Article
Chinese Science Bulletin
Crop Breeding
November 2010 Vol.55 No.33: 3768–3777
doi: 10.1007/s11434-010-4124-1

Development of gene-tagged molecular markers for starch synthesis-related genes in rice

TIAN ZhiXi[1*], YAN ChangJie[2*], QIAN Qian[3], YAN Song[2], XIE HuiLan[2], WANG Fang[2], XU JieFen[2], LIU GuiFu[1], WANG YongHong[1], LIU QiaoQuan[1], TANG ShuZhu[1], LI JiaYang[1†] & GU MingHong[2†]

[1] *Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China;*
[2] *Key Laboratory of Plant Functional Genomics, Ministry of Education of China; Jiangsu Key Laboratory of Crop Genetics and Physiology, Agricultural College of Yangzhou University, Yangzhou 225009, China;*
[3] *China National Rice Research Institute, Hangzhou 310006, China*

Received January 20, 2010; accepted July 27, 2010

Improving grain quality is an important goal in breeding new elite rice varieties, requiring effective tools for the identification of target genotypes. Molecular marker-aided selection (MAS), combined with conventional breeding approaches, enables us to precisely identify the individual genotypes that are associated with different grain quality features, which can dramatically improve the breeding efficiency. However, to date, the number of molecular markers used in MAS for grain quality improvement is still somewhat limited. In this study, based on our previous study that rice grain quality is strongly associated with starch synthesis in the endosperm, we developed 51 gene-tagged molecular markers according to sequence variations in 18 starch synthesis-related genes from 16 typical rice cultivars. These markers can discriminate the different alleles among rice germplasms. These novel markers will provide effective tools in improving grain quality via the breeding new elite rice varieties.

rice, starch synthesis-related gene, molecular marker, quality improvement

Citation: Tian Z X, Yan C J, Qian Q, et al. Development of gene-tagged molecular markers for starch synthesis-related genes in rice. Chinese Sci Bull, 2010, 55: 3768–3777, doi: 10.1007/s11434-010-4124-1

Rice is one of the most important crops, serving as the staple food for more than half of the world's population. High-yield and good grain quality have always been two major goals in rice breeding programs. Over the past three to four decades, the breeding of semi-dwarf varieties and the implementation of hybrid rice have dramatically increased rice yield, which has in turn promoted rice production on a large scale to meet the demands of consumers. However, progress in the improvement of rice grain qualities, especially in the cooking and eating qualities [1,2] of commercial varieties is relatively slow. The conflict between higher yield potential and lower grain quality has become increasingly prominent in recent years.

In a conventional breeding program, selection of target traits has been shown to be unpredictable. Successful selection of required traits depends largely on genetic variations in breeding populations, especially on the degree of favorable variation of target traits and the efficiency of directional selection [3]. Moreover, selection of target traits is mainly performed according to phenotypes, rather than genotypes. The existence of genetic effects, such as pleiotropic and multiple effects, together with regulation and modification of genes, will lead to inconsistencies between genotypes and corresponding phenotypes. Therefore, the selection for favored individuals by phenotypes is usually inaccurate, slow, and inefficient. However, application of molecular marker-aided selection (MAS) improves the selection efficiency and shortens the duration of breeding,

*These authors contributed equally to this work
†Corresponding authors (email: gumh@yzu.edu.cn, jyli@genetics.ac.cn)

© Science China Press and Springer-Verlag Berlin Heidelberg 2010

because it can be conducted at the genetic level directly in the segregation population [4]. Computer simulation analysis showed that the breeding efficiency by MAS is at least as twice as that through conventional breeding [5].

Rice grains are mainly composed of starch, which accounts for 90% of milled rice. Therefore, the structure and composition of grain starch are two main components affecting the grain quality [6]. Previous studies have shown that there are 5 types of enzymes involved in starch synthesis, including ADP-glucose pyrophosphorylase (AGPase), granule-binding starch synthase (GBSS), soluble starch synthase (SSS), starch branching enzyme (SBE) and debranching enzyme (DBE). Each enzyme has variant isoforms. Up to now, about 20 genes encoding starch synthesis enzymes have been identified, which function in different stages during starch synthesis [7–12]. Furthermore, allelic variations in starch synthesis-related genes (SSRGs) were also characterized, suggesting that the genetic control of starch synthesis is very complex. So far, the functions of SSRGs have mainly been obtained through the mutants or transgenic studies. Moreover, these studies mainly focused on a few SSRGs, for example GBSS encoded by *Wx* and starch branching enzymes encoded by *SBE1* and *SBE3* [13–21].

Previous studies have indicated that the allelic variation in the *Wx* gene is responsible for the variation of amylose content in rice grain endosperm; at least 5 different alleles (Wx^a, Wx^b, Wx^{op}, Wx^{in} and *wx*) of the *Wx* gene have been identified that exist in common *indica*, *japonica*, soft rice varieties from Yunnan of China, Lemont from America, and glutinous rice [22–24]. Compared with Wx^a, Wx^b has a G-to-T change at the 5′ splicing site of intron 1, leading to a significant decrease of splicing efficiency of the Wx^b pre-mRNAs. The base A_{762} in exon 4 of Wx^{op} was replaced by G_{762}, leading to an amino acid alteration of Asp to Gly. Similarly, base A_{1132} in Wx^{in} was replaced by C_{1132}, resulting in a change of amino acid from Tyr to Ser. As for the *wx* allele, a 23-bp fragment was found to be inserted in exon 2, resulting in a premature stop codon and generating a non-functional GBSS [24]. Gao et al. [25] cloned the *ALK* (*SSII3*) gene through a map-based cloning approach and found that G_{264} to C_{264} alteration downstream of the initial codon of *ALK* is responsible for the observed variation of rice gelatinization temperature (GT). Bao et al. [26] sequenced the fragments covering introns 6 and 7, exons 6 and 8, and part of the 3′-untranslated region of the *SSIIa* (*SSII-3*) gene from 30 varieties and association analysis indicated that the GT variation among 30 varieties could be mainly attributed to the alteration of GC to TT in exon 8. Han et al. [27] analyzed the allelic variations in *SBE1* and *SBE3* between *indica* and *japonica* species and found that the allelic variations in *SBE1* and *SBE3* correlated to the variation of Rapid Visco Analyzer (RVA) profile among 40 glutinous rice cultivars. Through a comprehensive association analysis, we found that the SSRGs have a decisive role to determine rice grain quality and the allelic diversities of these SSRGs contributed to the diverse arrays of rice eating and cooking qualities [28].

