Theor Appl Genet (2006) 113:1171–1183
DOI 10.1007/s00122-006-0355-6

ORIGINAL PAPER

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.)

J. S. Bao · H. Corke · M. Sun

Received: 13 September 2005 / Accepted: 21 June 2006 / Published online: 19 July 2006

Abstract The characteristics of starch, such as gelatinization temperature (GT), apparent amylose content (AAC), pasting temperature (PT) and other physicochemical properties, determine the quality of various products of rice, e.g., eating, cooking and processing qualities. The GT of rice flour is controlled by the *alk* locus, which has been co-mapped to the *starch synthase IIa* (*SSIIa*) locus. In this study, we sequenced a 2,051 bp DNA fragment spanning part of intron 6, exon 7, intron 7, exon 8 and part of 3′ untranslated region of *SSIIa* for 30 rice varieties with diverse geographical distribution and variation in starch physicochemical properties. A total of 24 single nucleotide polymorphisms (SNPs) and one insertion/deletion (InDel) were identified, which could be classified into nine haplotypes. The mean pairwise nucleotide diversity π was 0.00292, and Watterson's estimator θ was 0.00296 in this collection of rice germplasm. Tajima's D test for selection showed no significant deviation from the neutral expectation ($D = -0.04612$, $P > 0.10$). However, significant associations were found between seven of the SNPs and peak GT (T_p) at $P < 0.05$, of which two contiguous SNPs (GC/TT) showed a very strong association with T_p ($P < 0.0001$). With some rare exception, this GC/TT polymorphism alone can differentiate rice varieties with high or intermediate GT (possessing the GC allele) from those with low GT (possessing the TT allele). In contrast, none of these SNPs or InDel was significantly associated with amylose content. A further 509 rice varieties with known physicochemical properties (e.g., AAC and PT) and known alleles of other starch synthesizing genes were genotyped for the *SSIIa* GC/TT alleles. Association analysis indicated that 82% of the total variation of AAC in these samples could be explained by a (CT)*n* simple sequence repeat (SSR) and a G/T SNP of Waxy gene (*Wx*), and 62.4% of the total variation of PT could be explained by the GC/TT polymorphism. An additional association analysis was performed between these molecular markers and the thermal and retrogradation properties for a subset of 245 samples from the 509 rice varieties. The *SSIIa* GC/TT polymorphism explained more than 60% of the total variation in thermal properties, whereas the SSR and SNP of *Wx* gene explained as much as the *SSIIa* GC/TT of the total variation in retrogradation properties. Our study provides further support for the utilization of the GC/TT polymorphism in *SSIIa*. As shown in our study of 509 rice varieties, the GC/TT SNP could differentiate rice with high or intermediate GT from those with low GT in about 90% of cases. Using four primers in a single PCR reaction, the GC/TT polymorphism can be surveyed on a large scale.

Communicated by Q. Zhang

J. S. Bao (✉)
Institute of Nuclear Agricultural Sciences,
College of Agriculture and Biotechnology,
Zhejiang University, Hua Jiachi Campus,
Hangzhou 310029, People's Republic of China
e-mail: jsbao@zju.edu.cn

H. Corke
Department of Botany, The University of Hong Kong,
Hong Kong, People's Republic of China

M. Sun (✉) · J. S. Bao
Department of Zoology, The University of Hong Kong,
Pokfulam Road, Hong Kong, People's Republic of China
e-mail: meisun@hku.hk

Thus, this SNP polymorphism can be very useful in marker-assisted selection for the improvement of GT and other physicochemical properties of rice.

Introduction

Differences in starch properties of rice grains determine their eating, cooking and processing qualities. Apparent amylose content (AAC) and gelatinization temperature (GT) are two of the well-established parameters used to evaluate these qualities (Juliano 1998; Bergman et al. 2004). For example, low-GT rices have a softer texture than high-GT rices among freshly cooked waxy and low-AAC rices; and among intermediate- and high-AAC rices, intermediate-GT rices are softer than low-GT rices when freshly cooked, but they have similar hardness values on accelerated stalling (Perez et al. 1993). Different rice products require different types of rice with different starch properties, e.g., low GT is preferred in manufacturing rice breads and beer (Juliano 1998). Knowledge of the relationship between physicochemical properties and end-use qualities can direct breeding activities in selecting desired rice with unique qualities. However, lack of information on the genetic basis and related techniques for selection retards the breeding progress.

The GT of starch can be measured indirectly by alkali spreading value or directly by differential scanning calorimetry (DSC) as peak temperature (T_p) from the endotherm. The pasting temperature (PT) measured by viscoamylography such as using Rapid Visco Analyzer is correlated to GT. Genetically, GT is possibly controlled by one (McKenzie and Rutger 1983), two or three genes (Chang and Li 1991). Through quantitative trait locus (QTL) mapping, it has been reported that GT may be controlled either by the alkali degeneration gene (*alk*) (He et al. 1999; Bao et al. 2004; Fan et al. 2005) or by the *Wx* gene (Tan et al. 1999). Umemoto et al. (2002) reported that the *starch synthase IIa* (*SSIIa*) gene is located at the *alk* locus on chromosome 6 in the rice genome. Gao et al. (2003) reported the map-based cloning of the *alk* locus that encodes *SSIIa*, and found that nucleotide substitutions in the coding sequence of *SSIIa* may cause the alteration in GT. Jiang et al. (2004) cloned three isoforms of starch synthase II including *SSIIa* by cDNA library screening, finding that *SSIIa* was mainly expressed in endosperms. Chen et al. (2003) identified two separate mutations in *SSIIa* that were associated with low GT. Umemoto et al. (2004) studied three single nucleotide polymorphisms (SNPs) associated with GT using haplotype analysis. More recently, Nakamura et al. (2005) analyzed in detail the effect of amino acid replacement caused by these SNPs on the enzyme activity and on the amylopectin structure and GT, and the results indicated that two of the SNPs (at 4,198 and 4,229/4,330 bp as positioned in the present study) are essential for *SSIIa* activity and granule association.

Association mapping, also referred to as linkage disequilibrium (LD) mapping, is a popular method that has been widely used to test the relationship between specific sequence polymorphisms in candidate genes and phenotypic variation (Thornsberry et al. 2001; Flint-Garcia et al. 2003; Gupta et al. 2005), facilitating identification of actual functional polymorphisms within candidate gene and epistatic interactions (Hagenblad et al. 2004; Caicedo et al. 2004). LD studies conducted in plants have recently been summarized by Gupta et al. (2005), and examples using association mapping are still increasingly documented (Andersen et al. 2005; Simko et al. 2004; Szalma et al. 2005; Bundock and Henry 2004). Association analyses have also been conducted for the *SSIIa* gene in relation to GT, but only two or three SNPs were used in such studies (Chen et al. 2003; Umemoto et al. 2004), resulting in the same haplotypes consisting of rices with both high and low GT. Furthermore, previous studies began with alignment of *SSIIa* sequences to find SNPs among a few rices (Chen et al. 2003; Umemoto et al. 2004), e.g., the two genome-sequenced rice cultivars, 93-11 and Nipponbare (Chen et al. 2003). The few sequences studied unlikely represent all naturally occurring variations in this gene. Therefore, it is necessary to search for more SNPs from a different set of germplasm in order to find those that can differentiate high- and low-GT rice varieties. Furthermore, nucleotide diversity and LD in *SSIIa* have not been investigated.

