

Sequence variations in *OsAGPase* significantly associated with amylose content and viscosity properties in rice (*Oryza sativa* L.)

FU-HAO LU¹ AND YONG-JIN PARK^{1,2*}

¹ Department of Plant Resources, College of Industrial Sciences, Kongju National University, Yesan 340-802, Republic of Korea

² Legume Bio-Resource Center of Green Manure (LBRCGM), Kongju National University, Yesan 340-802, Republic of Korea

(Received 7 February 2012; revised 29 May 2012; accepted 22 June 2012)

Summary

Sequence-based variations in starch synthesis-related genes (SSRGs) exert a basic influence on the determination of eating quality in rice (*Oryza sativa* L.). This study aimed to investigate the relationship between the sequence variations from parts of 10 SSRGs and the amylose content (AC) plus rapid viscosity analysis (RVA) profiles in a heuristic rice core set by association mapping (AM). In total, 86 sequence variations were found in 10 sequenced amplicons, including 79 single-nucleotide polymorphisms (SNPs), six insertion-deletions (InDels) and one polymorphic simple sequence repeat (SSR). Among them, 61 variations were exon-based, of which 41 should lead to amino acid changes. Four subpopulations were revealed by population structure analysis based on 170 genome-wide SSR genotypes. The final AM showed a sum of four significant associations between three phenotypic indices and three sequence variations. An *ADP-glucose pyrophosphorylase small unit 1* (*OsAGPS1*) SNP (A–G) was significantly associated with increased AC ($P < 0.001$, $R^2 = 15.6\%$), while a 12-bp deletion of *AGPase large unit 4* (*OsAGPL4*) was significantly related to the decreased breakdown viscosity (BDV) ($P < 0.001$, $R^2 = 16.6\%$) in both general linear model (GLM) and mixed linear model (MLM). This study provides a new perspective of allele mining for breeding strategies based on marker-assisted selection.

1. Introduction

Eating quality is generally affected by cohesiveness, tenderness and gross of cooked rice, and could be mainly represented by amylose content (AC), gel consistency (GC), gelatinization temperature (GT) and protein content in rice (*Oryza sativa* L.) (Juliano *et al.*, 1965; Bett-Garber *et al.*, 2001). AC is related to the appearance and texture of rice and it affects cooking and eating quality. GC is responsible for softness in rice, whereas GT is a physical characteristic responsible for cooking time and the capacity of absorbing water when cooking (Sun *et al.*, 2006). Rice starch viscosity profile, also known as the rapid viscosity analysis (RVA) profile, could greatly reflect the physiochemical properties, palatability and cooking and eating quality by simulating the cooking process (Cai *et al.*, 2011), and was demonstrated to be closely

related to AC and GC (Hu *et al.*, 2004). AC and RVA profiles are important physical and chemical indices that were used to reflect the rice-eating quality (Shu *et al.*, 1998; Bao *et al.*, 1999, 2004; Roferos *et al.*, 2008).

Among the four main elements of eating quality, the former three (AC, GC and GT) are closely related to starch properties. Thus, one may reasonably propose that starch synthesis-related genes (SSRGs) play an important role in determining rice-eating quality. From the primary carbon accumulation in the Calvin cycle to the final products, many enzymes and regulators contribute to this intact pathway. Among them, ADP-glucose pyrophosphorylase (AGPase; EC 2.7.7.21) is a critical rate-limiting enzyme with an allosteric regulatory function (Kato *et al.*, 2007). Upstream enzymes mainly affect rice yield, whereas those downstream greatly influence rice quality (Bao *et al.*, 2012). Starch synthases (SSs) and starch debanching enzymes also affect starch components and structure, which further affect eating quality. Most enzymes in the starch pathway have many isoforms,

* Corresponding author: Dr Yong-Jin Park, Legume Bio-Resource Center of Green Manure (LBRCGM), Kongju National University, Yesan 340-802, Republic of Korea. Tel: +82 41 330 1201. Fax: +82 41 330 1209. E-mail: yjpark@kongju.ac.kr

which perform similar functions in a temporal or spatial way (Ohdan *et al.*, 2005). Different gene regulation mechanisms, different gene expression levels or different enzyme activities could lead to different eating quality. Good eating quality is a product of collaborative reactions regulated by multiple enzymes and regulators.

Currently, increasing attention is being paid to breeding strategies of rice with great eating quality (Tan *et al.*, 1999; Sun *et al.*, 2006; Lestari *et al.*, 2009). Due to natural mutation and selection, abundant sequence differences exist among multiple accessions from diverse resources (Jangsutthivorawat & Volkaert, 2009) that were characterized by molecular markers, such as single-nucleotide polymorphisms (SNPs), insertion-deletions (InDels) and simple sequence repeats (SSRs). Natural variation mining in these genes or regulator sequences will assist us to locate important alleles for eating quality using association mapping (AM) (Cardon & Bell, 2001). Tran *et al.* (2011) reported that a C/T transition in the tenth exon of the *Wx* gene showed a strong association with GC using a recombinant inbred line population. Sun *et al.* (2011) confirmed that several SNPs in *starch branching enzyme 1 (SBE1)*, *soluble SS 1 (SSS1)* and *SSS2a* were significantly associated with AC, alkali digestion value, setback viscosity (SBV), consistency viscosity and pasting temperature (PT). These SNPs provide a possibility to enrich key alleles in one cultivar using a crossing strategy.

To find out those sequence variations contributing to good eating quality, AM is a good candidate approach to detect significant correlations between genotypes and phenotypes in a specified or unrelated mapping population on the basis of linkage disequilibrium (Remington *et al.*, 2001; Zhang *et al.*, 2011; Zhao *et al.*, 2011). The most attractive aspect of AM lies in its ease when establishing mapping populations (Jin *et al.* 2010; Li *et al.*, 2011). The rice core set, also known as core collection, represents the largest genetic diversity within the smallest number of accessions (Brown, 1989). It has proved to be one of the most valuable resources for genetic research and molecular breeding, especially for allele mining (Zhang *et al.*, 2008; de Oliveira Borba *et al.*, 2010). Therefore, more sequence variations are expected to be present in worldwide diverse varieties. The discovery of those sequence variations or alleles with a partial contribution to phenotypes should be helpful in rice breeding using the marker-assisted strategy. In this study, we investigated sequence variations among 10 SSRGs, and evaluated the AC and RVA profiles in a heuristic core set of 166 accessions, aimed at finding some effective nucleotide sequence variations closely related to AC as well as RVA profiles, which would provide basic guidance for breeding good eating-quality rice.

