Theor Appl Genet (2009) 118:465–472
DOI 10.1007/s00122-008-0913-1

ORIGINAL PAPER

A causal C–A mutation in the second exon of *GS3* highly associated with rice grain length and validated as a functional marker

Chuchuan Fan · Sibin Yu · Chongrong Wang · Yongzhong Xing

Received: 30 June 2008 / Accepted: 9 October 2008 / Published online: 20 November 2008

Abstract Comparative sequencing of *GS3*, the most important grain length (GL) QTL, has shown that differentiation of rice GL might be principally due to a single nucleotide polymorphism (SNP) between C and A in the second exon. A total of 180 varieties representing a wide range of rice germplasm were used for association analysis between C–A mutation and GL in order to confirm the potential causal mutation. A cleaved amplified polymorphic sequence (CAPS) marker, SF28, was developed based on the C–A polymorphism in the *GS3* gene. A total of 142 varieties carried allele C with GL from 6.4 to 8.8 mm, while the remaining 38 varieties carried allele A with GL from 8.8 to 10.7 mm. Twenty-four unlinked SSR markers were selected to genotype 180 varieties for population structure analysis. Population structure was observed when the population was classified to three subpopulations. Average GL of either genotype A or genotype C within *japonica* among the three subpopulations had no significant difference from that in *indica*, respectively, although *indica* rice had longer grains on average than *japonica* in the 180 varieties. However, genotype C always had longer grain length on average than genotype A among three subpopulations. The mutation could explain 79.1, 66.4 and 34.7% of GL variation in the three subpopulations, respectively. These results clearly confirmed the mutation between C and A was highly associated with GL. The SF28 could be a functional marker for improvement of rice grain length.

Communicated by Y. Xu.

C. Fan · S. Yu · C. Wang · Y. Xing (✉)
National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research,
Huazhong Agricultural University, 430070 Wuhan, China
e-mail: yzxing@mail.hzau.edu.cn; yzhxing@hotmail.com

Introduction

Grain length plays an important role in determining the appearance and milling quality of rice, and affects grain weight which is one of the three components (number of panicles per plant, number of grains per panicle and grain weight) of grain yield. It is therefore an important agronomic trait in rice breeding (Luo et al. 2004). Although the preference for rice grain characteristics varies across different consumers, long and slender grain is generally preferred for *indica* rice by the majority of consumers in China, USA and most Asian countries (Unnevehr et al. 1992; Juliano and Villareal 1993).

Since rice grain length is quantitatively inherited (McKenzie and Rutger 1983), it is difficult for breeders to efficiently improve grain length using conventional selection methods. Thus, it should be particularly helpful for enhancing breeding efficiency to use markers closely linked to genes or major quantitative traits loci (QTL) for grain length in order to screen target genotypes directly in early generations. Independent studies across different genetic backgrounds and environments have identified a QTL with major effect on grain length around the centromeric region of chromosome 3 (Huang et al. 1997; Redoña and Mackill 1998; Tan et al. 2000; Xing et al. 2001; Aluko et al. 2004; Li et al. 2004; Wan et al. 2005), which provides a desirable target gene/QTL for improving rice grain length using a marker-assisted selection (MAS) strategy. However, genetic linkage between these molecular markers and the grain length gene, established by QTL studies, can be broken by genetic recombination; this limits the use of random DNA marker as a diagnostic tool (Rafalski and Tingey 1993). Functional markers (Andersen and Lübberstedt 2003), which are derived from polymorphic sites within genes affecting phenotypic trait variation, are highly predictive of phenotype

and overcome the problem of recombination/linkage. Recent knowledge generated in several agronomically important genes' isolation provides information on how to systematically develop functional markers that can be directly used with great reliability and efficiency to identify favorable alleles in a breeding program (Bormans et al. 2002; Ellis et al. 2002; Andersen and Lübberstedt 2003; Peter et al. 2004; Louis et al. 2005; Iyer-Pascuzzi and McCouch 2007).

Recently, a putative *GS3* gene controlling grain length on chromosome 3 has been cloned (Fan et al. 2006). *GS3* encodes 232 amino acids with a putative PEBP-like domain, a transmembrane region, a putative TNFR/NGFR family cysteine-rich domain and a VWFC module. Comparative sequencing showed that one common mutation in the second exon of *GS3* was occurred between short and long grain cultivars, which changed a cysteine codon (TGC) in the short-grain group to a termination codon (TGA) in the long-grain group. The premature termination resulted a 178-aa truncation in the C-terminus of the predicted protein. In theory, this polymorphic allele sequence provides us with an opportunity for the construction of a functional marker for grain length in rice. However, the common SNP was observed on the basis of the comparative sequencing of six genotypes, the relationship between C–A mutation and GL in rice germplasm is still not clear. In this study, a natural population consisted of 180 modern cultivars and landraces was collected for the C–A single nucleotide polymorphism genotyping and GL investigation. The goals of this study are to (1) design a cleaved amplified polymorphic sequence (CAPS) marker for genotyping of C–A mutation and (2) to study the association between GL and C–A mutation.