In this study, nucleotide diversity of *SSIIa* gene was characterized based on the sequence variation in 30 diverse rice varieties, and association tests were performed to investigate the relationships between the SNPs found in these rices and their GT values. Additionally, to verify the relationship between *SSIIa* GC/TT polymorphism and starch physicochemical properties, four primers were designed for amplifying the two continuous SNPs in a single PCR reaction for a total of 509 rice varieties with known alleles at other starch synthesizing gene loci and known amylose content and pasting properties (Bao et al. 2006b). The GC/TT polymorphism can differentiate rices with high or intermediate GT from those with low GT at a rate of about 90% correct prediction. A further marker association analysis was conducted for 245 of the rices with

known thermal and retrogradation properties (Bao et al. 2006c). The findings presented in this paper provide further support for the utility of association studies between molecular markers and physicochemical properties of starch, especially between the *SSIIa* GC/TT polymorphism and GT in rice on a large scale.

Materials and methods

Rice materials

For sequencing analysis of the *SSIIa* gene, 30 rice varieties with a wide geographical distribution were included in the study, e.g., IR64 (BP052), IR1552 (BP020) and Azucena (BP021) from the Philippines; Lemont (BP015), Sierra (BP038), and Bolivar (BP039) from the USA; IRGA 409 (BP037) from Brazil, and all others from China. The accession BP032 initially named as Nipponbare in our sample collection was shown to be a misnomer based on its sequence analysis. Therefore, we renamed the accession simply as ''Unknown'' in this study to distinguish it from the true cultivar Nipponbare.

These rice samples also differ drastically in AAC and GT (Table 1). Their alleles at other starch synthesizing gene loci have already been genotyped (Bao et al. 2006a), and a wide range of allelic diversity was found (Table 1). Even though some of the rice samples have the same alleles, their starch properties are different. Additionally, a common wild rice, *Oryza rufipogon* (BP580) was included for comparison in the sequencing analysis (Table 1).

To analyze the relationship between the GC/TT polymorphism and starch physicochemical properties, a total of 509 rice samples with known alleles at other starch metabolic gene loci (Bao et al. 2006a) and known amylose content and pasting properties (Bao et al. 2006b) were included for genotyping the GC/TT SNP of *SSIIa*. Because only 245 of the 509 rice samples were previously measured for the thermal and retrogradation properties (Bao et al. 2006c), further association analyses of the GC/TT and markers of other starch genes with the thermal and retrogradation properties were conducted for these 245 samples only.

Analysis of starch physicochemical properties

The starch physicochemical properties, including AAC, pasting properties, gel texture, thermal and retrogradation properties, have been reported elsewhere (see Bao et al. 2006b, c for detailed methods). The measurement of thermal properties of brown rice flour was performed with DSC as described by Bao et al. (2006c).

DNA extraction, PCR amplification, and sequencing

Total genomic DNA was extracted from five seedlings of each rice cultivar using the CTAB method (Doyle 1991). Two *SSIIa* sequences (AY423717 genomic sequence and AF419099 mRNA sequence) were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/), which were used as the query sequences to conduct BLAST search with genome sequences of 93-11 and Nipponbare to find 5′ and 3′ untranslated regions (UTRs). Based on the sequences of AP003509 (Nipponbare) and AAAA01011906 (93-11), PCR primers (Table 2) were designed with the GeneTool software 1.0 (BioTools, Inc., Edmonton, Alberta). Part of the *SSIIa* gene, i.e., from 2,705 to 4,851 bp spanning part of intron 6, exon 7, intron 7, exon 8 and part of 3′ UTR, was amplified in two separate PCR reactions with the primer combinations of F17/R6 and F7/R1 (Table 2), respectively. Part of the 3′ UTR was an extension from the sequence AY423717 (4,422 bp).

Each 30 μl amplification reaction consisted of 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 0.1% Triton X 100, 2 mM $MgCl_2$, 0.1 mM dNTPs, 200 nM primers, 0.5 unit of Taq polymerase, and 20 ng of genomic DNA. All amplifications were performed on a PTC-100 thermal cycler (MJ Research, Inc.) under the following conditions: 5 min at 94°C, followed by 45 s at 94°C, 60 s at 58°C, and 60 s at 72°C for 35 cycles, and 10 min at 72°C for a final extension.

The PCR product was purified with High Pure PCR Product Purification Kit (Roche Diagnostics GmbH, Germany). DNA sequencing was performed on an ABI 3100 automated sequencer following the manufacturer's instructions (Applied Biosystems, Inc.). Direct sequencing was performed using four primers (F7, F17, R1 and R6) and others primers (F18, R18 and R20) (Table 2). The rare SNPs were re-sequenced with additional independent PCR products. A total of 2,051 bp, from position 2,705 to 4,755, were used in the final analysis.

Statistical analysis

The nucleotide diversity (π, the average number of nucleotide difference per site between two sequences), Watterson's (1975) estimator of θ (θ_w), Tajima's D (statistic test of selection, Tajima 1989), and LD were estimated using DNASP 4.0 (Rozas and Rozas 1999).

Table 1 Rice genotypes used in sequencing analysis with references to the differences in molecular marker alleles, amylose content and gelatinization temperature

Code	Name	*Wx* SSR[a]	*Wx* SNP[b]	*SS1* SSR[c]	*SBE1* SSR[d]	*SBE1* STS[e]	*SBE3* SNP[f]	AAC[g]	T_p[h]
BP003	Jiayu 293	11	G	B	A	NI	C	25.6	75.2
BP005	Zhefu 504	18	T	B	A	NI	C	12.9	78.7
BP011	Youzaonuo	18	T	B	A	NI	C	2.1	79.5
BP015	Lemont	20	G	C	B	NI	C	23.0	75.7
BP020	IR1552	11	G	B	D	NI	C	26.3	67.7
BP021	Azucena	19	G	C	B	NI	G	20.9	74.1
BP032	''Unknown''	18	T	C	C	I	G	15.5	67.5
BP033	93-11	18	T	B	A	NI	C	16.9	68.7
BP034	Guangluai 4	11	G	A	A	NI	C	26.5	73.9
BP037	IRGA 409	11	G	A	A	NI	C	28.0	66.6
BP038	Sierra	11	G	C	B	NI	G	27.9	74.6
BP039	Bolivar	11	G	B	B	NI	C	24.7	76.3
BP049	Xieqingzao B	11	G	A	A	NI	C	29.1	74.8
BP052	IR64	17	G	C	A	NI	C	22.5	76.6
BP131	Zaoshuheidu	12	G	A	A	NI	G	28.3	74.6
BP133	Liuyuenuo	11	G	A	A	NI	G	30.7	63.2
BP139	Zhenzhuzao	11	G	A	A	NI	C	27.0	73.9
BP208	39-21	11	G	C	A	NI	C	28.9	67.0
BP216	Huangkenuo	17	T	A	A	NI	C	2.2	69.2
BP275	Wushizao	11	G	D	A	NI	G	26.7	73.8
BP289	Ganxiandahezi 3	21	T	C	B	NI	G	14.4	70.3
BP340	Xiushui 63	17	T	C	C	I	G	17.2	68.6
BP342	Zheda 413	17	T	C	C	I	G	2.0	70.3
BP419	Zounuo	17	T	C	C	I	G	2.0	78.3
BP442	Cisangarung	8	G	D	A	NI	C	29.7	64.1
BP508	Longtefu B	11	G	A	A	NI	C	29.1	66.3
BP553	Minghui 63	18	T	C	A	NI	G	12.0	79.5
BP573	Jiainuo	17	T	A	A	NI	C	1.7	69.4
BP574	Dalixiangnuo	11	G	B	A	NI	C	26.4	67.1
BP580[i]	Dongxiang wild rice	10	G	C	C	NI	G	-	64.3

