haplotypes in elite populations of tropical rice with phenotypic data specially generated for GWAS purposes (Begum et al., 2015). Besides the immediate advantage of using adapted germplasm, GWAS in local populations allows the discovery of adaptive alleles and allelic complexes, which may be locally common but globally rare (rare alleles) and therefore has the potential to unveil genetic variants that would otherwise be overlooked. Despite being potentially useful, rare alleles are often discarded by minor allele frequency filters when exploring natural variation (Jannink et al., 2001; Brachi et al., 2011; McCouch et al., 2016).

The main objective of this study was to identify genomic regions associated with rice grain quality in relevant adapted germplasm and to identify favorable haplotypes for selection. Specifically, we studied the genetic architecture of rice quality by conducting a GWAS analysis on a subtropical-adapted breeding population consisting of 637 elite rice lines representing two of the major subgroups of rice, *indica* and *tropical japonica*. Considering the extended linkage disequilibrium (LD) in this kind of population, we conducted a GWAS analysis that followed an appropriate analytical framework that involved a high-coverage genotyping strategy and a careful interpretation of population structure and phenotypic data. Candidate genes were predicted via an annotation approach. This strategy exploits existing breeding populations and historical phenotypic data, demonstrating that it is possible to use routine breeding data to perform haplotype selection.

MATERIALS AND METHODS

Plant Material

A total of 637 genotypes from the National Rice Breeding Program were used as the Uruguayan rice reeding GWAS population. The population included 324 *indica* lines, 310 *tropical japonica* lines, two *indica* cultivars [El Paso 144 (Yan et al., 2007) and INIA Olimar (Blanco et al., 2004)] and one *tropical japonica* cultivar [INIA Tacuarí (Blanco et al., 1993; Instituto Nacional de Semillas, 2017)]. Within the *indica* subpopulation, 228 genotypes originated from Instituto Nacional de Investigación Agropecuaria (INIA) breeding material and 98 from Fondo Latinoamericano de Arroz de Riego breeding material. All the individuals within a population, including the checks, share some level of common ancestry, and pedigree information was available from breeders.

Genotyping and Single Nucleotide Polymorphism Genotype Calling

Genotyping-by sequencing data were obtained for the 637 advanced inbred lines and cultivars. DNA was extracted from young leaf tissue from plants grown in the Biotechnology Unit in Las Brujas, Canelones, Uruguay. The extraction was conducted with the Qiagen DNeasy kit (www.qiagen.com/uy/, accessed 5 June 2018). Samples were submitted in 96-well plates and libraries were prepared using the protocol described by Elshire et al. (2011). Genotyping-by-sequencing library construction and sequencing were done at the Biotechnology Resource Center at Cornell University. Single nucleotide polymorphisms (SNPs) were called from fastq files via the TASSEL version 3.0 genotyping-by-sequencing pipeline (Glaubitz et al., 2014). Alignment to the Michigan State University Nipponbare reference genome version 7.0 (http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/, accessed 14 June 2018) was performed with Bowtie version 2 (Langmead and Salzberg, 2012). Imputation of missing data was performed with the FILLIN algorithm implemented in TASSEL version 5.0 (Bradbury et al., 2007; Swarts et al., 2014) for the *indica* and *tropical japonica* genotypes separately. The average imputation accuracy was approximately 94% for both *indica* and *tropical japonica* datasets. Single nucleotide polymorphism markers that had more than 50% missing data after imputation, along with monomorphic SNPs and SNPs with a minor allele frequency smaller than 1% were removed from the analysis.

Grain Quality Phenotyping

Rice lines were evaluated in the field located in Paso de la Laguna, Treinta y Tres, Uruguay (33°15′S, 54°25′W) during the growing seasons (October–March) in 2010–2011, 2011–2012, and 2012–2013 in replicated experiments. Adjusted means for each line were obtained with mixed models to include experimental design components and spatial corrections (Supplemental File S1) using the model in Eq. [1]:

$$Y_{ijk} = \mu + \lambda_i + \beta_{j(i)} + \gamma_k + \varepsilon_{ijk} \quad [1]$$

where Y_{ijk} is the response variable, μ is the overall mean or intercept, λ_i is a random variable associated with the *i*th trial with $\lambda_i \sim N(0, \sigma^2_\lambda)$, $\beta_{j(i)}$ is a random variable associated with the *j*th block nested within the *i*th trial with $\beta_{j(i)} \sim N(0, \sigma^2_\beta)$, γ_k is the effect of the *k*th genotype, and ε_{ijk} is the residual error with $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$.

Milling quality was evaluated by the yield after milling (YAM), the percentage of head rice recovery (PHR), and the percentage of chalky grain (GC). For YAM and PHR, 100 g of rough rice was dried to 13% moisture, hulled with a Satake Rubber Roll Huller (Satake Engineering Co., Ltd., Tokyo, Japan), milled with a Satake Grain Testing Mill (Model TM 05C, abrasive roll #36, Satake Engineering Co., Ltd.), and separated into broken and whole kernels using a thickness grader (Model TWSM, Satake Engineering Co. Ltd.) with an indented cylinder (cylinder indent sizes of 4.75 mm). The weight of grain recovered after milling and separating was used to calculate the percentage of total milled rice or YAM. The percentage of whole kernels recovered after separating was used as the PHR. For GC, a subsample of 50 g of total milled rice was visually inspected by analysts to determine GC. According to industry standards, whole or broken kernels were considered to be chalky when the area of chalk (core, white back, or belly) was larger than half of the kernel area.

Principal Component Analysis and Population Structure Analyses

Population structure was analyzed via principal component analysis (PCA) and a model-based clustering algorithm. The PCA analyses were performed with the imputed marker score matrix in R statistical software (R Core Team, 2017) using the package *rrBLUP* (Endelman, 2011). Based on the PCA results, clustering of *indica* or *tropical japonica* individuals was implemented with ADMIXTURE software version 1.23 (Alexander et al., 2009). The number of populations (*k*) was selected according to two main criteria: first, the lowest cross-validation error across a range of *k* values (i.e., *k* = 1–10); second, an ad hoc correspondence with pedigree information. The resulting probabilities of belonging to groups from ADMIXTURE were then plotted with the *barplot* function in the R statistical software (R Core Team, 2017) to obtain stacked bar charts.

Genome-Wide Association Study, LD Decay, and Haplotype Analysis

Genome-wide association studies were performed with mixed models to correct for population structure and genetic relationships. The most common mixed models for GWAS were compared: naive, kinship (Parisseaux and Bernardo, 2004), and eigenvalue (Price et al., 2006; Malosetti et al., 2007). The best model was selected on the basis of quantile–quantile plots (i.e., Schweder and Spjotvoll plots; Schweder and Spjotvoll, 1982) (Supplemental File S2). The kinship was the selected model, with:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{e}, \quad [2]$$

where $\boldsymbol{y}$ is the vector of phenotypic means, $\mathbf{X}$ is the molecular marker score matrix, β is the vector of marker allelic effects, $\mathbf{Z}$ is an incidence matrix, $\boldsymbol{u}$ is the vector of polygene background effects with $\text{Var}(\boldsymbol{u})$ being $2\mathbf{K}V_G$ ($\mathbf{K}$ is the matrix of kinship coefficients and V_G is the genetic variance), and ε is the residual error vector. A GWAS analysis for each rice subpopulation was performed in the R statistical software (R Core Team, 2017) with the *lmem.gwaser* package (Gutierrez et al., 2016). For QTL determination, the marker with the highest marker–trait association was chosen as an anchor and then, a sliding window of 1 Mb was used to identify all significant markers within that window. The window size was determined according to the LD decay in each chromosome (Supplemental File S3 and Supplemental File S4). A QTL was defined when three or more significant SNPs were found within the 1-Mb window, following Rosas et al. (2017). Given the level of genetic relatedness in our populations, markers in close proximity are in high LD, making it unlikely to have an isolated significant SNP. Therefore, isolated markers are more probably caused by genotyping or imputation error (Brandariz et al., 2016) than a true QTL. Choosing three markers for our threshold makes our analysis less likely to declare a false QTL. Linkage disequilibrium was computed as pairwise r^2 between all SNPs in the chromosome and then in a specific region, and limits between LD blocks were graphically assessed with the R package (R Core Team, 2017). The threshold level for calling a significant marker–trait associations was calculated by using a *p*-value corrected by multiple comparisons with the Li and Ji (2005) statistic at an α level of 0.05. The proportion of the total phenotypic variance explained (PVE) by each QTL was estimated by fitting a full multi-QTL model with all significant SNPs from all genomic regions involved for a trait in the *lme4* package (Bates et al., 2015) of R statistical software (R Core Team, 2017). Allelelic effects for each QTL were obtained with the *emmeans* package (Lenth, 2018) in R statistical software (R Core Team, 2017). Finally, the most prevalent haplotypes were identified with the *clusterhap* package (Quero et al., 2017) in R statistical software (R Core Team, 2017) and the phenotypic means for each haplotype were estimated (Supplemental File S5).

