Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities

Zhixi Tian,a,b,1, Qian Qian,a,1, Qiaoquan Liu,a,1, Meixian Yan,b, Xinfang Liu,1, Changjie Yan,a, Guifu Liu,a, Zhenyu Gao,b, Shuzhu Tang,a, Dali Zeng,a, Yonghong Wang,a, Jianming Yu,a,3, Minghong Gu,b,3, and Jiayang Lia,a,3

aState Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics, Chinese Academy of Sciences, Beijing 100101, China; bState Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China; and cKey Laboratory of Plant Functional Genomics, Ministry of Education, Yangzhou University, Yangzhou 225009, China

Communicated by Longping Yuan, China National Hybrid Rice Research and Development Center, Hunan, China, October 29, 2009 (received for review May 30, 2009)

More than half of the world’s population uses rice as a staple food for more than half of the world’s population. Successful application of hybrid production technology has greatly increased grain yield, but grain quality of hybrid rice remains to be improved. Grain quality has now become the primary consideration of rice customers and breeding programs. Cultivars with different grain qualities are also required for medicinal, ceremonial, or special production purposes. Tremendous efforts have been made to understand the genetic basis of the eating and cooking quality of rice (ECQ) (1–6), which is mainly affected by three physicochemical properties: amylose content (AC), gel consistency (GC), and gelatinization temperature (GT) (7–9), but so far only the Waxy gene (10) has been found to affect AC and GT. It is therefore still facing many challenging questions. For example, how many major and minor genes control grain ECQs? Are AC, GC, and/or GT controlled by one or multiple genes? What is the relationship among these genes? Is it possible to breed elite varieties with desired ECQ properties by genetic manipulation of AC, GC, and/or GT?

Given that starch comprises approximately 90% of the rice grain, genes involved in starch biosynthesis are naturally expected to affect ECQs. Starch biosynthesis is a complex system composed of multiple subunits or isoforms of four classes of enzymes: ADP-glucose pyrophosphorylase (AGP), starch synthase (SS), starch branching enzyme (SBE), and starch debranching enzyme (DBE) (12–14) (Fig. 1). Each enzyme plays a distinct role (15, 16), but presumably functions as part of a network. For example, the rice sugary mutant is defective in ISA1 (17). As parts of this complex synthesis pathway, genes controlling amylose synthesis also affect amylepectin formation (18), and amylepectin can give rise to amylose (19). These relationships jointly make the genetic dissection of ECQs difficult.

QTL mapping and cloning provides useful information of genetic loci, but usually it is difficult to isolate genes of minor effects that play a minor role or to elucidate a complex network due to narrow germplasms used in single experiments. Association mapping is a powerful tool for studying genetic loci involved in the inheritance of complex traits (20–22), and an initial analysis of kernel composition traits has revealed promising associations with starch biosynthesis genes in maize (23, 24). To gain a broader understanding of the functions of these SSRGs on rice grain ECQs, we carried out a candidate-gene association analysis approach, we found that genes related to starch synthesis cooperate with each other to form a fine regulating network that controls the eating and cooking quality and defines the correlation among these three properties. Genetic transformation results verified the association findings and also suggested the possibility of developing elite cultivars through modification with selected major and/or minor starch synthesis-related genes.
association mapping study and gene transformation verification, demonstrating that SSRGs form a fine network to control ECQs by regulating AC, GC, and/or GT and that rice with a desired ECQ can be achieved through genetic modification of one or more SSRGs in the regulating network.

Results
Diverse Panel of Rice Varieties Used for Analysis. To have enough statistical power to test an association, it is critical to choose various germplasms that span a full range of phenotypic variations. A diverse panel of rice varieties, including 33 indica and 37 japonica, were measured for AC, GC, and GT in 2 successive years with three replications each year, and various trait value combinations were observed (Fig. S1). These three properties are highly correlated: AC is negatively correlated with GC (−0.91) and GT value (−0.46), whereas GC is positively correlated with the GT value (0.50).