Data from Bao et al. (2006a, b, c)

[a]Number of (CT)*n* repeats of *Wx* gene

[b]T = AGTTATAC; G: AGGTATAC

[c]*SSS*-A: $(AC)_2$...TCC$(TC)_{11}$...$(TC)_5C(ACC)_{11}$; *SSS*-B: $(AC)_3$...TCT$(TC)_6$...$(TC)_4C(ACC)_9$; *SSS*-C: $(AC)_3$...TCT$(TC)_6$...$(TC)_4C(ACC)_8$; *SSS*-D: $(AC)_2$...TCC$(TC)_{10}$...$(TC)_4C(ACC)_9$

[d]*SBE*-A: CTCTCGGGCGA...$(CT)_{10}$; *SBE*-B: CTCTCGGGCGA...$(CT)_8$; *SBE*-C: $(CT)_8$; *SBE*-D: CTCTCGGGCGA...$(CT)_9$

[e]I = containing transposon insertion STS, NI = no transposon insertion STS

[f]G = ACTAGT; C = ACTACT

[g]AAC represents apparent amylose content

[h]T_p represents peak temperature measured by DSC

[i]AAC of BP580 was not measured, whereas the T_p was measured on the brown rice flour

Table 2 Primer sequences used to amplify *SSIIa* gene of rice

Primer[a]	Sequence (5′ → 3′)	T_m (°C)
F7 (3,517–3,540)	CTGGATCACTTCAAGCTGTACGAC	57
F17 (2,705–2,728)	CTTCGATACCTCTAGCAGCATTTC	54
F18 (3,112–3,134)	AGCAGGGACACGATAAACTCTTC	53
F22 (4,311–4,330)	CAAGGAGAGCTGGAGGGGGC	68
R1 (4,833–4,853)[b]	GCCGGCCGTGCAGATCTTAAC	57
R6 (3,698–3,717)	GTCGATGCCGTTCACGATGC	56
R18 (4,125–4,144)	TGGCGTAGAGCTGGTTGAGG	64
R20 (3,156–3,177)	TGAAACGGAGAAATTGCTGAAC	54
R21 (4,331–4,348)	ACATGCCGCGCACCTGGAAA	64

[a]*F* forward primer, *R* reverse primer, the number in parenthesis refers to the nucleotide positions in the accession AY423717

[b]This sequence is located at 3′ UTR of *SSIIa*, not included in AY423717, so the sequence numbers were extended from AY423717

Phylogenetic analysis was performed using PAUP*4.0 b10 (Swofford 2003) with wild rice BP580 as an outgroup. The branch support values of the neighbor-joining tree were obtained using 1,000 bootstrap replications.

Analysis of variance (ANOVA) was performed with the SAS System for Windows version 8 (SAS Institute Inc., Cary, NC, USA). Duncan's multiple range test was conducted for comparison of means. The general linear model was used in the analysis of associations between nucleotide polymorphisms or marker alleles and different starch physicochemical properties, which was performed with Tassel (Trait analysis by association, evolution and linkage) Version 2.0 software (E. S. Buckler IV, http://www.maizegenetics.net/bioinformatics/tasselindex.htm).

Results

Nucleotide diversity of *SSIIa*

The sequence data for 30 rice varieties were obtained from two DNA fragments amplified using separate PCR reactions. The total sequence length used in the final analysis was 2,051 bp long, positioned from base 2,705 to base 4,755 according to the sequence AY423717. The mean pairwise nucleotide diversity π was 0.00292, and Watterson's estimator θ was 0.00296. Tajima's (1989) test for selection indicated that these rices did not show a significant deviation from the neutral expectation ($D = -0.04612$ and $P > 0.10$). The diversity indices of the non-coding regions ($\pi = 0.00376$ and $\theta = 0.00409$) were higher than those of the coding regions ($\pi = 0.00223$ and $\theta = 0.00201$). But separate Tajima's D test for each region revealed no significant departure from the neutral expectation (0.32634 for non-coding regions and – 0.28229 for coding regions, $P > 0.10$). In the coding sequences, the expected synonymous sites were 274, and nonsynonymous sites were 854, and the π (0.00568) and θ (0.00552) of the synonymous sites were fivefold greater than those of the nonsynonymous sites ($\pi = 0.00111$ and $\theta = 0.00089$).

Linkage disequilibrium

The average squared allele frequency correlation (r^2) was 0.275, partially attributable to some rare SNPs in the sequences. Eight of the 25 SNPs found in the sequences were considered to be rare alleles because they appeared in these samples only once. However, the LD did not decay over 2 kb sequence (data not shown).

SNP haplotype

There were 24 SNPs and one insertion/deletion (InDel) in the *SSIIa* gene among the 30 rice samples sequenced (Table 3), averaging about one SNP or InDel per 82 bp. One SNP (at site 3,233 bp) has three substitutions, and the common nucleotide C was replaced by A in rice BP052 and by G in rice BP553, respectively. Except for this SNP, the sequence of BP553 was essentially the same as that of BP049 (Table 3). The frequency of nucleotide substitutions was higher in the noncoding regions (15 SNPs and one InDel) than in the coding regions (nine SNPs) (Table 3). The GC/TT (at sites 4,329 and 4,330) was the only locus that had a nearly even allele distribution in the 30 rices, 14 rices with the GC allele and 16 with the TT allele. As to other SNP sites, at most seven rices belong to the same SNP type.

Of the nine SNPs in the coding regions, six were silent substitutions, and the other three resulted in amino acid replacement. The first site was at the 3,799 bp, where the amino acid glycine encoded by GGC was replaced by serine encoded by AGC. The second site was at the 4,198 bp, where valine encoded by the common allele GTG was replaced by methionine encoded by ATG in rice BP289. The third site was at the 4,330 bp, the glycine–leucine encoded by GGGCTC was replaced by glycine–phenylalanine encoded by GGTTTC.