Materials and methods

Plant materials

A total of 180 rice varieties including 170 from core collection of Chinese rice germplasm, which included 203 accessions and represented about 70% of Chinese rice germplasm (Li et al. 2003), and ten varieties collected from other countries were used for marker validation (Table 1). Ten plants per cultivar were transplanted in a single row with 16.5 cm between plants and 26.4 cm between rows in the rice-growing season of 2005 on the experimental farm of Huazhong Agricultural University, Wuhan, China. Field management followed essentially the normal agricultural practice as described by Fan et al. (2006).

Measurements of grain length

Grain length was evaluated according to Fan et al. (2006). Ten randomly chosen grains of fully filled paddy rice from each plant were lined up length-wise along a vernier caliper to measure average grain length of each plant.

Primer design

Primer was designed using Primer Premier Version 5.0 (Premier Biosoft International, Palo Alto, CA). The genomic sequences of the gene *GS3* were obtained from the GenBank (accession number DQ355996).

DNA extraction, PCR and genotyping

Fresh leaves were harvested from field-grown plants and genomic DNA was extracted using the CTAB method (Murray and Thompson 1980). PCR was performed using 20–50 ng DNA, 10 mM Tris–HCl, 50 mM KCl, 0.1% Triton X-100, 1.8 mM $MgCl_2$, 0.1 mM dNTP, 0.2 μM SF28-F (5′-TGCCCATCTCCCTCGTTTAC-3′), 0.2 μM SF28-R (5′-GAAACAGCAGGCTGGCTTAC-3′) and 1 U *Taq* DNA polymerase, in a total volume of 20 μl. Using a Gene Amp PCR system 9700 thermocycler (Perkin Elmer Cetus), the PCR reactions were denatured at 94°C for 4 min, followed by 34 cycles of 94°C for 40 s, 55°C for 40 s and 72°C for 40 s. The final extension was at 72°C for 10 min. After PCR, 8 μl PCR product was digested with 1 U *Pst*I (TaKaRa, Dalian, China) according to the manufacturer's specification. The digested products were separated on 6% polyacrylamide denaturing gels, and DNA fragments were detected by silver staining (Bassam et al. 1991).

Population structure analysis

Twenty-four SSR (simple sequence repeat) markers, one each from the short and long arms of the 12 rice chromosomes were randomly selected for genotyping the 180 rice varieties according to the genetic map developed by Temnykh et al. (2001). The 24 markers were RM529, RM522, RM526, RM211, RM411, RM60, RM518, RM348, RM574, RM274, RM508, RM494, RM427, RM18, RM339, RM408, RM553, RM321, RM484, RM239, RM224, RM479, RM117 and RM463. The PCR was performed as described above and the PCR products were separated on 6% polyacrylamide denaturing gels to determine the alleles of each marker. Program Structure 2.0 (Princhard et al. 2000) was used to infer population structure using a burn-in of 10,000, run length of 100,000, and a model allowing for admixture and correlated allele frequencies. This approach estimated the genetic background matrix (Q); the proportion of an individual's genome that was contributed by each subpopulation. The number of subpopulations K from two to five was tested. Five independent runs for each K yielded consistent results. In addition,

Table 1 Related basic information of 180 tested rice accessions

Order	Variety	Sub	GL	SF28	Subp
1	Zhonglou1	J	8.3	C	2
2	Weiguo	J	6.9	C	3
3	Dandongludao	J	7.6	C	2
4	Xingguo	J	7.6	C	2
5	Laoguangtou83	J	6.7	C	2
6	Baimaodao	J	8.2	C	2
7	Muxiqiu	J	7.2	C	2
8	Laohuzhong	J	7.2	C	2
9	Youmangzaojing	J	7.1	C	2
10	Huangkezaonian	J	7.2	C	2
11	Baigedao	J	6.9	C	2
12	Cunsanli	J	9.6	A	2
13	Tieganwu	J	7.2	C	2
14	Liushizao	I	7.4	C	3
15	Qiuqianbai	I	7.6	C	3
16	Feidongtang	J	7.3	C	2
17	Jingxibai	I	8.0	C	1
18	Fanbopi	I	7.3	C	1
19	Sanbaili	I	8.3	C	3
20	Taishanluo	I	7.7	C	1
21	Aihechi	I	8.3	C	1
22	Aimi	I	8.1	C	3
23	Hongmisandan	J	6.6	C	2
24	Jinbaiyin	J	8.8	C	1
25	Minbeiwanxian	I	8.3	C	1
26	Lucaihao	I	7.0	C	3
27	Wukezhan	I	7.9	C	3
28	Yizhixiang	I	7.4	C	3
29	Yanshuichi	I	7.1	C	3
30	Shuyazhan	I	7.6	C	1
31	Simiao	I	7.3	C	1
32	Esiniu	I	7.9	C	3
33	Qimei	I	7.9	C	3
34	Sanlicun	J	9.4	A	1
35	Hengxianliangchun	I	8.0	C	3
36	Aizizhan	I	7.7	C	1
37	Hongailuo	I	7.9	C	1
38	Dawanluo	I	9.5	A	1
39	Zimi	J	9.4	A	1
40	Xianggu	J	8.1	C	1
41	Xiaohonggu	J	8.0	C	1
42	Qinghe	I	7.8	C	1
43	Wuzidui	J	7.1	C	2
44	Laozaogu	I	7.5	C	1
45	Bobayong1	J	8.0	C	2
46	Gongju73	J	8.1	C	3
47	Lengshuigu2	J	8.2	C	2