To avoid false association, population structure was evaluated with Admixture Model using simple sequence repeat (SSR) marker data. The resulting subpopulation membership percentage (Q-values) agreed with the subspecies classification. The majority of indica varieties have high AC, low GC, and low GT values, whereas the majority of japonica varieties show low AC, high GC, and high GT values (Fig. S2).

Nucleotide Analysis of SSRGs. Eighteen SSRGs were selected as candidate genes in this study (Table S1). The entire genomic regions of these genes were sequenced for 16 representative core varieties (Fig. S1). Nucleotide analysis showed that the diversities of the coding sequences were much lower than those of whole genes in all 18 SSRGs, and the diversities of the nonsynonymous substitution were lower than the synonymous except for GBSSI (Table S2). This result suggested that these SSRGs had likely undergone artificial selection during domestication and the linkage disequilibrium (LD) analysis provided further evidence (Fig. S3). Because most SSRGs showed a stable LD decay, we chose the most diverse SSRGs for the association analysis.

Genes Affecting Rice ECQs. The effect of SSRGs was tested against each ECO property (i.e., AC, GC, or GT) across the panel using a mixed model method (25). Association analysis with individual SSRGs revealed that two genes, Wx and SSI-3, mainly control AC. Wx sites have much smaller P-values (3 × 10−15 and 4 × 10−24) than SSI-3 (2 × 10−5), suggesting that Wx is likely the major gene affecting AC. Further analysis showed strong evidence of haplotype effects (P value 1 × 10−53); Wx-I, Wx-II, and Wx-III. Among these, Wx-I bears a loss-of-function mutation that results in glutinous rice varieties with extremely low AC, Wx-II shows a leaky phenotype that leads to a medium level of AC, and Wx-III functions as the wild type allele with high AC (Fig. 2A). To eliminate the effect of this major gene, a second model containing the Wx haplotype was used to search for evidence of minor genes (see Methods). The model detected five genes that affect AC additively with Wx: AGPlar, PUL, SSI, SSI-3, and SSI-3 (Table 1). Based on the association sites, each of the minor genes could also be classified into two haplotypes: haplotype I and haplotype II (Table S4). For each associated gene, the estimate AC of haplotype I is significantly higher than haplotype II under the control of the Wx haplotype (Table 1).

For the properties of GC, Wx was identified as the major gene with a haplotype P value of 2 × 10−19 (Table S5). Varieties with Wx-I show high GC values, those with Wx-II show medium GC values, and those with Wx-III have low GC values (Fig. 2B). As with AC, three minor genes, AGPiso, SBE3, and ISA, affect GC once the model accounted for the Wx haplotype (Table 1). Each of the minor genes was classified into two haplotypes: haplotype I and haplotype II (Table S4). The estimate GC value of haplotype I for each additive gene is significantly lower than haplotype II, while the Wx haplotype is controlled (Table 1).

For the properties of GT, SSI-3 has highly significant P value (5 × 10−24), suggesting that SSI-3 plays a major role in regulating GT. SSI-3 has two allelic states: SSI-3-1 and SSI-3-2. Varieties with SSI-3-1 have higher GT values than those with SSI-3-2 (Fig. 2C). Again, with the model controlling for SSI-3, a further search identified Wx, SBE3, ISA, and SSI-2 as minor genes that affect GT additively (Table 1). The estimate GT value of haplotype I of Wx is significantly lower than haplotypes II+III, while the SSI-3 haplotype is controlled (Table 1). Each of the other minor genes was also classified into two haplotypes: haplotype I and haplotype II (Table S4). The estimate GT value of haplotype I for each additive gene is significantly higher than haplotype II, while the SSI-3 haplotype is defined (Table 1).

