Theor Appl Genet (2010) 120:753–763
DOI 10.1007/s00122-009-1196-x

ORIGINAL PAPER

Cloning and characterization of a putative *GS3* ortholog involved in maize kernel development

Qing Li · Xiaohong Yang · Guanghong Bai · Marilyn L. Warburton · George Mahuku · Michael Gore · Jingrui Dai · Jiansheng Li · Jianbing Yan

Received: 6 June 2009 / Accepted: 17 October 2009 / Published online: 7 November 2009
© Springer-Verlag 2009

Abstract The *GS3* gene was the first identified gene controlling the grain size in rice. It has been proven to be involved in the evolution of grain size during domestication. We isolated the maize ortholog, *ZmGS3* and investigated its role in the evolution of maize grain size. *ZmGS3* has five exons encoding a protein with 198 amino acids, and has domains in common with the rice GS3 protein. Compared with teosinte, maize has reduced nucleotide diversity at *ZmGS3*, and the reduction is comparable to that found in neutrally evolving maize genes. No positive selection was detected along the length of the gene using either the Hudson–Kreitman–Aguadé or Tajima's *D* tests. Phylogenetic analysis reveals a distribution of maize sequences among two different clades, with one clade including related teosinte sequences. The nucleotide polymorphism analysis, selection test and phylogenetic analysis reveal that *ZmGS3* has not been subjected to selection, and appears to be a neutrally evolving gene. In maize, *ZmGS3* is primarily expressed in immature ears and kernels, implying a role in maize kernel development. Association mapping analysis revealed one polymorphism in the fifth exon that is significantly associated with kernel length in two environments. Also one polymorphism in the promoter region was found to affect hundred kernel weight in both environments. Collectively, these results imply that *ZmGS3* is involved in maize kernel development but with different functional polymorphisms and thus, possibly different mechanisms from that of the rice *GS3* gene.

Communicated by J. Snape.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1196-x) contains supplementary material, which is available to authorized users.

Q. Li · X. Yang · G. Bai · J. Dai · J. Li (✉) · J. Yan (✉)
National Maize Improvement Center of China,
Key Laboratory of Crop Genomics and Genetic Improvement (Ministry of Agriculture), China Agricultural University,
Yuanmingyuan West Road, Haidian, 100193 Beijing, China
e-mail: lijiansheng@cau.edu.cn

J. Yan
e-mail: yjianbing@cau.edu.cn; yjianbing@gmail.com

G. Bai
College of Agriculture, Xinjiang Agricultural University,
830052 Urumqi, Xinjiang, China

M. L. Warburton
USDA-ARS, Corn Host Plant Resistance Research Unit,
Box 9555, Mississippi State, MS 39762, USA

G. Mahuku · J. Yan
International Maize and Wheat Improvement Center (CIMMYT),
Apdo. Postal 6-641, 06600 Mexico D.F., Mexico

M. Gore
Department of Plant Breeding and Genetics, Cornell University,
175 Biotechnology Building, Ithaca, NY 14853, USA

Introduction

The domestication of wild plants to modern cultivated crops involved a suite of changes in morphological, physiological and biochemical traits, which are collectively known as the "domestication syndrome" (Doebley et al. 2006). One of the traits in the "domestication syndrome" is grain size. Under natural circumstances, plants tend to produce smaller seeds as opposed to producing larger seeds (Doebley et al. 2006). Grain size is an important component of grain yield and farmers or breeders tend to select plants with larger grain size. Therefore, genes involved in grain

size variation would be a logical target for artificial selection. Rice varieties show large variation in grain size, the genetic basis of which has been extensively investigated using quantitative trait loci (QTL) analysis. Several underlying genes have been cloned, including the first identified grain length gene, *GS3* (Fan et al. 2006, 2009; Li et al. 2004), and the grain width genes, *GW2* (Song et al. 2007), *GW5* (Wan et al. 2008; Weng et al. 2008) and *qSW5* (Shomura et al. 2008). Maize also has larger kernels than its progenitor, teosinte. However, map-based cloning of genes in maize has lagged behind that in rice due to the unavailability of a complete genome sequence before 2008, the large genome size (Arumuganathan and Earle 1991) and the highly repetitive sequence content (SanMiguel and Bennetzen 1998). To date, only one gene encoding glutamine synthetase isoenzyme has been shown to have an impact on maize kernel size by mutant analysis (Martin et al. 2006).

It is estimated that around 1,200 genes were involved in maize domestication and improvement from teosinte (Wright et al. 2005). Finding and validating the function of these genes is the key, not only to understand the mechanisms of maize domestication and improvement, but also to enhance the efficiency of future maize improvement, as many domestication-related traits are also highly important agronomic traits. Comparative genomics based on markers or sequence identities revealed macrosynteny and microsynteny relationships among different but related genomes (Devos 2005; Gale and Devos 1998; Salse et al. 2004). QTL controlling domestication-related or agronomically important traits in maize, rice and sorghum were identified simultaneously in orthologous regions (Paterson et al. 1995; Yan et al. 2004) and some homologous genes controlling traits of agronomic importance were shown to play similar roles in different species (Kojima et al. 2002; Yano et al. 2000). Genes cloned in rice and the conserved synteny relationship between the maize and rice genomes provide an opportunity to clone genes involved in maize domestication and improvement using comparative genomics.

The *GS3* gene on rice chromosome 3 was the first identified gene controlling grain size in rice (Fan et al. 2006; Li et al. 2004). In the small-grain varieties, *GS3* acts as a negative regulator in the size of the rice grain. In the large-grain varieties, a single nucleotide polymorphism (SNP) occurring in the second exon results in a stop codon, thus abolishing the protein's function. Recently, Fan et al. (2009) analyzed a panel of 180 rice varieties and found that 38 varieties contained this stop codon that was associated with larger grain length. The identification of *GS3* in at least ten different intra- and interspecific crosses with stable phenotypic expression suggested that it was an evolutionarily important gene and might have been involved in the domestication of rice by early farmers (Fan et al. 2006; Li et al. 2004). Takano-kai et al. (2009) confirmed this suggestion by nucleotide diversity analysis. They found 97% reduction in nucleotide diversity across the *GS3* gene in rice accessions carrying the mutant allele. Comparative genomics between rice and maize shows that the *GS3* region on rice chromosome 3 is highly similar to the short arm of maize chromosome 1 (Salse et al. 2004; www.tigr.org). Interestingly, a QTL for kernel weight was detected in a cross between maize and its progenitor, *Zea mays* ssp. *parviglumis* in this corresponding region, suggesting that a homologous maize gene to rice *GS3* may be associated with domestication (Doebley et al. 1994; Li et al. 2004; Supplementary Table S1; Supplementary Fig. S1). In addition, a QTL cluster affecting many traits including grain yield was also identified in the short arm of maize chromosome 1 around SSR marker umc1169 (Ma et al. 2007; Yan et al. 2006; Supplementary Table S1; Supplementary Fig. S1). The objectives of this study were to (1) isolate a *GS3*-like gene in maize, (2) investigate if the maize *GS3*-like gene co-localizes with the previously identified QTL for kernel weight and the QTL cluster around SSR marker umc1169, (3) elucidate whether the maize *GS3*-like gene has been a target of selection, and (4) determine if the maize *GS3*-like gene has a similar function as the rice *GS3* gene.