2. Materials and methods

(i) Plant material

Morphological and molecular genetic diversities of rice (*O. sativa* L.) were evaluated in a heuristic core set of 166 accessions (Chung & Park, 2009). This core set collection was selected by the advanced maximization strategy implemented through a modified heuristic algorithm using Powercore program (Kim *et al.*, 2007) from 4406 worldwide rice accessions, which were obtained from Genebank of the Rural Development Administration (RDA), Republic of Korea. Plants were grown in the Kongju National University farm field in a natural environment during 2009. Young plant leaves were collected at the grain-filling stage for each accession and stored at -80°C until genomic DNA extraction. The mature seeds for each accession were harvested in bulk for phenotypic data analysis. By removing those accessions with lots of missing genotype and phenotypic data ($>25\%$), 104 accessions originated from 28 countries over the world in the heuristic core set were finally chosen for association analysis in this study (see Supplementary Table S1 available at <http://journals.cambridge.org/GRH>).

(ii) Phenotypic performance

AC determination was performed using a method based on colorimetric analysis described by Cho *et al.* (2008). Starch viscosity profiles were determined on a Rapid Visco Analyzer (serial-3; Newport Scientific, Warriewood, NSW, Australia). A total of 3-g starch from each accession was dispersed in 25 ml of distilled water in the viscometer test canister. The time-temperature regimen used was set up as follows: idle temperature 50°C for 1 min, heated from 50 to 95°C in 4.7 min, then held at 95°C for 2.5 min. The sample was subsequently cooled down to 50°C in 3.7 min, followed by 2 min at 50°C . The profiles of RVA parameters, including peak viscosity (PV), trough viscosity (TV), breakdown viscosity ($\text{BDV} = \text{PV} - \text{TV}$), final viscosity (FV), setback viscosity ($\text{SBV} = \text{FV} - \text{PV}$) and PT were investigated for the core set of 166 accessions. For all traits, the Kolmogorov-Smirnov test was applied to determine a normal distribution at a significance of 0.05.

(iii) Genotype data

Genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method. A total of ten genes (six AGPase units, GBSSI, pullulanase, SSSIIa and one SSSIIa-like gene) were selected based on their important roles in the starch synthesis in rice endosperm. The amplicons were chosen by

Table 1. Information on genes and primers investigated

| Primer ID | Forward primer | Reverse primer | T_m | Gene locus | Amplicon (bp) |
|-------------|----------------------------|---------------------------|-------|------------------|---------------|
| OsAGPL1 | ctttactgcacaaaagcctatcaact | gttaaaggctattgactttgttacc | 58 | LOC_Os05g50380 | 552 |
| OsAGPL2 | cgttgtaactgtgaatgaataatc | acacatgcatcttatgttgataag | 58 | LOC_Os01g44220 | 805 |
| OsAGPL3 | cttgattgttttctggtgttg | ttctagtcaactgcatggctatag | 58 | LOC_Os03g52460 | 710 |
| OsAGPL4 | caaatgcatgacgaaatgtgctg | acctgcacctcttagaaccttc | 64 | LOC_Os07g13980 | 521 |
| OsAGPS1 | cctacagcatgaactgagagaag | tgtactcctgttgattcactcttc | 64 | LOC_Os09g12660 | 824 |
| OsAGPS2b | tatgttctacatgaataaggaattcc | tcacctataaattagaggatactgc | 58 | LOC_Os08g25734.2 | 377 |
| OsGBSSI | caaaccttaacaattcaattcagtg | cgtacctgtctgcaaccttgatc | 58 | LOC_Os06g04200 | 723 |
| OsPUL3 | tcagatgccctagctctgttgat | cagttgcataaatcaagtagacag | 58 | LOC_Os04g08270 | 834 |
| OsSSIIaChr6 | tcgcttcggttcggttcggct | cgtcgcgtctgggatgacgaac | 64 | LOC_Os06g12450 | 823 |
| OsSSIIaChr7 | tgcgtcgacgcaagggtcggttaa | ggtgacgacgacaaggttcttcg | 64 | LOC_Os07g13880 | 764 |

T_m , melting temperature.

Gene locus: from rice genome annotation project (Pseudomolecules version 6.1, <http://rice.plantbiology.msu.edu/>).

searching NCBI nucleotide database using blastn (<http://blast.ncbi.nlm.nih.gov>), and the primers were designed using the corresponding genomic sequences retrieved from NCBI (Table 1). PCR was performed using H-Taq DNA polymerase (Solgent, Daejeon, Republic of Korea) on a GenePro Thermal Cycler (Bioler, Hangzhou, China). Reaction components were as follows: 2.5 μ l of H-Taq buffer (10 \times), 0.5 μ l of dNTP (10 mM), 1.0/1.0 μ l of forward/reverse primers (10 pmol), 0.25 μ l of H-Taq DNA polymerase, 1.0 μ l of genomic DNA (50 ng/ μ l) and 18.75 μ l of ddH₂O. Cycling profiles were set as follows: denaturation at 95 °C for 15 min; 40 cycles of denaturation at 95 °C for 20 s, specific annealing temperature (Table 1) for 40 s and extension at 72 °C for 1 min; and a final extension at 72 °C for 3 min. PCR products were purified using the PCR Purification Kit (Solgent, Daejeon, Republic of Korea), and then sent to Solgent Co. Ltd (Daejeon, Republic of Korea) for sequencing based on a ABI 3730 sequencing platform (Applied Biosystems, Foster City, CA, USA). Rare alleles were re-sequenced with additional independent PCR products.

(iv) Candidate sequence variation discovery and genetic analysis

To discover the sequence variations, Sequence Analysis software (v. 5.4; Applied Biosystems) was used for base calling and sequence masking and Variant Reporter (v. 1.1; Applied Biosystems) was employed to align the sequences and find variations. Basic statistics were calculated using the PowerMarker genetic analysis package (v. 3.25; Liu & Muse, 2005) to measure the diversity at each locus, including variation frequency, gene diversity and polymorphism information content (PIC). The observed heterozygosity, expected heterozygosity and Shannon's information index were calculated using the POPGENE software (v. 1.32; Yeh & Boyle, 1997).

Previously, we characterized this heuristic core set using 170 genome-wide SSR markers (see Supplementary Table S1). To reduce spurious deviations, a dendrogram was constructed from a distance matrix based on the SSR genotypes using the unweighted pair group method and the arithmetic averages (UP-GMA) algorithm embedded in the PowerMarker program, and edited using the MEGA5 program (Tamura *et al.*, 2011).