Interestingly, Wx and SSI-3 are diversified and cooperated in affecting AC and GT. Besides acting as a major gene controlling GT, SSI-3 also affected AC as a minor gene. Consistent with its effect on GT, SSI-3-I contributed to higher AC under the same Wx background, whereas SSI-3-II led to lower AC (Fig. 2D). This means that varieties with the SSI-3-1 allele would show higher AC and GT values, and varieties with the SSI-3-II allele would have lower AC and GT values. On the other hand, Wx also cooperated with SSI-3 to affect GT. Wx-I not only led to lower AC but also decreased GT value under the SSI-3-1 background, whereas Wx-II and Wx-III increased GT value (Fig. 2E). Therefore, we concluded that the effect of either Wx or SSI-3 on AC and GT values falls into a consistent pattern.

To understand why AC and GT values are negatively correlated, we checked the haplotype combination of Wx and SSI-3 in the
panel and found a natural grouping pattern of each haplotype combination when AC and GT values were plotted (Fig. 2F). The number of varieties within each haplotype combination was different. With 35 varieties belonging to Wx-II/SSII-3-I (middle AC and high GT values) and 15 varieties belonging to Wx-III/SSII-3-II (high AC and low GT values), 71% of the germplasms were accounted for. At the same time, varieties with Wx-III/SSII-3-I haplotype had high AC and high GT values, and those with Wx-II/SSII-3-II had low AC and low GT values. When correlation was conducted collectively across the panel, the correlation coefficient became −0.46 (P value, 7 × 10⁻²).

**Network Controlling Grain ECQs.** These results strongly suggest that SSRGs, which determine the rice grain ECQs by affecting AC, GC, and GT, form a fine network (Fig. 3) with the following features. First, Wx and SSII-3 are central in determining grain ECQs by affecting all three properties: AC, GC, and GT. Wx functions as the sole major gene for both AC and GC but as a minor gene affecting GT, whereas SSII-3 is the sole major gene controlling GT but as a minor gene affecting AC and GC. Second, two genes affect two properties simultaneously: both ISA and SBE3 affect GC and GT. Third, several minor genes are specific for each property: SSII-2, AGPlar, PUL, and SSII-2 for AC, AGPlar for GC, and SSIV-2 for GT. Fourth, the correlations among AC, GC, and GT were caused by the joint action of these associated genes and unequal haplotype combination. Summarizing the associated genes into the starch biosynthesis pathway, we can see that different SSRGs affect different characters at the different starch biosynthesis stage (Fig. S4). For example, at the step from Glc1P to ADPG, AC is mainly affected by AGPlar, whereas GC is mainly affected by AGPlar.

**Verification by Transgenic Tests.** To validate the effects of major and minor genes identified with the association analysis, we carried out transgenic studies by introducing Wx and SSII-3, the two major genes, and SBE3, one minor gene, into different background rice cultivars. To determine how the properties were modified when Wx-II or Wx-III were altered to Wx-I (a loss-of-function mutation), an antisense Wx RNA under control of the rice Wx promoter (Fig. 4A, a) was introduced into a Wx-II variety, Wuyunjing 7 (WY, medium AC, medium GC value, and high GT value) and a Wx-III variety, Longtefu B (LTF, high AC, very low GC value, and high GT value). We randomly chose two representatives, WY-WxA and LTF-WxA, from 58 stable transgenic lines (Fig. 4A, b). Protein analyses showed that grains of the two transgenic lines produced nearly undetectable Wx protein (Fig. 4A, c), indicating that the Wx expression was successfully inhibited in transgenic plants. Compared with untransformed controls, AC was decreased significantly (Fig. 4A, d), and the GC values were increased remarkably for the homozygous transgenic lines (Fig. 4A, e). As expected, the grain GT value levels of the two transgenic lines were also decreased (Fig. 4A, f).