Materials and methods

Genetic mapping

Rice genes in the region surrounding *GS3* on chromosome 3 were retrieved from TIGR (www.tigr.org) based on the fine mapping results reported by Li et al. (2004) and Fan et al. (2006), and used to blast against three databases containing maize sequences (www.plantgdb.org; www.maizegdb.org; www.ncbi.nlm.nih.gov, updated in March 2006). Primers were designed based on the identified maize homologous sequences using Primer 3 (http://frodo.wi.mit.edu/) or Primer Premier 5.00 (PREMIER Biosoft International) and used to screen for polymorphisms between two maize inbred lines, Zong3 and 87-1, which had already been used to develop a recombinant inbred line (RIL) population to map QTL affecting grain yield (Ma et al. 2007). Polymorphic primers were mapped onto the existing RIL genetic map using MAPMAKER/EXP3.0 (Lincoln et al. 1992).

Sequencing procedure

Thermal asymmetric interlaced PCR (TAIL-PCR) was performed to clone the promoter sequence of the maize *GS3*-like gene according to the procedure established by

Liu et al. (1995). For molecular evolution analysis, the *GS3* promoter regions, coding regions and partial intron regions (approximately 2,400 bp in total) were sequenced for a set of 26 lines, including ten elite inbred lines, eight landraces and eight teosinte lines (Table 1), to yield a total of 104 sequences. The ten inbred lines consist of materials from the five main maize heterotic groups in China (Teng et al. 2004). For association mapping analysis, the promoter region, the fifth exon and 3′ UTR were sequenced across a panel of 121 maize inbred lines. All primer pairs used in this study are listed in Supplementary Table S2.

PCR products were first sequenced directly. For ambiguous results, the PCR products were cloned into a pGEM-T easy vector (Promega, Madison, WI, USA) and one randomly selected clone was sequenced. An initial alignment was done with the multiple sequence alignment program MUSCLE (Edgar 2004) to find singletons, which generally result either from PCR/sequencing errors or from sequence variation. Re-amplification was carried out for lines with detected singletons, and the products were either sequenced directly for homozygous lines or cloned into a pGEM-T easy vector (Promega) for heterozygous lines. For confirmation, at least three clones per line per locus were sequenced.

Table 1 Plant materials used in this study

Type	ID	Name	Heterotic group[a]
Inbred line (10)			
Zea mays ssp. *mays*	MAI1	Ye515	Lancaster
	MAI2	Mo17	Lancaster
	MAI3	HuangC	Reid
	MAI4	B73	Reid
	MAI5	Chang7-2	TangSPT
	MAI6	HZS	TangSPT
	MAI7	87-1	Tem-tropic I
	MAI8	P178	Tem-tropic I
	MAI9	Zi330	Zi330
	MAI10	Zong3	Zi330
Landrace (8)			
Zea mays ssp. *mays*	LAN1	04K5686	–
	LAN2	BZN	–
	LAN3	MZ	–
	LAN4	HSBN	–
	LAN5	303WX	–
	LAN6	SW92E114-15-1	–
	LAN7	QTHHSBTS	–
	LAN8	BEM	–
Teosinte (8)			
Zea mays ssp. *parviglumis*	TEO1	PI384064	–
	TEO2	Ames21785	–
	TEO3	PI566688	–
Zea mays ssp. *huehuetenangensis*	TEO4	PI441934	–
Zea mays ssp. *mexicana*	TEO5	PI384060	–
	TEO6	Ames8083	–
	TEO7	PI566673	–
Zea diploperennis[b]	TEO8	Ames21890	–

[a] Information from Teng et al. (2004)

[b] Only used in the HKA test as the outgroup species

Analyses of the sequences

Three parameters of nucleotide diversity (the expected heterozygosity per nucleotide site π, number of segregating sites S and number of haplotypes h) were calculated. Tajima's *D* statistic (Tajima 1989) and the Hudson–Kreitman–Aguadé (HKA) test (Hudson et al. 1987) were used to investigate evidence for past selection. For the HKA test, *Z. diploperennis* was used as the outgroup species and two neutral genes (*adh1* and *glb1*) as the controls (Tenaillon et al. 2004; White and Doebley 1999). All sequence analyses were conducted using DnaSP, version 4.00 (Rozas and Rozas 1999). Alignments of the nucleotide sequences were done using the multiple sequence alignment software MUSCLE (Edgar 2004) and subsequently refined manually. The phylogeny was generated using the neighbor-joining method, Kimura two-parameter distance and pairwise deletion analysis in MEGA, version 3.1 (Kumar et al. 2004). Robustness of the constructed phylogenetic tree was tested with 1,000 bootstrap repetitions.

RNA extraction and real-time quantitative reverse transcription PCR (qRT-PCR)

Expression analysis was conducted with RNA from tissues obtained at 12 different developmental stages of inbred line 87-1, including leaf, seedling shoot, seedling root, tassel from plants with 15 expanded leaves, silk and husk from ears 0 days after pollination (DAP), immature ears from plants with 14 and 18 expanded leaves and kernels harvested at four stages (0 DAP, 10 DAP, 15 DAP, 20 DAP). All plant materials were frozen in liquid nitrogen and stored at −70°C until use. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and digested with RNase-free DNase (Promega). Complementary DNA (cDNA) was synthesized using MMLV retro-transcriptase and an oligo (dT) primer (Promega). qRT-PCR was performed with Ex Taq premix (Takara Shuzo, Kyoto, Japan). Three replicates were performed to calculate the average and standard deviation of expression levels. All experiments were performed according to the manufacturer's instructions. The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001) was employed to calculate *ZmGS3* gene expression

levels with the housekeeping gene ubiquitin as the endogenous control and ears from plants with 18 expanded leaves as the reference tissue.