Identification of Candidate Genes

A literature survey and a genome annotation pipeline were used to search for putative causal candidate genes.

The numbers of genes located within a defined QTL were retrieved from the Michigan State University public gene annotation database (http://rice.plantbiology.msu.edu/downloads_gad.shtml, accessed 14 June 2018) via an in-house script (Supplemental File S6). The Plant Metabolic Network was used to assign a function to each gene (Zhang et al., 2010). OryzaCyc was used to search for plant metabolic pathway functionality. Gene function was further explored by studying the metabolic pathways where the encoded enzymes were involved. This was analyzed with the Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto, 2000), which is a collection of pathway maps. The literature survey was focused on major genes involved in starch synthesis known as starch synthesis-related genes (SSRGs) (Zeng et al., 2017). The SSRGs include genes for ADP-glucose pyrophosphorylase, granule-bound starch synthase, starch synthase, branching enzyme, debranching enzyme, starch phosphorylase, disproportionating enzyme, and glucose 6-phosphate translocator. After identification of the candidate genes, we analyzed the presence of SNPs in its coding sequence in the Nipponbare reference genome (Michigan State University Nipponbare reference genome version 7.0). When a SNP was found within the coding sequence, the amino acid sequence of the encoded protein with both versions of the SNP was determined. To do this, Mega version 6 software was used (Tamura et al., 2013).

RESULTS

Population Structure and Phenotypic Variation

Two main subpopulations were observed in the PCA corresponding to the *indica* and *tropical japonica* subpopulations (Fig. 1a). The first two principal components explained more than 78.2% of the total genotypic variance. Within the *indica* subpopulation, two subgroups were identified with a model-based algorithm corresponding to the two distinct origins of lines coming from the breeding programs at the INIA or the Fondo Latinoamericano de Arroz de Riego (Fig. 1a). Five subgroups were identified within the *tropical japonica* subpopulation (Fig. 1a) on the basis of the clustering algorithm and pedigree information. The *tropical japonica* subpopulation is defined as a multiparent cross where the lines were derived from 12 parents, and each of the five subgroups was comprised of half-sib families. The *indica* subpopulation had lower GC on average, with smaller variance than the *tropical japonica* subpopulation (Fig. 1b), whereas estimates of YAM and PHR were similar in both subpopulations. In the *tropical japonica* subpopulation, the three grain quality traits were correlated: YAM and PHR were positively correlated, but GC was negatively correlated with YAM and PHR. On the other hand, no correlation among traits was observed for the *indica* subpopulation (Fig. 1b).

Quantitative Trait Loci Identified by GWAS

Genome-wide association studies for every targeted trait were performed separately for *indica* and *tropical japonica* because of the population structure. When different mixed models for GWAS were compared, the kinship model using the realized relationship matrix estimated from the marker data was the best model (Supplemental File S2). We therefore used this model to map QTL following the analytical framework outlined in Supplemental File S5. We identified a total of 22 putative QTL in the two subpopulations (Fig. 2). These QTL were identified with a multiple comparison test at a p-value threshold of 8.5×10^{-4} for *tropical japonica* and 1.3×10^{-3} for *indica* subpopulation, according to the calculated Li and Ji (2005) threshold. In the *indica* subpopulation, five putative QTL were identified for GC, one for YAM, and six for PHR (Supplemental File S7), whereas in *tropical japonica*, three QTL were identified for GC, five for YAM, and two for PHR (Supplemental File S8; Fig. 2). We did not find any QTL shared between the *indica* and *tropical japonica* subpopulations. There is a genomic region on chromosome 6 (26,894,513–29,480,530 bp) of *tropical japonica*, where three putative QTL mapped for all quality traits evaluated (qYAM.j.6.1, qPHR.j.6.1, and qGC.j.6.2). Markers on these QTL were in high LD (Fig. 3a). The PVE for each QTL was above 30%, which means all of them are large-effect QTL (Table 1). On the other hand, all QTL in the *indica* subpopulation had lower PVE and only two QTL had a PVE above 10% (qPHR.i.2.1 and qPHR.i.3.1).

Genomic Regions and Haplotype Analysis

Of the 22 marker–trait associations detected in this study, eight SSRGs (Zeng et al., 2017) were identified within or in the flanking regions of nine QTL (Table 2). By using a genome annotation approach, we also identified eight candidate genes in five putative QTL observed in our subpopulations (Supplemental File S9). Beside starch metabolism, these candidate genes are involved in cell wall formation and degradation. Our study provides hypothetical candidate genes that should be further studied to elucidate whether they have a functional role in grain quality.

Three putative QTL (qYAM.j.6.1, qPHR.j.6.1, and qGC.j.6.2) were colocated in a region on chromosome 6 in the *tropical japonica* subpopulation. We identified three candidate genes within the QTL region involved in starch metabolism and three involved in cell wall formation (Supplemental File S9). These genes include two α-glucosidases that are part of the first pathway of starch degradation. A functional mutation (genotyping-by-sequencing SNP S6_28101061; A > G) in one of the α-glucosidase genes, *LOC_Os06g46340*, alters the protein sequence (p.Glu45Gly) and is noteworthy because the derived allele codes for an amino acid belonging to a different group and would probably result in a conformational change in the protein product.

Besides these candidate genes, we found that one SSRG, specifically a starch branching enzyme I gene, *OsBEI* (*LOC_Os06g51084*; Ohdan et al., 2005) is located next to the QTL interval on chromosome 6. Rice has two branching enzyme families, BEI and BEII, coded by one (*OsBEI*) and two genes (*OsBEIIa* and *OsBEIIb*), respectively. *OsBEI* is the only BEI gene in rice and it is known to be involved in amylopectin structure (Ohdan et al., 2005). We detected two

