DOI: 10.1534/genetics.109.103002

Evolutionary History of *GS3*, a Gene Conferring Grain Length in Rice

Noriko Takano-Kai,*,[1] Hui Jiang,[†,1,2] Takahiko Kubo,*,[3] Megan Sweeney,[†,4] Takashi Matsumoto,[§] Hiroyuki Kanamori, Badri Padhukasahasram,[††] Carlos Bustamante,[††] Atsushi Yoshimura,* Kazuyuki Doi*,[5] and Susan McCouch[†,6]**

**Faculty of Agriculture, Kyushu University, Fukuoka 812-8581, Japan, [†]Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853, [§]Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan, **Institute of Society for Techno-Innovation of Agriculture, Forestry, and Fisheries, Tsukuba, 305-0854, Japan and [††]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853*

Manuscript received March 27, 2009
Accepted for publication May 24, 2009

ABSTRACT

Unlike maize and wheat, where artificial selection is associated with an almost uniform increase in seed or grain size, domesticated rice exhibits dramatic phenotypic diversity for grain size and shape. Here we clone and characterize *GS3*, an evolutionarily important gene controlling grain size in rice. We show that *GS3* is highly expressed in young panicles in both short- and long-grained varieties but is not expressed in leaves or panicles after flowering, and we use genetic transformation to demonstrate that the dominant allele for short grain complements the long-grain phenotype. An association study revealed that a C to A mutation in the second exon of *GS3* (A allele) was associated with enhanced grain length in *Oryza sativa* but was absent from other Oryza species. Linkage disequilibrium (LD) was elevated and there was a 95.7% reduction in nucleotide diversity (θ_π) across the gene in accessions carrying the A allele, suggesting positive selection for long grain. Haplotype analysis traced the origin of the long-grain allele to a *Japonica*-like ancestor and demonstrated introgression into the *Indica* gene pool. This study indicates a critical role for *GS3* in defining the seed morphologies of modern subpopulations of *O. sativa* and enhances the potential for genetic manipulation of grain size in rice.

SEED size and seed number are the major determinants of crop yield in both the cereals and the grain legumes. Seed size was also a target of artificial selection during domestication, where large seeds are generally favored due to ease of harvesting and enhanced seedling vigor (Harlan *et al.* 1972). In rice, traits related to grain size and appearance have a large impact on market value and play a pivotal role in the adoption of new varieties (Champagne *et al.* 1999; Juliano 2003). However, different grain quality traits are prized by different local cultures and cuisines and, unlike other cereals such as wheat, barley, and maize that are sold largely in processed forms, the physical properties of rice grains are immediately obvious to consumers (Fitzgerald *et al.* 2009). Thus, rice offers a unique opportunity to investigate the genetics and evolutionary history of seed size and shape.

Cultivated rice (*Oryza sativa*) was domesticated in Asia from the wild progenitor *O. rufipogon* Griff. and/or *O. nivara* Sharma (Ishii *et al.* 1988; Oka 1988; Dally and Second 1990; Nakano *et al.* 1992; Chen *et al.* 1993). Classical studies of the subpopulation structure of *O. sativa* have identified two primary subspecies or varietal groups, namely *Indica* and *Japonica* (Oka 1988; Wang and Tanksley 1989; Sun *et al.* 2002). Studies that have dated the divergence between the *Indica* and the *Japonica* groups indicate that it predates rice domestication by at least 100,000 years (Ma and Bennetzen 2004; Vitte *et al.* 2004; Zhu and Ge 2005), suggesting that at least two genetically distinct gene pools of *O. rufipogon* were cultivated and subsequently domesticated.

Isozyme and DNA studies revealed that there is additional genetic structure within these two groups, with three subpopulations composing the *Japonica* varietal group (*temperate japonica*, *tropical japonica*, and *aromatic*, written all in lowercase) and two subpopulations composing the *Indica* group (*indica* and *aus*) (Second 1985; Glaszmann 1987; Garris *et al.* 2005; Caicedo *et al.* 2007). While there is great diversity of

Sequence data from this article have been deposited with the DDBJ/GenBank Data Libraries under accession nos. AB488612–AB488665.

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.109.103002/DC1.

[1]These authors contributed equally to this work.

[2]*Present address:* Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan 610041, P. R. China.

[3]*Present address:* Plant Genetics Laboratory, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan.

[4]*Present address:* Department of Plant Sciences, University of Arizona, Tucson, Arizona, 85721.

[5]*Present address:* Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan.

[6]*Corresponding author:* Cornell University, 162 Emerson Hall, Ithaca, NY 14853-1901. E-mail: srm4@cornell.edu

seed size and shape both within and between the different subpopulations of *O. sativa*, each subpopulation is popularly associated with a characteristic seed morphology. *Temperate japonica* varieties are known for their short, round grains, *indica* and *aus* for slender grains, and within the *aromatic* subpopulation [hereafter referred to as *Group V* varieties, according to the isozyme group designation (Glaszmann 1987)] the group of *basmati* varieties is highly valued for their very long, slender grains (Juliano and Villareal 1993). Identification of the genes that control the range of seed size variation in rice will offer opportunities to study the evolutionary history and phenotypic diversification of the five subpopulations within *O. sativa* and also provide valuable targets for genetic manipulation.

In rice, four genes contributing to seed or grain size have been identified and characterized. The first, *grain size 3* (*GS3*), was isolated from an *indica* × *indica* population and found to encode a novel protein with several conserved domains including a phosphatidylethanolamine-binding protein (PEBP)-like domain, a transmembrane region, a putative tumor necrosis factor receptor/nerve growth factor receptor (TNFR/NGFR) family domain, and a von Willebrand factor type C (VWFC) domain (Fan *et al.* 2006). A second gene, *grain weight 2* (*GW2*), was found to encode an unknown RING-type protein with E3 ubiquitin ligase activity (Song *et al.* 2007). The third, *grain incomplete filling 1* (*GIF1*), encodes a cell-wall invertase required for carbon partitioning during early grain filling (Wang *et al.* 2008). Finally, the recently characterized *seed width 5* (*SW5*) has no apparent homolog in the database but was shown to interact with polyubiquitin in a yeast two-hybrid assay; thus it likely acts in the ubiquitin–proteasome pathway to regulate cell division during seed development (Shomura *et al.* 2008; Weng *et al.* 2008).

Many genes controlling seed size have also been identified in Arabidopsis and tomato, providing a framework for assembling the genetic pathway that determines this trait in dicotyledonous plants (Chaudhury *et al.* 2001; Jofuku *et al.* 2005; Ohto *et al.* 2005; Sundaresan 2005; Schruff *et al.* 2006; Yoine *et al.* 2006; Roxrud *et al.* 2007; Li *et al.* 2008; Xiao *et al.* 2008; Orsi and Tanksley 2009; Zhou *et al.* 2009). Several of these genes show maternal control by regulating endosperm and/or ovule development (Garcia *et al.* 2003; Jofuku *et al.* 2005; Li *et al.* 2008; Ohto *et al.* 2005; Xiao *et al.* 2008).