Table 1 continued

Order	Variety	Sub	GL	SF28	Subp
48	Nanxiongzaoyou	I	8.0	C	1
49	Baikehualuo	I	8.3	C	3
50	Heidu4	I	7.8	C	3
51	Chikeluo	I	8.3	C	2
52	Lamujia	J	7.8	C	2
53	Banjiemang	I	7.3	C	2
54	Ziluo	I	9.3	A	1
55	Yuyanluo	J	7.7	C	2
56	Huangpiluo	J	7.2	C	2
57	Mowanggunei	J	7.4	C	1
58	Nangaogu	I	8.2	C	1
59	Xinmaxian	I	6.9	C	2
60	Jixieluo	J	8.2	C	2
61	Cunguluo	I	9.4	A	2
62	Youzhan	J	7.5	C	1
63	Yangkeluo	I	7.5	C	2
64	Xiaobaimi	I	8.0	C	3
65	Maguzi	J	8.2	C	2
66	Laohongdao	J	6.9	C	2
67	Taidongludao	J	8.6	C	2
68	Taizhongxianxuan2	I	9.6	A	1
69	Mengjiagao1	J	7.0	C	1
70	Mengjiading2	I	8.9	A	1
71	Baoerfu	I	8.4	C	3
72	Maweizhan	I	8.0	C	1
73	ZhuzhenB	I	8.3	C	3
74	Chaoyang1B	I	7.3	C	3
75	Qiyuexian	I	6.9	C	1
76	Dongtingwanxian	I	8.5	C	3
77	Zegu	I	8.2	C	1
78	Xuanenchangtan	I	8.0	C	3
79	Bawangbian1	J	7.5	C	3
80	Baikezaohe	I	8.3	C	3
81	Xuguluo	J	7.5	C	3
82	Mugualuo	J	7.3	C	2
83	Hongqi5	J	7.8	C	2
84	Wanlixian	I	8.2	C	3
85	L301B	I	8.8	A	3
86	AnnongwanjingB	J	6.4	C	2
87	Jinnante43B	I	7.9	C	3
88	Zaoshunonghu6	J	7.1	C	2
89	Qingsiai16B	I	7.6	C	1
90	Zhenshan97B	I	7.8	C	3
91	Jiangnongzao1	I	10.0	A	1
92	JinhuB	J	6.8	C	2
93	LimingB	J	6.9	C	2
94	Funingzipi	J	7.7	C	2