The combination of these 25 SNPs and InDel resulted in a total of nine haplotype groups (Table 3). Three of the haplotypes, HP1, HP3 and HP7 included eight, four, and 12 rice varieties, respectively, whereas each of the other six haplotype groups contained only one variety (Table 3).

With wild rice BP580 as the outgroup, the rest of rice samples were placed into two major clades in the neighbor-joining tree (Fig. 1), corresponding to the high- or intermediate-GT (the top clade) and low-GT (the bottom clade) classes. The nine haplotypes were placed into different clades or subclades (Table 3 and Fig. 1). With rare exceptions, the low-GT rices had the TT allele, whereas the high-GT rices had the GC allele at sites 4,329/4,330. The subclade including BP039, BP131, BP139, BP275, BP052, BP553 and BP049 was well separated from the rest of high-GT rice varieties with strong bootstrap support (100%). However, there were three samples (BP133, BP289 and BP419) that did not follow the general GC/TT rule. For these three samples, we measured their GT with DSC again and the results confirmed that the original measurements of GT were true.

Table 3 Nucleotide polymorphisms in the *SSIIa* gene associated with AC and GT and comparison of AC and GT among the three common haplotype groups

Haplotype[a]	Intron 6		Exon 7		Intron 7												Exon 8						3′ UTR		AAC[d]	T_p[d]
	2,739[b]	2,781	2,920	2,955	3,003	3,164	3,168	3,170	3,176	3,206	3,222	3,233	3,239	3,338	3,414	3,437	3,799	3,834	3,903	3,918	4,198	4,329–4,330	4,682[c]	4,699[c]		
HP1 (8)	C	C	C	G	G	A	C	C	C	T	C	C	C	G	G	–	A	G	T	C	G	GC	A	A	21.2[a]	74.4[a]
HP2 (1)	C	C	C	G	G	A	C	C	C	T	C	C	C	G	G	–	A	G	T	C	A	GC	A	A	14.4	70.3
HP3 (4)	A	C	C	G	G	C	C	T	C	T	C	C	A	A	T	A	G	A	G	T	G	GC	G	G	26.6[a]	74.7[a]
HP4 (1)	A	C	T	G	T	C	C	T	A	T	C	C	C	G	G	–	G	G	G	C	G	GC	G	G	29.1	74.8
HP5 (1)	A	**T**	C	G	G	C	C	T	C	**C**	C	**A**	A	A	T	A	G	G	G	C	G	GC	G	G	22.5	76.6
HP6 (1)	A	C	T	G	T	C	C	T	A	T	C	**G**	A	A	T	A	G	G	G	C	G	GC	G	G	12	79.5
HP7 (12)	C	C	C	G	G	A	C	C	C	T	C	C	C	G	G	–	A	G	T	C	G	TT	A	A	17.4[a]	68.5[b]
HP8 (1)	C	C	C	G	G	A	**G**	C	C	T	C	C	C	G	G	–	A	G	T	C	G	TT	A	A	17.2	68.6
HP9 (1)	A	C	C	**A**	G	A	C	C	C	T	**T**	C	C	G	G	–	A	G	T	C	G	TT	A	A	–	64.3
AAC[e]	0.06	0	0	0	0	0.06	0	0.06	0	0	0	0.03	0.03	0.03	0.03	0.03	0.06	0.08	0.06	0.08	0.01	0.06	0.06	0.06		
GT	0.10	0.04	0.09	0.08	0.09	0.21*	0.01	0.21*	0.09	0.04	0.08	0.13	0.19	0.19	0.19	0.19	0.21*	0.06	0.21*	0.06	0	0.45****	0.21*	0.21*		

*Significant difference at $P < 0.05$ level

****Significant difference at $P < 0.0001$ level

[a]The number in parenthesis is the number of rices in this haplotype group. HP1 haplotype group includes BP003, BP005, BP011, BP015, BP021, BP034, BP038 and BP133; HP2 includes only BP289; HP3 includes BP039, BP131, BP139 and BP275; HP4 includes only BP049; HP5 includes only BP052; HP6 includes only BP553; HP7 includes BP020, BP032, BP033, BP 037, BP208, BP216, BP342, BP419, BP442, BP508, BP573, and BP574; HP8 includes only BP340, and HP9 includes only BP580

[b]The nucleotide site position refers to the sequence order of the gene accession AY423717; the letters in bold indicate rare SNPs

[c]These two loci are located at 3′ UTR that was not included in AY423717, so the sequences are extended from AY423717

[d]Only HP1, HP3 and HP7 were compared for AC and T_p, different letters indicate significant difference at $P < 0.05$ level

[e]Association of nucleotide polymorphisms with AAC and GT (r^2)