(v) AM

The hypothesis behind associating those sequence variations with AC and RVA profiles in the presence of population structure was tested using a general linear model (GLM) and a mixed linear model (MLM), as described by Yu *et al.* (2006), in the TASSEL program (v. 2.1; Bradbury *et al.*, 2007). To reduce possible spurious associations, the population structure matrix (Q matrix) and the relative kinship matrix (K matrix) were calculated based on the 170 genome-wide SSR data (Supplementary Table S1). The Q matrix at $k=4$ was identified by running the Structure program (v. 2.3.3; Hubisz *et al.*, 2009) following a method described previously (Falush *et al.* 2003), whereas the Loiselle kinship coefficients (Loiselle *et al.*, 1995) between accessions were estimated using the SPAGeDi program (v. 1.2 g; Hardy & Vekemans, 2002). Only markers with a variation frequency of 5% or more were included in the association analysis. For GLM, to account for type I error bias, the P -value was adjusted for multiple tests with 1000 permutations using a procedure proposed by Whitt & Buckler (2003). The P -value determined whether a locus was associated with a marker, and the R_{sq_marker} evaluated the magnitude of the locus effects. For MLM, a correction for multiple testing was applied using the false discovery rate (FDR) method to confirm the association significance (Storey & Tibshirani, 2003).

Table 2. Descriptive statistics for AC and starch viscosity in a core set of 104 accessions

| | AC | PV | TV | BDV | FV | SBV | PT |
|----------|----------|----------|---------|----------|---------|---------|-------|
| Mean | 19.14 | 212.89 | 146.10 | 66.79 | 252.85 | 39.96 | 75.26 |
| Range | 2.8 | 246.9 | 234.9 | 111.8 | 385.7 | 190.3 | 18.1 |
| Min | 18.1 | 85.8 | 13.5 | 19.3 | 20.5 | -65.3 | 68.1 |
| Max | 20.9 | 332.8 | 248.4 | 131.2 | 406.2 | 125.0 | 86.1 |
| StDev | 0.54 | 43.84 | 35.73 | 21.06 | 51.15 | 33.70 | 4.86 |
| Skewness | 0.921 | -0.138 | -0.247 | 0.627 | -0.544 | -0.645 | - |
| Kurtosis | 0.820 | 0.436 | 1.584 | 0.647 | 4.040 | 0.794 | - |
| AC | | | | | | | |
| PV | -0.413** | | | | | | |
| TV | -0.411** | 0.879** | | | | | |
| BDV | -0.163 | 0.590** | 0.134 | | | | |
| FV | -0.388** | 0.759** | 0.911** | 0.034 | | | |
| SBV | -0.051 | -0.149 | 0.238* | -0.715** | 0.531** | | |
| PT | 0.034 | -0.377** | -0.163 | -0.508** | -0.064 | 0.393** | |

AC, amylose content; PV, peak viscosity; TV, trough viscosity; BDV, breakdown viscosity; FV, final viscosity; SBV, setback viscosity; PT, pasting temperature; StDev, standard deviation.

**Correlation is significant at the 0.01 level (two-tailed).

*Correlation is significant at the 0.05 level (two-tailed).

3. Results

(i) Statistics and distribution of phenotype data

Detailed information of phenotypic data means, ranges and correlations for amylose and RVA parameters (PV, TV, BDV, FV, SBV and PT) are summarized in Table 2. Among the correlations between each other, more than half of them were nominally significant at $P < 0.01$. Notably, PV showed significant positive correlations with TV (0.879) and FV (0.759). The correlation (0.911) between FV and TV confirmed the most significantly positive, whereas BDV showed the most significant negative correlation (-0.715) with SBV. The distribution of all traits except for PT conformed to a classical Gaussian distribution at a significance cut-off of 0.05 in the Kolmogorov–Smirnov test, which was proved to be suitable for AM analysis (Fig. 1 and Table 2).

(ii) Sequence variation discovery in SSRGs

Among the 10 amplicons sequenced, the sequenced length ranged from 377 (*OsAGPS2*) to 834 bp (*OsPUL3*). Overall, 86-sequence variations were discovered from nine amplicons of the ten genes sequenced, including 79 SNPs, six InDels and one SSR. No SNP loci were found in the *OsAGPL1* gene amplicon. Specifically, 25 and 61 variations were separately found in the intronic and exonic parts of the nine amplicons. Among the 61 exonic variations, 41 (47.67% in 86) in exon regions should have resulted in amino acid changes by analysing the genetic code changes. Most SNPs were A/G (39.53%) and C/T (34.88%) transition substitutions (Table 3). Notably, 47 of the 86 total sequence variations were

discovered from the *OsSSIIaChr7* gene on chromosome 7. InDels were found in the first exon and first intron of *OsAGPL4*, 5-untranslated region of *OsAGPS2* and the first exon of *OsSSIIchr6*, in addition to two InDels in the *OsGBSSI* gene. Notably, the two InDels in the *OsGBSSI* were found separately in first exon (23 bp insertion) and first intron (16 bp deletion), both of which were confirmed to be rare variations (variation frequency < 0.05). In particular, the 23 bp InDel was expected to result in a frame shift during translation process. The polymorphic SSR appeared in the second exon of *OsSSIIchr6*.

(iii) Sequence polymorphisms

Each sequence variation locus revealed two variation types through a total of 86 variation loci among the 104 rice accessions (see Supplementary Table S2 available at <http://journals.cambridge.org/GRH>). According to the distribution of variation loci, rare variations (frequency < 0.05) accounted for 21.51%, while the common ($0.05 < \text{frequency} < 0.2$) and abundant variations (frequency > 0.2) made up 6.39 and 72.10%, respectively. These results indicated the presence of a relatively large proportion of rare and abundant variations, and most variations were of a low frequency among the rice studied (Fig. 2). The frequency of heterozygous alleles in one individual was rare, but it existed. The averaged observed heterozygosity was obviously lower than the averaged expected heterozygosity, indicating the presence of heterozygosity deficit probably due to forces such as inbreeding. The genetic diversity for all variation loci varied from 0.0190 to 0.4936, with an average of 0.2256. PIC values ranged from 0.0189 to 0.3718, and

Table 3. Summary of variation patterns discovered from ten starch synthesis-related genes

| Gene_ID | Summary | | | Detail | | | | | | | |
|-------------------|------------|----------|---------|--------|-----|-----|-----|-----|-----|-------|-----|
| | Intron Var | Exon Var | AAc Var | A/C | A/G | A/T | C/G | C/T | G/T | InDel | SSR |
| <i>OsAGPL1</i> | 0 | 0 | 0 | – | – | – | – | – | – | – | – |
| <i>OsAGPL2</i> | 2 | 0 | 0 | – | 2 | – | – | – | – | – | – |
| <i>OsAGPL3</i> | 1 | 0 | 0 | – | – | – | – | 1 | – | – | – |
| <i>OsAGPL4</i> | 1 | 2 | 1 | – | 1 | – | – | – | – | 2 | – |
| <i>OsAGPS1</i> | 1 | 3 | 2 | 1 | 1 | 1 | – | – | 1 | – | – |
| <i>OsAGPS2</i> | 3 | 3 | 2 | – | 2 | – | 1 | 2 | – | 1 | – |
| <i>OsGBSSI</i> | 1 | 1 | 1 | – | – | – | – | – | – | 2 | – |
| <i>OsPUL3</i> | 8 | 5 | 4 | 2 | 2 | – | 4 | 4 | 1 | – | – |
| <i>OsSSIIchr6</i> | 3 | 5 | 4 | 1 | 1 | – | 1 | 3 | – | 1 | 1 |
| <i>OsSSIIchr7</i> | 5 | 42 | 27 | – | 25 | – | – | 20 | 2 | – | – |
| Total | 25 | 61 | 41 | | | | | | | | |