Numerous studies have identified rice QTL associated with grain weight and grain length [www.gramene.org (Ni *et al.* 2009)]. Ten of these studies identified a seed size QTL located in the pericentromeric region of rice chromosome 3, using both inter- and intraspecific crosses (Li *et al.* 1997; Yu *et al.* 1997; Redona and Mackill 1998; Xiao *et al.* 1998; Kubo *et al.* 2001; Moncada *et al.* 2001; Brondani *et al.* 2002; Xing *et al.* 2002; Thomson *et al.* 2003; Li *et al.* 2004). In interspecific crosses, the wild accessions always contributed the dominant allele for small seed size at this locus. Comparative mapping of QTL controlling seed weight in rice, maize, and sorghum further suggested that orthologous seed size genes at this locus might be associated with domestication in all three crops (Paterson *et al.* 1995).

In the current study, we used positional cloning and transformation to demonstrate that the *GS3* gene underlies both the *gw3.1* QTL (Thomson *et al.* 2003; Li *et al.* 2004) and the *lk3* QTL (Kubo *et al.* 2001). In transformation experiments, we demonstrated for the first time that the dominant allele for small grain size complements the long-grain phenotype and we characterized the spatial expression patterns of the gene at different developmental stages. We undertook an association analysis to examine the relationship between the alleles at *GS3* and the observed variation for grain length/size in both wild and cultivated rice. Finally, we examined sequence haplotypes across the *GS3* region to look for evidence of selection and to identify the origin of the mutation leading to increased grain length in *O. sativa*.

MATERIALS AND METHODS

Fine mapping of *gw3.1* and *lk3*: A total of 4148 BC_5F_2 individuals derived from a cross between Jefferson × *O. rufipogon* and 1641 BC_3F_3 plants derived from a cross between Asominori × IR24 were screened for recombinants within target regions defined by the QTL *gw3.1* and *lk3*, respectively. Plants from the Jefferson × *O. rufipogon* population were grown in 2-inch-deep pots in the Guterman greenhouse of Cornell University. Plants from the Asominori × IR24 population were grown at University Farm, Kyushu University, Japan. Once informative recombinants were identified, seedlings of recombinants and parental controls were transplanted and allowed to set seeds. The seed length phenotype was evaluated using seeds harvested from primary panicles. Seeds from the *gw3.1* population were measured using a digital caliper as described (Li *et al.* 2004) and seed lengths from the *lk3* population were determined by visual observation. Progeny testing was conducted as necessary.

Additional DNA markers were designed and used to detect recombination break points. Simple sequence repeat (SSR) markers were designed using the SSRIT tool [(Temnykh *et al.* 2000) http://www.gramene.org/db/markers/ssrtool] and the publicly available Nipponbare genome sequence (http://rice.plantbiology.msu.edu/). SSR and indel markers were amplified using standard PCR protocols. Cleaved amplified polymorphic sequence (CAPS) markers were designed on the basis of Nipponbare sequences, PCR products were sequenced, and appropriate restriction enzyme were chosen; CAPS products were run on 2% agarose gels. Sequences of all primers are available in supporting information, Table S1.

Complementation test and expression analysis of *GS3*: The dominant C allele at *GS3* was subcloned as a 7-kb *Xmn*I fragment of a Nipponbare bacterial artificial chromosome (BAC) (OSJNBa0002D18) into the pPZP2H-lac binary vector (Fuse *et al.* 2001) (Figure 1). This construct was introduced into the chromosome segment substitution line AIS22 by Agrobacterium-mediated transformation (Toki 1997). AIS22 was constructed via backcrossing using Asominori (a short-seeded *temperate japonica* cultivar that is easy to transform) as

the recurrent parent and IR24 as the donor of the long seed allele (Kubo *et al.* 2002).

Total RNA was extracted from samples using the Aurum total RNA minipurification kit (Bio-Rad, Hercules, CA). *GS3* cDNA templates were generated from total RNA samples using the Revertra-Ace kit (TOYOBO, Osaka, Japan). The reverse-transcribed (RT)–PCR products of the *GS3* gene were sequenced. The sequences of the 5′ and 3′ ends of the cDNA were determined using the SMART RACE cDNA Amplification kit (Clontech, Mountain View, CA).

To detect the differential expression of *GS3* in wild-type and mutant alleles, total RNA was extracted from 3-cm-long young panicles of cv. Asominori, AIS22, and cv. Nipponbare. To detect the time and spatial expression of *GS3*, total RNA was isolated from panicles at different developmental stages and leaves of Asominori for RT–PCR. *GS3* was amplified using the primer pairs TGAGATCAAAACTAGCTACTACCAGCTAGA and CATGGCAATGGCGGCGGCGCCCCGGCCCAA. As controls, actin cDNA was amplified using the primer pairs TCCATCTTGGCATCTCTCAG and GTACCCGCATCAGG CATCTG with fewer PCR cycles.

To confirm the time and spatial expression of *GS3*, a fragment that contained 1030 bp of the promoter region of *GS3* was amplified from genomic DNA of Nipponbare, cloned, and fused with the beta-glucuronidase (GUS) gene and transferred into the *temperate japonica* variety, Nipponbare. The panicles, leaves, and leaf sheaths of the transgenic plants were stained to detect GUS activity as described by Takeda *et al.* (2003).

Plant materials for the analysis of association, linkage disequilibrium, haplotype diversity, and gene sequence: Information about the germplasm used in this study is listed in Table S2. This study included 235 accessions of *O. sativa* from 30 countries (75 *indica*, 36 *aus*, 15 *Group V*, 34 *temperate japonica*, 64 *tropical japonica*, and 11 admixed varieties), 79 accessions of *O. glaberrima* (Semon *et al.* 2005), 10 *O. barthii*, 4 *O. longistaminata*, 12 *O. glumaepatula*, 12 *O. meridionalis*, 18 *O. spontanea*, and 266 *O. rufipogon/O. nivara/O. spontanea*. Accessions were surveyed to determine the frequency of the A allele at *GS3* and a subset was used for gene sequencing and haplotype analysis. The subpopulation identities of the *O. sativa* accessions were as determined previously (Garris *et al.* 2005); accessions new to this study were genotyped using 50 well-distributed SSRs (http://www.gramene.org/microsat/index.html) and analyzed using STRUCTURE as reported by Garris *et al.* (2005). Additional information is provided in Table S2.

DNA extraction, PCR, and sequencing: DNA was extracted from leaf samples using a modified potassium acetate–SDS protocol (Dellaporta *et al.* 1983). PCR was conducted using modified PCR protocols described previously (Garris *et al.* 2005) with the annealing temperature at 58°. For sequencing, 3 μl of diluted PCR product was treated with 3 μl 1:3 Exo-SAP (containing 3 units exonuclease I and 1.6 units shrimp alkaline phosphatase diluted with 1× PCR buffer) and incubated at 37° for 45 min followed by 80° for 20 min. Sequencing was performed using both forward and reverse primers to ensure accuracy on ABI Prism 3700/3100 DNA analyzers (Applied Biosystems, Foster City, CA). Sequences were aligned using the Codon Code program (Codon Code, Dedham, MA). The ends of fragments were trimmed to remove low-quality sequences. Heterozygous sites were identified by visual inspection of chromatograms for double peaks; the singletons and ambiguous sites were resequenced as necessary.