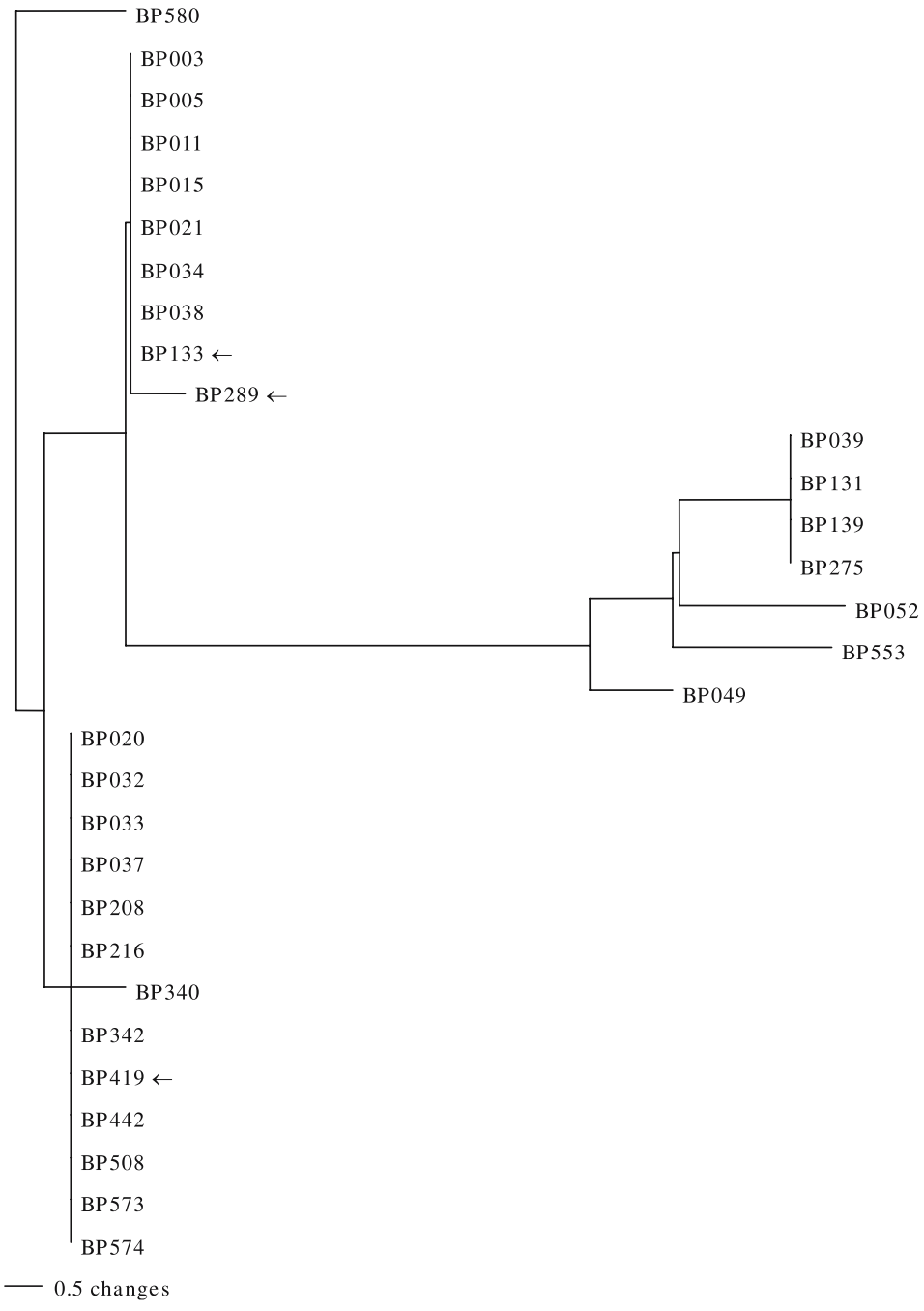

Fig. 1 Neighbor-joining tree based on the *SSIIa* gene sequences of 30 rice varieties. *Arrow* indicates the exception of gelatinization temperature to the GC/TT rule. See Table 1 for the name and GT value of each rice sample

Association test

The AAC and GT of three haplotype groups with more than one accession (Haplotypes 1, 3, and 7) were compared (Table 3). Haplotypes 1 and 3 had higher peak GT (T_p) than Haplotype 7 ($P < 0.05$), whereas no difference in AAC was found among the three haplotype groups (Table 3). None of the total 25 SNPs and InDel was significantly associated with AAC, whereas seven SNPs were significantly associated with T_p at $P < 0.05$, of which the two contiguous SNPs (GC/TT at sites 4,329–4,330) showed a very strong association with T_p ($P < 0.0001$) (Table 3).

Validation of the GC/TT SNP in association with starch gelatinization temperature

Because the GC/TT SNPs in the 30 rices were found to be significantly correlated with their GT, the confronting two-pair primers (Domon et al. 2004) were subsequently designed to amplify *SSIIa* GC/TT in a single PCR reaction (Fig. 2 and Table 2) for a total of

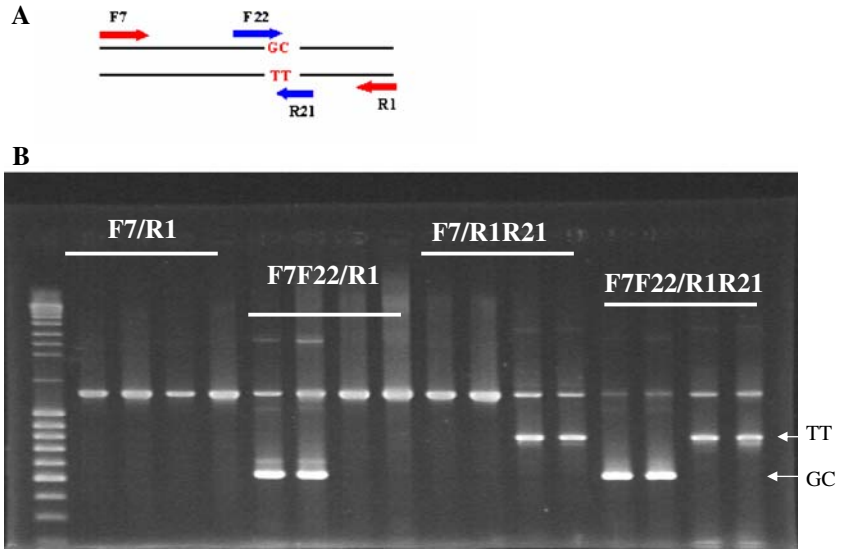

Fig. 2 Strategy to genotype the GC/TT SNPs. **a** Schematic representation of PCR with two-pair confronting primers (see Table 2 for primer sequences). **b** Gel image showing different PCR products amplified by different combinations of the primers

Table 4 Association of the marker alleles of starch synthesizing genes with apparent amylose content (AAC) and pasting temperature (PT) of rice starch (n = 509)

Marker	AAC		PT	
	r^2	$P > F$	r^2	$P > F$
Wx SNP	0.827	3.99E-194	0.214	3.72E-28
Wx SSR	0.819	1.05E-177	0.252	1.10E-26
*SSI*SSR	0.437	3.03E-62	0.156	2.16E-18
SSIIa SNP	0.271	1.59E-36	0.624	4.02E-109
SBE1 SSR	0.169	5.02E-20	0.171	2.40E-20
SBEI STS	0.155	3.42E-20	0.163	3.43E-21
SBE3 SNP	0.033	4.32E-05	0.040	5.32E-06
Total	0.892	2.82E-173	0.770	3.73E-103

See Table 1 for marker abbreviations

509 rices with known alleles of other starch metabolic genes and physicochemical properties (e.g., AAC, and PT which is another measurement of the GT) (Bao et al. 2006b). A total of 348 rice accessions had the GC SNP allele while 161 accessions had the TT SNP allele. In the 334 breeding lines, 176 and 158 had the GC SNP and TT SNP allele, respectively. In the 172 landraces, only two had the TT SNP whereas all others had the GC SNP allele. The wide rice sample (BP580) had the TT allele.

Associations analyses between starch physicochemical properties and the GC/TT SNP marker together with other molecular markers indicated that although all the molecular markers were significantly associated with AAC at $P < 0.0001$, most of the variation in AAC among the 509 rices could be explained by *Wx* SNP ($r^2 = 0.827$) and *Wx* SSR ($r^2 = 0.819$) (Table 4). Similarly, all the molecular markers were significantly associated with PT at $P < 0.00001$, but the *SSIIa* GC/TT polymorphism alone explained 62.4% of the total variation in PT, much higher than the variation explained by any other marker, such as *Wx* SSR (Table 4). Similar to the findings reported above from the association study between the GC/TT alleles and starch GT in the 30 sequenced rices, there were

Table 5 Association (r^2) of marker alleles of starch synthesizing genes with thermal and retrogradation properties of rice starch (n = 245)

Marker[a]	T_o	T_p	T_c	ΔH_g	ΔH_r	R%
Wx SNP	0.078	0.051	0.030	0.000	0.509	0.553
Wx SSR	0.130	0.090	0.057	0.030	0.539	0.581
*SSI*SSR	0.078	0.053	0.031	0.037	0.286	0.306
SSIIa SNP	0.660	0.659	0.624	0.202	0.596	0.590
SBE1 SSR	0.152	0.117	0.082	0.077	0.155	0.160
*SBEI*STS	0.144	0.109	0.073	0.067	0.150	0.155
SBE3 SNP	0.048	0.026	0.009	0.005	0.039	0.042
Total	0.851	0.840	0.812	0.445	0.872	0.890

[a]See Table 1 for marker abbreviations; T_o onset temperature, T_p peak temperature, T_c conclusion temperature, ΔH_g enthalpy of gelatinization, ΔH_r enthalpy of retrogradation and R% percentage of retrogradation

exceptions to the general correlation between GC/TT and PT: 25 of the 509 genotyped rice samples had the GC allele but with low PT.