In addition to genetic complexity, rice grain quality also varies with environmental stimuli, leading to difficulties in conventional breeding selection. MAS will be an effective approach for the improvement of grain quality. However, few operable molecular markers have been explored. Therefore, development of applicable molecular markers based on allelic variation would be a valuable contribution to improve the grain eating and cooking quality through an MAS approach.

In this study, a series of typical *indica* and *japonica* cultivars with distinct grain qualities were selected, 18 SSRGs from each cultivar were sequenced, and more than 50 molecular markers were developed. In addition, the potential of these markers in multiple allele identification was assessed based on 64 different varieties.

1 Materials and methods

1.1 Materials

A total of 64 rice varieties were used in this study, including 26 *indica*, 27 *japonica*, 7 glutinous varieties, 1 Africa cultivar and 3 wild varieties (provided by the International Rice Research Institute). Details are listed in Table 1. All the varieties were planted in the experimental field of Yangzhou University, Yangzhou and the experimental farm of the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing with conventional water and fertilizer management. Leaves of each variety were harvested for DNA extraction.

1.2 SSRG sequences

The full-length cDNA sequences of 18 SSRGs were obtained by searching the NCBI database (http://www.ncbi.nlm.nih.gov), and the corresponding *indica* and *japonica* genomic sequences were determined by the Blastn program (http://www.ncbi.nlm.nih.gov/BLAST/Genome/PlantBlast.shtml).

1.3 DNA extraction and sequencing

DNA extraction was performed according to the CTAB method [29]. To sequence the genomic DNA of SSRGs, each individual gene was divided into ~1 kb fragments with average 100 bp overlap between adjacent fragments. *Taq* polymerase was from Takara Company and the DNA Gel Extraction kit from the Shanghai Shenneng Bocai Company. The DNA fragments were sequenced with a 3700 DNA sequencer (ABI) and the full-length gene sequences were assembled by DNAStar. The sequences were aligned with MEGA3, followed by cluster analysis of the SSRGs from different varieties. Molecular markers for each individual gene were developed based on the sequence differences among alleles of different varieties. Primers for the polymerase chain reaction (PCR) were designed by the Primer Premier 5.0 software and synthesized by Sangon

Table 1 List of 64 varieties used in this study

Variety	Origin	Type[a)]	Variety	Origin	Type	Variety	Origin	Type
Nipponbare	Japan	*J*	Xiangjing 49	China	*J*	R6547	China	*I*
Chunjiang06	China	*J*	Dijing	Japan	*J*	R084	China	*I*
Wuyunjing 7	China	*J*	Eyi 105	China	*J*	Qimiaoxiang	China	*I*
9308	China	*J*	Wuxiangjing 3	China	*J*	Ezhong4	China	*I*
Boro	China	*J*	93–11	China	*I*	Milyang 50	Korea	*I*
Bulu	China	*J*	Guichao 2	China	*I*	Shuiyuan 262	Korea	*I*
Nongken 57	Japan	*J*	Longtefu	China	*I*	Yuexiangzhan	China	*I*
Lemont	America	*J*	Minghui 63	China	*I*	Newrex	America	*I*
KetanNangka	Indonesia	*J*	Taichung Native 1	China	*I*	Zhenzhuai	China	*I*
Balilla	Italy	*J*	Zhenshan 97	China	*I*	Jiangzhouxiangnuo	China	G
Guihuahuang	China	*J*	Dular	India	*I*	Suyunuo	China	G
Liaojing 5	China	*J*	Aman	India	*I*	Taihunuo	China	G
Yanjing 2	China	*J*	Tjereh	India	*I*	Xiangzhunuo	China	G
Sidao 10	China	*J*	Nantehao	China	*I*	Zhuzhongnuo	China	G
Zhendao 88	China	*J*	Nanjing 11	China	*I*	Zinuo	China	G
Hanfeng	China	*J*	Maweizanzhan	China	*I*	Majinnuo	China	G
O2428	China	*J*	Yanhui 559	China	*I*	*O. glaberrima*		A
Lunhui 422	China	*J*	Milyang 46	Korea	*I*	*O. rufipogon*		W
Cpslo17	America	*J*	Chaofengzhan	China	*I*	Dongxiang wild rice	China	W
Zhonghua 11	China	*J*	Teqing	China	*I*	*O. perennis*		W
Koshihikari	Japan	*J*	E32	China	*I*			
Xiaoqianyou	Japan	*J*	Yangfuxian 2	China	*I*			

a) *I*, *indica*; *J*, *japonica*; G, glutinous rice; A, Africa cultivar; W, wild species.

Biotech (Shanghai) Co., Ltd.

1.4 DNA fragments amplification and digestion

The PCR reaction mixture (total volume 20.0 μL) contained 20 ng DNA template, 50 pmol primers, 0.25 mmol/L dNTP, 1×GCI buffer, and 1 U *Taq* DNA polymerase. The program for the PCR amplification of DNA fragments contained one denaturing cycle (94°C, 4 min), 45 amplification cycles (94°C, 4 min; 58°C, 1 min, 72°C, 1 min), and a final extension at 72°C for 10 min. The enzymatic digestion mixture (20 μL) of DNA fragments contained 1 μg PCR products, 1×buffer, and 5 U restriction enzyme (BioLab). The DNA was digested at appropriate temperature for 12–16 h.