Table 1 continued

Order	Variety	Sub	GL	SF28	Subp
95	Longhuamaohu	J	8.6	C	2
96	Gaoyangdian	J	8.3	C	2
97	Shuiyuan300	J	6.9	C	2
98	Dianyui409B	I	9.6	A	1
99	Baoxie123B	I	10.0	A	1
100	80B	I	10.0	A	1
101	Baoxie7B	I	9.2	A	1
102	Gzhenshan97B	J	8.2	C	3
103	Yenicanghua	J	8.2	C	2
104	88B	I	9.7	A	1
105	Jiabala	J	7.3	C	3
106	Gu154	I	9.3	A	1
107	IR661-1	I	8.0	C	3
108	PeiC122	J	6.7	C	2
109	Jing7623	J	7.5	C	2
110	Ninghui21	J	7.5	C	2
111	76–1	J	8.2	C	2
112	Huhui628	J	7.0	C	2
113	Xiangdao	I	8.0	C	3
114	TeqingXuanhui	I	9.3	A	1
115	Xianghui91269	I	8.8	A	1
116	Hanma	I	8.1	C	3
117	Xibaizhan	J	7.9	C	3
118	Mamagu	I	8.0	C	3
119	Nantiangangjiu	J	7.0	C	2
120	Meihualuo	I	8.5	C	3
121	Zhongnong4	I	7.8	C	1
122	Honggu	I	8.0	C	2
123	Wenxiangluo	I	7.9	C	1
124	Guangluai4	I	7.0	C	3
125	Aijiaonante	I	7.6	C	3
126	Liusha1	I	7.3	C	3
127	Chenwan3	J	7.0	C	3
128	Erjiunan1	I	7.7	C	3
129	Nanjing11	I	7.8	C	3
130	Guihuahuang	J	7.3	C	2
131	Sujing2	J	7.0	C	2
132	Chenduai3	I	7.4	C	3
133	Aimakang	I	8.0	C	3
134	Shufeng101	I	9.9	A	1
135	Taizhong65	J	7.0	C	2
136	Taizhongzailai1	I	7.4	C	3
137	Lixingjing	J	7.9	C	2
138	Guichao2	I	7.2	C	3
139	Ergangai	I	7.3	C	3
140	Guanglai15	I	7.9	C	3
141	Hongwan1	I	7.4	C	1
142	Xiangaizao10	I	7.9	C	3
143	Xiangwanxian1	I	8.1	C	1
144	Nante	I	8.0	C	3
145	Xiushui115	J	7.1	C	2
146	Yangdao2	I	8.8	A	1
147	Huke3	I	7.9	C	3
148	Aituogu151	I	7.0	C	3
149	Zhonghua8	J	7.3	C	2
150	Suidao1	J	7.8	C	2
151	Liaojing287	J	7.1	C	2
152	Huangsiguizhan	I	7.7	C	3
153	Zaoshuxianghei	I	9.8	A	1
154	Momi	I	10.2	A	3
155	Jinyou1	I	8.4	C	1
156	Jing87-304	J	7.1	C	2
157	Xiangwanxian3	I	10.0	A	1
158	Xiangzaoxian7	I	7.8	C	1
159	Zhenxian232	I	9.0	A	1
160	Zaoxian240	I	7.3	C	3
161	Dangyu5	J	9.1	A	1
162	Chenlongshuijin	I	10.7	A	1
163	9311	I	9.5	A	1
164	ZS97	I	7.8	C	1
165	MH63	I	10.1	A	1
166	ZH11	J	7.5	C	2
167	Sanqishi	J	9.2	A	2
168	Qitoubaigu	J	8.3	C	3
169	Benbanggu	J	8.3	C	2
170	JWR221	J	9.6	A	1
171	Ballina	J	7.2	C	2
172	Nipponbare	J	7.3	C	2
173	BASMATI	I	10.0	A	2
174	IR24	I	9.7	A	1
175	Lemont	J	9.9	A	2
176	KDL105	I	10.1	A	1
177	SLG	I	10.2	A	2
178	Largue	I	9.3	A	2
179	H94	I	10.1	A	1
180	NYZ	I	10.3	A	2

SF28 genotypes detected by marker SF28

Sub subspecies, *I indica*, *J japonica, GL* grain length (mm), *A* genotype of Minghui 63 (long grain allele), *C* genotype of Chuan 7 (short grain allele), *Subp* subpopulation classified by population structure analysis

F test was carried out for each of the subpopulations and the overall population.

Results

Development of an allelic specific marker for grain length assay

Our previous study reported there was a single nucleotide mutation detected in the second exon of the *GS3* gene between two different grain-length groups, which changed the sequence CTGCAG in the short-grain group to CTGAAG in the long-grain group and accounted for the grain length transformation (Fan et al. 2006). A primer pair, SF28, was designed to amplify a 136 bp segment of the second exon of the *GS3* gene (Fig. 1). The CTGCAG/CTGAAG polymorphism resulted in a restriction site change of *Pst*I. PCR products could be digested with *Pst*I and produced 110 bp and 26 bp sized bands for the genotypes of CTGCAG. Otherwise only a 136 bp segment was left due to the lack of *Pst*I site in genotypes of CTGAAG.

Genotyping varieties with the CAPS marker

The 180 varieties, showed a wide variation in grain length that ranged from 6.4 to 10.7 mm (Table 1), were genotyped with the CAPS marker SF28. As expected, all varieties generated fragments approximately 136 bp in size. After digestion by *Pst*I, the PCR products of 142 varieties including 73 *indica* varieties and 69 *japonica* varieties could produce a 110 bp fragment and a 26 bp fragment, while the PCR products of the remaining 38 varieties including 31 *indica* varieties and seven *japonica* varieties could not be digested due to lack of the *Pst*I site (Table 2; Fig. 2). It could be concluded that all varieties are homozygous at the loci of marker SF28, and the 142 varieties did correspond to the homozygous genotype of CTGCAG (homozygous 'C' in brief) while the 38 varieties represented the homozygous genotype of CTGAAG (homozygous 'A' in brief). The A-allele was observed in 36% of *indica* and 9% of *japonica*.