Table S1 provides a list of all primers used for PCR and sequencing. Sequencing of the *GS3* gene spanned the five exons and four introns and also included 781 bp upstream and 16 bp downstream of the gene. The upstream border was set at −781 bp because the genomic fragment used for transformation experiments extended 781 bp upstream and showed good complementation (Figure 1B); thus we believe our sequencing incorporates the promoter region.

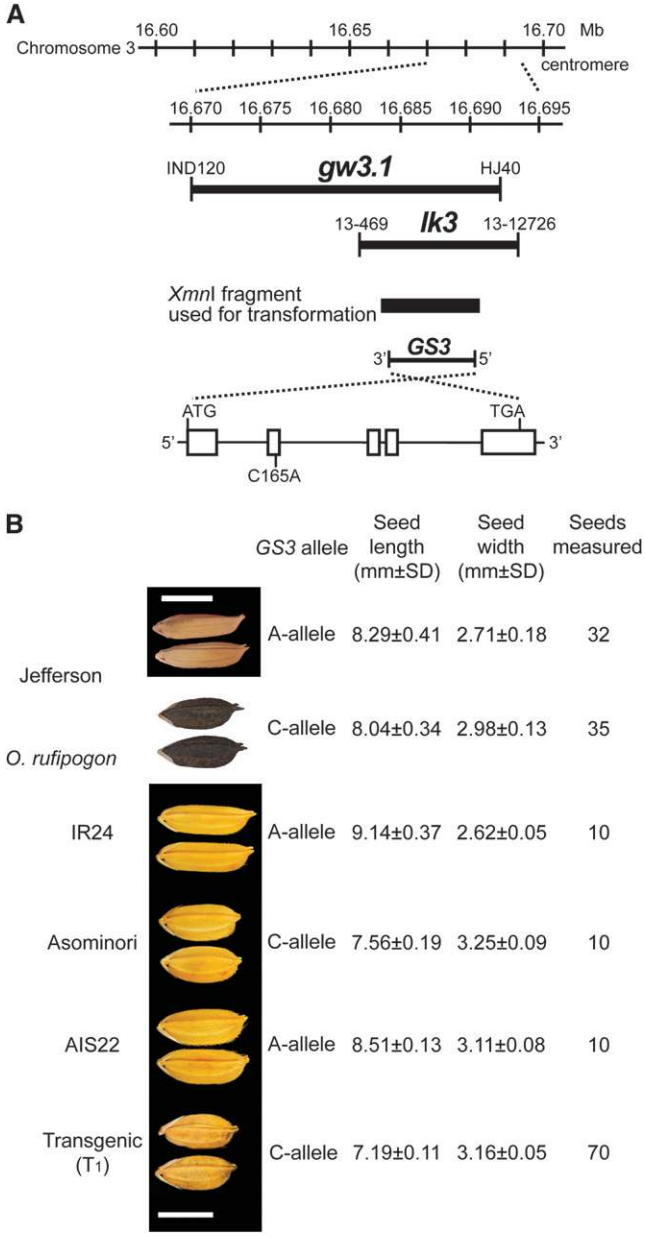

Figure 1.—Fine mapping of grain size QTL from two independent populations. (A) Location of fine-mapped regions on the short arm of rice chromosome 3. Megabase positions are based on TIGR pseudomolecule v.5; SSR and indel markers delineate fine mapping of *gw3.1* in a Jefferson × *O. rufipogon* population and *lk3* in an Asominori × IR24 population; the *Xmn*1 fragment contains the *GS3* gene used to transform AIS22; and the *GS3* gene model shows 5′–3′ orientation. (B) Seed phenotypes (photos) and mean length and width for parents of two mapping populations: Jefferson × *O. rufipogon* and IR24 × Asominori. AIS22, *japonica* line carrying the A allele used for transformation; T1 line, transformed with the C allele. Bar, 5 mm.

Association analysis: For association analysis, 18–226 seeds from 157 accessions of *O. sativa* and 12–65 seeds from 162

TABLE 1

Seed length and width of the transgenic plants in two independent T_1 lines

	Seed length			Seed width		
Line	+ transgene	No transgene	*P*	+ transgene	No transgene	*P*
T_1 1	7.2 ± 0.11 (7)	8.2 ± 0.12 (4)	<0.01	3.2 ± 0.00 (7)	2.9 ± 0.10 (4)	<0.01
T_1 2	7.2 ± 0.20 (6)	8.3 ± 0.20 (4)	<0.01	3.1 ± 0.10 (6)	2.9 ± 0.10 (4)	<0.01

Means (millimeters ± SD) of the plants with and without transgenes were presented and compared by *t*-test. Parentheses indicate number of plants. The average of the measurement of 10 seeds was used as the value of each plant.

accessions of *O. rufipogon/nivara/spontanea* (Table S2) were analyzed for seed length and seed width using the Winseedle scanner and software system (http://regentinstruments.com/). One CAPS marker, SHJ210 was designed to identify the functional mutation at *GS3*. PCR products amplified from the primer pair SHJ210F (GCTTGATTTCCTGTGCTATTAGG AG) and SHJ210R (CTCAAAAAGCTTGCACGATACTATG) were digested with the restriction enzyme *Pst*I and run on 2% agarose gels. Seed size and seed weight data analyses were conducted using JMP software (SAS Institute, Cary, NC).

Sequence analysis of the *GS3* gene: A total of 6.57 kb of DNA within the *GS3* gene was sequenced in 54 diverse accessions of *O. sativa* (Table S2). Three pairs of PCR primers were designed to amplify overlapping genomic regions within the gene and internal primers within each amplicon were designed to sequence the PCR products (Table S1). Sequences were assembled and aligned using the Sequencher program (Gene Codes, Ann Arbor, MI).

Nucleotide diversity (θ_π) was calculated using the DnaSP program (Rozas *et al.* 2003). Haplotypes were extracted using the same program after removing low-frequency alleles (<5%) and noninformative indels [*i.e.*, poly(A)] to reduce the complexity.

Haplotype and genetic diversity analysis across the *GS3* region: Extended haplotypes (EH) spanning a 66.2-kb region flanking *GS3* were used to distinguish the *indica* or *japonica* origin of the long-grain allele on the basis of analysis of 172 accessions of *O. sativa* and 39 accessions of *O. rufipogon*. The haplotypes were constructed from SNPs and indels (frequencies >5%) identified in six 500-bp reads within a 66.2-kb region flanking *GS3*. Population genetic analyses were conducted using DnaSP 4.1 (Rozas *et al.* 2003).