To further confirm the functions of Wx, we also generated homozygous transgenic lines that overexpress Wx under the control of the rice glutelin (GlI) promoter that specifically expresses in the endosperm (Fig. 4B, a) by transforming WY and Guanlingxiangnuo (GLXN), a Wx-1 variety with extremely low AC, high GC value, and medium GT value (Fig. 4B, d–f). We randomly chose two representatives, WY-WxO and GL-WxO,
from 75 stable transgenic lines (Fig. 4B, b). The Wx protein levels in these transgenic lines were increased significantly (Fig. 4B, c), indicating that the transgenes were strongly expressed. As expected, the AC of the homozygous transgenic grains was increased dramatically compared with the nontransgenic grains (Fig. 4B, d). Consistent with the previous experiment, the GC value of the transgenic grains was decreased remarkably (Fig. 4B, e). As in the antisense Wx RNA experiment, the GT of the transgenic grains was also affected (Fig. 4B, f).

To confirm the effect of SSII-3 on ECOs, we isolated the entire SSII-3 gene including its fully functional promoter region from Shuangkezao (SKZ) (Fig. 4C, a). SKZ is a SSII-3-II haplotype variety whose grains have an extremely low GT value, high AC, and a medium GC value. We introduced the SKZ SSII-3 gene into Cbao, a SSII-3-I haplotype variety with a high GT value, medium AC, and a medium GC value. We chose ALK-T1 and ALK-T2 randomly as representatives of the five transgenic lines (Fig. 4C, b); the transgene expression levels were significantly increased (Fig. 4C, c). Comparing with the untransformed SKZ, the grain ECOs of the homozygous transgenic lines showed a dramatic decrease in the GT value (Fig. 4C, d). In the grains of the two transgenic lines, the AC levels were significantly decreased (Fig. 4C, e). But the GC values were increased significantly in one transgenic line and slightly in the other (Fig. 4C, f).

To test the effect of minor genes on grain ECOs, we performed transgenic studies by repressing the expression of SBE3 in the variety WY by RNA interference (RNAi) under the control of the Gt1 promoter (Fig. 4D, a). Among the 57 stable homozygous transgenic lines, three representative SBE3 RNAi lines, sSBE3–1 to sSBE3–3, were randomly chosen and compared with untransformed controls (Fig. 4D, b). Protein blotting analyses showed that the expression of SBE3 was entirely inhibited in all three SBE3 RNAi lines (Fig. 4D, c, loading control, see Fig. S5). Compared with untransformed controls, the AC levels were substantially increased in the homozygous transgenic lines, whereas the GC and GT values were significantly decreased (Fig. 4D, d–f).
In addition to exploring the effects of the transgenic approach on grain ECQs, we also carefully investigated and compared traits of agronomic importance between transgenic lines and their wild type controls. In a normal field trial of Wx transgenic rice lines (Table S6), the data showed minimal effect on the major agronomic traits of transgenic lines when expression of the major gene Wx was inhibited or enhanced. Similar results were found between transgenic rice lines with other transgenes and their untransformed wild type.

**Discussion**

**Elucidation of SSRGs Functions in Regulating Rice ECQs with a Diverse Background.** Over the past decades, properties of AC, GC, and GT of rice grains have been extensively studied, but no consensus has been reached. For example, AC has been reported to be governed only by Wx (3, 5, 26, 27), by several gene loci (28, 29), or even by some unidentified non-Wx genes (2). The discrepancies in different reports were due to the limitation of the QTLs mapping approach, which depends on haplotype configurations of both parents in a specific study (30, 31). Through association analysis, our study provides a clearer picture of how the allelic diversities at SSRGs collectively regulate rice ECQs through the starch biosynthetic network. Our study provides strong evidences that Wx not only affects AC, as reported in refs. 10 and 32, but also regulates GC as a major gene and GT as a minor one. Furthermore, our results also clearly demonstrated that SSII-3 plays an essential role not only in controlling GT, but also AC and GC. In addition, we also showed that some other SSRGs affect additively AC, GC, and GT as minor genes and all these associated genes form a fine complex network controlling ECQs of rice grains. Moreover, besides the minor genes that function in regulating these grain properties, different haplotype combinations of major genes are present in rice germplasms. The correlations between these ECO properties, as demonstrated in the current study, add another layer of complexity to the process of unraveling genetic architecture of the ECQs.