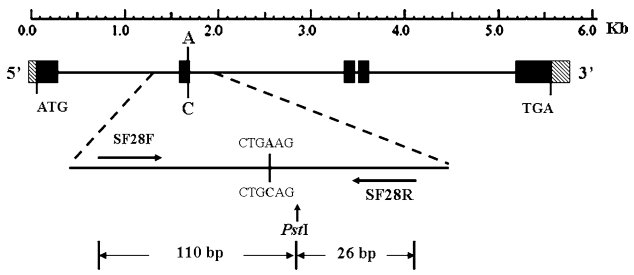

Fig. 1 Organization of the *GS3* gene and relative positions of PCR primers for amplifying a segment of 136 bp and *Pst*I restriction site in PCR product. The positions of coding regions (*black boxes*), 5′ and 3′ UTR (*hatched boxes*), translation start codon (*ATG*), translation stop codon (*TGA*), the SNP (*C/A*) in the second exon between two grain-length groups are indicated

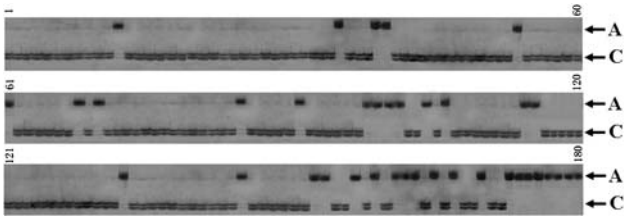

Fig. 2 Marker assay of the 180 rice varieties by the CAPS marker SF28. The two alleles of *A* and *C* are as indicated. The order of samples is the same as Table 1

Grain length variation between C–Allele and A-allele

The 180 tested varieties were assigned to two subspecies: *indica* rice and *japonica* rice. Overall, the average GL of *indica* rice (8.3 ± 0.9 mm) was longer than *japonica* (7.7 ± 0.8 mm) at the level of $P = 0.05$. For the 142 varieties with homozygous 'C' genotype, the GL of 73 *indica* rice varieties ranged from 6.9 to 8.5 mm with an average of 7.8 ± 0.4 mm while the GL of 69 *japonica* rice varieties ranged from 6.4 to 8.8 mm with an average of 7.5 ± 0.6 mm (Table 2). The difference in GL between subspecies within genotype C was not significant. Out of the 38 cultivars with homozygous 'A' genotype, 31 *indica* rice varieties have GL ranging from 8.8 to 10.7 mm with an average of 9.6 ± 0.5 mm. Seven *japonica* rice varieties have GL ranging from 9.1 to 9.9 mm with an average of 9.6 ± 0.3 mm. The difference in GL between *indica* and *japonica* within genotype A was also not significant (Table 2).

However, a significant difference in GL between two genotypes C and A was observed (Table 2; Fig. 3). The GL of the 142 cultivars with homozygous 'C' genotype ranged from 6.4 to 8.8 mm with an average of 7.7 ± 0.5 mm; while the GL of the 38 cultivars with homozygous 'A' genotype varied from 8.8 to 10.7 mm with an average of 9.6 ± 0.5 mm.

Table 2 Descriptive statistics of grain length (GL, mm) for 180 rice varieties in the study

Genotype	Subspecies	No. of varieties	Grain length (mm)	
			Range	Mean ± SD
A	*Indica*	31	8.8–10.7	9.6 ± 0.5
	Japonica	7	9.1–9.9	9.6 ± 0.3
C	*Indica*	73	6.9–8.5	7.8 ± 0.4
	Japonica	69	6.4–8.8	7.5 ± 0.6
Whole	*Indica*	104	6.9–10.7	8.3 ± 0.9
	Japonica	76	6.4–9.9	7.7 ± 0.8

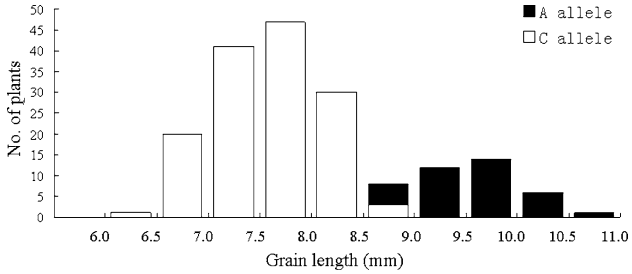

Fig. 3 Frequency distribution of grain length for the 180 rice varieties. Two genotype classes assessed by the marker SF28 are as indicated

Genetic structure of 180 rice varieties

All SSR markers were polymorphic in the 180 rice cultivars, generating 2–11 alleles with 4.5 alleles on average, which indicated a wide range of genetic diversity in the germplasm. The 180 varieties represented a broad range of the cultivated rice germplasm including landraces, primitive cultivars, modern elite cultivars and parents of hybrid rice. High likelihoods of the population structure were observed when the number of populations was set to three when running the program Structure 2.0 (Table 1). Subpopulation 1 contained 58 varieties including 48 *indica* and ten *japonica* varieties; Subpopulation 2 contained 66 varieties including 56 *japonica* and ten *indica* varieties; Subpopulation 3 contained 56 varieties including 47 *indica* and nine *japonica* varieties.