Field trials, trait evaluation and statistical analysis

A total of 121 diverse maize lines, including 88 elite inbred lines widely used in Chinese maize breeding programs and 33 high-oil lines, were used in the present study. The population structure of these lines had been estimated using 82 SSR markers that are distributed evenly across the maize genome, and the familial relatedness was estimated using 884 SNP markers (Dr. Yang unpublished data). All 121 lines were planted at the Agronomy Farm of China Agricultural University in the spring of 2007 (Beijing, BJ, E 116°46′, N 39°92′) and in the winter of 2007 (Hainan, HN, E 108°56′, N 18°09′). A randomized complete-block design with two replications was employed, in which each genotype was planted in a plot with 11 plants in a 3-m long row with 0.6-m between rows. Normal agronomic practices were used in field management, and the third to the ninth plants in each row were self-pollinated to avoid any xenia effects from foreign pollen. At maturity, the self-pollinated ears were harvested manually and air dried to grain moisture of 13%. Seeds from the middle of at least three ears per plot were shelled and bulked for trait measurement. Twenty randomly selected kernels were used to measure kernel length (KL), kernel width (KW) and kernel thickness (KT). Three samples of 100 kernels each were weighed, and the mean was used to measure hundred kernel weight (HKW). The average value of the two replications for all traits was used for final data analysis. Association analysis was performed with TASSEL 2.0.1 (Bradbury et al. 2007) using the mixed linear model (MLM) controlling for both population structure (Q) and relative kinship (K) (Yu et al. 2006).

Results

Comparative cloning of the *GS3*-like gene in maize

Among the 18 predicted rice genes in the region surrounding *GS3*, nine had homologous sequences in the maize databases ($E < -10$, identity >80%). Five of the nine homologous maize sequences were mapped as a cluster on chromosome 1 between SSR markers umc1169 and umc2112, with a different order compared to the rice loci (Fig. 1a). Interestingly, this region overlapped with the kernel weight QTL mapped by Yan et al. (2006) and Doebley et al. (1994) (Supplementary Table S1; Supplementary Fig. S1). The other four maize sequences homologous to rice genes were mapped to different chromosomes (Fig. 1a), indicating gene level exceptions in the orthologous relationship between maize chromosome 1 and rice chromosome 3.

The protein sequence of the rice *GS3* gene was used to blast against the maize high throughput genomic sequences database (www.ncbi.nlm.nih.gov). Three continuous stretches of nucleotide sequences with two interrupted intervals on BAC clone AC209328 had similarity with the

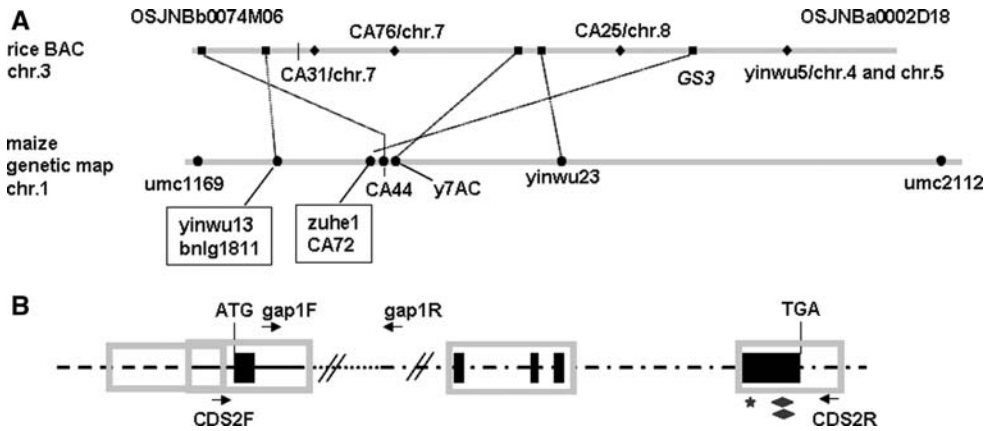

Fig. 1 **a** Collinear relationship between the rice *GS3* region and the corresponding maize region. *Squares* represent rice genes mapped onto maize chromosome 1. *Diamonds* indicate rice genes mapped onto other maize chromosomes, with the primers' name and their locations on the maize genetic map being given above or below the *diamonds*. All rice genes were from two overlapping BAC clones (OSJNBb0074M06 and OSJNBa0002D18). *Dots* represent maize markers used for genetic mapping, and *primers* in the *open boxes* represent co-segregating markers. **b** Structure of the *ZmGS3* gene. *Filled boxes* represent exons. *Black solid line* and *dash-dotted line* represent sequences from GSS clone-CC638768 and BAC clone-AC209328, respectively. *Dotted line* indicates sequences obtained through PCR amplification with the gap1 primer pair. *Dashed line* indicates sequences obtained with TAIL-PCR. CDS2 primer pair was used to amplify the putative cDNA sequence of the *ZmGS3* gene. *Diamonds* indicate two overlapping TNFR/NGFR family cysteine-rich domains. *Star* indicates the transmembrane domain which was predicted using InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/). The four sequenced regions (parts A, B, C and D, see "Results") are depicted as *gray open boxes*

rice GS3 protein over exons 2, 3 and 4 ($E = 0.38$, 0.009 and 0.009, respectively; identity = 57, 73 and 55%, respectively). Although the E values are slightly high, the three interrupted nucleotide sequences reflect the exact gene structure of rice *GS3*. Subsequent analysis revealed a gap in the BAC, which may contain the first exon of the maize homolog. Another blast search of the maize genomic survey sequences (GSS) database identified a clone, GSS-CC638768, with homology to rice exon 1 (Fig. 1b). Two primer sets, CA72 and zuhe1, designed from the homologous sequences of *GS3* in the BAC and the GSS, respectively, co-segregated in the RIL population derived from Zong3 and 87-1 (Fig. 1a). This strongly suggests that the two sequences represent different parts of the maize *GS3*-like gene. The gap between these two sequences was closed by PCR amplification (Fig. 1b; Supplementary Fig. S2a), thus, giving a final assembly comprising the entire maize *GS3* genomic sequence, which was termed *ZmGS3* (GenBank accession no. FJ797615). Because GSS-CC638768 has only 213 bp 5′ to the putative translational start site (TSS), TAIL-PCR was employed to extend the genomic sequence upstream from the TSS, and a band of about 1,200 bp was amplified (Fig. 1b; Supplementary Fig. S2b), which had a 400 bp overlap with GSS-CC638768 and thus yielded a total of 1,000 bp upstream of the TSS. Blast searches with *ZmGS3* did not identify other homologous sequences in the maize genome (updated in December 2008), suggesting that *ZmGS3* may represent a single-copy gene in maize. Based on this suggestion and the collinear relationship found with rice, we concluded that *ZmGS3* is the ortholog of the rice *GS3* gene.