Grain quality has long been considered in rice production. Starch comprises ~90% of milled rice, and its structure determines grain quality properties. The genes associated with grain quality properties would have been subjected to artificial selection during breeding and production. Our results indicated that the diversities of the coding sequences were much lower than those of the whole genes and the diversities of the nonsynonymous substitution were significantly lower than the synonymous in the associated genes. In addition, the associated genes were from different classes of enzymes. We propose that the SSRGs affect the starch structure at different starch synthesis steps during starch development and further affect the grain quality.

In this study, we showed that the negative correlation between AC and GT values was a result of unequal trait values for different haplotype combinations of Wx and SSII-3 and unequal frequencies of these haplotypes in rice germplasm. Because Wx and SSII-3 are closely located on chromosome 6, it is possible that these two genes may have undergone coselection during domestication. This is consistent with our previous work on the marker-assisted selection, in which it was found that under the same SSII-3 background, when Wx-III was substituted for Wx-II, not only were AC and GC changed, but the GT value was also down-regulated (33).

**Association Analysis and Genetic Transformation Provide Critical Information to Rice Breeding.** Improving the grain quality of hybrid rice requires long term effort from many breeding programs. Genetic approaches to breeding rice varieties with desired grain ECQs, either by genetic engineering or molecular marker-assisted selection, will definitely improve the efficiency of rice ECQs breeding. Our results indicate that the selection for any single gene in rice breeding would be insufficient because modification of one gene may lead to changes in all three properties of ECQs. Rather, future network-based multitrait selection would more efficiently improve ECQ properties. Moreover, we demonstrated that it is possible to breed varieties with desired grain ECQs by genetic modification without adversely affecting general agronomic performance. Our genetic characterization of widely used breeding germplasms provides essential information for future marker-assisted selection.

In summary, we demonstrated in this study that a combination of association analysis and genetic transformation can be exploited to understand and manipulate economic traits of relevance to consumers. Findings described herein should also facilitate investigation of other complex agronomic traits in rice and are applicable to other important crops.

**Methods**

**Measurement of Grain ECQs.** Mature rice grains were milled after being harvested, air dried, and stored at room temperature for 3 months. Their ECQs were measured according to methods reported in ref. 5. GT was evaluated as the alkali spreading value (ASV). Grain ECQs were independently analyzed in two successive years with three repeats each year.

**Identification of SSRGs.** DNA sequences of rice SSRGs were searched and retrieved from all available DNA databases in 2005 with Blastn and analyzed with CLUSTAL W (34). A total of 50 full-length cDNA sequences in rice were identified, representing 18 SSRGs from AGP, SS, SBE, and DBE gene families (Table S1).

**Statistic Analysis.** Population structure was evaluated by the program STRUCTURE (35) with 49 SSR markers distributed on 12 chromosomes in rice (SI Experimental Procedures) using admixture model. The relative kinship (K) matrix was calculated with the software package SPAGeDi (36). Diversities of genomic DNA and CDS sequences were analyzed with TASSEL (37) and DnaSP (38) separately. The LD was evaluated with TASSEL (37). Association analysis followed the unified mixed model reported in ref. 25, using SAS 9.0 (39). Models with both one-single-gene and two-single-gene effects and interaction terms were applied.

**Additional Experimental Procedures.** A detailed description of plant materials, DNA preparation and sequencing (primer sequences are shown in Table S7), SSR primers used for population structure evaluation, and protein blotting analysis can be found in the SI Experimental Procedures.

**ACKNOWLEDGMENTS.** We thank Dr. Guiai Jiao at China National Rice Research Institute for measuring rice grain ECQs and Guojun Dong for assistance in growing plant materials. This work was supported by the Ministry of Science and Technology of China Grants 2006AA10A1 and 2005CB12, United States Department of Agriculture Cooperative State Research, Education and Extension Service (2006-35300-17155), and the National Natural Science Foundation of China Grant 30530470.