2 Results

2.1 Determination of SSRGs

A total of 50 full-length cDNAs belonging to 4 starch synthesis gene families, namely, ADPase, starch synthase (SS), SBE, and DBE, were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov). ADPase is composed of 2 subunits, SS contains 2 types (GBSS and SSS) with multiple isoforms, SBE has at least 3 isoforms (SBE1, SBE3 and SBE4), and DBE can be classified into isoamylase (ISA) and limit dextrin enzyme (Pullulanase, PUL). Each isoform of the above enzymes is encoded by a distinct gene [6]. Therefore, the 50 full-length cDNAs from the NCBI database were grouped into 18 genes, which were located on nine chromosomes and referred to as starch synthesis-related genes (SSRGs). The genomic sequences corresponding to each cDNA from *indica* variety 93-11 and *japonica* variety Nipponbare were obtained through the Blastn program (http://www.ncbi.nlm.nih.gov/BLAST/Genome/PlantBlast.shtml). Detailed information of the sequences was presented in our previous report [28].

2.2 Gene sequencing and development of molecular markers

To determine the differences among SSRGs at the genomic level, 16 *indica, japonica,* and glutinous cultivars with dramatic differences in grain quality were used. They were Nipponbare, Chunjiang 06, Wuyunjing 7, Suyunuo, Taihunuo, Jiangzhouxiangnuo, Guixiangsinuo, 93-11, 9308, Minghui 63, Longtefu, Guichao 2, Taichung Native 1, Zhenshan 97, Kasalath and Zhonghan 3, respectively. The genomic sequences, covering 1–2 kb 5′ and 3′-untranslation region, of the 18 SSRGs from each variety were sequenced and aligned to identify the allelic sites among the 16 varieties. Each of the 18 SSRGs could be classified into 2–6 types. Among them, *SSV-1*, *SSV-2*, *SBE1*, *SBE3* and *ISA* can be classified into 2 allelic types, *AGPiso* into 5 types, *SSIII-1* into 6 types, and the rest into 3–4 types. To discriminate the different alleles of each SSRG locus, 31 core

markers were developed based on sequence polymorphism sites (marked with * in Table 2), and a further 20 markers were designed based on other mutation sites. Primer sequences for these molecular markers are listed in Table 2.

Table 2 Primer sequences of SSRGs markers[a)]

Gene name	Marker name	Primer sequence (5′→3′)	Marker type
AGPlar	AGPlar M1*	[F] CGTTCAGGTTCAGGCAATCA	STS
		[R] GGAAGGGTGGTGATGTGGAG	
	AGPlar M2	[F] GCGTGAACTGAACATCCATCT	CAPS (*Tsp*45 I[#])
		[R] GGTTCAAGCCTTCAGGTCAG	
AGPiso	AGPiso M1*	[F] CAATGAGAAGGGCTGGTGAG	CAPS (*Bst*X I)
		[R] CGAAATAACCAAATGATACCAGAT	
	AGPiso M2*	[F] CAATCGCTGCCATCGGTTG	STS
		[R] TTCCACATCGTTAGGTACACG	
	AGPiso M3	[F] TGGAATGGGAACTCTATTATTGG	CAPS (*Eco*R I)
		[R] TCCCAACCTCTACCTTCAAATG	
AGPsma	AGPsma M1*	[F] TCTATTCTCAGCCCTCCAACC	STS
		[R] GTGTGTTTAGAGGTGCTTTTCG	
	AGPsma M2	[F] TACGCTATGCTCTTGAAAC	STS
		[R] TATCTTCCCAGTAACCATCA	
Wx	Wx M1*	[F] CACAGCAACAGCTAGACAACCAC	STS
		[R] CACGACGACGGAGGGGAAC	
	Wx M2*	[F] GACAAGGCAAGAATGAGTGACA	CAPS (*Nhe* I)
		[R] GTGCCGACGACCTTGATGG	
	Wx M3	[F] GCTTCACTTCTCTGCTTGTG	CAPS (*Acc* I)
		[R] ATGATTTAACGAGAGTTGAA	
GBSSII	GBSSII M1*	[F] TTGCTGCGAATTATCTGCG	STS
		[R] ACCTCCTCCCACTTCTTTGC	
	GBSSII M2*	[F] CAACCCAGTGCGTCCCTCTA	CAPS (*Eco*R I)
		[R] CTTGGGCTCAGAATGGCAG	
SSI	SSI M1*	[F] GGTAGGGTAGGTCAATCTGGC	CAPS (*Nru* I)
		[R] ATAGAGAAGACAATGTGGCAACC	
	SSI M2*	[F] CTTCTATCCATTCCTTAATCCCA	STS
		[R] ATGCTATTGATGTTAAGAGGGC	
	SSI M3	[F] GACCCACCTCGCTATCTGTTG	CAPS (*Apa* I)
		[R] GGAAACACCAGACATCAACCAG	
SSII-1	SSII-1 M1*	[F] CACCCCACCGTTCTACTATGC	STS
		[R] TCCATAGTTTCATTGAGATTGCTC	
	SSII-1 M2*	[F] CAAGTTGGTGACGATAGTGATGA	CAPS (*Age* I)
		[R] AACAGAGCCTCCATTACCTTTAC	
	SSII-1 M3	[F] AGAGATCAAATCGTGGAAC	STS
		[R] TGGAGTGAAGTAGTGGAAT	
	SSII-1 M4	[F] ATCTTTAGACGATTAGCG	STS
		[R] AAGTCACAAGTAGAAGGG	
SSII-2	SSII-2 M1	[F] AGATTTGAACTCAGGACTTGGTG	STS
		[R] TCTATGGGCTCTATCCTTACTAGG	
	SSII-2 M2*	[F] CGCTCGTTGCCTAGCTAGC	STS
		[R] GGCGAGGAAGCGATTGCC	
	SSII-2 M3	[F]ACAGTATGTTTGCCTCAGCG	STS
		[R] GTAAATCCACCCAGCCAGTC	
SSII-3	SSII-3 M1*	[F] CCAATACCGTAAACTAGCGACTATG	STS
		[R] TACAGGTAGAATGGCAGTGGTG	
	SSII-3 M2*	[F] GGTTCTCGGTGAAGATGGC	CAPS (*Ban* II)
		[R] GTGGTCCCAGCTGAGGTCC	
	SSII-3 M3	[F] AACTGACTCATACACGGATAACG	CAPS (*Nhe* I)
		[R] CACGCACGAACGGAAACC	
SSIII-1	SSIII-1 M1*	[F] AAGAAGGGAAGGGAGTCAGC	SSR
		[R] GCCATCTCCATTGCCAGC	
	SSIII-1 M2*	[F] CAAGCAATGATTCAGGCACA	CAPS (*Eco*R I)
		[R] GGAGACAGGAGCAAAAGGC	