For extended haplotype homozygosity (EHH) analysis, haplotypes were constructed from 269 SNPs and indels identified from 14 500-bp reads (Table S1) that spanned a 7-Mb region around *GS3* (5 Mb downstream and 2 Mb upstream) in the same 172 accessions of *O. sativa* and 39 accessions of *O. rufipogon* as described above. The Fastphase program (Scheet and Stephens 2006) was used to fill in missing data to allow haplotype reconstruction across the target regions.

RESULTS

Fine mapping of *gw3.1* and *lk3*: To determine whether the *GS3* gene, which had previously been cloned from an *indica* × *indica* mapping population, was responsible for seed length and/or seed weight in crosses involving *japonica* cultivars, we fine mapped two QTL that had been previously mapped to the pericentromeric region of rice chromosome 3. The QTL *gw3.1* (Thomson *et al.* 2003) was fine mapped to a 22-kb region using a population derived from a cross between a long-seeded *tropical japonica* (cv. Jefferson) and an accession of *O. rufipogon*, and the QTL *lk3* (Kubo *et al.* 2001) was fine mapped to a 12-kb region in a cross between a short-seeded *temperate japonica* (cv. Asominori) and a long-seeded *indica* (cv. IR24) (Figure 1). The *GS3* gene was contained within both fine-mapped regions and was polymorphic in the second exon (C165A) (Fan *et al.* 2006) in both pairs of mapping parents. These results were consistent with the hypothesis that a single mutation in *GS3* was responsible for the seed size/seed weight QTL in the pericentromeric region of chromosome 3 in both *indica* and *japonica* genetic backgrounds. As documented by Li *et al.* (2004), seed size (caryopsis with hull) was highly correlated with grain size (caryopsis without hull, or brown rice grain; $R^2 = 0.975$) and we therefore refer to seed or grain size interchangeably in this article.

Complementation test: To confirm that the *GS3* gene was causally responsible for variation in seed or grain size in rice, we transformed a line containing the recessive C165A mutation conferring long grain (hereafter referred to as the A allele) with a 7-kb *Xmn*I fragment containing the dominant short-grain allele (the C allele) from cv. Nipponbare (Figure 1A). The recipient line used in this work was the chromosome segment substitution line, AIS22 (Kubo *et al.* 2002). AIS22 was genetically identical to the short-grained cultivar, Asominori, a *temperate japonica* cultivar that is easy to transform, except that it contained an introgression from the long-grained *indica* cultivar, IR24, across the *GS3* region on chromosome 3. Due to this introgression, the grains of AIS22 were significantly longer (12.6%) and thinner (4.3%) than those of Asominori ($P < 0.01$) (Figure 1B). No significant difference for grain thickness was observed between Asominori and AIS22. When AIS22 was transformed with the dominant, functional C allele at *GS3*, 16 of 32 T_0 plants showed a short-grain phenotype (data not shown). To further confirm cosegregation between grain length and the transgene, we obtained two independent T_1 families from two T_0 plants and observed segregation patterns. The grain length of T_1 individuals containing the transgene was significantly shorter than that of the T_1 plants that did not inherit the transgene (Figure 1B; Table 1). We thus concluded that

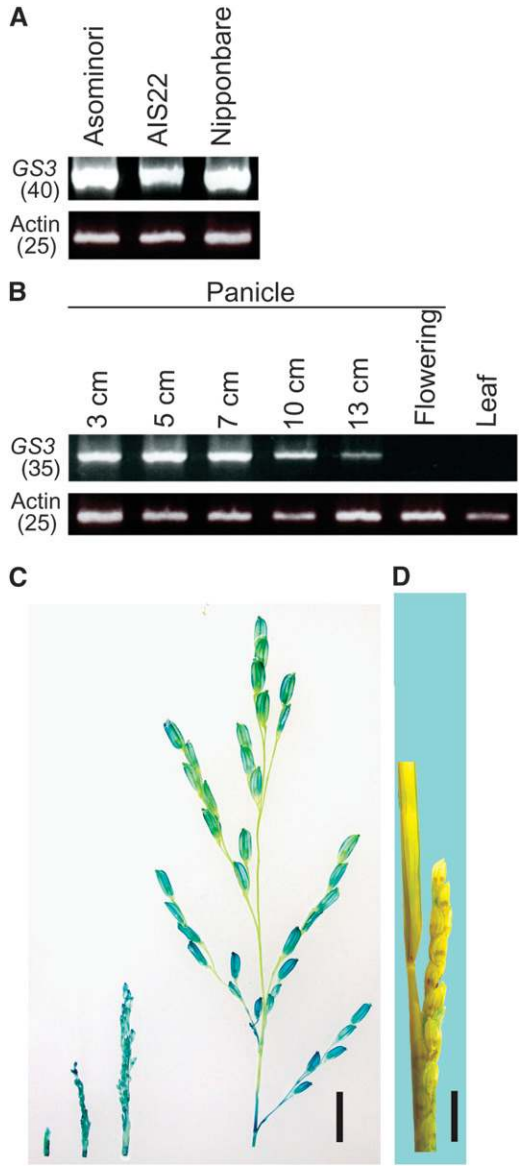

FIGURE 2.—Expression analysis of *GS3* using RT–PCR and genetic transformation. (A) RT–PCR of *GS3* from Asominori, AIS22, and Nipponbare using young panicles (~3 cm) for RNA extraction. Numbers in parentheses indicate cycles of PCR. Actin cDNA was amplified as a control. (B) RT–PCR of *GS3* from Asominori using panicles ~3, 5, 7, 10, and 13 cm and after heading and a flag leaf for RNA extraction. (C and D) *GUS* staining of transgenic plants transformed with the *GS3* promotor:*GUS* fusion construct showing panicles before flowering (C) and a panicle at flowering with a flag leaf (D). Bars, 1 cm.

the wild-type allele for short grains complemented the recessive long-grained phenotype.

mRNA expression of *GS3*: To determine in which tissues and developmental stages the gene was expressed, we examined the mRNA expression of *GS3* using RT–PCR. As seen in Figure 2A, *GS3* mRNA was expressed strongly in 3-cm-long panicles in both short-grained varieties (Asominori and Nipponbare) and in the long-grained chromosome substitution line (AIS22). As shown in Figure 2B, expression levels remained high as panicles developed from 3 to 7 cm in length and then decreased as they elongated past 10 cm. At flowering, mRNA expression levels were below detection in the panicles and *GS3* mRNA was not detected in leaf tissue. This suggested that the *GS3* gene regulates grain size in rice during the early phases of panicle development while spikelets are elongating. Further experiments using a *GS3* promotor::*GUS* fusion confirmed this pattern of expression (Figure 2, C and D). Because the genomic DNA fragment used for the complementation test contained only 781 bp upstream from the start codon, we used a 1030-bp promoter for the *GS3* promoter::*GUS* fusion. Another construct containing a 1984-bp promoter gave the same results. GUS expression was observed in panicles up until ~5 days before heading (Figure 2C), but the signal was not detected in either flowering panicles or leaves (Figure 2D).