Intron Var, number of variation in introns; Exon Var, number of variations in exons; AAc Var, number of variations presumed to result in amino acid changes in the exon parts; InDel, insertion-deletion; SSR, simple sequence repeat.

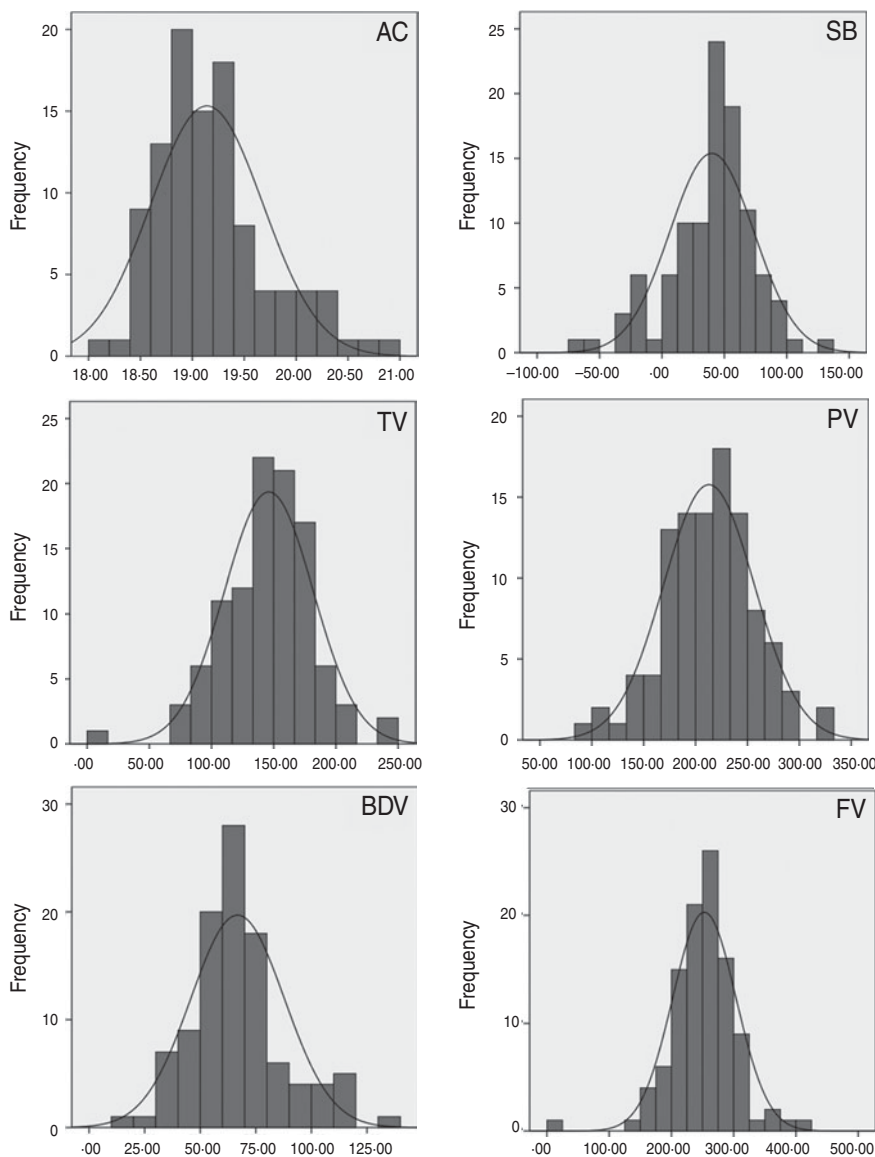


Fig. 1. Gaussian distribution of eating-quality evaluation data for AC and RVA profiles.

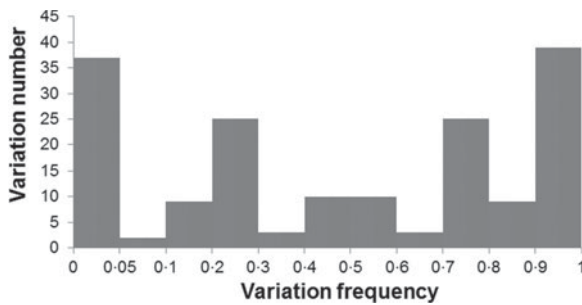


Fig. 2. Variation frequency distribution in 104 rice accessions studied.

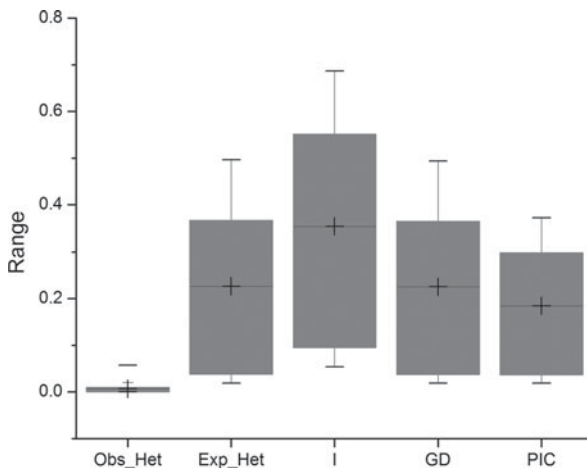


Fig. 3. Boxplot presentation of sequence variation statistics. Obs_Het, observed heterozygosity; Exp_Het, expected heterozygosity; I, Shannon's information index; GD, genetic diversity; PIC, polymorphism information content (PIC).

averaged 0.1848 (Fig. 3). The broad ranges of genetic diversity and Shannon's information index represented the multiple-level inheritable variations in the selected rice accessions.