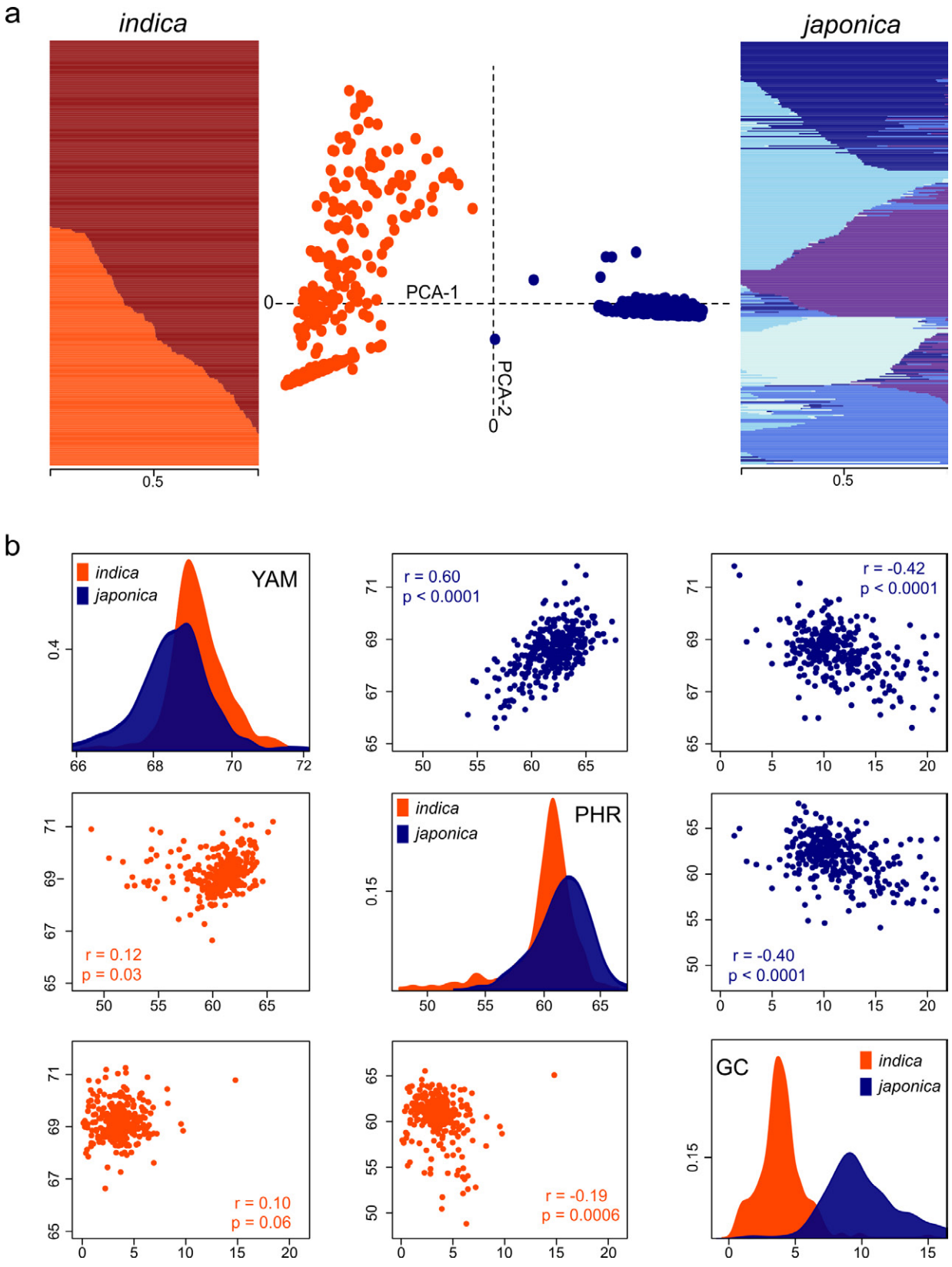

Figure 1. Genetic structure and phenotypic variation in locally adapted rice populations. (a) The central image shows the first two principal components separating *indica* (red, n = 326) and tropical japonica (blue, n = 311) individuals; the left-hand image shows the two genetic subgroups within *indica*; the right-hand image shows the five genetic subgroups within *tropical japonica*. (b) Scatterplot matrix for grain quality traits showing density plots for each trait [yield after milling (YAM), percentage of head rice recovery (PHR), and percentage of chalky grain (GC)] in the diagonal, scatterplots between traits for *tropical japonica* (blue, n = 311) above the diagonal, and scatterplots between traits for *indica* (red, n = 326) below the diagonal. Numbers inside the scatterplots indicate Pearson's correlation between pairs of traits and their *p*-values. Each scatterplot displays two variables with the *x* and *y* axes corresponding to the variables in the diagonal.

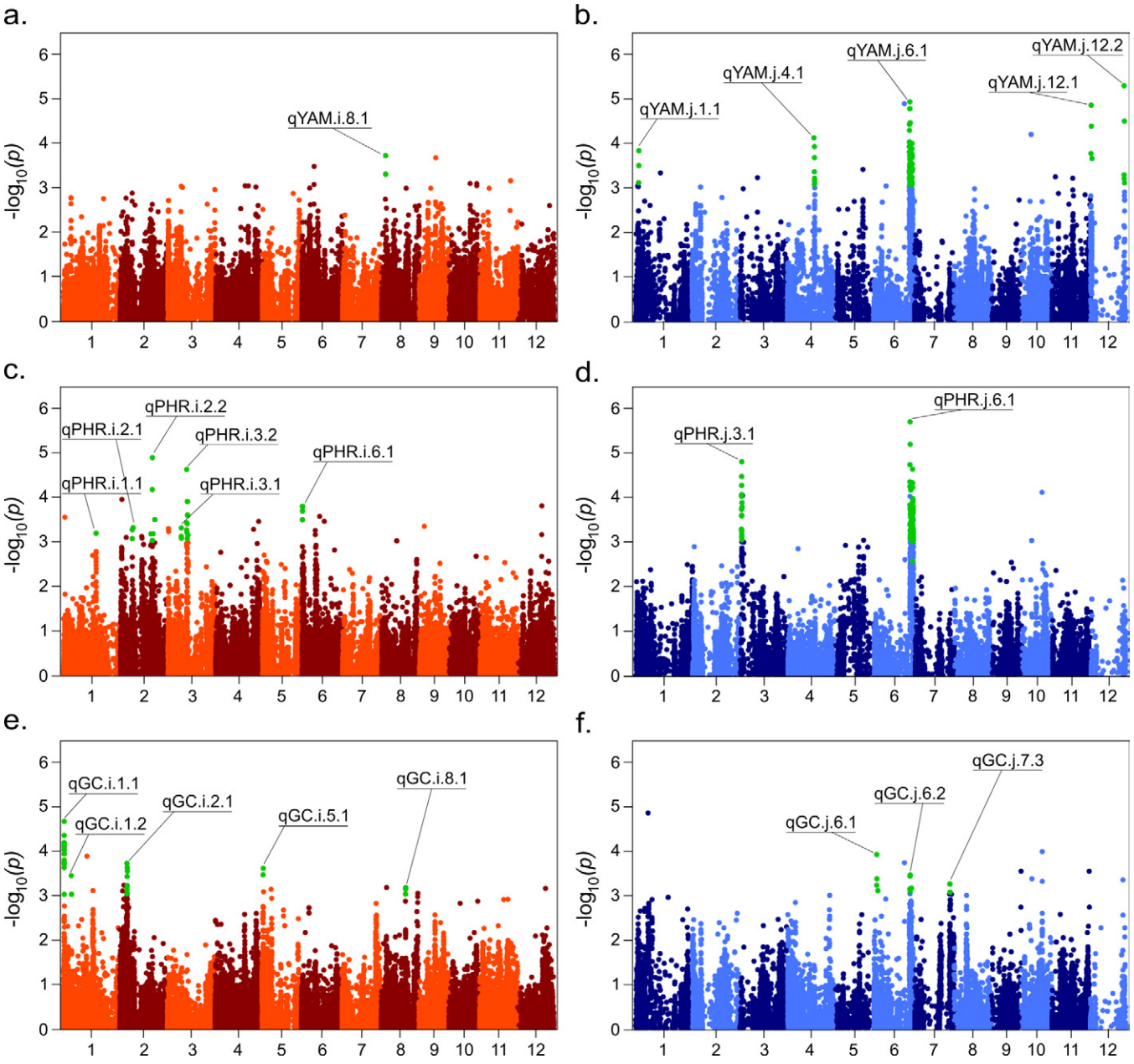

Figure 2. Genome-wide association study (GWAS) for rice physical grain quality in two breeding subpopulations. Manhattan plots of the *indica* subpopulation (red, n = 326) for yield after milling (YAM) (a), percentage of head rice recovery (PHR) (c), and percentage of chalky grain (GC) (e) and of the *tropical japonica* subpopulation (blue, n = 311) for YAM (b), PHR (d), and GC (f). The x-axis shows single nucleotide polymorphism (SNP) positions along chromosomes; the y-axis shows the –log10 of the P-value for each marker–trait association. Associations above the significance threshold [calculated via the method of Li and Ji (2005) with multiple test correction] are shown in green.