(*To be continued*)

(*Continued*)

Gene name	Marker name	Primer sequence (5′→3′)	Marker type
	SSIII-1 M3*	[F] CAAATCAACTGTAAGTGCTGGAG	CAPS (*Nde* I)
		[R] GAGAACGGAGAAAATGGCAT	
SSIII-2	SSIII-2 M1*	[F] AAGTCCTTCGGCTTACTATTCC	CAPS (*Xba* I)
		[R] GGAGAAGGAACATAACAGGGAC	
	SSIII-2 M2*	[F] GAACTTGTGCCTTAAGCTGACTG	STS
		[R] GGAATAGTAAGCCGAAGGACTT	
SSIV-1	SSIV-1 M1*	[F]CATTGTGTCTTGAAGTCTGTGCT	CAPS (*Nde* I)
		[R] CGATGGGTTAGTGCTGTGG	
SSIV-2	SSIV-2 M1*	[F] CTTCTGATTGATGGTTGGTTGC	CAPS (*Sph* I)
		[R] GGAAGAATAATCTCTACTAGGTGGC	
	SSIV-2 M2	[F] TTCCCTTGGTGGTGCGTG	STS
		[R] TAAAGCGTTCCGACAGTA	
	SSIV-2 M3	[F] TCAAGTATGGTTTACCTATG	CAPS (*Eco*72 I)
		[R] TTTCCCAATGACTTCTAA	
SBE1	SBE1 M1*	[F] TGCTACATAACACGCATACAAAGT	STS
		[R] AGACAAAAGCGAAAGGTAATGAG	
	SBE1 M2	[F] GTGGGGAAAACAAGTAAGTCTG	STS
		[R] AGTTCCATCAGAAGAATCAGGG	
	SBE1 M3	[F] GGAAATGGGAGTCGCC	STS
		[R] CGAAGAAACCACGCTCA	
	SBE1 M4	[F] ATTGTTGCTGAAGATGTTT	CAPS (*Taq* I)
		[R] ACGGTTGATGGTAGGTG	
SBE3	SBE3 M1*	[F] AAGGTTAGCATTGGTTGGTGAG	STS
		[R] TCTCCTTGAACAGCGACAGC	
	SBE3 M2	[F] GTGGGGTTCTCAACTTAGC	STS
		[R] CATCAGCATTGTTAGGCAG	
SBE4	SBE4 M1*	[F] CACCAATTATATTAGCGTGCTCC	STS
		[R] CGTGGCTCTTGGCTCTCTTG	
	SBE4 M2*	[F] CCATCACCTCAAATACATCACTC	STS
		[R] AGACTGGAATGCCCCTTAGG	
ISA	ISA M1*	[F] ATAGATGCTAATGTGATGTGGC	STS
		[R] TGGTATAGGCACAACCGTAGA	
	ISA M2	[F] ACAAGCACACGACACCTA	CAPS (*Hin*d III)
		[R] CAACAAACCAAACTCATT	
	ISA M3	[F] TGTGGGAATACCTTCAACTG	STS
		[R] ATAAAACCCTTACAGGCTTG	
PUL	PUL M1*	[F] AGAGAAGGAGAAAGAAGTGGAGAC	STS
		[R] GTCCAAACTGAATCACTCAATCG	
	PUL M2*	[F] GACAACCGTCCGCTTTAGTTTC	STS
		[R] GCATTTGAGAGGGTTTGGATTC	
	PUL M3*	[F] CTGTATGGACTGAGTAGTCGATGG	STS
		[R] TGAGCCTCATCTGCCAGAGT	
	PUL M4	[F] TACACCATCCTCACTACCA	STS
		[R] GCAACATCTAAAACACCAA	
	PUL M5	[F] ATTGGCATTTGTAAGTTTC	STS
		[R] CAATCTTGGTTTTATCCTG	
	PUL M6	[F] ATTTAACTGTATGGACTGAG	STS
		[R] GATACCAACCAAACAAGA	

a) * Indicates the core markers, and # represents the enzymes used for CAPS.

2.3 Verification of molecular markers

To determine whether the developed molecular markers were operable, each marker was verified by PCR amplification. The results showed that the lengths of the amplified DNA fragments were as expected. Moreover, the amplified products could be digested by corresponding designed restriction enzymes. Thus, different alleles could be identified by individual markers or marker combinations. For example, the cluster analysis showed that *AGPlar*, encoding the AGPase large subunit, has 3 different allele types in the 16 varieties. Type I is present in *japonica* rice varieties (and in *japonica* glutinous varieties), Type II in Longtefu and 93-11 (although the sequences of the alleles were not fully identical, the similarity was very high, so they were grouped into one type), and Type III in Minghui 63 and 5 other varieties (in which the allele sequences were also not identical, but had high similarity) (Figure 1A, a). These 3 allele types could be easily differentiated with the AGPlar M1 marker (Figure 1A, b). Similarly, cluster analysis showed that *AGPiso* could be grouped into 5 allele types (Figure 1B, a). When AGPiso M1, a Cleaved Amplified Polymorphisms (CAPS) marker, was applied to the 16 varieties, a two-band pattern was obtained: allele types I, IV and V had the same pattern, but types II and III showed another pattern (Figure 1B, b). Furthermore, type I could be separated from IV and V by the marker AGPiso M2. However, types IV and V, and types II and III could not be discriminated by the above two markers, although sequence differences existed (Figure 1B, c). *AGPsma* was classified into 4 allele types (Figure 1C, a) and the AGPsma M1 could separate 3 allele types (Figure 1C, b), but types I and II could not be differentiated. Due to the lack of appropriate PCR-based markers, DNA sequencing is needed to distinguish types I and II. Studies have shown that the sequence variation in the *Wx* gene, encoding GBSSI, is significantly associated with variation in the amylose content in rice grain endosperms [22,28,30]. Cluster analysis showed that 3 different alleles existed at the *Wx* locus (Figure 1D, a), and that type III was specific to glutinous rice varieties. Sequence analysis showed that a 23-bp insertion occurs in the alleles of all the glutinous rice, resulting in a premature stop codon, which results in very low amylose content in glutinous rice varieties [30]. Therefore, Wx M1 was developed according to this sequence variation to distinguish glutinous rice varieties from the others (Figure 1D, b). In addition, the accumulated data indicated that the G-to-T mutation in intron 1 of *Wx* was strongly associated with the amylose content of the rice grain endosperm [21], and a CAPS marker was developed to identify varieties with medium amylose content [31], which could separate type I from II. Sequence analysis showed that another polymorphic site, which was highly linked to this site, could be used to develop a CAPS marker. Thus, a new marker, Wx M2, was developed to discriminate types I and II, and the patterns could be easily identified. The use of the combination of these two markers can easily distinguish different *Wx* alleles.