Exactly a half of 48 *indica* varieties in Subpopulation 1 carried alleles C and A, respectively. The average GL is significant different between genotypes A and C. Six and four *japonica* varieties carried alleles C and A, respectively. Accordingly, genotypes A and C had different GL (Table 3). There was no significant different between the average GL of *indica* and *japonica* of genotype C and the same case held true for genotype A. Overall, average GL of 7.8 mm in the 30 C genotypes was much shorter than that of 9.6 mm in the 28 A genotypes. In Subpopulation 2 and 3, the same situation was observed. Additionally, no difference in average GL was detected within genotypes A or C across the three Subpopulations and between *japonica* and *indica*.

Association between GL and SF28

The C–A mutation was highly associated with GL among all the 180 varieties. It can explain 72.0% of GL variation (Table 4). In Subpopulations 1, 2 and 3, high association was also observed between the C–A mutation and GL. It can explain 79.1, 66.4 and 34.7% of GL variation, respectively.

Discussion

A causal mutation for grain length

It was reported that grain size had played important roles in evolution of cereal crops (Paterson et al. 1995; Li et al. 2004). The grains of wild relatives are usually small and round in shape, which is often favored under natural selection because such a phenotype is ideal for high fecundity and advantageous for dispersal by natural vectors. Domestication has greatly increased the diversity of grain shape and size together with many other changes, as consequences of physiologic response and adaptation to diverse natural environments and human needs. Thus, long grain is the product of artificial selection. In the present study, 142 of the 180 varieties had the 'C' genotype of the *GS3* C–A mutation which showed short grain, and only 38 cultivars with longer grains had the 'A' genotype. The lower frequency (21%) of the 'A' genotype might indicate that it was a mutant allele consistent with the long grain phenotype which occurred during domestication. Rice consists

Table 3 Grain length (GL, mm) of C/A allele in different clusters

Subspecies	Genotypes	Items	Sub 1	Sub 2	Sub 3	Whole
Indica	C	Accessions	24	5	45	74
		Mean ± SD	7.8 ± 0.4a[a]	7.6 ± 0.6a	7.8 ± 0.4a	7.8 ± 0.4a
	A	Accessions	24	5	2	31
		Mean ± SD	9.6 ± 0.8b	9.8 ± 0.6b	9.5 ± 1.0b	9.7 ± 0.5b
Japonica	C	Accessions	6	53	9	68
		Mean ± SD	7.8 ± 0.6a	7.5 ± 0.6a	7.6 ± 0.5a	7.5 ± 0.6a
	A	Accessions	4	3	0	7
		Mean ± SD	9.4 ± 0.2b	9.6 ± 0.4b	–	9.5 ± 0.3b
Whole	C	Accessions	30	58	54	142
		Mean ± SD	7.8 ± 0.4a	7.5 ± 0.6a	7.8 ± 0.4a	7.7 ± 0.5a
	A	Accessions	28	8	2	38
		Mean ± SD	9.6 ± 0.5b	9.7 ± 0.4b	9.5 ± 1.0b	9.6 ± 0.5b

Sub 1, 2, and 3 subpopulations 1, 2 and 3; *Whole* all the population of 180 varieties

[a] Within a column and within a row, means followed by different letters are significantly different at $P = 0.01$ according to Duncan test

Table 4 Associate analysis of grain length and C/A alleles

	F	*P*	R^2 (%)
Whole	459.8	0.00000	72.0
Sub 1	215.6	0.00000	79.1
Sub 2	126.4	0.00000	66.4
Sub 3	29.0	0.000002	34.7

Sub 1, 2, and 3 subpopulations 1, 2 and 3; *Whole* all the population of 180 varieties, R^2 coefficient of determination is calculated in proportion of genotype sum of squares in total sum of squares

mainly of two subspecies: *japonica* and *indica*. *Indica* is known for its large variation of grain shape, frequently long and slender, whereas *japonica* is often recognized for its short and round grains (Luo et al. 2004). In this study, only seven *japonica* cultivars carried A-allele. However, all *japonica* cultivars with long GL always carried A-allele without an exceptional. This result is in agreement with the fact that *japonica* rice with long GL is hard to find in nature. Among the 180 accessions, *indica* has a longer GL than *japonica* on average. However, the C and A alleles exist in both *indica* and *japonica* indicating the origin of this mutation of prior to the *indica–japonica* differentiation. Hence, grain length differentiation is not the result of *indica* and *japonica* differentiation.