SNPs (S6_30900078 and S6_30900838) within the *OsBEI* gene that are in high LD with most of the significant SNPs located within the QTL (Fig. 3b), with an average r^2 of 0.63, 0.65 and 0.83 between *OsBEI* and qYAM.j.6.1, qPHR.j.6.1, and qGC.j.6.2, respectively. Furthermore, we observed three clearly differentiated haplotypes present in ~90% of the lines and a few minor recombinant haplotypes (data not shown). For YAM and PHR, the groups of individuals carrying the H1 and H2 haplotypes had significantly higher phenotypic means than H3, while for GC the groups of individuals carrying the H1 and H2 haplotype had significantly lower phenotypic means than H3 (Fig. 3b). For all traits, H1 and H2 were different from each other by only one SNP (S6_30900078), which is located within the *OsBEI* gene. Considering the phenotypic mean of theses haplotypes, this polymorphism had no effect on PHR and GC, but it had an effect on YAM (Fig. 3b).

Other SSRGd were found linked to be to QTL identified by GWAS in the *tropical japonica* subpopulation (Table 2). The SSRG *OsBEIIa* (*LOC_Os04g33460*) is located within the QTL qYAM.j.4.1 and all markers within the QTL are in perfect LD with one SNP located within the *OSBEIIa* gene (Fig. 4a). This QTL region has three major haplotypes with different phenotypic means (Fig. 4b).

Table 1. Proportion of the phenotypic variance explained (PVE) by each putative quantitative trait locus (QTL) identified from the genome-wide association study analysis of a large rice population in both *indica* (i) and *tropical japonica* (j) subpopulations showing the single nucleotide polymorphisms (SNP) with the most significant marker–trait associations.

QTL	SNP	PVE
qGC.i.1.1	S1_1066894	4.78
qGC.i.1.2	S1_6427540	4.49
qGC.i.2.1	S2_5409930	7.29
qGC.i.5.1	S5_1053773	6.30
qGC.i.8.1	S8_18577900	0.84
qPHR.i.1.1	S1_24735280	0.70
qPHR.i.2.1	S2_9639305	10.34
qPHR.i.2.2	S2_24210614	0.34
qPHR.i.3.1	S3_10247958	15.94
qPHR.i.3.2	S3_14353133	9.95
qPHR.i.6.1	S6_366796	1.81
qYAM.i.8.1	S8_2999273	7.18
qGC.j.6.1	S6_2140954	2.78
qGC.j.6.2	S6_27519265	35.27
qGC.j.7.1	S7_26562521	3.63
qYAM.j.1.1	S1_3059047	2.65
qYAM.j.4.1	S4_19985917	8.90
qYAM.j.6.1	S6_27313987	34.26
qYAM.j.12.1	S12_566928	7.89
qYAM.j.12.2	S12_25469756	8.46
qPHR.j.3.1	S3_1354379	5.63
qPHR.j.6.1	S6_27037310	37.40

Table 2. Putative candidate starch synthesis-related genes (SSRGs) that are located within or close to quantitative trait loci (QTL) identified from the genome-wide association study analysis of a large rice population in both *indica* and *tropical japonica* subpopulations.

QTL	Reference gene locus	SSRG name	SSRG product
qPHR.i.1.1	*LOC_Os01g44220*	*OsAGPL2*	ADP-glucose pyrophosphorylase large subunit 2
qPHR.i.6.1	*LOC_Os06g04200*	*OsGBSSI*	Granule-bound starch synthase I
qYAM.i.8.1	*LOC_Os08g09230*	*OsSSIIIa*	Starch synthase IIIa
qYAM.j.4.1	*LOC_Os04g33460*	*OsBEIIa*	Starch branching enzyme IIa
qGC.j.6.1	*LOC_Os06g06560*	*OsSSI*	Starch synthase I
qGC.j.6.2	*LOC_Os06g51084*	*OsBEI*	Starch branching enzyme I
qYAM.j.6.1	*LOC_Os06g51084*	*OsBEI*	Starch branching enzyme I
qPHR.j.6.1	*LOC_Os06g51084*	*OsBEI*	Starch branching enzyme I
qGC.j.7.1	*LOC_Os07g43390*	*OsDEP1*	Disproportionating enzyme I
qGC.j.7.1	*LOC_Os07g46790*	*OsDEP2*	Disproportionating enzyme II

On the other hand, the SSRG *OsSSI* (*LOC_Os06g06560*) is in perfect LD with the QTL qGC.j.6.1 (Supplemental File S10) and the only two disproportionation enzyme genes in rice genome, DP1 (*LOC_Os07g43390*) and DP2 (*LOC_Os07g46790*; Lu and Sharkey, 2004), flank the QTL qGC.j.7.1 (Table 2). All SNPs within this QTL are in complete LD with the SNPs flanking both genes (Fig. 4a). On the other hand, for qGC.j.7.1, there is a difference of ~2% of GC between the two major haplotypes (Fig. 4c).

We could identify candidate genes in two *indica* QTL (Supplemental File S9): a gene for a fructose-bisphospate aldolase (*LOC_Os01g02880*), which is associated with starch metabolism, in qGC.i.1.1 and a gene encoding an arabinofuranosidase (*LOC_Os02g10190*). In addition, in the QTL q.GC.i.2.1, one of the SNPs in this gene generates a mutation at position 845, where an alanine is changed to a threonine (p.Ala845Thr). Again, as occurred for *LOC_Os06g46340*, both amino acids belong to different groups.

DISCUSSION

By using *indica* and *tropical japonica* breeding populations and historical phenotypic data from 3 yr of replicated and balanced experiments on grain quality, we found enough diversity to map relevant quantitative traits for rice quality and to identify haplotypes with significant differences on rice quality traits. The use of breeding populations for GWAS has some advantages over the use of diverse populations, including immediate application to breeding programs (Kraakman et al., 2004) because it can identify loci that are segregating in the population (Langridge et al., 2001) and that have a reduced genetic background (Langridge et al., 2001) and QTL × environment interactions (Malosetti et al., 2004; Mathews et al., 2008; Gutiérrez et al., 2015). The use of larger structured populations might increase mapping resolution for detecting global QTL (McCouch et al., 2016). However, we decided against a global test with individuals from both subpopulations because: (i) when SNPs were called for all individuals, fewer SNPs that passed our data curation process were identified; (ii) most SNPs were in opposite phases, being monomorphic for one subpopulation and polymorphic for the other, reducing the power for marker–trait associations; (iii) the phenotype was associated with population structure at least for one of our traits (i.e., GC), which would also reduce the power of our statistical tests; and (iv) because of the general challenges of properly controlling for population structure. Therefore, we mapped within subpopulations.

Quality phenotypic traits were correlated in *tropical japonica*. This finding was consistent with previous studies where GC was associated with an increase in grain breakage and therefore a decrease in the percentage of head rice recovery (Lisle et al., 2000; Zhou et al., 2015). In the case of *indica* subpopulation, the low level of correlation among the phenotypic variables could be related to the most diverse grain morphology being found in this population. These findings point toward selection as one possible explanation for the strong correlation in *tropical japonica* instead of pleiotropy. Although a genomic region responsible for all traits was found in *tropical japonica*, the genetic mechanism behind it could be a series of linked genes maintained through selection.