Blast searches with *ZmGS3* genomic sequences failed to detect matching maize ESTs (updated in March 2007), but matching ESTs from other cereals were identified, including two from sorghum, BI098398 and BI098087, which overlapped with each other and can be assembled into a singleton of 1,139 bp with a definite polyA tail. The high sequence similarity between the sorghum singleton and the maize genomic sequences enabled us to identify the putative intron–exon boundaries in maize. Primers that amplify a region from 32 bp 5′ of the putative TSS to 169 bp 3′ of the predicted stop codon were used in RT-PCR with total RNA isolated from immature ears, and a single band with the expected size of 800 bp was obtained (Fig. 1b; Supplementary Fig. S2c). Subsequent sequence analysis confirmed this fragment as the homologous maize EST of rice *GS3*. The *ZmGS3* transcriptional unit consists of five exons that encode a predicted protein of 198 amino acids with an estimated molecular weight of 21,146 D (GenBank accession no. FJ797616).

The predicted protein sequence of *ZmGS3* had several known domains, including one transmembrane domain and two overlapping tumor necrosis factor receptor (TNFR)/nerve growth factor receptor (NGFR) family cysteine-rich domains, which were also found in the rice GS3 protein sequence (Fig. 1b; Supplementary Fig. S3). However, the C-terminus von Willebrand factor type C domain and the N-terminus phosphatidylethanolamine-binding protein-like domain in the rice GS3 protein sequence were not found in the ZmGS3 protein sequence (Supplementary Fig. S3). In addition, alignment of the protein sequences of *GS3*-like genes from several higher plants revealed that the N-terminus was more conserved than the C-terminus, which might imply that it is performing a critical function (Supplementary Fig. S4).

Nucleotide diversity of *ZmGS3* in maize inbreds, landraces and teosintes

To examine levels of genetic diversity in *ZmGS3*, a panel of 25 lines (Table 1) was used to compare the genetic diversity between maize elite inbreds and landraces, and between all maize lines and teosintes. Sequences of four different regions within the *ZmGS3* gene (hereafter labeled parts A, B, C and D) ranging from 420 to 648 bp in length were obtained using the sequence of inbred line B73 as a reference. Part A corresponds to a region from 632 to 47 bp upstream of the putative TSS. Part B incorporates exon 1 plus 189 bp upstream and 281 bp downstream sequences. Part C begins from exon 2 and extends to intron 4. Part D contains the fifth exon and 80 bp of the 3′ UTR (Fig. 1b). Together, these four parts have a length of 1,763 bp (length of aligned sequences without gaps).

Among these four parts, 69 and 107 SNPs were identified in maize and teosinte, respectively, with an average of one SNP every 26 bp (in maize) and 16 bp (in teosinte) (Table 2). The number of SNPs was comparable between maize inbreds and landraces for the four parts, but teosinte contained more SNPs than all maize lines in three out of the four parts. As a further indication of teosinte diversity, every teosinte line represented a unique haplotype (h), while two or more maize lines often shared the same haplotype (Table 2). In parts A, B, and D, the average π over all maize lines was lower than that in teosinte (Table 2; Fig. 2), although more maize lines were sampled. Maize accounted for only 36, 37 and 54% of the genetic diversity levels observed in teosinte in parts A, B and D, respectively, in agreement with the diversity levels preserved in other neutrally evolving loci (Table 3). In part C, maize owns more diversity than teosinte, which may be because most of this region is composed of introns (Fig. 1b). Sliding-window analysis showed that nucleotide diversity was not equally distributed in either maize or teosinte, with exons containing less genetic variation (Fig. 2).

Table 2 Sequence diversity statistics for *ZmGS3* in maize and teosinte

Type	*N*	Part A (475 bp)[a]			Part B (380 bp)[a]			Part C (497 bp)[a]			Part D (411 bp)[a]			Total (1,763 bp)[a]
		S	π	*h*	*S*	π	*h*	*S*	π	*h*	*S*	π	*h*	*S*
All maize lines	18	14	0.00959	7	15	0.01268	6	35	0.03064	5	5	0.00336	5	69
Inbreds	10	13	0.01029	5	10	0.01333	5	31	0.03116	3	4	0.00357	3	58
Landraces	8	8	0.00820	4	10	0.01034	3	33	0.03291	4	3	0.00269	4	54
All teosintes[b]	7	38	0.02637	7	37	0.03459	7	25	0.01878	7	7	0.00626	7	107

N number of sampled lines, *S* number of segregating sites, π the expected heterozygosity per nucleotide site, *h* number of haplotypes

For the regions referred by Parts A, B, C and D, see "Results" and Fig. 1

[a] Length of the alignments in which all sequences contain bases, excluding gaps

[b] Including the species *Zea mays* ssp. *mexicana*, *Zea mays* ssp. *parviglumis* and *Zea mays* ssp. *huehuetenangensis*

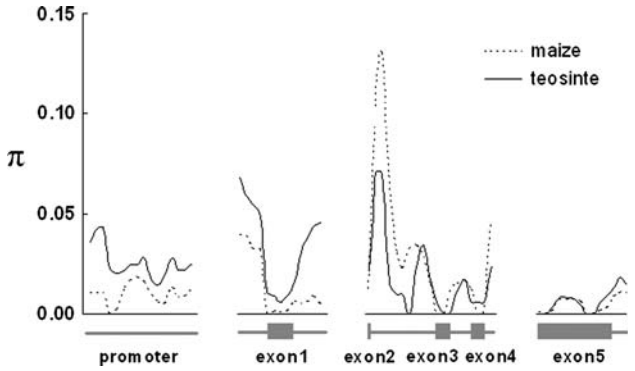

Fig. 2 Sliding-window analysis of nucleotide diversity in ten maize inbred lines, eight maize landraces and seven *Zea mays* teosinte lines (window length = 50 sites for part C (exons 2–4; Fig. 1b) and 100 sites for other three parts, step size = 25 sites)

Tests for selection at *ZmGS3*

Tajima's *D* test and the HKA test were conducted for each of the four sequences from *ZmGS3* to investigate evidence of selection during domestication and improvement. For Tajima's *D* test, negative values are consistent with directional selection. The HKA test relies on divergence information relative to the outgroup species and requires sequence information of neutrally evolving reference genes and target genes both in the tested materials and in the outgroup species. These two tests have successfully identified genes with a clear role in maize domestication, such as the gene controlling lateral branching, *tb1* and the gene controlling glume architecture, *tga1* (Table 3). For Tajima's *D* test, parts A and C gave positive values, parts B and D gave negative but not significant values, indicating that no selection had occurred on *ZmGS3*. The HKA test also gave consistently negative results for selection (Table 3).