In self-pollinated species, both breeding system and domestication history had large effects on rice population structure. Although inter-subspecies hybrid showed partial sterility, the exchange of genetic makeup between *indica* and *japonica* could be accomplished by consecutive backcross in some extent. Thus, there were probably some intermediate rice between *japonica* and *indica* (Sano and Morishima 1992), and such kind of intermediate is expected in nature. In ecogeographical terms, *indica* is primarily known as lowland rices in tropical and subtropical Asia, while *japonica* is typically grown in temperate East Asia. However, *indica* and *japonica* can be grown in subtropical region, which provides more chance to communicate between *indica* and *japonica*. In the study, all the investigated varieties were classified to three subpopulations, each consisted of *japonica* and *indica*. Subpopulations 1 and 3 mainly consisted of *indica* (83 and 84%), which were mainly and originally grown in tropical and subtropical region, respectively. However, subpopulation 2 mainly consisted of *japonica* (85%). In the same subpopulation, GL in either genotype C or A had no difference between *japonica* and *indica*, respectively. GL in genotypes C or A had no difference within *indica*, *japonica* or between *japonica* and *indica* among the three subpopulations. However, GL was significantly different between genotypes C and A within each subpopulations and in the overall population. C–A mutation was highly associated with GL under consideration of population structure. The mutation explained most of GL variation in all the three subpopuations; hence the C–A mutation would be the causal mutation of grain length.

A functional marker for grain length improvement

Studies on several cloned genes indicated functional differences in trait performance mainly caused by SNP (Peng et al. 1999; Takahashi et al. 2001; Liu et al. 2002; Toshiyuki et al. 2003). Rice GL was greatly controlled by the major effect QTL (*GS3*) located in the pericentromeric region of rice chromosome 3. A single mutation, leading to a premature stop codon in the second exon of the putative *GS3* gene, is the most likely cause of grain elongation (Fan et al. 2006). In the study, significant association between GL and C–A mutation further confirmed the substantial contribution of C–A mutation to GL. Hence, development of markers based on the causal SNPs can directly differentiate alleles conferring long and short grains. However, current SNP maker validation, which is mainly detected by gene chip, fluorescence polarization and mass spectrometry technology (Gut 2001; Kim et al. 2003), is very expensive and traditional breeders are unable to utilize the technology. CAPS markers can be easily genotyped with very ordinary methods such as running agarose gel or polyacrylamide denaturing gels, which is a routine technique in most laboratories. Breeders only provide plant tissues to the lab for *GS3* allele identification and then get desired plants within 3 days. The CAPS marker SF28 assay exhibits a perfect consistency between grain length variation and genotypes. In the overall population, genotype A has long GL ranging from 8.8 to 10.7 mm, and C from 6.3 to 8.8 mm. Several varieties, whose grain lengths are around 8.8 mm, shared different genotypes A or C. This can be explained by other genomic regions that may also underlie grain length with minor effects (Tan et al. 2000; Xing et al. 2001; Aluko et al. 2004; Wan et al. 2005; Amarawathi et al. 2008). However, this CAPS marker provides the prediction with 100% accuracy to the long (>9.0 mm) and short grain (<8.5 mm) genotypes (Fig. 3). Thus the CAPS marker has practical utility in variety testing based on the presence or absence of specific alleles for rice breeders. This simple robust and economical marker genotyping method combinded with rapid DNA extraction protocol will simplify and streamline MAS for grain length in rice breeding.

Moreover, the consumer-preferred long grain is controlled by the recessive alleles of *GS3*. Its heterozygote and dominant homozyogte always expresses the same short grain phenotype which is unexpected in the process of grain improvement. Therefore it is impossible to select the genotype with long grains in the generation of hybrid based on phenotypic selection. By contrast, SF28, derived from functional motifs within *GS3* gene causally involved in GL

variation, can reliably target the heterozygous genotype which could generate recessive homozygous plant by selfing. With this marker, we only need to focus on a few heterozygote plants in each generation in the backcross breeding process of direct improvement of grain length. Several breeders have utilized the marker in China and highly recommend it (private communication).

In summary, this study further confirmed the common C–A mutation of *GS3* between short and long grain rice in a large number of rice accessions, which was firstly found among six cultivars by Fan et al. (2006). Meanwhile, high association was found between the C–A mutation and GL in both the whole population and three subpopulations. Moreover, the C–A mutation was validated as a functional marker, SF28, which could be directly used to assist selection for GL improvement without the risk of recombinant.