We used genotyping by sequencing to identify more than 44,000 polymorphic SNPs in *tropical japonica* and more than 92,000 SNPs in *indica* (Supplemental File S11). This genotyping strategy generated a relatively high density of SNPs that were appropriate for association mapping and avoided

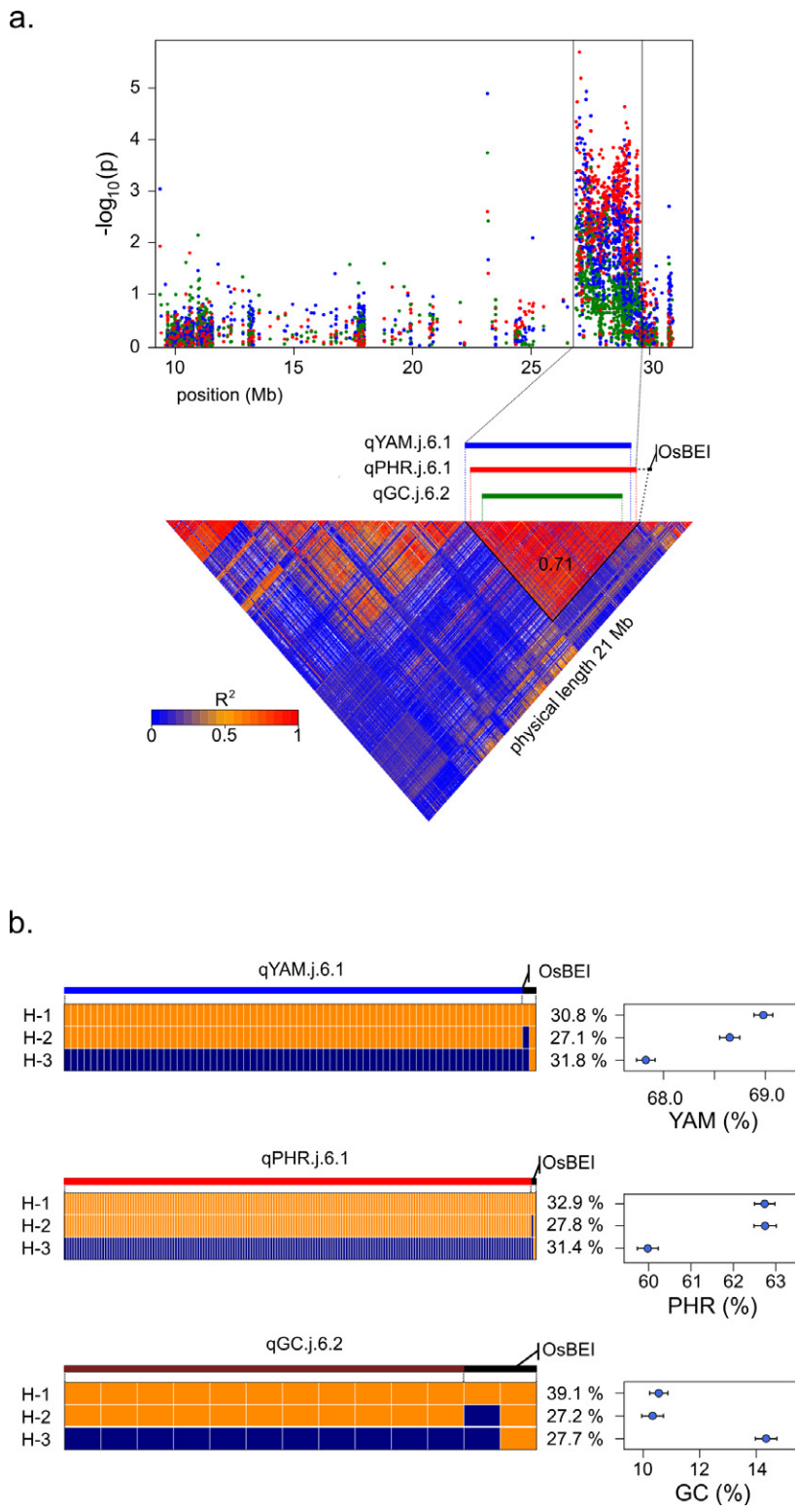

Figure 3. Genome-wide association study (GWAS) hits on chromosome 6 in the *tropical japonica* subpopulation in relation to *OsBEI*, a gene encoding Starch Branching Enzyme 1. (a) Manhattan plot for the three quality traits [yield after milling (YAM), represented in blue, percentage of head rice recovered (PHR) in red, and percentage of chalky grain (GC) in green] in the *tropical japonica* (n = 311) subpopulation showing significant marker–trait associations in the same region on chromosome 6. Linkage disequilibrium (LD) plot within the region showing the LD among the quantitative trait locus (QTL) region and the *OsBEI* gene is shown below the Manhattan plots. The number within the LD block is the average r^2 among all pairwise marker combinations within the window. (b) Predominant haplotypes, with the percentage of individuals carrying each haplotype. Haplotypes are defined by all significant single nucleotide polymorphisms (SNPs) for each trait within the QTL region and SNPs within and flanking the *OsBEI* gene. Each SNP allele is represented with a different color: orange or blue. At the right of each haplotype, a dot-plot showing the adjusted phenotypic means and SE for lines carrying each haplotype is shown.