Phylogenetic analysis

Assuming that only a single favorable ancestral haplotype of domestication-related genes became fixed by selection in the past, maize sequences would be expected to form a single clade during phylogenetic analysis (Clark et al. 2004; White and Doebley 1999). As shown in Fig. 3, phylogenetic analysis of combined sequences of the four parts of *ZmGS3* from the 25 *Z. mays* lines revealed that the maize lines clustered into two groups. In group I, maize lines either fell within clades of nearly identical sequences (like MAI2, MAI7 and MAI8) or among teosinte lines (like MAI3). Group II only contained maize lines. The teosinte line TEO4, belonging to *Z. mays* ssp. *huehuetenangensis*, was more distantly related to the other 24 lines, in agreement with known phylogenetic analyses (Fukunaga et al. 2005). The distribution of maize lines in two groups revealed divergence of *ZmGS3* among maize lines.

ZmGS3 is primarily expressed in unfertilized ears

Real-time qRT-PCR was performed to determine the expression pattern of *ZmGS3*. As shown in Fig. 4, expression was most intense in growing regenerative tissues, including tassel, silk, ears and kernels. The highest expression was seen in immature ears of plants with 18 expanded leaves, and expression level in kernels collected at 0 DAP was comparable. However, mRNA levels decreased rapidly after pollination and became nearly undetectable in kernels collected at 15 DAP. In contrast to regenerative tissues, transcript levels of *ZmGS3* were low in vegetative tissues, including husk, seedling shoot and seedling root, and was absent from leaves.

Genetic polymorphisms in *ZmGS3* are associated with variations in maize kernel traits

In rice, one stop codon occurring in the second exon led to increased grain size, primarily grain length (Fan et al. 2006; Takano-kai et al. 2009). However, in the 18 sequenced maize lines, the first four exons were completely conserved and only the fifth exon showed polymorphisms (Fig. 5a). Furthermore, analysis of *ZmGS3* cDNA sequences from

Table 3 Tests for selection at the *ZmGS3* gene and examples of genes with evidence of selection

Locus	Length (bp)	HKA (*P* value)	Tajima's *D*	π_m/π_t	References
ZmGS3-part A	514	0.977	1.16	0.36	This study
ZmGS3-part B	441	0.385	−0.38	0.37	This study
ZmGS3-part C	495	0.983	1.41	1.63	This study
ZmGS3-part D	414	0.468	−0.16	0.54	This study
tga1-exon 2	463	0.220	0.25	0.35	Wang et al. (2005)
tga1-exon 3	712	0.938	−0.83	0.44	Wang et al. (2005)
***tga1*-promoter**	739	≪0.001	−2.02*	0.05	Wang et al. (2005)
***tb1*-1.7 kb**	601	<0.001	−0.92	0.01	Clark et al. (2004)
***tb1*-7.1 kb**	841	<0.001	−1.48	0.06	Clark et al. (2004)

For the regions referred by parts A, B, C and D, see results and Fig. 1; loci that are subjected to selection are in bold

m maize, *t* teosinte

* $P < 0.05$

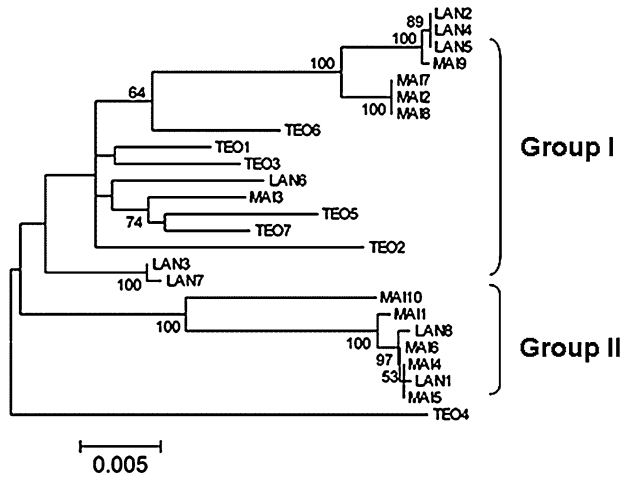

Fig. 3 Phylogenetic analysis of *ZmGS3* in ten maize inbred lines, eight maize landraces and seven *Zea mays* teosinte lines. Maize lines are classified within two groups, Group I and Group II. Numbers at the branches are percentages based on 1,000 bootstrap repetitions, bootstrap values >50% are given. The *scale bar* indicates the number of nucleotide substitutions per position. *MAI* maize, *TEO* teosinte, *LAN* landrace

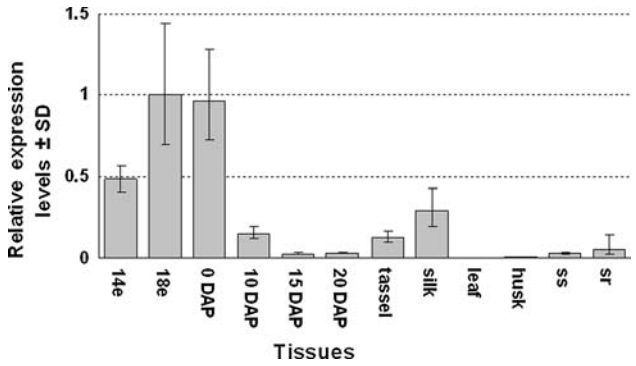

Fig. 4 Expression pattern of *ZmGS3* based on real-time qRT-PCR analysis. 14e and 18e refer to immature ears from plants with 14 and 18 expanded leaves, respectively. 0 DAP, 10 DAP, 15 DAP and 20 DAP refer to kernels harvested at 0, 10, 15 and 20 days after pollination, respectively. *ss* seedling shoot, *sr* seedling root. Ubiquitin was used as the endogenous control

another 55 maize lines failed to identify any polymorphisms in the first four exons (data not shown). Thus, only the fifth exon and a stretch of 600 bp sequences from the promoter region were further analyzed and inspected for possible polymorphisms causing phenotypic variation using association mapping analysis.