TABLE 2

***GS3* allele frequency in Oryza species (A) and subpopulations in *O. sativa* (B)**

A. Oryza species	No. of accessions: Total	No. of accessions: A allele	A-allele frequency (%)
O. sativa	235	81	34
O. nivara/ O.rufipogon / O. spontanea	284	14	4
O. glaberrima	79	0	0
O. barthii	10	0	0
O. longistaminata	4	0	0
O. glumaepatula	12	0	0
O. meridionalis	12	0	0

B. Subpopulations in *O. sativa*	Varietal group	No. of accessions: Total	No. of accessions: A allele	A-allele frequency (%)
O. sativa		235	81	34
tropical japonica	*Japonica*	64	39	61
temperate japonica	*Japonica*	34	2	6
aromatic (Group V)	*Japonica*	15	7	47
indica	*Indica*	75	28	37
aus	*Indica*	36	1	3
admixture	—	11	4	36

***GS3* allele frequencies in diverse germplasm accessions:** The frequency of the A allele conferring long grain was evaluated in 322 wild accessions, 235 accessions of *O. sativa*, and 79 accessions of *O. glaberrima*. The A allele was observed in 34% of *O. sativa* and 4% of *O. rufipogon/O. nivara/O. spontanea*, while it was observed in none of the other accessions (Table 2A, Table S2).

When *GS3* allele frequencies were compared among the five subpopulations of *O. sativa*, highly significant ($P < 0.001$) differences were observed (Table 2B). The A allele was observed at relatively high frequency in

TABLE 3

Comparison of means and variances for seed morphology between *O. sativa* and *O. rufipogon* accessions (means ± SD)

	O. sativa: $n = 157$	*O. rufipogon*: $n = 162$	Means		Variances	
			F ratio[a]	*P*[a]	*F* ratio[b]	*P*[b]
Seed length (mm)	8.36 ± 1.04	8.49 ± 0.51	2.12	0.1465	43.44	<0.0001
Seed width (mm)	2.96 ± 0.41	2.43 ± 0.27	192.77	<0.0001	26.91	<0.0001
SL/SW ratio	2.89 ± 0.62	3.53 ± 0.40	123.98	<0.0001	31.24	<0.0001
Seed weight (g)[c]	23.21 ± 4.82	15.75 ± 3.16	269.07	<0.0001	16.56	<0.0004

SL, seed length; SW, seed width.
[a] *F* ratio and probability based on one-way ANOVA.
[b] *F* ratio and probability based on O'Brien's test for equal variance.
[c] Seed weight calculated as seed weight/1000 seeds.

tropical japonica (61%) and *Group V* varieties (47%), but at low frequency in *temperate japonica* (6%). Within the *Indica* subspecies or varietal group (capitalized when referring to varietal group), the A allele was present at moderately high frequency in *indica* (37%) (lowercase when referring to subpopulation) but it was almost entirely absent from *aus* (3%).

Phenotypic variation for seed morphology in wild and cultivated rice: Using a subset of the wild ($n = 162$) and cultivated materials ($n = 157$) described above, no significant differences were observed in average seed length ($P = 0.1465$) between the wild and the cultivated groups, although seeds of *O. sativa* were significantly wider and heavier than seeds of *O. rufipogon* ($P < 0.001$) (Table 3). Despite similar average seed lengths, the variance around the mean for seed length was significantly greater in *O. sativa* than in *O. rufipogon* ($P < 0.0001$). The coefficients of variation were also significantly greater in *O. sativa* for seed width, seed length/width ratio, and seed weight (Table 3).

Association between *GS3* genotype and seed morphology: When the two genotypic classes (A allele and C allele) of *O. rufipogon* were compared, there was no significant difference in seed length, width, length/width ratio, or weight (Table 4). In contrast, *O. sativa* accessions carrying the A allele ($n = 51$) had significantly longer and thinner seeds than accessions carrying the wild-type C allele ($n = 106$) ($P < 0.001$) (Table 4; Figure S1). This suggested that *GS3* affects seed morphology through interactions with factors in the genetic background that differ between *O. sativa* and its wild progenitor.

When the seed lengths of C-allele accessions of *O. sativa* were compared, there were significant differences among the subpopulations (Table 5). This provides evidence that genetic factors in addition to *GS3* contribute to the variation in seed length in the different subpopulations of *O. sativa*. Paradoxically, where the A-allele accessions of *O. sativa* were all significantly longer than C-allele accessions, but there were no longer significant differences among the five subpopulations. This suggests that the A allele of *GS3* masks the differences in seed length among the subpopulations that were detectable in accessions carrying the wild-type allele.

When the five subpopulations of *O. sativa* were considered individually, the association between *GS3* alleles and seed length was significant in every case, and R^2 values indicated that the A allele explained 57% of the phenotypic variation for seed length in *Group V*, 27% in *indica*, 22% in *temperate japonica*, 15% in *tropical japonica*, and 13% in *aus* (Table 4). *GS3* was significantly associated with seed width only in the *indica* subgroup ($R^2 = 0.26$), and it was associated with length/width ratio in *Group V* ($R^2 = 0.58$) and *indica* ($R^2 = 0.38$) (Table S3). The association between *GS3* alleles and seed weight was not significant in any of the individual subpopulations.

Sequence and haplotype variation at *GS3*: Sequencing of the *GS3* gene from 54 diverse accessions of

TABLE 4

Comparison of means and variances for seed morphology between the C and A alleles at *GS3* in *O. rufipogon* and *O. sativa*

	C allele[a]	A allele[b]	*F* ratio	*P*[c]
Seed length (mm)				
O. rufipogon	8.49 ± 0.50	8.49 ± 0.65	0.41	0.52
O. sativa	7.94 ± 0.90	9.21 ± 0.75	9.86	<0.0001
Seed width (mm)				
O. rufipogon	2.43 ± 0.27	2.36 ± 0.27	0.41	0.52
O. sativa	3.03 ± 0.39	2.82 ± 0.40	9.86	<0.0001
SL/SW ratio				
O. rufipogon	3.53 ± 0.40	3.63 ± 0.39	0.39	0.53
O. sativa	2.67 ± 0.52	3.34 ± 0.57	53.96	<0.0001
Seed weight (g)[d]				
O. rufipogon	15.72 ± 3.14	16.53 ± 3.62	0.43	0.5
O. sativa	22.52 ± 4.61	24.63 ± 4.98	6.93	0.01

[a] C allele: *O. rufipogon*, $N = 155$; *O. sativa*, $N = 106$.
[b] A allele: *O. rufipogon*, $N = 7$; *O. sativa*, $N = 51$.
[c] *P*-values calculated using the *F*-statistics.
[d] Calculated as seed weight/1000 seeds.