Because only 245 of the 509 rice samples were previously used to measure the thermal and retrogradation properties with DSC (Bao et al. 2006c), further association analyses of molecular markers with these physicochemical properties were conducted for the 245 varieties only (Table 5). For the thermal properties, including onset temperature (T_o), peak temperature (T_p) and conclusion temperature (T_c), the *SSIIa* SNP explained more than 60% of the total variation, which were much higher than any other molecular markers (explaining < 15% each). For enthalpy of gelatinization (ΔH_g), even though *SSIIa* SNP only explained 20% of the total variation, it was still much higher than the variation explained by other individual markers. However, for the retrogradation properties, i.e., enthalpy of retrogradation (ΔH_r) and percentage of retrogradation (R%), *Wx* SNP and *Wx* SSR explained as much of the total variation as *SSIIa* SNP, suggesting that these two parameters may be controlled by both *Wx* and *SSIIa*.

Discussion

Analysis of genetic diversity of candidate gene can provide information on nucleotide polymorphism and LD, and can be used for association mapping and for inferring selection, domestication and demographic patterns (Buckler IV and Thornsberry 2002; Flint-Garcia et al. 2003; Gupta et al. 2005; Kim and Nielsen 2004). The discovery of SNPs of candidate gene has been reported in many crop species, such as maize (Remington et al. 2001; Andersen et al. 2005; Szalma et al. 2005) and rice (Garris et al. 2003).

Compared to the *Wx* gene sequence of rice (Olsen and Purugganan 2002), the genetic diversity indices of the *SSIIa* gene in this study are lower than those of *Wx* gene of nonwaxy rice ($\pi = 0.0045$ and $\theta = 0.0040$), but much higher than those of *Wx* gene of waxy rice ($\pi = 0.0003$ and $\theta = 0.0012$). The genetic diversity ($\pi = 0.00927$ and $\theta = 0.00789$) of the *Wx* gene of wild barley (Morrell et al. 2005) is greater than that of the same gene in rice (Olsen and Purugganan 2002). The Tajima's *D* test provides no evidence for selection on the *Wx* gene of nonwaxy rice (Olsen and Purugganan 2002) or on the *SSIIa* gene in this study, but selection is detected acting on the *Wx* gene of waxy rice ($D = -2.336$ and $P < 0.0001$) (Olsen and Purugganan 2002). The analysis of genetic diversity of starch metabolic genes in maize has also provided strong evidence for selection on three (*ae1*, *bt2* and *su1*) of the six genes studied (Whitt et al. 2002). The *SSIIa* plays an important role in the elongation of short chains of DP < 10 that leads to the formation of intermediate chains of amylopectin, which is especially responsible for gelatinization of starch (Umemoto et al. 2002; Gao et al. 2003; Jiang et al. 2004; Nakamura 2002). However, the lack of detectable selection on the *SSIIa* gene in the present study, either in the coding or noncoding region, may not be equated to a lack of directional selection for traits related to GT of starch in previous rice breeding programs. As shown in the present study, the nearly even frequency distribution of the GC (52.7%) and TT (47.3%) alleles in the 334 breeding lines is in sharp contrast with the rarity of the TT allele (1.2%) in the 172 landraces, suggesting directional selections for either high- or low-GT rice in breeding programs (see below for more discussion on the relationship between the GC/TT alleles and GT).

The frequency of nucleotide polymorphism in the 30 rice samples is about one SNP per 82 bp, higher than a previous report of 1 SNP/100 bp (Garris et al. 2003). This could be due to different samples included in the studies. In the coding regions, the frequency is one SNP per 125 bp, whereas in the noncoding regions, the frequency increases to one SNP per 58 bp. This is consistent with the prediction of purifying selection against deleterious or slightly deleterious nucleotide substitutions in the coding region. It has been reported that the frequency of SNP polymorphism changes in different genomic regions and in different samples (e.g., Garris et al. 2003).

Although the *SSIIa* SNPs in the 30 rice samples represent nine haplotypes, six of the nine haplotypes are each represented by only one sample. The presence of a few haplotypes shared by multiple individuals is indicative of LD, i.e., SNPs at one locus are not randomly assorted with alleles at another locus (Buckler IV and Thornsberry 2002; Flint-Garcia et al. 2003; Gupta et al. 2005; Rafalski 2002). This kind of LD should be expected in the present study, as the sources of our rice samples are diverse and obviously they cannot be equated with individuals of a large random mating population.

In association mapping, alleles of the candidate gene can be tested for association with a particular phenotype. The variation predicted to have a functional consequence, such as causing changes in amino acid or in the level of gene expression, should be the first choice for the association analysis. It is possible to use LD to identify actual functional polymorphisms within the candidate gene. There are four nonsynonymous mutations at the base positions 264, 1,810, 2,209, and

2,340/2,341 in the exons of the *SSIIa* gene (Nakamura et al. 2005). The latter three correspond to the SNPs at sites 3,799, 4,198 and 4,329/4,330, respectively, in the present study (Table 3). Umemoto et al. (2004) genotyped 65 rice cultivars for the first three SNPs and found four haplotypes, but no unique GT or amylopectin chain-length distribution was found for the haplotypes. More recently, Umemoto and Aoki (2005) genotyped the GC/TT SNPs for the same rice cultivars, and found that the SNPs at sites 4,198 and 4,329/4,330 (as positioned in this study) are strongly associated with chain-length distribution of amylopectin and the GT of rice flour. This finding was further supported by Waters et al. (2006) and by the present study. A detailed biochemical analysis by *SSIIa* gene shuffling experiment revealed that the SNPs at sites 4,198 and 4,329/4,330 are essential for the optimal *SSIIa* activity and granule associations, and that active *SSIIa* in rice plays a specific role in the synthesis of the long B1-chains by elongating short A and B_1 chains (Nakamura et al. 2005). Thus, the enzyme activity in rice with the GC allele is higher than in rice with the TT allele (Umemoto and Aoki 2005; Nakamura et al. 2005). Further work is necessary to find out whether other SNP site(s) exist inside or outside the *SSIIa* gene that could be responsible for the difference between high- and intermediate-GT in rice with the same GC allele background.