(iv) Population structure

A model-based clustering method was performed using our previous 170 genome-wide SSR genotyping data in 104 accessions (Supplementary Table S1) to infer the exact population number using Structure program. Estimated likelihood [$\ln P(D)$] values for a given k in five independent runs showed consistent results, but the $\ln P(D)$ distribution did not indicate a clear plateau for the true k . Therefore, another *ad hoc* quantity Δk was employed to overcome the difficulties in interpreting the real k value (Evanno *et al.*, 2005). The ΔK showed a clear peak at the true value of $k=4$ (Fig. 4a), which was in accordance with previous data. A sum of 93 accessions was assigned to four subpopulations (Fig. 4b), of which each shared at least 75% ancestry, leaving 11 accessions as admixtures. The size of Pop2 and Pop3 were relatively

small, including only 8 and 19 individuals, respectively. This was supported by the alpha parameter ($\alpha=0.0464$), indicating that most accessions in each subpopulation originated from one primary ancestor, except the 11 admixed individuals (Falush *et al.*, 2003; Ostrowski *et al.*, 2006). The genetic distance-based UPGMA tree supported the model-based population structure assignment (Fig. 4c).

(v) Association analysis

Marker-trait associations ($P<0.001$) for traits were evaluated in 50 variation loci after removing 36 loci with rare frequency using the program TASSEL 2.1. A total of four associations were detected between three loci and three trait indices in GLM and/or MLM (Table 4). Notably, one 12-bp InDel (TCCGCCGCCGCC) in the first exon of the *OsALPL4* gene were significantly associated with BDV at $P<0.001$ in both GLM and MLM, with an interpretation of up to 16.6% of the trait variations. Moreover, one SNP with an A/G transition, which should result in the encoded amino acid from asparagine to aspartic acid, from the 33rd nucleotide of *OsAGPS1*'s third exon was significantly associated with two trait indices (AC at $P<0.001$ for both models, and FV at $P<0.001$ for MLM model only), explaining up to 15.6% of the phenotype variation. Some sequence variations in introns were also detected to be significant in association with some traits. A 8-bp InDel (cgcgtgcg) in the first intron of *OsAGPL4* gene was closely associated with BDV at $P<0.001$ in GLM and MLM, mostly because this InDel showed exactly the same insertion- and deletion-types with the 12-bp InDel in the first exon described above in each investigated accession.

(vi) Allelic effects

Based on the population structure revealed by the 170 genome-wide SSR genotyping data, those 104 accessions investigated were assigned into four subgroups, which was consistent with our previous data using 130 core set accessions. With regard to the variations with highly significant associations ($P<0.001$), the allelic effect showed a subpopulation-specific pattern to some extent in Pop2 and Pop3, mostly because of the small population sizes. In *OsAGPL4*, a 12-bp deletion between the 106th and 117th nucleotide (TCCGCCGCCGCC) that encodes four amino acids (Ser-Ala-Ala-Ala) at the first exon was strongly associated with a dramatic decrease in BDV, and those accessions with the deletion were all assigned only to Pop1 (Fig. 5a). In *OsAGPS1*, a transition from A to G at the 133rd nucleotide of the third exon that changed the amino acid from asparagine to aspartic acid was closely related to a slight AC increase

Table 4. Associations between sequence variations and phenotype

| Gene | Location | Nucleotide variations | AAC | Trait | P_GLM | P_MLM | Rsqr_Marker |
|----------------|--------------------|-----------------------|-----------|-------|----------|----------|-------------|
| <i>OsAGPLA</i> | 1st exon 106–117th | del(TCCGCCGCCGCC) | del(SAAA) | BDV | 9.99E−04 | 2.64E−05 | 0.1656 |
| <i>OsAGPLA</i> | 1st intron 36th | del(cgcgtgcg) | – | BDV | 9.99E−04 | 2.64E−05 | 0.1656 |
| <i>OsAGPSI</i> | 3rd exon 133rd | AAT→GAT | N→D | AC | 9.99E−04 | 2.03E−04 | 0.1555 |
| | | | | FV | 6.00E−03 | 9.05E−04 | 0.1236 |

AAC, types of amino acid changes; P_GLM, adjusted *P*-values with 1000 permutations; P_MLM, *P*-values that were significant in FDR test; amino acid code: S, serine; A, alanine; N, asparagine; D, aspartic acid.