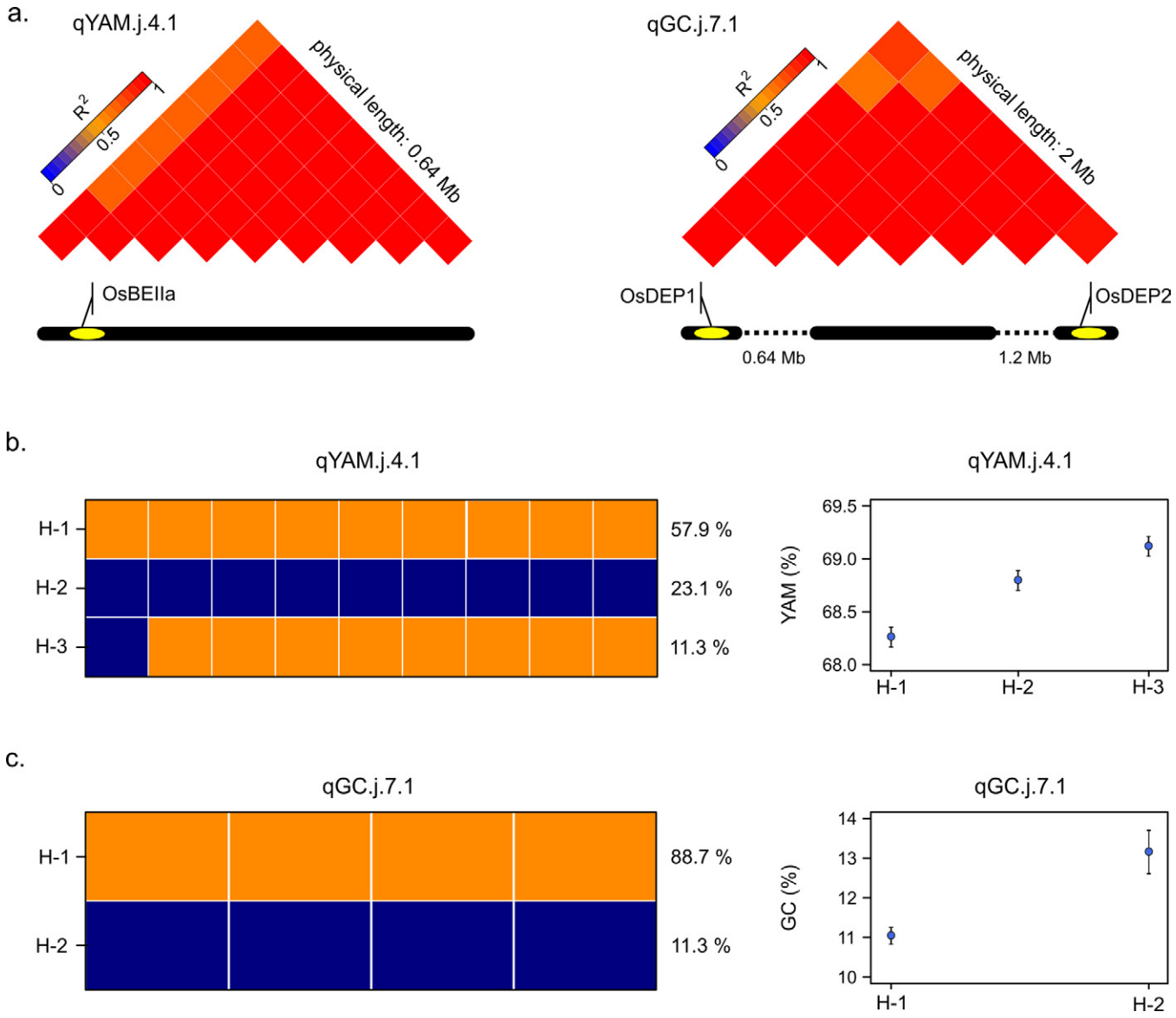

Figure 4. Genome-wide association study (GWAS) in the *tropical japonica* subpopulation in relation to starch synthesis-related genes (SSRGs). (a) Linkage disequilibrium (LD) plot across the qYAM.j.4.1 quantitative trait locus (QTL) showing the LD within the QTL region and single nucleotide polymorphisms (SNPs) located within the SSRG *OsBEIIa* (left) and the LD plot across the QTL qGC.j.7.1 showing the LD among the QTL region and SNPs located within the SSRG genes *DPEI* and *DPEII* (right). Predominant haplotypes with the percentage of individuals carrying each haplotype are shown for (b) qYAM.j.4.1 and (c) qGC.j.7.1. Haplotypes are defined by all significant SNPs within the QTL region. Each SNP allele is represented with a different color: orange or blue. At the right of each haplotype scheme, a dot-plot showing the adjusted phenotypic means and SE for lines carrying each of the haplotypes is shown.

the ascertainment bias that is inherent to the use of fixed genotyping arrays (Heslot et al., 2013). Combined with our analytical strategy that involved a combination of GWAS, LD determination, haplotype identification, and candidate gene identification, this provided good candidate QTL for marker-assisted selection strategies. With the high quality of genome annotation available for rice (National Agriculture and Food Research Organization, 2017), there are several studies that have successfully integrated genetic, genomic (Fitzgerald et al., 2009; Tian et al., 2009) and, in some cases, metabolomic (Heuberger et al., 2010; Yamakawa and Hakata, 2010; Calingacion et al., 2012; Chen et al., 2014; Okazaki and Saito, 2016) information to begin to define the molecular mechanisms underlying important grain quality traits such as grain chalkiness and milling properties. In this study, we identified genes involved in two cellular processes, starch metabolism and cell wall formation or degradation. Many previous studies have related starch content to rice grain quality traits (Su, 2000; Vandeputte and Delcour, 2004; Ashida et al., 2009; Bao, 2014; Lin et al., 2014; Zhao et al., 2015; Zeng et al., 2017), particularly with GC (Ashida et al., 2009) and PHR (Gallant et al., 1997; Patindol and Jabe-Wang, 2003; Yamakawa and Hakata, 2010). Here, we identified five SSRGs and four candidate genes involved in starch metabolism that were physically colocated with our QTL. Once these genotype–phenotype associations are validated, the SNPs will provide breeders

with a valuable tool for early generation selection for several economically valuable grain quality traits, similar to the use of the *waxy* SNP markers currently used for amylose content selection in rice (Kharabian-Masouleh et al., 2012).

In addition to starch content, rice grain quality has been associated with the polysaccharide composition of the cell wall in different grain tissues, such as the starchy endosperm, the aleurone layer, and the transfer cells (Burton and Fincher, 2014; Lin et al., 2014), all affecting grain consistency. The milling process removes the cell walls in the aleurone layer and therefore aleurone cell walls are associated with YAM performance (Burton and Fincher, 2014). In this study, we identified putative QTL with regions coding for three glycosyl hydrolase genes that are involved in heteroxylan backbound formation, modifications, and degradation (Strohmeier et al., 2004; Numan and Bhosle, 2006; Mitchell et al., 2007; Eklöf and Brumer, 2010) and an arabinofuranosidase gene. Arabinofuranose is one of the constituents of hemicellulose, the major component of the rice wall endosperm (Shibuya and Iwasaki, 1978; Shibuya and Nakane, 1984; Numan and Bhosle, 2006).

The candidate genes colocated with putative QTL identified in this work, especially the SSRGs, provide interesting targets for follow-up studies to enhance our understanding of the genetics of grain quality in rice. On the other hand, the results of this study can also improve the accuracy of the genomic selection models used to estimate breeding values and help implement a pragmatic genomic selection strategy in breeding programs.

CONCLUSIONS

In this work, we were able to find putative QTL associated with rice grain quality. The use of locally adapted germplasm with narrow genetic variance provided an opportunity to map subtle phenotypic differences that are likely to be overlooked with a more diverse germplasm panel. From the breeding perspective, the haplotypes provide an opportunity to examine whether substitution of alleles across one particular region of the QTL contributes positively or negatively to the mean performance of each trait. In addition, the use of a locally adapted population of elite breeding materials allows for immediate application in breeding programs, including marker-assisted introgression of favorable genomic regions conferring rice quality traits or targeted genome editing as the basis for future genetics experiments and breeding applications.

Supplemental Information

Supplemental File S1. Phenotypic means for YAM, PHR, and GC of each *indica* and *tropical japonica* rice lines. This is a table of adjusted phenotypic means for each line after experimental design and spatial corrections were performed.

Supplemental File S2. Quantile–quantile (QQ) plots for the GWAS model comparison. Three GWAS models were compared via QQ-plots of observed and expected *p*-values: naive, kinship, and eigen.

Supplemental File S3. Linkage disequilibrium decay for all chromosomes in the *indica* subpopulation.

Supplemental File S4. Linkage disequilibrium decay for all chromosomes in the *tropical japonica* subpopulation.

Supplemental File S5. Analytical framework for QTL and candidate gene identification. This is a description of the analytical procedure followed in this study from genotyping-by-sequencing and phenotypic data to the identification of QTL, haplotypes, and candidate genes.

Supplemental File S6. Annotated genes within QTL regions. This is a list of all genes retrieved from the Michigan State University public gene annotation database with an in-house script.

Supplemental File S7. Significant SNPs and QTL identified in *indica* subpopulation. *P*-values and chromosomal localization for all significant SNPs within QTL for YAM, PHR, and GC are shown. The QTL nomenclature specifies the trait, the subpopulation *indica*, the chromosome, and a correlative number.

Supplemental File S8. Significant SNPs and QTL identified in the *tropical japonica* subpopulation. *P*-values and chromosomal localization for all significant SNPs within QTL for YAM, PHR, and GC are shown. The QTL nomenclature specifies the trait, the subpopulation *japonica*, the chromosome, and a correlative number.

Supplemental File S9. Candidate genes involved in starch metabolism and cell wall formation found within or close to QTL identified through GWAS.

Supplemental File S10. Linkage disequilibrium and haplotype analysis for QTL in relation to SSRGs. For all QTL located near a SSRG, the possible LD between the QTL and SNP within or next to the genes was determined. The LD is expressed as the recombination rate. The haplotypes for the QTL genomic region and the phenotypic mean of lines carrying each haplotype in the mapping population are also shown.

Supplemental File S11. Summary of genotyping-by-sequencing (GBS) results, showing the total number of markers after each filter in the GBS pipeline was applied.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

ACKNOWLEDGMENTS

This work was funded by INIA (Project L2 AZ_12), Agencia Nacional de Investigación Agropecuaria (ANII-POS_NAC_2012_1_8560), and a Fulbright Fellowship and Beachell-Bourlag International PhD Scholarship to EM. We acknowledge Gonzalo Zorrilla, Marco Dalla Rizza, and Omar Borsani for their continuous support for this work. We thank the molecular biology laboratory group and the field team of INIA. We also thank the anonymous reviewers for their comments that substantially improved the manuscript.