On the other hand, Nakamura et al. (2005) also showed that the SNP at site 4,198 was crucial for *SSIIa* activity. The enzyme was inactive when the site had the A SNP allele (coding for methionine), resulting in a low GT of starch, no matter which allele at sites 4,229/4,330 (GC/TT) was present. However, the A SNP allele was rare in most rice samples. The frequency of A at site 4,198 was 1 in 30 rice samples in the present study, 9 in 180 in the study of Chen et al. (2003), and 5 in 65 in the study of Umemoto et al. (2004). In contrast, our study and the previous studies (Chen et al. 2003; Umemoto et al. 2005; Waters et al. 2006) all showed that the GC/TT polymorphism was high compared to other SNPs, contributing to its strong relationship with starch GT in the association analysis (Table 3). However, the SNP at site 4,198 should not be overlooked, especially for the low-GT rice with the GC allele. As shown in the present study, 25 of the 509 rice samples had low PT values but with the GC allele; these rice samples could be further genotyped for this SNP.

Waters et al. (2006) showed that other cereal crops, such as wheat, barley and maize, all captured the same amino acid associated with high-GT as rice, supporting that the related GC allele in rice is most likely the ancestral or wild type and the TT allele represents mutations. Although selection of low-GT rice in breeding programs could contribute to a high frequency of the TT allele in the breeding lines, polymorphism at the SNP sites could also exist in the wild rice. The wild rice included in this study happened to have the TT allele. When it was used as an outgroup for phylogenetic analysis, the separation of the GC clade of rices from the TT clade could not be interpreted as *SSIIa* changed from TT to GC in the course of evolution of the gene (Fig. 1). As we only randomly selected one wild rice sample for inclusion in this study, further investigations that include more samples of wild rice are necessary to fully understand the GC/TT evolution of *SSIIa* gene in rice. Similarly, only one sequence of wheat, barley and maize was used in the analysis by Waters et al. (2006), it is not known whether polymorphisms exist at the corresponding nucleotide sites in these cereals because each of them also has a wide genetic diversity in GT.

It should be pointed out that population stratification may give spurious associations between a candidate marker and a phenotype in association mapping (Flint-Garcia et al. 2003; Freedman et al. 2004). Recently, empirical methods to detect stratification based on genotypes of unlinked markers have been described (Pritchard et al. 2000). Rice is well known for its divergence between the *indica* and *japonica* subspecies. The deep genetic structure in rice may be influenced by the natural history of ancestral populations prior to domestication, as well as by the autogamous breeding system and complexity of the breeding practices exercised by humans (Garris et al. 2005). Even in well designed studies, modest amounts of population stratification can still exist (Freedman et al. 2004), indicating the difficulties in detecting and controlling population stratification as a source of false positive associations (Hagenblad et al. 2004).

Because of the difficulties in obtaining information on population structure of the rice varieties included in our association analysis, we have used alleles of six other starch metabolic genes and another physicochemical trait of starch (amylose content) as controls or references (Tables 3, 4, 5). The association of *Wx* SSR and *Wx* SNP with AAC is easy to understand because it is the granule-bound starch synthase encoded by *Wx* that controls the synthesis of amylose (Ayres et al. 1997). QTL mapping has consistently confirmed that *Wx* is a major locus contributing to the variation of AAC in rice grain (He et al. 1999; Tan et al. 1999; Bao et al. 2004; Fan et al. 2005). The present study has shown that *Wx* SSR and *Wx* SNP are the most important in controlling the genetic basis of AAC, agreeing with those results from QTL mapping

(Tables 4, 5). Similarly, QTL mapping has detected the *alk* or *SSIIa* locus for GT (He et al. 1999; Umemoto et al. 2002; Gao et al. 2003; Bao et al. 2004). The present association analysis also indicated that the *SSIIa* GC/TT explained more than 60% of the total variation of the phenotypes (Tables 4, 5). The variation that could not be explained by the GC/TT SNPs may result, at least partially, from the difference between high- and intermediate-GT genotypes that cannot be differentiated by the GC/TT alone.

The retrogradation property evaluated by enthalpy of retrogradation (ΔH_r) and percentage of retrogradation (R%) is another interesting case. In a previous QTL mapping study of a population consisting of 107 recombinant inbred lines, both ΔH_r and R% were found to be significantly correlated with AAC and GT, but only the QTL close to the *Wx* locus was identified (Bao et al. 2003). Among the rice varieties used in this study, a previous correlation analysis also showed that these traits were significantly correlated with both AAC and GT (Bao et al. 2006c). However, the present association study indicates that both *Wx* and *SSIIa* SNP loci are important in controlling these traits, each explaining > 50% of the total variation for both traits (Table 5). This finding suggests that association mapping allow for finer mapping or higher resolution than the conventional QTL mapping. Another complicating factor could be the sample size used in the studies, as more rice varieties were used in the present study than in the previous study. Because all the markers we analyzed are from the genes involved in starch synthesis, it is not surprising that other gene loci also have statistically significant associations with the traits examined, but their contributions are relatively small when compared to the *Wx* SSR and *SSIIa* GC/TT SNPs (Tables 4, 5). Larkin and Park (2003) also showed that *SS1* microsatellite had a lesser and additive effects on AAC and paste viscosity characteristics, but they suggested that this locus may primarily be a linkage effect since it is only 5–10 cM distance from the *Wx* locus (Tanaka et al. 1995). Our previous QTL mapping indicated that other starch synthesizing genes, such as *SBE1* and *SBE3*, can be detected for AAC and GT associations in the population derived from parents with similar AAC, e.g., IR64 and Azucena (Bao et al. 2002).

SNPs and InDels are very useful genetic markers for LD analysis and association studies, and also for marker-assisted selection (MAS) (Rafalski 2002). Some of the SNPs discovered in genes and expressed sequence tags (EST) are functional markers that may causally affect phenotypic trait variation (Andersen and Lübberstedt 2003). These functional DNA markers are superior to other markers owing to their complete linkages with the alleles at the trait loci (Andersen and Lübberstedt 2003). It can be directly used in MAS without prior mapping if a relationship between the marker alleles and phenotypic variation has been established. The present study showed that the relationship between the GC/TT marker alleles and GT holds in 90% (27/30 sequenced rices) to 94% (479/509 genotyped rices) rice samples. This functional GC/TT SNP marker can therefore be used in diagnostic analysis to predict whether the rice's starch GT is high or low. It can also be used for MAS in rice breeding programs if the parents have different GC/TT alleles as well as divergent GTs. Our design of two-pair confronting primers facilitates a large-scale genotyping of the co-dominant GC/TT alleles in a single PCR reaction (Fig. 2). This PCR-based approach can be easily conducted for MAS, which circumvents the step of restriction enzyme digestion of the site containing the GC/TT marker.

Acknowledgments This project was supported in part by grants from National Natural Science Foundation of China (30300227), International Atomic Energy Agency (12847/R0) and the Hong Kong Research Grants Council. The authors thank A. M. McClung at USDA-ARS, Rice Research, U.S., Liqing Wei at Chinese Academy of Agricultural Science, Jiankun Xie at Jiangxi Academy of Agricultural Sciences, Guoqing Tu and Jieyun Zhuang at China Rice Research Institute, Qinglong Liu and Junmin Wang at Zhejiang Academy of Agricultural Sciences, Wenyue Chen at Hangzhou Institute of Agricultural Sciences, Xiuru Xu and Honghua Chen at Wenzhou Institute of Agricultural Sciences, and Shengquan Shen at Zhejiang University, for kindly supplying the rice materials.