Taken together, the 31 core molecular markers developed in this study can distinguish the different alleles of each SSRG locus (Figure 1), and these markers can be directly applied to select target alleles for the improvement of rice grain quality.

In addition to the 31 core markers, an additional 20 markers were also developed based on the sequence variation among different varieties. Our study showed that these 20 markers could also be applied to distinguish different alleles in the rice germplasm collections. For example, AGPlar M2 was developed according to the A-to-G mutation in the promoter region of *AGPlar*, which could separate the alleles of Guichao 2, Taichung Native 1, and Zhenshan 97 from other varieties (Figure 2(a)). The Wx M3 marker was designed to target the G-to-T mutation in the first intron of *Wx*, and could distinguish types I and II alleles, just as marker Wx M2 did (Figure 2(b)). SSI M3 was a CAPS marker based on the T-to-C mutation in the second exon of *SSI*, and could separate the alleles of Zhenshan 97 and Longtefu from 14 other varieties (Figure 2(c)). These results indicated that these markers could not only be applied in the selection of target alleles, but also show potential in the exploitation of multiple alleles.

2.4 Application of molecular markers in identifying multiple alleles

To further understand allelic variations in SSRGs in rice germplasms, a total of 60 typical *indica* and *japonica* varieties, 3 wild species, and 1 African rice variety from different rice production areas, including China, India, America, Korea and Japan, were genotyped using the developed markers. The results showed that many novel alleles, whose patterns were different from those of the 16 sequenced varieties, were found in the three wild species and the African rice *O. glaberrima*, by markers AGPsma M1, SSII-1 M3, SSII-1 M4, SBE1 M2, ISA M3, and PUL M3. However, no novel allele was found in the other 60 varieties. This result strongly suggests that the alleles of SSRGs in cultivars have been artificially selected, and some alleles have been eliminated in the process of domestication. In contrast, those alleles that are absent in cultivars but remain in wild rice species, need to be exploited.

The haplotypes of each SSRG were analyzed among the 60 cultivars (Table 3). Except for *AGPlar, AGPsma, GBSSII, SSIII-2*, and *SSIV-1*, novel haplotypes that were not present in the 16 sequenced varieties were found in the 60 cultivars for the remaining 13 SSRGs. Among them, more novel haplotypes were found at the *ISA, SBE3*, and *SSII-1* loci, each of which contained 3 novel haplotypes. As for the varieties harboring novel haplotypes, Aman, Boro, and Dular from India, and Lemont and Cpslo 17 from America showed more haplotypes. These results suggest that the developed