REFERENCES

Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664. doi:10.1101/gr.094052.109

Ashida, K., S. Lida, and T. Yasui. 2009. Morphological, physical, and chemical properties of grain and flour from chalky rice mutants. Cereal Chem. 86:225–231. doi:10.1094/CCHEM-86-2-0225

Bao, J. 2014. Genes and QTLs for rice grain quality improvement. In: J. Bao, editor, Rice—Germplasm, genetics and improvement. InTech, London, UK. p. 239–278.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixedeffects models using lme4. J. Stat. Softw. 67:1–48. doi:10.18637/jss.v067.i01

Begum, H., J.E. Spindel, A. Lalusin, T. Borromeo, G. Gregorio, J. Hernandez, et al. 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):e0119873. doi:10.1371/journal.pone.0119873

Blanco, P., F. Molina, F. Pérez de Vida, S. Avila, A. Lavecchia, C. Marchesi, et al. 2004. INIA Olimar: Characterization and performance in season 2003–2004. (In Spanish) Arroz 38:40–48.

Blanco, P., F. Pérez de Vida, and M. Piriz. 1993. INIA Tacuarí: New early and high yielding rice variety. (In Spanish) Bol. Divulg. INIA 31:5–10.

Brachi, B., G.P. Morris, and J. Borevitz. 2011. Genome-wide association studies in plants: The missing heritability is in the field. Genome Biol. 12:232. doi:10.1186/gb-2011-12-10-232

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Brandariz, S.P., A. González Reymúndez, B. Lado, M. Malosetti, A.A.F. Garcia, M. Quincke, et al. 2016. Ascertainment bias from imputation methods evaluation in wheat. BMC Genomics 17(1):773. doi:10.1186/s12864-016-3120-5

Burton, R.A., and G.B. Fincher. 2014. Evolution and development of cell walls in cereal grains. Front. Plant Sci. 5:456–470. doi:10.3389/fpls.2014.00456

Calingacion, M.N., C. Boualaphanh, V.D. Daygon, R. Anacleto, R. Sackville, R.S. Hamilton, et al. 2012. A genomics and multi-platform metabolomics approach to identify new traits of rice quality in traditional and improved varieties. Metabolomics 8:771–783. doi:10.1007/s11306-011-0374-4

Chen, W., Y. Gao, W. Xie, L. Gong, K. Lu, W. Wang, et al. 2014. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat. Genet. 46:714–721. doi:10.1038/ng.3007

Eklöf, J.M., and H. Brumer. 2010. The *XTH* gene family: An update on enzyme structure, function, and phylogeny in xyloglucan remodeling. Plant Physiol. 153:456–466. doi:10.1104/pp.110.156844

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. doi:10.1371/journal.pone.0019379

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255. doi:10.3835/plantgenome2011.08.0024

Famoso, A., K. Zhao, R.T. Clark, C.W. Tung, M.H. Wright, C. Bustamante, et al. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7(8):e1002221. doi:10.1371/journal.pgen.1002221

Fitzgerald, M.A., S.R. McCouch, and R.D. Hall. 2009. Not just a grain of rice: The quest for quality. Trends Plant Sci. 14:133–139. doi:10.1016/j.tplants.2008.12.004

Food and Agriculture Organization of the United Nations. 2009. Increasing crop production sustainably. The perspective of biological processes. Food and Agricultural Organization of the United Nations. http://www.fao.org/docrep/012/i1235e/i1235e00.htm (accessed 5 June 2018).

Gallant, D.J., B. Bouchet, and P.M. Baldwin. 1997. Microscopy of starch: Evidence of a new level of granule organization. Carbohydr. Polym. 32:177–191. doi:10.1016/S0144-8617(97)00008-8

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2):E90346 doi:10.1371/journal.pone.0090346

Godfray, C.H., J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, et al. 2010. Food security: The challenge of feeding 9 billion people. Science 327:812–818. doi:10.1126/science.1185383

Gore, M.A., J.M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, B.L. Hurwitz, et al. 2009. A first-generation haplotype map of maize. Science 326:1115–1117. doi:10.1126/science.1177837

Gutiérrez, L., S. Germán, S. Pereyra, P.M. Hayes, C.A. Pérez, F. Capettini, et al. 2015. Multi-environment multi-QTL association mapping identifies disease resistance QTL in barley germplasm from Latin America. Theor. Appl. Genet. 128:501–516.

Gutierrez, L., G. Quero, S. Fernandez, and S. Brandariz. 2016. lmem.gwaser: Linear mixed effects models for genome-wide association studies. R package version 0.1.0. https://CRAN.R-project.org/package=lmem.gwaser (accessed 17 Jan. 2017).

Heslot, N., J. Rutkoski, J. Poland, J.L. Jannink, and M.E. Sorrells. 2013. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS One 8:e74612. doi:10.1371/journal.pone.0074612

Heuberger, A.L., M.R. Lewis, M.H. Chen, M.A. Brick, E.L. Leach, and E.P. Ryan. 2010. Metabolomic and functional genomic analyses reveal varietal differences in bioactive compounds of cooked rice. PLoS One 5:e12915. doi:10.1371/journal.pone.0012915

Hsiaoping, C. 2005. Rice is life: Scientific perspectives for the 21st century. In: K. Toriyama, K.L. Heong, and B. Hardy, editors, Proceedings of the World Rice Research Conference, Tokyo and Tsukuba, Japan. 4–7 Nov. 2004. IRRI, Philippines. p. 497–499.

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42:961–967. doi:10.1038/ng.695

Huang, X., Y. Zhao, X. Wei, C. Li, A. Wang, Q. Zhao, et al. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. 44:32–39. doi:10.1038/ng.1018

Instituto Nacional de Semillas. 2017. Registro nacional de cultivares y registro de propiedad de cultivares. Instituto Nacional de Semillas. http://www.inase.uy/EvaluacionRegistro/ (accessed 14 June 2018).

Jannink, J.L., M. Bink, and R. Jansen. 2001. Using complex plant pedigrees to map valuable genes. Trends Plant Sci. 6:337–342. doi:10.1016/S1360-1385(01)02017-9

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28:27–30. doi:10.1093/nar/28.1.27

Kharabian-Masouleh, A., D. Waters, R. Reinke, R. Ward, and R. Henry. 2012. SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. Sci. Rep. 2:557–566. doi:10.1038/srep00557

Kraakman, A., R. Niks, P. Van den Berg, P. Stam, and F.A. Van Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446. doi:10.1534/genetics.104.026831

Kump, K.L., P.J. Bradbury, R.J. Wisser, E.S. Buckler, A.R. Belcher, and M.A. Oropeza-Rosas. 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43:163–168. doi:10.1038/ng.747

Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359. doi:10.1038/nmeth.1923

Langridge, P., E.S. Lagudah, T.A. Holton, R. Appels, P.J. Sharp, and K.J. Chalmers. 2001. Trends in genetic and genome analyses in wheat: A review. Crop Pasture Sci. 52:1043–1077. doi:10.1071/AR01082

Lenth, R.V. 2018. Least-squares means: The R package lsmeans. J. Stat. Softw. 69:1–33. doi:10.18637/jss.v069.i01

Li, J., and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95:221–227. doi:10.1038/sj.hdy.6800717

Lin, Z., X. Zhang, X. Yang, G. Li, S. Tang, S. Wang, et al. 2014. Proteomic analysis of proteins related to rice grain chalkiness using iTRAQ and a novel comparison system based on a notched-belly mutant with white-belly. BMC Plant Biol. 14:163–179. doi:10.1186/1471-2229-14-163

Lisle, A., M. Martin, and M. Fitzgerald. 2000. Chalky and translucent rice grains differ in starch composition and structure and cooking properties. Cereal Chem. 77:627–632. doi:10.1094/CCHEM.2000.77.5.627

Lu, Y., and T. Sharkey. 2004. The role of amylomaltase in maltose metabolism in the cytosol of photosynthetic cells. Planta. 218:466–473. doi:10.1007/s00425-003-1127-z

Mace, E.S., C.H. Hunt, and D.R. Jordan. 2013. Supermodels: Sorghum and maize provide mutual insight into the genetics of flowering time. Theor. Appl. Genet. 126:1377–1395. doi:10.1007/s00122-013-2059-z

Malosetti, M., D. Bustos-Korts, M.P. Boer, and F.A. van Eeuwijk. 2016. Predicting responses in multiple environments: Issues in relation to genotype× environment interactions. Crop Sci. 56(5):2210–2222. doi:10.2135/cropsci2015.05.0311