References

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trend Plant Sci 8:554–560

Andersen JR, Schrag T, Melchinger AE, Zein I, Lubberstedt T (2005) Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays L.*). Theor Appl Genet 111:206–217

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. Theor Appl Genet 94:773–781

Bao JS, Wu YR, Hu B, Wu P, Cui HR, Shu QY (2002) QTL for rice grain quality based on a DH population derived from parents with similar apparent amylose content. Euphytica 128:317–324

Bao JS, Corke H, He P, Zhu LH (2003) Analysis of quantitative trait loci for starch properties of rice based on an RIL population. Act Bot Sin 45:986–994

Bao JS, Sun M, Zhu LH, Corke H (2004) Analysis of quantitative trait locus for some starch properties in rice (*Oryza sativa* L.), thermal properties, gel texture, swelling volume. J Cereal Sci 39:379–385

Bao JS, Corke H, Sun M (2006a) Microsatellites, single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.). Theor Appl Genet (in press)

Bao JS, Shen SQ, Sun M, Corke H (2006b) Analysis of genotypic diversity in the starch physicochemical properties of nonwaxy rice: apparent amylose content, pasting viscosity and gel texture. Starch/Starke 58:259–267

Bao JS, Sun M, Corke H (2006c) Analysis of genotypic diversity in starch thermal and retrogradation properties in nonwaxy rice. Carbohydr Polym 63:online, 10.1016/j.carbpol.2006.05.011

Bergman CJ, Bhattacharya KR, Ohtsubo K (2004) Rice end-use quality analysis. In: Champagne ET (ed) Rice chemistry and technology. AACC, St. Paul, MN, pp 415–472

Buckler IV ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the *lsa* gene of barley. Theor Appl Genet 109:543–551

Caicedo A, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. Proc Natl Acad Sci USA 101:15670–15675

Chang TT, Li CC (1991) Genetics and breeding. In: Luh BS (ed) Rice Production, 2nd edn. Van Nostrand Reinhold, New York, pp 32–101

Chen MH, Bergman CJ, Fjellstrom RG (2003) SSSIIa locus genetic variation associated with alkali spreading value in international rice germplasm. In: Proceedings of the plant and animal genomes conference, vol XI. p 314

Doyle JJ (1991) DNA protocols for plants-CTAB total DNA isolation. In: Hewitt GM (ed) Molecular techniques in taxonomy. Springer, Berlin Heidelberg New York, pp 283–293

Domon E, Yanagisawa T, Saito A, Takeda K (2004) Single nucleotide polymorphism genotyping of the barley waxy gene by polymerase chain reaction with confronting two-pair primers. Plant Breed 123:225

Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang QF (2005) The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. Theor Appl Genet 110:1445–1452

Flint-Garcia SA, Thornsberry JM, Buckler IV ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. Nat Genet 36:388–393

Gao ZY, Zeng DL, Cui X, Zhou YH, Yan M, Huang D, Li JY, Qian Q (2003) Map-based cloning of the ALK gene, which controls the GT of rice. Sci China C-Life Sci 46:661–668

Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics 165:759–769

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa L.* Genetics 169:1631–1638

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future propects. Plant Mol Bio 57:461–485

Hagenblad J, Tang C, Molitor J, Werner J, Zhao K, Zheng H, Marjoram P, Weigel D, Nordborg M (2004) Haplotype structure and phenotypic associations in the Chromosomal regions surrounding two Arabidopsis thaliana flowering time loci. Genetics 168:1627–1638

He P, Li SG, Qian Q, Ma YQ, Li JZ, Wang WM, Chen Y, Zhu LH (1999) Genetic analysis of rice grain quality. Theor Appl Genet 98:502–508

Jiang HW, Dian WM, Liu FY, Wu P (2004) Molecular cloning and expression analysis of three genes encoding starch synthase II in rice. Planta 218:1062–1070

Juliano BO (1998) Varietal impact on rice quality. Cereal Food World 43:207–211, 214–216, 218–222

Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. Genetics 167:1513–1524

Larkin PD, Park WD (2003) Association of *waxy* gene single nucleotide polymorphisms with starch characteristics in rice (*Oryza sativa* L.). Mol Breed 12:335–339

McKenzie KS, Rutger JN (1983) Genetic analysis of amylose content, alkali spreading score, and grain dimensions in rice. Crop Sci 23:306–319

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare ssp. spontaneum*) despite high rate of self-fertilization. Proc Natl Acad Sci USA 102:2442–2447

Nakamura Y (2002) Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. Plant Cell Physiol 43:718–725

Nakamura Y, Francisco Jr PB, Hosaka Y, Sato A, Sawada T, Kubo A, Fujita N (2005) Essential amino acid of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties. Plant Mol Biol 58:213–227

Olsen KM, Purugganan MD (2002) Molecular evidence on the origin and evolution of glutinous rice. Genetics 162:941–950

Perez CM, Villareal CP, Juliano BO, Biliaderis CG. (1993) Amylopectin-staling of cooked nonwaxy milled rices and starch gels. Cereal Chem 70:567–571

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic association in the maize genome. Proc Natl Sci USA 98:11479–11484

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Simko L, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004) Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor Appl Genet 208:217–224

Swofford D (2003) Phylogenetic analysis using parsimony, Beta Version 4.0b7. Sinauer Associates, Inc

Szalma SJ, Buckler IV ES, Snook ME, McMullen MD (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor Appl Genet 110:1324–1333

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tan YF, Li JX, Yu SB, Xing YZ, Xu CG, Zhang Q (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. Theor Appl Genet 99:642–648

Tanaka KI, Ohnishi S, Kishimoto N, Kawasaki T, Baba T (1995) Structure, organization, and chromosomal location of the gene encoding a from of rice soluble starch synthase. Plant Physiol 108:677–683

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler IV ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Umemoto T, Aoki N (2005) Single-nucleotide polymorphisms in rice *starch synthase IIa* that alter starch gelatinization and starch association of the enzyme. Funct Plant Biol 32:763–768

Umemoto T, Yano M, Satoh H, Shomura A, Nakamura Y (2002) Mapping of a gene responsible for the difference in amylopectin structure between japonica-type and indica-type rice varieties. Theor Appl Genet 104:1–8

Umemoto T, Aoki N, Lin HX, Nakamura Y, Inouchi N, Sato Y, Yano M, Hirabayashi H, Maruyama S (2004) Natural variation in rice *starch synthase IIa* affects enzyme and starch properties. Funct Plant Biol 31:671–684

Waters DE, Henry RJ, Reinke RF, Fitzgerald MA (2006) Gelatinization temperature of rice explained by polymorphisms in starch synthase. Plant Biotechnol J 4:115–122

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Whitt SR, Wilson LM, Tenaillon MI, Gaut B, Buckler IV ES (2002) Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci USA 99:12959–12962