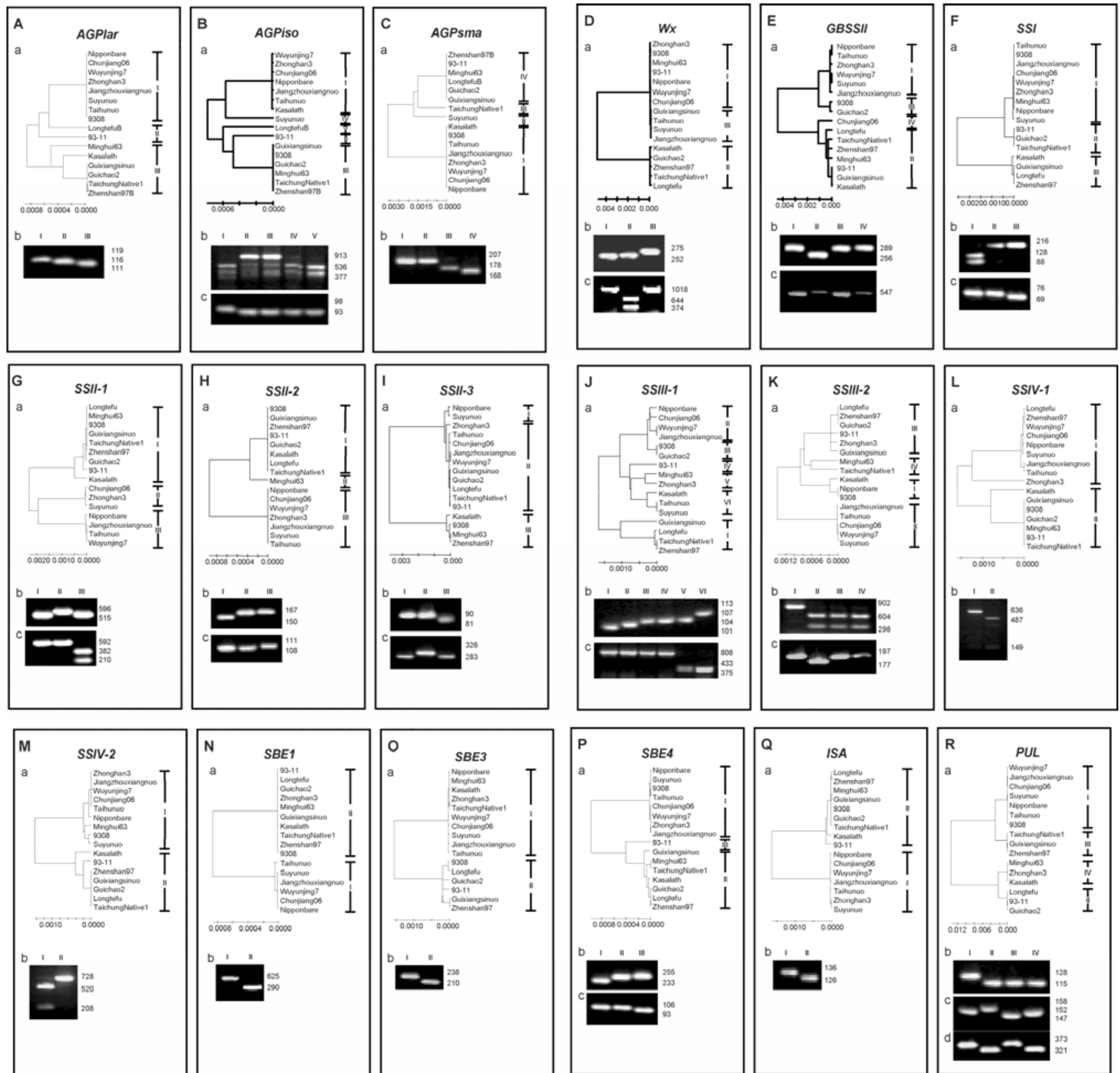

Figure 1 Genotyping analysis for SSRGs among different varieties by molecular markers. Each box in the figure shows the analysis result of one SSRG. A, Genotype analysis of *AGPlar*. a, Clustal analysis of *AGPlar* in 16 representative varieties; (b) the detection of marker AGPlar M1. B, Genotype analysis of *AGPiso*. a, Clustal analysis of *AGPiso* in 16 representative varieties; b, detection of marker AGPiso M1; c, detection of marker AGPiso M2. C, Genotype analysis of *AGPsma*. a, Clustal analysis of *AGPsma* in 16 representative varieties; b, detection of marker AGPsma M1. D, Genotype analysis of *Wx*. a, Clustal analysis of *Wx* in 16 representative varieties; b, detection of marker Wx M1; c, detection of marker Wx M2. E, Genotype analysis of *GBSSII*. a, Clustal analysis of *GBSSII* in 16 representative varieties; b, detection of marker GBSSII M1; c, detection of marker GBSSII M2. F, Genotype analysis of *SSI*. a, Clustal analysis of *SSI* in 16 reprehensive varieties, b, detection of marker SSI M1; c, detection of marker SSI M2. G, Genotype analysis of *SSII-1*. a, Clustal analysis of *SSII-1* in 16 representative varieties; b, detection of marker SSII-1 M1; c, detection of marker SSII-1 M2. H, Genotype analysis of *SSII-2*. a, Clustal analysis of *SSII-2* in 16 representative varieties; b, detection of marker SSII-2 M1; c, detection of marker SSII-2 M2. I, Genotype analysis of *SSII-3*. a, clustal analysis of *SSII-3* in 16 representative varieties; b, detection of marker SSII-3 M1; c, detection of marker SSII-3 M2. J, Genotype analysis of *SSIII-1*. a, Clustal analysis of *SSIII-1* in 16 representative varieties; b, detection of marker SSIII-1 M; c, detection of marker SSIII-1 M2. K, Genotype analysis of *SSIII-2*. a, Clustal analysis of *SSIII-2* in 16 representative varieties; b, detection of marker SSIII-2 M1; c, detection of marker SSIII-2 M2. L, Genotype analysis of *SSIV-1*. a, Clustal analysis of *SSIV-1* in 16 representative varieties; b, detection of SSIV-1 M1. M, Genotype analysis of *SSIV-2*. a, Clustal analysis of *SSIV-2* in 16 representative varieties; b, detection of marker SSIV-2 M1. N, Genotype analysis of *SBE1*. a, Clustal analysis of *SBE1* in 16 representative varieties; b, detection of marker SBE1 M1. O, Genotype analysis of *SBE3*. a, Clustal analysis of *SBE3* in 16 representative varieties; b, detection of marker SBE3 M1. P, Genotype analysis of *SBE4*. a, Clustal analysis of *SBE4* in 16 representative varieties; b, detection of SBE4 M1; c, detection of marker SBE4 M2. Q, Genotype analysis of *ISA*. a, Clustal analysis of *ISA* in 16 representative varieties; b, detection of marker ISA M1. R, Genotype analysis of *PUL*. a, Clustal analysis of *PUL* in 16 representative varieties; b, detection of marker PUL M1; c, detection of marker PUL M2; d, detection of marker PUL M3. The Roman number on the right side in each of the (a) indicates the allelic type.

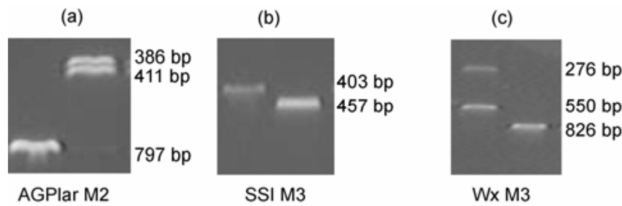

Figure 2 Molecular marker detection in different varieties. (a) AGPlar M2; (b) SSI M3; (c) Wx M3.

markers in the present study are able to be applied not only in distinguishing different alleles of SSRGs, but also in exploiting novel alleles using the marker combinations of the target gene. The novel markers not only provide a powerful tool for identifying novel alleles, but can also be used in the improvement of rice grain quality.

The information concerning different categories of SSRGs among the 64 rice varieties clearly demonstrated that new alleles exist in addition to those found in the sequenced varieties, which lays a solid foundation for characterizing the function of SSRGs and for developing new varieties in rice breeding programs.