Multiple sequence alignment analysis revealed a total of seven polymorphisms with minor allele frequency >0.05 (Fig. 5b) in the fifth exon and 3′ UTR. Site 47 is a 3 bp insertion/deletion (InDel) polymorphism coding the amino acid valine. Sites 79, 80 and 81 together constitute one amino acid codon. Sites 79 and 80 are in complete linkage disequilibrium (LD, $r^2 = 1$), thus they were combined for subsequent association analysis. Site 172 is a large InDel that contains several amino acids. For maize lines that have the insertion, several SNPs were found, thus giving a total of five haplotypes. Because haplotypes CCC and CCCCC have allele frequencies <0.05, they were eliminated from subsequent association analysis to avoid statistical false positives. An InDel (site 381) and a SNP (site 397) were also identified in the 3′ UTR. Association analysis revealed that the haplotype rich in cysteine (site 172) is significantly associated with kernel length in both environments at the $P = 0.05$ level (Table 4), with 5.77% of the phenotypic variation being explained in Beijing and 5.21% in Hainan. The other remaining five sites showed no significant association with any of the four traits at the $P = 0.05$ level.

In the promoter region, we identified a total of 43 polymorphisms with minor allele frequency >0.05. These include 24 SNPs and 19 InDels. Association analysis with the MLM model controlling for both population structure and individual relatedness revealed 18 polymorphisms that are associated with at least one of the four yield-related traits. Two sets of the polymorphisms are in high LD. One set includes five sites, site P307, site P325, site P363, site P437 and site P627 (mean $r^2 = 0.93$), the other includes site P141, site P366, site P386, site P388, site P531 and site P561 (mean $r^2 = 0.90$), thus only one polymorphism in each of the two sets were presented (Table 4, for a full list of

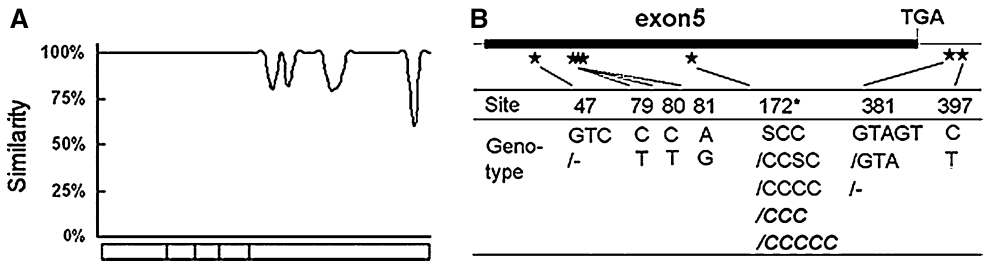

Fig. 5 **a** Diversity of predicted amino acids in 10 maize inbred lines and eight landraces. The five exons are indicated by *open boxes* at the bottom. The similarity was averaged over every five aligned amino acids. **b** Nucleotide diversity in the fifth exon and the 3′ UTR of *ZmGS3* among 121 maize inbred lines. Site is given according to the position of the corresponding polymorphisms in the aligned sequences. For site 172 (*), the amino acid sequences are given (C cysteine, S serine), and the two haplotypes with allele frequency <0.05 are in *bold italics*. For the other sites, the nucleotide sequences are given. – indicates a deletion

Table 4 Association analysis of four yield-related traits based on polymorphisms in the promoter, the fifth exon and 3′ UTR

Site (genotype)[a]	Frequency	KL		KW		KT		HKW	
		07BJ	07HN	07BJ	07HN	07BJ	07HN	07BJ	07HN
172 (1/2/3)[b]	6/53/55	0.010	0.033	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
P120 (0/4/18)	62/33/23	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.020
P143 (-/A/T)	19/62/37	n.s.	n.s.	8.928E-04	n.s.	n.s.	n.s.	0.009	n.s.
P172 (0/5/6/8)	19/24/13/62	n.s.	n.s.	0.002	n.s.	n.s.	n.s.	0.015	n.s.
P299 (0/1)	14/104	n.s.	n.s.	0.029	n.s.	n.s.	n.s.	0.024	n.s.
P311 (0/6/9)	19/76/23	n.s.	n.s.	0.007	n.s.	n.s.	n.s.	n.s.	0.018
P319 (-/G/T)	23/12/83	n.s.	n.s.	0.009	n.s.	n.s.	n.s.	0.018	0.017
P325 (A/C)	23/95	n.s.	n.s.	0.037	n.s.	n.s.	n.s.	n.s.	0.005
P366 (0/3)	17/101	n.s.	n.s.	0.001	n.s.	n.s.	n.s.	0.008	n.s.
P635 (0/2/6)	24/37/56	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.039

KL kernel length, *KW* kernel width, *KT* kernel thickness, *HKW* hundred kernel weight, *07BJ* Beijing (year 2007), *07HN* Hainan (year 2007), *n.s.* not significant at $P = 0.05$ level

[a] The nucleotides are given. -, deletion. For site 172, 1/2/3, SCC/CCSC/CCCC (C cysteine, S serine). The numbers in other parentheses indicate the number of deleted nucleotides. The prefix P indicates polymorphisms from the promoter region

[b] If including the two rare haplotypes (CCC/CCCCC; Fig. 5b), site 172 is still significantly associated with KL-07BJ ($P = 0.0193$), KL-07HN ($P = 0.0496$)

associations, see Supplementary Table S3). Contrary to the associations in the fifth exon and 3′ UTR, only associations with kernel width and HKW were identified. Most of the effect can only be detected in one environment, indicating the instability of gene function. Site P319 are associated with HKW in both environments (Table 4), and the phenotypic variation explained by this polymorphism is 6.29% in Beijing and 7.73% in Hainan.