TABLE 5

Comparisons among *O. sativa* subpopulations for seed length between accessions carrying the A *vs.* C allele at *GS3* (mean ± SD)

Comparison	C allele: seed length (mm)	*N*	Level[b]	A allele: seed length (mm)	*N*	Level[b]	R^2	*F* ratio	*P*[a]
tropical japonica	8.47 ± 0.85	14	A	9.18 ± 0.83	27	A	0.15	6.64	0.01
temperate japonica	7.64 ± 0.59	27	BC	9.10 ± 2.21	2	A	0.22	7.79	0.01
aromatic (*Group V*)	7.21 ± 1.33	8	C	9.55 ± 0.54	6	A	0.57	16.06	<0.0001
indica	8.23 ± 0.86	24	AB	9.11 ± 0.46	15	A	0.27	14.48	<0.0001
aus	7.92 ± 0.89	33	ABC	9.94 ± 0.64	1	A	0.13	4.93	0.03

[a] Probability calculated using the *F*-statistics.
[b] Multiple means comparisons based on a Tukey–Kramer honestly significant difference test. Levels not connected by the same letter are significantly different.

O. sativa identified a total of 86 SNPs and 28 indels in the 6.57 kb of aligned sequenced DNA (Table S4). Of these changes, 2 SNPs and 1 indel were in exons. Other than the C165A SNP in exon 2 (described above), neither of the other 2 polymorphisms in *GS3* were associated with a significant difference in seed size, suggesting that they were not causally responsible for the phenotype.

***Japonica* origin of the C165A mutation:** Given the reported magnitude of the *Indica–Japonica* differentiation in rice ($F_{st} = 0.47$) (Garris *et al.* 2005; Caicedo *et al.* 2007; Kovach *et al.* 2007), we were interested to determine whether *GS3* haplotypes showed evidence of divergent *Indica* and *Japonica* ancestry and, if so, whether we could use ancestral differences to determine the origin of the C165A mutation.

Using the 70 SNPs/indels identified with allele frequency >5% in the 6.57 kb of *GS3* sequence, we constructed a total of 14 gene haplotypes from 54 *O. sativa* accessions (Figure 3A). Considering only wild-type (C-allele) accessions (33 accessions), a total of 11 ancestral gene haplotypes (GH1–11) were observed. To determine whether these gene haplotypes could be assembled into distinct clusters, we evaluated them using STRUCTURE and found the best-resolved clusters at $K = 2$ (Pritchard *et al.* 2000). Eighty-eight percent of the accessions from one cluster were from the *Japonica* varietal group, while 86% of accessions from the other cluster were from the *Indica* varietal group, defining the ancestral *Japonica* and *Indica* gene haplotype groups (Jap_GH and Ind_GH). Jap_GH contained 2 haplotypes (GH1 and GH2); the haplotypes differed by only 1 SNP in intron 2 (indicated in yellow in Figure 3A). Ind_GH contained 7 haplotypes (GH5–GH11). In addition, 2 accessions (haplotypes GH3 and GH4) were classified as admixed because they shared ancestry with both Jap_GH and Ind_GH. The admixed accessions were recombinant haplotypes (Figure 3A).

Three gene haplotypes carrying the A allele were identified (GH12–GH14) (Figure 3). GH12 clustered with Jap_GH, while GH13 and GH14 (recombinants) clustered with the admixed group. There were no A-allele haplotypes that clustered with Ind_GH. GH12, found in 90% of A-allele accessions, was identical to GH1 (C allele) across the entire sequence of the *GS3* gene, except for the functional C165A mutation (Figure 3A). Thus, we conclude that GH1, a *Japonica* haplotype, was the immediate ancestor of the C165A mutation. Our results demonstrated that all three A-allele haplotypes GH12, GH13, and GH14 are common by descent across the critical region of *GS3* containing the functional SNP.

To test the hypothesis that a single, *Japonica*-derived mutation in *GS3* was responsible for long grain in both the *Indica* and *Japonica* varietal groups, we examined a larger number of accessions (56 A-allele and 116 C-allele accessions of *O. sativa*) across a broad genomic region flanking *GS3*. A total of 30 extended haplotypes (EH1–30) were observed among the wild-type (C-allele) accessions, while only 4 extended haplotypes (EH31–34) were observed among the A-allele accessions. The extended C-allele haplotypes could be assembled into three distinct ancestral groups, corresponding to the composite *Japonica* varietal group (Jap_EH) and the two divergent subpopulations that compose the *Indica* varietal group, *indica* (Ind_EH) and *aus* (Aus_EH) (Figure 3B).

Almost all of the A-allele accessions (91%) were found to carry the EH31 haplotype, which clustered with Jap_EH. This was entirely consistent with the situation described above for the gene haplotype that carried the A-allele, GH12. A single A-allele accession carried haplotype EH32, which differed from EH31 at a single SNP, and this accession also clustered with Jap_EH. Haplotypes EH33 and EH34 were represented by one and three accessions, respectively, and these four accessions were found to be recombinant types that clustered with Ind_GH. Despite clustering with Ind_GH, both EH33 and EH34 carry a region of *Japonica*-like DNA near the *GS3* gene (Figure 3B). These results demonstrate that all accessions with the A allele carry a genomic region flanking the *GS3* gene that is closely related to ancestral *Japonica*. We can therefore conclude that this derived mutation in the *GS3* gene conferring