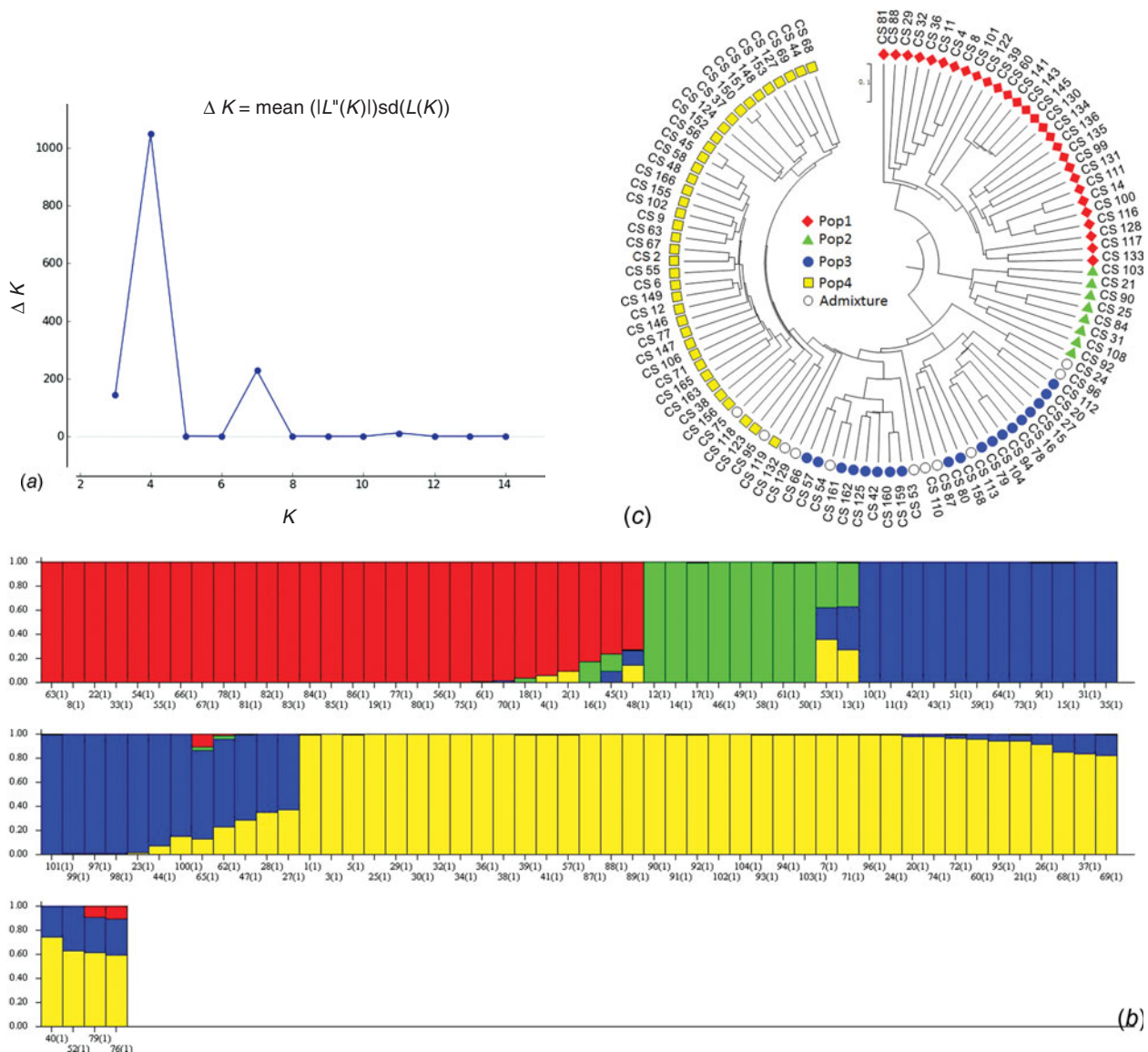


Fig. 4. Model-based structure and genetic distance based UPGMA dendrogram of the 104 rice accessions based on previous 170 genome-wide SSR markers used to build Q-matrix. (a) Determination of the *ad hoc* quantity ΔK ; (b) Model based ancestries of each accession; (c) UPGMA dendrogram based on a genetic distance matrix.

(Fig. 5b) and a significant FV decrease (Fig. 5c). According to the differences observed in AC and FV, those accessions with A and G seemed to be present only in Pop 3 and Pop 1 plus Pop 2, respectively, and then mixed in Pop 4.

4. Discussion

Eating quality represents an important index reflecting the rice quality, which was greatly controlled or regulated by those genes involved in starch and

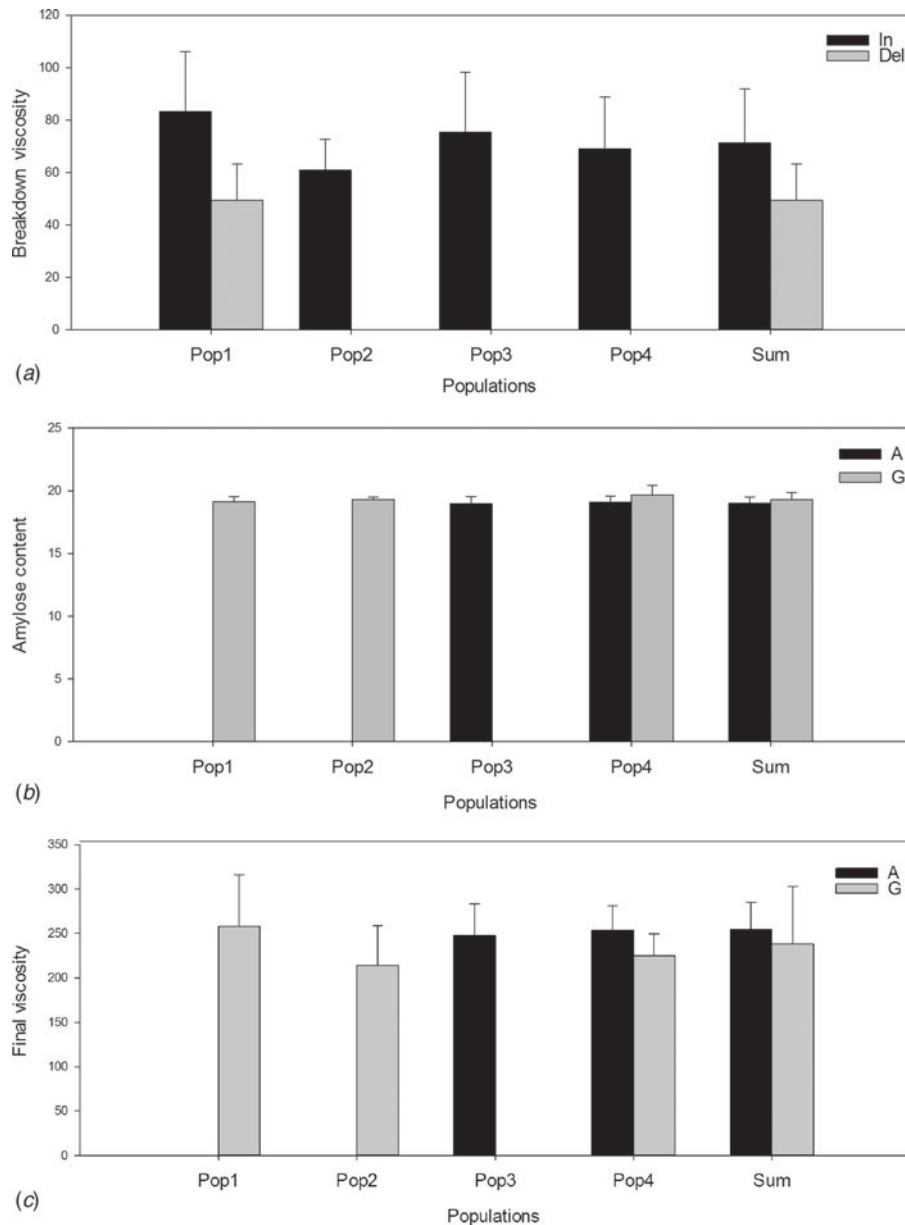


Fig. 5. Allele effect of those associations in four subpopulations with a significance level of 0.001. (a) Allelic effect of an insertion–deletion from *OsAGPL4* on BDV; allelic effect of an SNP from *OsAGPS1* on AC (b) and FV (c).

protein metabolism network. Each gene may play a partial role in determining eating quality by affecting carbohydrate partitioning, whereas some gene(s) may exert main-effect forces on the eating quality, such as the *Wx* gene, but the contributions from other enzymes could not be excluded. In the present study, we characterized 86 sequence-based variations, including SSRs, SNPs and InDels, from the 10 selected SSRG amplicons in the heuristic core set, and then evaluated their association relationship to eating quality indices (i.e. AC and RVA properties).