Discussion

Feasibility of gene cloning methods used in the present study

The availability of abundant maize whole genomic (mainly from B73) and EST sequences in public databases provides an opportunity for gene cloning. However, limitations such as gaps in the genomic sequences, and the poor representation of genes with low expression levels or tissue-specific expression patterns still exist. This can be partially complemented by sequences from High Cot (Whitelaw et al. 2003) and methylation filtration (Palmer et al. 2003) sequencing. In addition, sequences from close relatives of maize, such as sorghum (Bedell et al. 2005; Paterson et al. 2009) and rice (Goff et al. 2002; Yu et al. 2002), can also be utilized to close the gaps and determine gene structure. Cloning of *ZmGS3* in this study integrated the methods mentioned above. Using rice sequences as a bridge, unrelated BAC and GSS sequences were identified, which contained sequences from different parts of *ZmGS3*. Divergence between rice and maize sequences did not allow the direct use of rice sequences to determine the exon–intron boundaries of maize *ZmGS3*, so the more closely related EST

sequences from sorghum (Kellogg 1998) was used to provide the needed information. The RT-PCR product in this study and the recently released maize cDNA sequence in the database (www.ncbi.nlm.nih.gov; Alexandrov et al. 2009) confirmed the feasibility of this method. It is apparent that this method will help to improve the quality of sequences generated from ongoing maize whole genome sequencing project using BAC by BAC strategy (Wilson 2008).

The maize *GS3* gene was mapped to a region showing collinear relationship with the rice *GS3* region, indicating that we have obtained the orthologous gene of *GS3* in maize. Recently, Huang et al. (2009) isolated another rice grain yield gene, *DEP1*, which shares some homology with the *GS3* gene. To verify that we have indeed isolated the orthologous gene of *GS3* in maize, we also obtained the two homologous genes of *DEP1* in maize. Multiple sequence alignment and phylogenetic analysis suggested that *ZmGS3* and *OsGS3* were more closely related than with the *DEP1* genes, further supporting the orthologous relationship between them (Supplementary Fig. S5).

ZmGS3 is involved in maize kernel development with mechanisms differing from rice *GS3*

In this study, *ZmGS3* was mapped to a region containing kernel weight QTL found in two different populations, one derived from two maize inbred lines and the other from a cross between maize and its proposed progenitor *parviglumis* teosinte (Supplementary Table S1; Supplementary Fig. S1). This suggested that *ZmGS3* might be the underlying gene for these kernel weight QTL and may have been involved in maize kernel weight improvement during domestication. Association mapping was employed to examine whether *ZmGS3* influences these kernel weight QTL. Indeed, we identified polymorphisms in the sequenced promoter region of *ZmGS3* that were associated with HKW. This implied that *ZmGS3* might be the underlying gene for the previously identified evolutionary QTL. However, our selection test did not support *ZmGS3* as a gene underlying selection (Table 3). One possible explanation is that the identified QTL region is very large and multiple genes may be responsible for the QTL region. As shown in Supplementary Fig. S1, the QTL region identified by Doebley et al. spanned Bins 1.02–1.07 and the QTL mapped by Yan et al. spanned Bins 1.04–1.05. Nevertheless, *ZmGS3* is involved in maize kernel development as evidenced by expression pattern and association analyses. *ZmGS3* is primarily expressed in unfertilized ears and decreases rapidly after pollination, suggesting a role in kernel development, as it does in rice (Fan et al. 2006, 2009). Results from association analysis further support this suggestion, as both polymorphisms in the fifth exon and the promoter region were shown to be significantly associated with at least one of the four yield-related traits (Fig. 5b; Table 4; Supplementary Table S3).

Recently, the rice *GS3* gene was shown to be highly expressed in young panicles but is not expressed in leaves (Takano-kai et al. 2009). The *ZmGS3* gene has a similar expression pattern, but differences still exist. While the gene expression is below detection in rice panicles at flowering, the maize gene shows a relatively high expression in maize kernels at silking (Fig. 4). Whether this expression divergence has impact on variation in kernel traits is unknown.

The mutation associated with phenotypic variation for grain size appears to be different between maize and rice. Cloning of the *GS3* gene in rice revealed that a single SNP occurring in the second exon changed a cysteine codon (TGC) in the small-grain varieties to a termination codon (TGA) in the large-grain varieties (Fan et al. 2006). The function of this SNP was further confirmed using association analysis across a panel of 180 diverse rice varieties (Fan et al. 2009). Also it is this same SNP that accounts for seed morphology variation in another panel of cultivated rice varieties (Takano-kai et al. 2009). However, this functional mutation was not found in the maize panel used in this study, which implies that this SNP appeared after the maize–rice divergence. Besides, the size of *GS3* effect might also be different between rice and maize. In rice, *GS3* is a major gene for grain length and it can explain up to 72% of the variation in grain length in the analyzed 180 rice varieties (Fan et al. 2006, 2009). However, our results showed that the effect of *ZmGS3* was marginally significant with the *P* values ranging between 0.05 and 0.001 (Table 4; Supplementary Table S3), and the phenotypic variation explained by the identified polymorphism is less than 8%, indicating *ZmGS3* is only a minor gene for variations in maize kernel traits, at least in our association panel. One possible explanation for this marginal association is the different genetic architecture of complex quantitative traits in outcrossing species like maize and self-fertilizing species like rice, as it does in the case of flowering time discovered by Buckler et al. (2009). Another explanation may be the small size of our association panel. It is well known that genetic variations found in the testing materials are the basis for association analysis (Yamasaki et al. 2005). The relatively small size of our association panel may disable us to identify some important genetic variations, including corresponding null alleles as the rice *GS3* gene leading to larger grain size, and thus, may limit the power of association analysis. Therefore, association analysis in another panel with larger size may help resolve these problems and lead to improved understanding of the gene function (Myles et al. 2009). Nevertheless, based on the association results in our panel composed of elite inbred lines used in

China, we cannot find any variations that could completely eliminate the negative function of the ZmGS3 protein as the stop codon mutation in the large-grain varieties of rice, indicating that knocking out *ZmGS3* in future breeding program could potentially result in improved yield in maize.

ZmGS3 was not a target of selection

Grain size is an evolutionarily and agriculturally important trait, so genes involved in its variation could be targets of selection during domestication and improvement. QTL studies suggest that homologous genes may be involved in convergent domestication in different species (Paterson et al. 1995). As studies concluded that the rice *GS3* gene might be involved in the domestication of grain size (Fan et al. 2006; Li et al. 2004; Takano-kai et al. 2009), we investigated evidence for selection on *ZmGS3*. Contrary to our expectation, evidence from different lines of investigation support that *ZmGS3* has not been subjected to selection. Compared with teosinte, nucleotide polymorphism of *ZmGS3* retained in maize is similar to that of neutrally evolving genes (Wang et al. 2005; White and Doebley 1999). In the phylogenetic tree, maize sequences are grouped into two different clades, with one clade containing both teosintes and maize lines, in agreement with other unselected genes (Clark et al. 2004; White and Doebley 1999). Finally, both Tajima's *D* and HKA tests showed no evidence of selection during maize domestication and improvement for all four sequenced regions of *ZmGS3*. These results may also be reasonable, as recently cloned QTL for domestication-related traits, such as shattering (Konishi et al. 2006; Li et al. 2006; Lin et al. 2007) and grain width in rice (Shomura et al. 2008; Song et al. 2007; Wan et al. 2008; Weng et al. 2008), revealed that even in the same species, different target genes or different alleles are involved in domestication of the same trait. The marginal associations of *ZmGS3* with kernel size suggest that the absence of favorable polymorphisms which could eliminate the negative regulation of grain size by the ZmGS3 protein was the reason why selection did not occur. Collectively, these results imply that in currently used maize breeding materials, *ZmGS3* might not be a major domestication- or improvement-related gene influencing kernel size. However, based on the results of expression and association analyses, *ZmGS3* may still play an important role in maize kernel development, and may yet hold potential for yield improvement in maize.