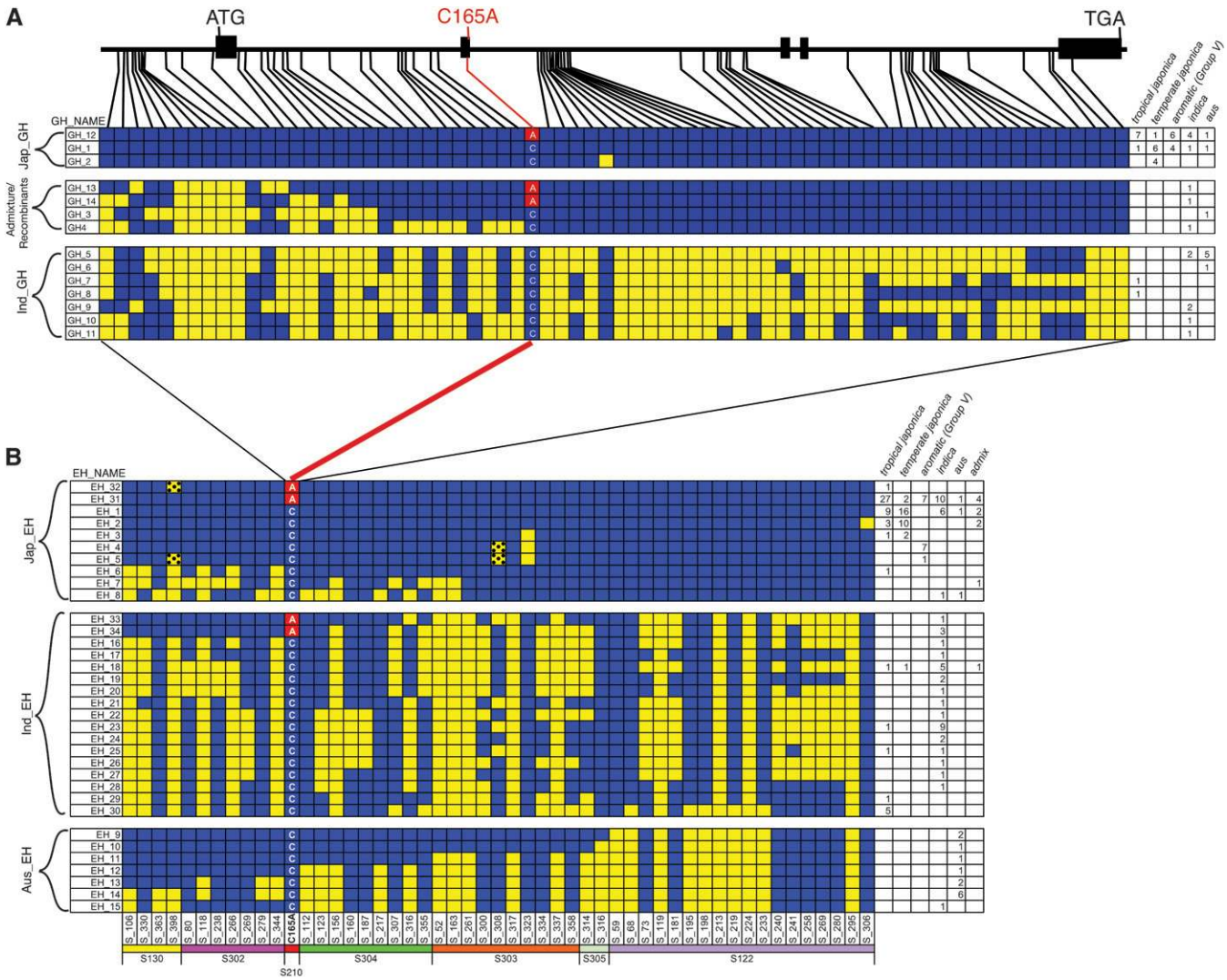

FIGURE 3.—Haplotype data across the *GS3* gene region. (A) The gene model for *GS3* containing five exons and comprising ~6.2 kb is shown horizontally along the top; the position of the C165A mutation is shown in red; and SNP positions with *GS3* are connected by lines to the haplotype table below. Rows in the table correspond to gene haplotypes (GH), which are organized into groups on the basis of STRUCTURE analysis; columns indicate the distribution of polymorphisms at each SNP position; blue indicates the common Group 1 SNP allele; yellow indicates a variant SNP allele; and red highlights the presence of the A allele for long grain at the C165A SNP. The number and subpopulation identity of varieties carrying each haplotype are indicated in columns to the right. (B) Extended haplotype (EH) table corresponding to a 66.2-kb region flanking the *GS3* gene; the position of the C165A mutation within the extended haplotype is indicated in red; colors are the same as described in A above; stippled yellow squares indicate the presence of a third allele; and SNP names and primers used to detect them (Table S1) are shown along the *x*-axis at the bottom.

long grain arose only once in the *Japonica* gene pool or in a *Japonica*-like ancestor and was disseminated through introgression into the *Indica* gene pool during the process of rice domestication.

Origin of the C165A mutation within the *Japonica* varietal group: To determine the origin of the A allele within the *Japonica* varietal group, we examined polymorphism data in regions flanking *GS3* from wild-type (C-allele) Jap_EH accessions, looking for alleles that could differentiate between the *temperate japonica*, *tropical japonica*, and *Group V* subpopulations.

An informative SNP that distinguished all 8 ancestral (C-allele) *Group V* accessions from the other two *Japonica* subpopulations was identified 11 kb upstream of *GS3* (Figure 3B). When this polymorphism was assayed in the 211 diverse accessions, it was not found in any other subpopulations of *O. sativa* or in any of the wild accessions in our study. None of the varieties carrying the A allele, even the *Group V* accessions, contained this polymorphism. Therefore, we concluded that the C165A mutation did not originate in the *Group V* subpopulation, but must have been introgressed from a different *Japonica* ancestor. We were unable to determine whether the mutation originated in *temperate* or *tropical japonica* due to the lack of polymorphism in the *GS3* region that could distinguish these subpopulations.

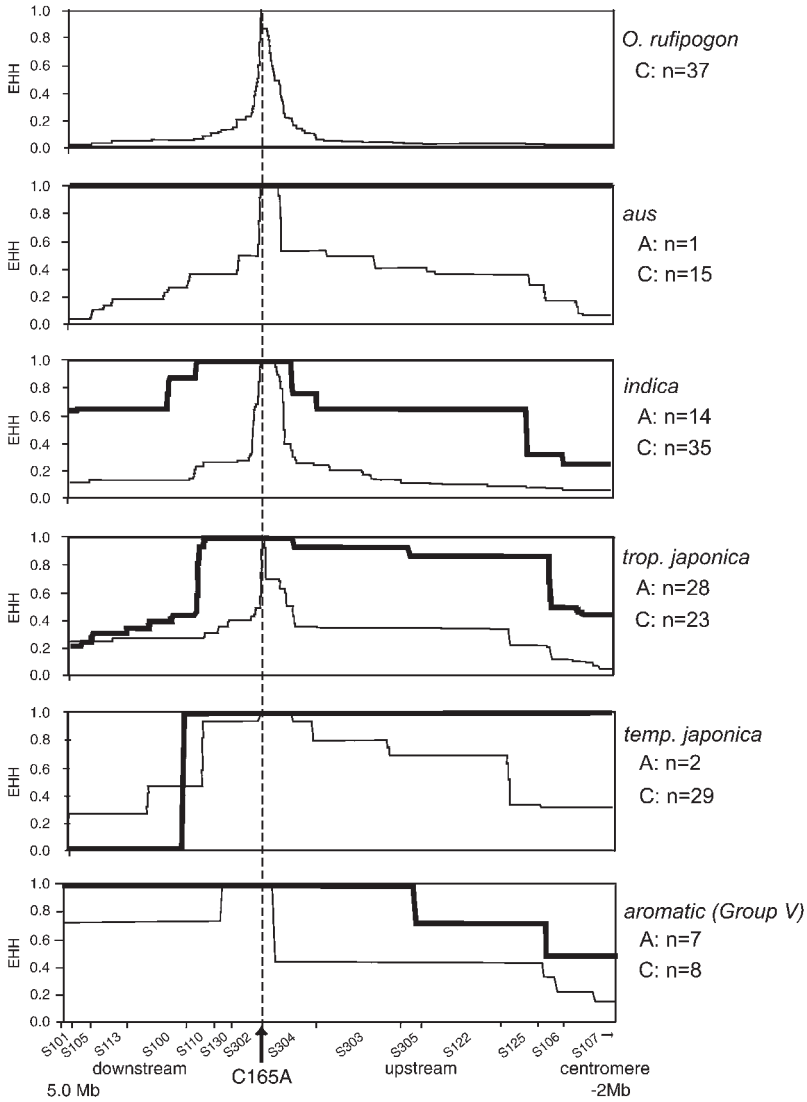

FIGURE 4.—EHH across the *GS3* genomic region for *O. rufipogon* and the five varietal groups of *O. sativa*. EHH values for individuals containing the C allele (short grain) or the A allele (long grain) at *GS3* are indicated by thin or thick lines, respectively. The arrow indicates the relative position of the C165A mutation; S designations along the *x*-axis indicate primers (Table S1) used to detect polymorphisms; and tick marks along the *x*-axis indicate correspondence between polymorphisms and amplicons.