A large proportion (47.67%) of sequence variations detected in the 10 amplicons investigated was expected to cause amino acid changes, which should have potentially contributed to the relative phenotypes

among these core set accessions. However, a large proportion of the total detected variations notably confirmed to be of low frequency, accounting for rare alleles. The effect of rare alleles in the association population mean would be minimal, whereas it could be detected in a biparental population. This portion of rare variations with modest effects could be difficult to detect because they explain only a trivial fraction of the variance in a trait (Hirschhorn & Daly, 2005). Therefore, to reduce spurious associations, this portion of rare alleles in the population was treated as missing data, and excluded from association tests (Brescghello & Sorrells, 2006; Liu *et al.*, 2010; Lu *et al.*, 2010). Furthermore, those variations discovered from *OsSIIchr7* comprised over half of the total

86 variations in a 764-bp amplicon, indicating that it is most probably a non-functional gene, despite the fact that it was annotated as a soluble SS-like protein (protein ID: BAC16084.1).

Most previous studies found that AC is greatly determined by *Wx*, *SSI*, *SSII-3*, *SSIII-2* and *PUL* (Roferos *et al.*, 2008; Tian *et al.*, 2009), while RVA profiles were mainly controlled by the *Wx* gene, encoding GBSSI enzyme on rice chromosome 6 (Bao *et al.*, 1999, 2003). However, in the present study, two InDels from *OsGBSSI* gene was declared to be rare variations, and should be excluded during AM. Huang *et al.* (2010) reported a transition SNP (T/C) around 1 kb away from *OsGBSSI* gene was significantly associated with AC. Based on their resequencing data of 950 rice accessions (<http://202.127.18.221/RiceHap2/>), only three sequence variations with major frequency were found in the fifth, eighth and ninth exons, indicating that the coding region of *OsGBSSI* gene was rather conservative. It is reasonable for the difficulties in finding some sequence variations with major frequency in the *OsGBSSI* coding region.

Instead, an SNP from *OsAGPS1* and an InDel from *OsAGPL4* were identified as significantly associated with AC and RVA profiles. It is partly supported by the important role of AGPase in starch synthesis in the previous reports. AC could be significantly affected by *AGPlar* at the step from Glc-1-P to ADP-Glc catalysed by AGPase (Tian *et al.*, 2009). Kato *et al.* (2010) investigated the nucleotide variations in *OsAGPS2* and *OsAGPL2*, and found that alleles at *OsAGPS2* may associate with the grain-filling degree and AGPase activity in the extra-heavy panicle type of rice. Bao *et al.* (2012) also reported an InDel in *OsAGPL4* and an SNP in *OsAGPL2* were associated with grain weight in two environments and in one environment, respectively, by investigating the nucleotide variations in the *OsAGP* genes.

This paper was dedicated to unveil those desired natural sequence variations in SSRGs significantly associated with eating quality indicators for rice breeding using AM strategy. It does not conflict with those previous studies determining the main-effect gene(s) for eating quality using genome-wide mapping strategy. Undoubtedly, some SSRGs, especially the *OsGBSSI* gene, are necessary to be conserved through the long evolution history to ensure the plant progenies' survival (Mason-Gamer *et al.*, 1998). The significant associations between gene-based variations in *AGPase* subunit genes and eating quality do not exclude the importance of other genes including the *OsGBSSI* (Cui *et al.*, 2006).

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2010-0022616).

5. Declaration of interest

None.

6. Supplementary material

The supplementary Tables S1 and S2 are available online at <http://journals.cambridge.org/GRH>.

References

- Bao, J. S., Corke, H., He, P. & Zhu, L. H. (2003). Analysis of quantitative trait loci for starch properties of rice based on an RIL population. *Acta Botanica Sinica* **45**, 986–994.
- Bao, J. S., He, P., Xia, Y. W., Chen, Y. & Zhu, L. H. (1999). Starch RVA profile parameters of rice are mainly controlled by *Wx* gene. *Chinese Science Bulletin* **44**, 2047–2051.
- Bao, J. S., Lu, Y., Yang, F., Zhang, G., Shao, Y. F., Corke, H. & Sun, M. (2012). Nucleotide polymorphisms in *OsAGP* genes and their possible association with grain weight of rice. *Journal of Cereal Science* **55**, 312–317.
- Bao, J. S., Shen, S. Q. & Xie, J. K. (2004). A study on the improvement of eating quality of early *indica* rice with the assistance of RVA. *Molecular Plant Breeding* **2**, 49–53.
- Bett-Garber, K. L., Champagne, E. T., McClung, A. M., Moldenhauer, K. A., Linscombe, S. D. & McKenzie, K. S. (2001). Categorizing rice cultivars based on cluster analysis of amylose content, protein content and sensory attributes. *Cereal Chemistry* **78**, 551–558.
- 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**, 2633–2635.
- Bresegghello, F. & Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* **172**, 1165–1177.
- Brown, A. H. D. (1989). Core collections: a practical approach to genetic resources management. *Genome* **31**, 818–824.
- Cai, Y. X., Liu, C. X., Wang, W., Zhang, H. X., Zhang, Z. J., Yang, J. & Tang, H. Z. (2011). Dynamic differences of the RVA profile and gel consistency in two rice varieties with similar apparent amylose content during grain filling. *Scientia Agricultura Sinica* **44**, 2439–2445.
- Cardon, L. R. & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics* **2**, 91–99.
- Cho, Y. H., Lee, S. Y., Kim, S. M., Yu, J. W., Lee, J. R., Hong, H. C., Kim, J. B., Ma, K. H., Kwon, T. R., Kang, H. K., Lee, G. A., Gwag, J. G., Kim, T. S. & Park, Y. J. (2008). Amylose, tocopherol, free sugar and fatty acid content in selected mutant lines of *Oryza sativa* cv. Shindongjin. *Journal of Crop Science and Biotechnology* **11**, 83–90.
- Chung, J. W. & Park, Y. J. (2009). Genetic diversity and population structure of Korean rice core collection. *Korean Journal of International Agriculture* **21**, 282–288.
- Cui, Y. H., Wu, J. G., Shi, C. H., Littell, R. C. & Wu, R. L. (2006). Modelling epistatic effects of embryo and endosperm QTL on seed quality traits. *Genetics Research* **87**, 61–71.
- de Oliveira Borba, T. C., Brondani, R. P. V., Bresegghello, 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**, 515–524.

- Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611–2620.
- Falush, D., Stephens, M. & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587.
- Hardy, O. J. & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618–620.
- Hirschhorn, J. N. & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95–108.
- Hu P. S., Zhai H. Q., Tang S. Q. & Wan J. M. (2004). Rapid evaluation of rice cooking and palatability quality by RVA profile. *Acta Agronomica Sinica* **30**, 519–524.
- Huang, X. H., Wei, X. H., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C. Y., Zhu, C. R., Lu, T. T., Zhang, Z. W., Li, M., Fan, D. L., Guo, Y. L., Wang, A. H., Wang, L., Deng, L. W., Li, W. J., Lu, Y. Q., Weng, Q. J., Liu, K. Y., Huang, T., Zhou, T. Y., Jing, Y. F., Li, W., Lin, Z., Buckler, E. S., Qian, Q., Zhang, Q. F., Li, J. Y. & Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* **42**, 961–967.
- Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322–1332.
- Jangsutthivorawat, S. & Volkaert, H. (2009). Nucleotide polymorphism of starch synthesis genes in Thai rice. *Kasetsart Journal: Natural Sciences* **43**, 154–163.
- Jin, L., Lu, Y., Xiao, P., Sun, M., Corke, H. & Bao, J. S. (2010). Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theoretical and Applied Genetics* **121**, 475–487.
- Juliano, B. O., Onate, L. U. & Mundo, A. M. D. (1965). Relation of starch composition, protein content, and gelatinization temperature to cooking and eating qualities of milled rice. *Food Technology* **19**, 116–121.
- Kato, T., Shinmura, D. & Taniguchi, A. (2007). Activities of enzymes for sucrose-starch conversion in developing endosperm of rice and their association with grain filling in extra-heavy panicle types. *Plant Production Science* **10**, 442–450.
- Kato, T., Taniguchi, A. & Horibata, A. (2010). Effects of the alleles at and on the grain filling in extra-heavy panicle type of rice. *Crop Science* **50**, 2448–2456.
- Kim, K. W., Chung, H. K., Cho, G. T., Ma, K. H., Chandrabalan, D., Gwag, J. G., Kim, T. S., Cho, E. G. & Park, Y. J. (2007). PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* **23**, 2155–2162.
- Lestari, P., Ham, T. H., Lee, H. H., Woo, M. O., Jiang, W. Z., Chu, S. H., Kwon, S. W., Ma, K. H., Lee, J. H., Cho, Y. C. & Koh, H. J. (2009). PCR marker-based evaluation of the eating quality of japonica rice (*Oryza sativa* L.). *Journal of Agricultural and Food Chemistry* **57**, 2754–2762.
- Li, X. B., Yan, W. G., Agrama, H., Jia, L. M., Shen, X. H., Jackson, A., Moldenhauer, K., Yeater, K., McClung, A. & Wu, D. X. (2011). Mapping QTLs for improving grain yield using the USDA rice mini-core collection. *Planta* **234**, 347–361.
- Liu, K. & Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128–2129.
- Liu, L. H., Wang, L. X., Yao, Y., Zheng, Y. L. & Zhao, C. P. (2010). Association mapping of six agronomic traits on chromosome 4A of wheat (*Triticum aestivum* L.). *Molecular Plant Breeding* **1**, 1–10.
- Loiselle, B. A., Sork, V. L., Nason, J. & Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**, 1420–1425.
- Lu, Y. L., Zhang, S. H., Shah, T., Xie, C. X., Hao, Z. F., Li, X. H., Farkhari, M., Ribaut, J. M., Cao, M. J., Rong, T. Z. & Xu, Y. B. (2010). Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proceedings of the National Academy of Sciences USA* **107**, 19585–19590.
- Mason-Gamer, R. J., Weil, C. F. & Kellogg, E. A. (1998). Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molecular Biology and Evolution* **15**, 1658–1673.
- Ohdan, T., Francisco, P. B., Jr., Sawada, T., Hirose, T., Terao, T., Satoh, H. & Nakamura, Y. (2005). Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *Journal of Experimental Botany* **56**, 3229–3244.
- Ostrowski, M. F., David, J., Santoni, S., McKhann, H., Reboud, X., Le Corre, V., Camilleri, C., Brunel, D., Bouchez, D., Faure, B. & Bataillon, T. (2006). Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Molecular Ecology* **15**, 1507–1517.
- Remington, D. L., Ungerer, M. C. & Purugganan, M. D. (2001). Map-based cloning of quantitative trait loci: progress and prospects. *Genetics Research* **78**, 213–218.
- Roferos, L. T., Butardo, V. J. M., Fitzgerald, M. A. & Juliano, B. O. (2008). Association between alleles of the *Waxy* gene and traits of grain quality in Philippine seed board rice varieties. *Philippine Agricultural Scientist* **91**, 334–337.
- Shu, Q. Y., Wu, D. X., Xia, Y. W., Gao, M. W. & McClung, A. (1998). Relationship between RVA profile character and eating quality in *Oryza sativa* L. *Chinese Agricultural Sciences* **31**, 25–29.
- Storey, J. D. & Tibshirani, R. (2003). Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences USA* **100**, 9440–9445.
- Sun, M. M., Abdula, S. E., Lee, H. J., Cho, Y. C., Han, L. Z., Koh, H. J. & Cho, Y. G. (2011). Molecular aspect of good eating quality formation in japonica rice. *PLoS One* **6**, e18385.
- Sun, S. Y., Hao, W. & Lin, H. X. (2006). Identification of QTLs for cooking and eating quality of rice grain. *Rice Science* **13**, 161–169.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739.
- Tan, Y. F., Li, J. X., Yu, S. B., Xing, Y. Z., Xu, C. G. & Zhang, Q. F. (1999). The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. *Theoretical and Applied Genetics* **99**, 642–648.
- Tian, Z. X., Qian, Q., Liu, Q. Q., Yan, M. X., Liu, X. F., Yan, C. J., Liu, G. F., Gao, Z. Y., Tang, S. Z., Zeng, D. L., Wang, Y. H., Yu, J. M., Gu, M. H. & Li, J. Y. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities.

- Proceedings of the National Academy of Sciences USA* **106**, 21760–21765.
- Tran, N., Daygon, V., Resurreccion, A., Cuevas, R., Corpuz, H. & Fitzgerald, M. (2011). A single nucleotide polymorphism in the *Waxy* gene explains a significant component of gel consistency. *Theoretical and Applied Genetics* **123**, 519–525.
- Whitt, S. R. & Buckler, E. S. (2003). Using natural allelic diversity to evaluate gene function. In *Plant Functional Genomics* (ed. E. Grotewold), pp. 123–139. New York: Humana Press.
- Yeh, F. C. & Boyle, T. J. B. (1997). Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany* **129**, 157.
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S. & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**, 203–208.
- Zhang, J. N., Hao, C. Y., Ren, Q., Chang, X. P., Liu, G. & Jing, R. L. (2011). Association mapping of dynamic developmental plant height in common wheat. *Planta* **234**, 891–902.
- Zhang, Q. F., Li, J. Y., Xue, Y. B., Han, B. & Deng, X. W. (2008). Rice 2020: a call for an international coordinated effort in rice functional genomics. *Molecular Plant* **1**, 715–719.
- Zhao, K. Y., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., Norton, G. J., Islam, M. R., Reynolds, A., Mezey, J., McClung, A. M., Bustamante, C. D. & McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications* **2**, 467.