Malosetti, M., C. Van der Linden, B. Vosman, and F.A. Van Eeuwijk. 2007. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175:879–889. doi:10.1534/genetics.105.054932

Malosetti, M., J. Voltas, I. Romagosa, S.E. Ullrich, and F.A. Van Eeuwijk. 2004. Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139–145. doi:10.1023/B:EUPH.0000040511.46388.ef

Mathews, K.L., M. Malosetti, S. Chapman, L. McIntyre, M. Reynolds, R. Shorter, et al. 2008. Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor. Appl. Genet. 117(7):1077–1091. doi:10.1007/s00122-008-0846-8

Maurer, A., V. Draba, Y. Jiang, F. Schnaithmann, R. Sharma, E. Schumann, et al. 2015. Modeling the genetic architecture of flowering time in barley through nested association mapping. BMC Genomics 16:290. doi:10.1186/s12864-015-1459-7

McCouch, S.R., M.H. Wright, C.W. Tung, L.G. Maron, K.L. McNally, M. Fitzgerald, et al. 2016. Open access resources for genome wide association mapping in rice. Nat. Commun. 7:11346. doi:10.1038/ncomms11346

Mitchell, R., P. Dupree, and P. Shewry. 2007. A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. Plant Physiol. 144:43–53. doi:10.1104/pp.106.094995

Mohanty, S. 2013. Rice facts. Trends in global rice consumption. Rice Today 12: 44–45.

Numan, M., and N. Bhosle. 2006. Alpha-*l*-arabinofuranosidases: The potential applications in biotechnology. J. Ind. Microbiol. Biotechnol. 33:247–260. doi:10.1007/s10295-005-0072-1

Ohdan, T., P.B. Francisco, T. Sawada, T. Hirose, T. Terao, H. Satoh, et al. 2005. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. J. Exp. Bot. 56:3229–3244. doi:10.1093/jxb/eri292

Okazaki, Y., and K. Saito. 2016. Integrated metabolomics and phytochemical genomics approaches for studies on rice. Gigascience 5:11. doi:10.1186/s13742-016-0116-7

Parisseaux, B., and R. Bernardo. 2004. In silico mapping of quantitative trait loci in maize. Theor. Appl. Genet. 109:508–514. doi:10.1007/s00122-004-1666-0

Patindol, J., and Y. Jabe-Wang. 2003. Fine structures and physicochemical properties of starches from chalky and translucent rice kernels. J. Agric. Food Chem. 51:2777–2784. doi:10.1021/jf026101t

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909. doi:10.1038/ng1847

Quero, G., S. Simondi, V. Bonnecarrere, and L. Gutierrez. 2017. Clusterhap: Clustering genotypes in haplotypes. R package version 0.1.0. The R Foundation. https://CRAN.R-project.org/package=clusterhap (accessed 5 June 2018).

R Core Team. 2017. R: A language and environment for statistical computing. The R Foundation. https://www.R-project.org (accessed 5 June 2018).

Rosas, J.E., S. Martínez, P. Blanco, F. Pérez de Vida, V. Bonnecarrere, G. Mosquera, et al. 2017. Resistance to multiple temperate and tropical stem and sheath diseases of rice. Plant Genome 11:170029. doi:10.3835/plantgenome2017.03.0029

Schweder, T., and E. Spjotvoll. 1982. Plots of *P*-values to evaluate many tests simultaneously. Biometrika 69:493–502. doi:10.1093/biomet/69.3.493

Shibuya, N., and R. Nakane. 1984. Pectic polysaccharides of rice *Oryza sativa* cultivar norin-29 endosperm cell walls. Phytochemistry 23:1425–1429. doi:10.1016/S0031-9422(00)80479-3

Shibuya, N., and T. Iwasaki. 1978. Polysaccharides and glycoproteins in rice endosperm cell-wall. Agric. Biol. Chem. 42:2259–2266. doi: 10.1271/bbb1961.42.2259

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, et al. 2015. Genomic selection and association mapping in rice *(Oryza sativa)*: Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 11(2):e1004982. doi:10.1371/journal.pgen.1004982

Strohmeier, M., M. Hrmova, M. Fischer, A.J. Harvey, G.B. Fincher, and J. Pleiss. 2004. Molecular modeling of family GH16 glycoside hydrolases: Potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the *Poaceae*. Protein Sci. 13:3200–3213. doi:10.1110/ps.04828404

Su, J. 2000. Starch synthesis and grain filling in rice. In: K. Gupta and N. Kaur, editors, Carbohydrate reserves in plants—Synthesis and regulation. Developments in Crop Sci. 26. Elsevier, Amsterdam. p. 107–124.

Swarts, K., H. Li, J.A. Romero Navarro, D. An, M.C. Romay, S. Hearne, et al. 2014. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7:1–12. doi:10.3835/plantgenome2014.05.0023

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729. doi:10.1093/molbev/mst197

National Agriculture and Food Research Organization. 2017. The rice annotation project. National Agriculture and Food Research Organization. http://rapdb.dna.affrc.go.jp (accessed 5 June 2018).

Tian, F., P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, et al. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. 43:159–162. doi:10.1038/ng.746

Tian, Z., Q. Qian, Q. Liu, M. Yan, X. Liu, and C. Yan, et al. 2009. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. Proc. Natl. Acad. Sci. 106:21760–21765. doi:10.1073/pnas.0912396106

Vandeputte, G., and J. Delcour. 2004. From sucrose to starch granule to starch physical behaviour: A focus on rice starch. Carbohydr. Polym. 58:245–266. doi:10.1016/j.carbpol.2004.06.003

Yamakawa, H., and M. Hakata. 2010. Atlas of rice grain filling-related metabolism under high temperature: Joint analysis of metabolome and transcriptome demonstrated inhibition of starch accumulation and induction of amino acid accumulation. Plant Cell Physiol. 51:795–809. doi:10.1093/pcp/pcq034

Yan, W., J.N. Rutger, R.J. Bryant, H.E. Bockelman, R.G. Fjellstrom, M.H. Chen, et al. 2007. Development and evaluation of a core subset of the USDA rice germplasm collection. Crop Sci. 47:869–876. doi:10.2135/cropsci2006.07.0444

Yu, J., J. Holland, M. McMullen, and E. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551. doi:10.1534/genetics.107.074245

Yu, Y., R.A. Wing, and J. Li. 2013. Grain quality. In: Q. Zhang and R.A. Wing, editors, Genetics and genomics of rice. Springer, New York. p. 237–254.

Zader, A. 2011. Technologies of quality: The role of the Chinese state in guiding the market for rice. EASTS 5:461–477. doi:10.1215/18752160-1458155

Zeng, D., Z. Tian, Y. Rao, G. Dong, Y. Yang, L. Huang, et al. 2017. Rational design of high-yield and superior-quality rice. Nat. Plant. 3:17031. doi:10.1038/nplants.2017.31

Zhang, P., K. Dreher, A. Karthikeyan, A. Chi, A. Pujar, R. Caspi, et al. 2010. Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. Plant Physiol. 153:1479–1491. doi:10.1104/pp.110.157396

Zhao, K., C.W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, and A.H. Price. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat. Commun. 2:467. doi:10.1038/ncomms1467

Zhao, X., L. Zhou, K. Ponce, and G. Ye. 2015. The usefulness of known genes/QTLs for grain quality traits in an *indica* population of diverse breeding lines tested using association analysis. Rice (N. Y.) 8:29–41. doi:10.1186/s12284-015-0064-3

Zhou, L., S. Liang, K. Ponce, S. Marundon, G. Ye, and X. Zhao. 2015. Factors affecting head rice and chalkiness in *indica* rice. Field Crops Res. 172:1–10. doi:10.1016/j.fcr.2014.12.004

Zhu, Z., F. Zhang, H. Hu, A. Bakshi, M.R. Robinson, J.E. Powell, et al. 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48:481–487. doi:10.1038/ng.3538