3 Discussion

Rice is a polymorphic crop with a long history of cultivation. In China, the history of rice cultivation can be traced back more than 7000 years. Over the long period of evolution and differentiation in different regions, ecological environments and even different seasons, the variations among species have become extremely complicated, which provides a solid foundation for developing new varieties to meet consumer demand through cross-breeding.

In general, rice eating and cooking quality (ECQ) is mainly reflected by three indices, namely, amylose content (AC), gelatinization temperature (GT) and gel consistency (GC). In recent years, starch RVA profile has also been considered as another criterion for evaluating rice ECQ. The relationship between the RVA profile and rice ECQ has been clarified to some degree, but some basic questions are still not fully understood. Due to the existence of abundant variations of grain quality in rice germplasm resources, the indices mentioned above can only roughly reflect the physical and chemical properties of rice, but not the feature of rice quality in nature. In the long process of breeding rice varieties, rice ECQ is usually not evaluated until the varieties are released due to the difficulties in measuring ECQ properties, which in turn makes the improvement of ECQ even slower and more difficult.

Rapid progress has been made recently in the genetic analysis of the 4 above-mentioned indices of rice ECQ, and chromosome locations of the genes/QTLs related to the grain quality have been determined [32–39]. Some of the relationships between the observed genetic variation and the ECQ have been defined (such as for *Wx*). However, almost all of the genes or QTLs identified in previous studies were obtained from a specific cross, which cannot provide a clear picture to fully understand the allelic variation among the existing cultivars. Thus the genetic network among SSRGs/

Table 3 Novel alleles detected in the 64 varieties

Gene	No. of haplotypes expected	No. of haplotypes detected	Varieties carrying novel haplotypes
AGPlar	3	3	
AGPiso	3	4	Teqing, Cpslo17, Nongken 57
AGPsma	4	4	
Wx	3	5	Aman, Bulu, Guihuahuang, Chaofengzhan
GBSS II	2	2	
SSI	3	4	Dular, Boro, Maweizhan
SS II-1	3	6	Dular, Boro, Nantehao, Nanjing 11, Maweizhan, E32, Zinuo, Eyi 105, Ketan Nangka, Cpslo17
SS II-2	3	5	Nongken 57, Lunhui 422, Cpslo17, Zhonghua 11
SS II-3	3	5	Lemont, Tjereh, Newrex
SSIII-1	5	6	Milyang 46, Yangfuxian 2, Lemont
SSIII-2	3	3	
SSIV-1	2	2	
SSIV-2	3	4	Aman, Milyang 46, Cpslo17
SBE1	3	6	Teqing, Zhendao 88, Hanfeng, Dular, Aman, Boro, Lemont, Ketan Nangka, Guihuahuang, Sidao 10, Zhuzhongnuo, Xiangzhunuo, Zinuo, Majinnuo, Cpslo17
SBE3	2	3	Boro, Aman, Bulu, Tjereh, Lemont
SBE4	3	4	Hanfeng, Zhuzhongnuo, Zinuo, Majinnuo, Lunhui 422, 02428, Xiangjing49
ISA	2	5	Dular, Xiangzhunuo, E32, Majinnuo
PUL	4	6	Dular, Lemont, Guihuahuang, Zhendao 88, Newrex, Zhonghua 11, Koshihikari, Dijing, Eyi 105, Wuxiangjing 3, Aman, Nanjing 11, Yangfuxian 2, Shuiyuan 262, Yuexiangzhan, Ketan Nangka, Nongken 57, Ballila, Lunhui 422

QTLs and their influences on grain quality remained unclear. Therefore, it is powerless to guide the quality improvement based on the results obtained in these studies [40].

In this study, we developed 51 gene-tagged markers according to the sequence variations in SSRGs among different cultivars. These markers can be used to define the genotypes of SSRGs loci from different varieties and to evaluate the genetic effects of SSRGs on the grain quality. Two to three markers, on average, were developed for each SSRG locus, tagging the different mutation sites of the target gene. Therefore, the availability of these markers enables us to exploit more alleles. The genotyping of 64 varieties by the newly developed markers indicated that there are multiple allelic differentiations for most of the SSRGs loci. We found novel alleles for 13 SSRGs loci, which were different from those in the 16 varieties used for the sequence analysis. These results suggested that the markers we developed could be used by MAS for the improvement of rice grain quality, and also have the potential to identify favorable multiple alleles in the rice germplasm resource.

Our results also demonstrated that the varieties from different countries or different ecotypes might carry different alleles, and that varieties from the same cultivating region carry similar genotypes for SSRGs loci. For example, some novel alleles can be identified in varieties from India and America. These results suggested that the rice ECQ trait had been subjected to artificial selection over a long period of cultivation and that more novel alleles might exist in wild rice species. As a result, more attention should be paid to wild species, in addition to the cultivated germplasms, to find favorable multiple alleles,

This work was supported by the National High Technology Research and Development Program of China (2006AA10Z118, 2006AA10A102, 2005CB120804), National Natural Science Foundation of China (30530470), Fok Ying Tung Young Teachers' Fund (10103) and National Postdoctoral Fellowship (20060390966).