Acknowledgments This work was supported in part by grants from the National Program on the Development of Basic Research, National Program on the Development of Hi-tech Research (2007AA10Z103) and the National Natural Science Foundation of China.

References

Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard HJ (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O.* glaberrima. Theor Appl Genet 109:630–639

Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR, Singh NK (2008) Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). Mol Breeding 21:49–65

Bassam BJ, Anolles GC, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83

Bormans CA, Richard RB, Kephart DD, McClung AM, Park WD (2002) Analysis of a single nucleotide polymorphism that controls the cooking quality of rice using a non-gel based assay. Euphytica 128:261–267

Ellis MH, Spielmeyer W, Gale KR, Rebetzke GJ, Richards RA (2002) "Perfect" markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theor Appl Gene 105:1038–1042

Fan CC, Xing YZ, Mao HL, Lu TT, Han B, Xu CG, Li XH, Zhang QF (2006) *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

Gut IG (2001) Automation in genotyping single nucleotide polymorphisms. Hum Mutat 17:475–492

Huang N, Parco A, Mew T, Magpantay G, McCouch S, Guiderdoni E, Xu JC, Subudhi P, Angeles ER, Khush GS (1997) RFLP mapping of isozymes, RAPD, and QTLs for grain shape, brown planthopper resistance in a doubled-haploid rice population. Mol Breed 3:105–113

Iyer-Pascuzzi AS, McCouch SR (2007) Functional markers for *xa5*-mediated resistance in rice (*Oryza sativa*, L.). Mol Breed 19:291–296

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends plant sci 8:554–560

Juliano BO, Villareal CP (1993) Grain quality evaluation of world rice. International Rice Research Institute, Manila

Kim S, Ruparel HD, Gilliam TC, Ju JY (2003) Digital genotyping using molecular affinity and mass spectrometry. Nat Rev Genet 4:1001–1008

Li J, Thomson M, McCouch SR (2004) Fine mapping of a grain-weight quantitative trait locus in the pericentrometric region of rice chromosome 3. Genetics 168:2187–2195

Li ZC, Zhang HL, Cao YS, Qiu ZE, Wei XH, Tang SX, Yu P, Wang XK (2003) Studies on the sampling strategy for primary core collection of Chinese ingenious rice. Acta Agron Sin 29:20–24

Liu J, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Natl Acad Sci USA 99:13302–13306

Luo YK, Zhu ZW, Chen N, Duan BW, Zhang LP (2004) Grain types and related quality characteristics of rice in China. Chinese J Rice Sci 18:135–139

Louis MTB, Robert JH, Jin QS, Russell FR, Daniel LEW (2005) A perfect marker for fragrance genotyping in rice. Mol Breed 16:279–283

McKenzie KS, Rutger JN (1983) Genetic analysis of amylose content, alkali spreading score, and grain dimensions in rice. Crop Sci 23:306–313

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Paterson AH, Lin YR, Li ZK, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256–261

Peter E, Brian R, Graham S (2004) Allele-specific markers within the barley stem rust resistance gene (*Rpg1*). Barley Genet Newsl 33:7

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

Rafalski JA, Tingey SV (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet 9:275–280

Redoña ED, Mackill DJ (1998) Quantitative trait locus analysis for rice panicle and grain characteristics. Theor Appl Genet 96:957–963

Sano R, Morishima H (1992) Indica-Japonica differentiation of rice cultivars viewed from variation in key characters and isozymes with special reference to landraces from the Himalayan hilly areas. Theor Appl Genet 84:266–274

Takahashi Y, Shomura A, Sasaki T, Yano M (2001) *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the subunit of protein kinase CK2. Proc Natl Acad Sci 98:7922–7927

Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang Q (2000) Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. Theor Appl Genet 101:823–829

Temnykh S, DeCklerck G, Lukashova A, Lipovich L, Cartinhiur S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Toshiyuki K, Shizo O, Nobuhiko M, Yoshimitsu T, Yoshiki K, Shoichi S, Yukoh H, Hidemasa I, Naoto N (2003) Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). Plant J 37:315–325

Unnevehr LJ, Duff B, Juliano BO (1992) Consumer demand for rice grain quality. International Rice Research Institute, Manila, and International Development Research Center, Ottawa

Wan XY, Wan JM, Weng JF, Jiang L, Bi JC, Wang CM, Zhai HQ (2005) Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. Theor Appl Genet 110:1334–1346

Xing YZ, Tan YF, Xu CG, Hua JP, Sun XL (2001) Mapping quantitative trait loci for grain appearance traits of rice using a recombinant inbred line population. Acta Bot Sin 43:840–845