Acknowledgments The authors are grateful to Dr. Qifa Zhang for kindly providing the rice *GS3* information before the results' publication. This research was supported by the National Natural Science Foundation of China (30500322) and National Hi-Tech Research and Development Program of China.

References

Alexandrov NN, Brover VV, Freidin S, Troukhan ME, Tatarinova TV, Zhang HY, Swaller TJ, Lu YP, Bouck J, Flavell RB, Feldmann KA (2009) Insights into corn genes derived from large-scale cDNA sequencing. Plant Mol Biol 69:179–194

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, Jones J, Flick E, Rohlfing T et al (2005) Sorghum genome sequencing by methylation filtration. PLoS Biol 3(1):e13

Bradbury PJ, Zhang ZW, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc Natl Acad Sci USA 101(3):700–707

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155–162

Doebley J, Bacigalupo A, Stec A (1994) Inheritance of kernel weight in two maize-teosinte hybrid populations: implications for crop evolution. J Hered 85:191–195

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127(7):1309–1321

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acid Res 32(5):1792–1797

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

Fan CC, Yu SB, Wang CR, Xing YZ (2009) A causal C–A mutation in the second exon of *GS3* highly associated with rice grain length and validated as a functional marker. Theor Appl Genet 118:465–472

Fukunaga K, Hill J, Vigouroux Y, Matsuoka Y, Sanchez GJ, Liu K, Buckler ES, Doebley J (2005) Genetic diversity and population structure of teosinte. Genetics 169:2241–2254

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95(5):1971–1974

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296(5565):92–100

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009) Natural variation at the *DEP1* locus enhances grain yield in rice. Nat Genet 41(4):494–497

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Kellogg EA (1998) Relationships of cereal crops and other grasses. Proc Natl Acad Sci USA 95(5):2005–2010

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 43(10):1096–1105

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science 312(5778):1392–1396

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinf 5(2):150–163

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

Li CB, Zhou AL, Sang T (2006) Rice domestication by reducing shattering. Science 311(5769):1936–1939

Lin ZW, Griffith ME, Li XR, Zhu ZF, Tan LB, Fu YC, Zhang WX, Wang XK, Xie DX, Sun CQ (2007) Origin of seed shattering in rice (*Oryza sativa* L.). Planta 226(1):11–20

Lincoln S, Daly M, Lander E (1992) Constructing genetic maps with MAPMAKER/EXP3.0. Whitehead Institute, Cambridge

Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J 8(3):457–463

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408

Ma XQ, Tang JH, Teng WT, Yan JB, Meng YJ, Li JS (2007) Epistatic interaction is an important genetic basis of grain yield and its components in maize. Mol Breed 20(1):41–51

Martin A, Lee J, Kichey T, Gerentes D, Zivy M, Tatout C, Dubois F, Balliau T et al (2006) Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. Plant Cell 18:3252–3274

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell. doi:10.1105/tpc.109.068437

Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU, Dike S, Bastide M, Martienssen RA, McCombie WR (2003) Maize genome sequencing by methylation filtration. Science 302(5653):2115–2117

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(5231):1714–1718

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15(2):174–175

Salse J, Piégu B, Cooke R, Delseny M (2004) New *in silico* insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. Plant J 38:396–409

SanMiguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 82:37–44

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet 40(8):1023–1028

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39(5):623–630

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

Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, Kanamori H, Padhukasahasram B, Bustamante C, Yoshimura A, Doi K, McCouch S (2009) Evolutionary history of *GS3*, a gene conferring grain length in rice. Genetics (published articles ahead of print).doi:10.1534/genetics.109.103002

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 21(7):1214–1225

Teng WT, Cao JS, Chen YH, Liu XH, Jing XQ, Zhang FJ, Li JS (2004) Analysis of maize heterotic groups and patterns during past decade in China. Sci Agric Sin 37(12):1804–1811

Wan XY, Weng JF, Zhai HQ, Wang JK, Lei CL, Liu XL, Guo T, Jiang L, Su N, Wan JM (2008) Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5. Genetics 179:2239–2252

Wang H, Nussbaum-Wagler T, Li BL, Zhao Q, Vigouroux Y, Faller M, Bomblies K, Lukens L, Doebley JF (2005) The origin of the naked grains of maize. Nature 436:714–719

Weng JF, Gu SH, Wan XY, Gao H, Guo T, Su N, Lei CL, Zhang X, Cheng ZJ, Guo XP, Wang JL, Jiang L, Zhai HQ, Wan JM (2008) Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. Cell Res 18:1199–1209

White SE, Doebley JF (1999) The molecular evolution of *terminal ear1*, a regulatory gene in the genus Zea. Genetics 153:1455–1462

Whitelaw CA, Barbazuk WB, Pertea G, Chan AP, Cheung F, Lee Y, Zheng L, Heeringen S et al (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302(5653):2118–2120

Wilson R (2008) Sequence and assembly of the maize B73 genome. Abstract of 50th annual maize genetics conference, 27 February–1 March 2008, Washington, DC, p 18

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell 17:2859–2872

Yan JB, Tang H, Huang YQ, Zheng YL, Li JS (2004) Comparative analyses of QTL for important agronomic traits between maize and rice. Acta Genet Sin 31(12):1401–1407

Yan JB, Tang H, Huang YQ, Li JS, Zheng YL (2006) Quantitative trait loci mapping and epistatic analysis for grain yield and yield components using molecular markers with an elite maize hybrid. Euphytica 149:121–131

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell 12:2473–2484

Yu J, Hu SN, Wang J, Wong G, Li SG, Liu B, Deng YJ et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296(5565):79–92

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2):203–208