1 Mo H D. Quality improvement of rice grain in China (in Chinese). Sci Agri Sin, 1993, 26: 8–14
2 Cheng S H, Min S K. Rice varieties in China: Status and Prospects (in Chinese). China Rice, 2000, 1: 13–16
3 Li Z K. Strategies for molecular rice breeding in China (in Chinese). Mol Plant Breed, 2005, 3: 603–608
4 Feng J C. Application of Marker-assisted Selection in rice breeding (in Chinese). Chin Agric Sci Bull, 2006, 22: 43–47
5 Bernatzky R, Tanksley S D. Restriction fragments as molecular markers for germplasm evaluation and utilization. In: Brown A H D, Frandel O H, Marshall D R, et al, eds. The Use of Plant Genetic Resources. Cambridge: Cambridge University Press, 1989. 353–362
6 Juliano B O. Criteria and test for rice grain quality. In: Juliano B O, ed. Rice Chemistry and Technology. Minnesota: American Association of Cereal Chemists, 1985. 443–513
7 Myers A M, Morell M K, James M G, et al. Recent progress toward understanding biosynthesis of the amylopectin crystal. Plant Physiol, 2000, 122: 989–997
8 Smith A M. The biosynthesis of starch granules. Biomacromolecules, 2001, 2: 335–341
9 Nakamura Y. Towards a better understanding of the matabolic system for amylopectin biosynthesis in plants: Rice endosperm as a model tissue. Plant Cell Physiol, 2002, 43: 718–725
10 Martin C, Smith A M. Starch biosynthesis. Plant Cell, 1995, 7: 971–985
11 Ball S, Guan H P, James M, et al. From glycogen to amylopectin: A model for the biogenesis of the plant starch granule. Cell, 1996, 86: 349–352
12 James M G, Denyer K, Myers A M. Starch synthesis in the cereal endosperm. Curr Opin Plant Biol, 2003, 6: 215–222
13 Baba T, Nishihara M, Mizuno K, et al. Identification, cDNA cloning, and gene expression of soluble starch synthase in rice (*Oryza sativa* L.) immature seeds. Plant Physiol, 1993, 103: 565–573
14 Kawasaki T, Mizuno K, Baba T, et al. Molecular analysis of the gene encoding a rice starch branching enzyme. Mol Gen Genet, 1993, 237: 10–16
15 Kubo A, Fujita N, Harada K, et al. The starch-debranching enzymes isoamylase and pullulanase are both involved in amylopection biosynthesis in rice endosperm. Plant Physiol, 1999, 121: 399–409
16 Fujita N, Yoshida M, Asakura N, et al. Functional and characterization of starch synthase I using mutants in rice. Plant Physiol, 2006, 140: 1070–1084
17 Dian W, Jiang H, Wu P. Evolution and expression analysis of starch synthase III and IV in rice. J Exp Bot, 2005, 56: 623–632
18 Jiang H, Dian W, Liu F, et al. Molecular cloning and expression analysis of three genes encoding starch synthase II in rice. Planta, 2004, 218: 1062–1070
19 Nakamura Y, Yamanouchi H. Nucleotide sequence of a cDNA encoding starch-branching enzyme, or Q-enzyme I, from rice endosperm. Plant Physiol, 1992, 99: 1265–1266
20 Nakamura Y, Francisco P B Jr, Hosaka Y, et al. Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between *japonica* and *indica* rice varieties. Plant Mol Biol, 2005, 58: 213–227
21 Wang Z Y, Wu Z L, Xing Y Y, et al. Nucleotide sequence of rice *Waxy* gene. Nucleic Acids Res, 1990, 18: 5898
22 Wang Z Y, Zheng F Q, Shen G Z, et al. The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. Plant J, 1995, 7: 613–622
23 Sano Y. Different regulation of *Waxy* gene expression in rice endosperm. Theor Appl Genet, 1984, 68: 4567–4573
24 Mikami I, Uwatoko N, Ikeda Y, et al. Allelic diversification at the wx locus in landraces of Asianrice. Theor Appl Genet, 2008, 116: 979–989
25 Gao Z Y, Zeng D L, Cui X, et al. Map-based cloning of the *ALK* gene, which controls the gelatinization temperature of rice. Sci China Ser C-Life Sci, 2003, 46: 661–668
26 Bao J S, Corke H, Sun M. Nucleotide diversity in starch synthase IIa and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (*Oryza sativa* L.). Theor Appl Genet, 2006, 113: 1171–1183
27 Han Y P, Xu M L, Liu X Y, et al. Genes coding for starch branching enzymes are major contribution to starch viscosity characteristics in waxy rice (*Oryza sativa* L.). Plant Sci, 2004, 166: 357–364
28 Tian Z X, Qian Q, Liu Q Q, et al. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualityes. Proc Natl Acad Sci USA, 2009, 106: 21760–21765
29 Rogers S O, Bendich A J. Extraction of DNA from plant tissues. In: Gelvin S B, Schilperoort R A, eds. Plant Molecular Biology Manual. Boston: Kluwer Academis Publishers, 1988. A1~10
30 Wanchana S, Toojinda T, Tragoonrung S, et al. Duplicated coding sequence in the waxy allele of tropical glutinous rice (*Oryza sativa* L.). Plant Sci, 2003, 165: 1193–1199
31 Cai X L, Liu Q Q, Tang S Z, et al. Development of a molecular marker for screening the rice cultivars with intermediate amylose content in *Oryza sativa* subsp. *indica* (in Chinese). J Plant Physiol Mol Biol, 2002, 28: 137–144
32 He P, Li S G, Qian Q, et al. Genetic analysis of rice grain quality. Theor Appl Genet, 1999, 98: 502–508
33 Li J, Xiao J, Grandillo S, et al. QTL detection for rice grain quality

traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. Genome, 2004, 47: 697–704

34 Lanceras J C, Huang Z L, Naivikul O, et al. Mapping of genes for cooking and eating qualities in *Thai jasmine* rice (KDML105). DNA Res, 2000, 7: 93–101

35 Umemoto T, Yano M, Satoh H, et al. Mapping of a gene responsible for the difference in amylopectin structure between *japonica*-type and *indica*-type rice varieties. Theor Appl Genet, 2002, 104: 1–8

36 Wan X Y, Wan J M, Su C C, et al. QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. Theor Appl Genet, 2004, 110: 71–79

37 Septiningsih E M, Trijatmiko K R, Moeljopawiro S, et al. Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theor Appl Genet, 2003, 107: 1433–1441

38 Tan Y F, Li J X, Yu S B, et al. 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. Theor Appl Genet, 1999, 99: 642–648

39 Aluko G, Martinez C, Tohme J, et al. QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. Theor Appl Genet, 2004, 109: 630–639

40 Tanksley S D, Nelson J C. Advanced backcross QTL analysis: A method for the simultanouse discovery and transfer of valuable QTLs from unadapted germplasm into elite breedng lines. Theor Appl Genet, 1996, 92: 191–203