Evidence for selection at *GS3*: *O. sativa* accessions carrying the C165A mutation had a nucleotide diversity (θ_π) of 0.0002 across the 6.57 kb of *GS3* sequence compared to $\theta_\pi = 0.00464$ in wild-type C-allele accessions. This 95.7% reduction in diversity was consistent with positive selection on the A allele at this locus. We next investigated the extent of linkage disequilibrium (LD) at *GS3* associated with A- and C-allele accessions in the different varietal groups. In *O. rufipogon,* we observed rapid, symmetrical decay of LD around *GS3* in wild-type C-allele accessions, indicating a lack of selection at *GS3* in the wild progenitor (Figure 4). A similar pattern was observed for C-allele accessions in the *aus*, *indica*, and *tropical japonica* subpopulations of *O. sativa*, with slightly slower LD decay in *temperate japonica* and in *Group V* accessions (consistent with a more intense domestication bottleneck in the latter two groups). In contrast, EHH extended over a larger region in all A-allele accessions, indicating an extended region of LD around *GS3* in these accessions (Figure 4). The patterns of LD and the marked reduction in θ_π observed in all A-allele accessions are indicative of strong positive selection for the derived allele conferring long grain at *GS3*. It is noteworthy that the patterns of EHH for A-allele accessions in *indica* and *tropical japonica* were almost identical, suggesting a very similar selection regime in these two groups.

Gene flow between *O. sativa* and *O. rufipogon* at *GS3*: Seven accessions of *O. rufipogon* carried the homozygous C165A mutation at *GS3*. To determine whether these wild accessions represented A-allele ancestors, or whether the A-allele had been transmitted as an introgression from an *O. sativa* cultivar, we first compared the nucleotide diversity (θ_π) of the seven A-allele and seven randomly selected C-allele accessions of *O. rufipogon* on the basis of concatenated sequences from the 66.2-kb region around the *GS3* gene. The A-allele accessions had $\theta_\pi = 0$ while the C-allele accessions had $\theta_\pi = 0.00883$. This is similar to the reduction of genetic diversity (θ_π) in A-allele *vs.* C-allele accessions of *O. sativa* in the *GS3* gene region (0.0002 *vs.* 0.00464, respectively). Further, when the A-allele haplotypes of

O. sativa (EH31–32) were compared with A-allele haplotypes found in *O. rufipogon*, all seven wild accessions contained identical EH31 haplotypes. If the wild materials harbored an ancestral, predomestication version of the *GS3* A allele, we would expect to see greater diversity around *GS3* in wild A-allele compared to cultivated A-allele accessions. However, because no sequence polymorphism was found in the regions flanking *GS3* in A-allele accessions of *O. rufipogon*, and because all of the wild accessions harbored the same EH31 haplotype found in *O. sativa*, we concluded that the A alleles in these wild accessions were the result of recent introgression events from *O. sativa* to *O. rufipogon*.

DISCUSSION

***GS3* and the domestication process:** Unlike maize (Doebley *et al.* 1994) and wheat (Dubcovsky and Dvorak 2007), rice domestication was not accompanied by a unidirectional increase in seed size, but rather by increased size variation, with artificial selection for seeds that were both longer and shorter and fatter and thinner than those of its wild progenitor, *O. rufipogon* (Morishima *et al.* 1992). The fact that the A allele had no effect on grain length in *O. rufipogon* suggests that it was not selected early in the domestication process. Other mutations were necessary before the phenotypic effect of the A allele could be observed and we infer that these other mutations accumulated in *O. sativa* prior to selection on the *GS3* locus. As such, the C165A allele can be considered a "diversification allele" because it would have enhanced the variation observed in *O. sativa* and contributed to the differentiation of the subpopulations within the cultivated gene pool. Evidence from this study suggests that once it attracted the attention of humans, it became a target of artificial selection and was introgressed from a *Japonica* ancestor(s) into the *Indica* gene pool. Interactions between the A allele and diverse factors that distinguish the genetic backgrounds of the *Indica* and *Japonica* varietal groups would have generated novel grain morphologies and expanded the range of variation observed in *O. sativa*.

Phenotypic impact of the A allele within the *Japonica* and *Indica* varietal groups: In this study, the A allele at *GS3* contributes significantly to grain length in *tropical japonica*, *indica*, and *Group V* varieties, and it confers unique grain morphologies in each of the subpopulations. In the *indica* and *Group V* backgrounds, the A allele gives rise to long, slender grains that differ in appearance from the long, bold grains of *tropical japonica*. It will be of interest to determine whether the same genes confer slender grain in the *indica* and the *Group V* genetic backgrounds, or whether different alleles are responsible for grain width in these divergent subpopulations.

Gene flow from cultivar to wild: The C165A mutation is not found in the African cultivated species, *O. glaberrima*, or in any of the wild species examined, except where it is associated with recent gene flow from *O. sativa*. Its presence in a few accessions of *O. rufipogon* is similar to the situation reported for the nonshattering allele, *sh4* (Li *et al.* 2006) and the *badh2.1* allele [M. Kovach (Cornell University), personal communication), except for the fact that the A allele at *GS3* had no discernible phenotypic effect in the wild material.

Origin within the *Japonica* and *Indica* varietal groups: The history of the C165A mutation in *GS3* suggests that it arose in a *Japonica* ancestor and moved into the *Indica* varietal group through introgressive hybridization. This pattern is reminiscent of the *rc* mutation for white pericarp and the Wx^b mutation for glutinous rice that also arose in *Japonica* and became widely disseminated in the *Indica* gene pool over the course of rice domestication (Yamanaka *et al.* 2004; Kovach *et al.* 2007; Sweeney *et al.* 2007).

In the case of *GS3*, the A allele for long grain is associated with strong positive artificial selection in *tropical japonica*, where it attained the highest allele frequency (61%) of any varietal group within *O. sativa*, while it is virtually absent from *temperate japonica*. This suggests that it is likely to have arisen in the *tropical japonica* group where its presence may be used as a marker to help distinguish *tropical japonica* varieties from their close relatives in the genetically narrower *temperate japonica* group.

Function of *GS3* and the genetic pathway in seed size control: Longer seed length in *O. sativa* is at least partially due to relaxed constraint on seed elongation mediated by the recessive C165A mutation in the *GS3* gene. Results from both RT–PCR and GUS expression in transgenic plants showed that mRNA expression of *GS3* begins during early panicle development (~3–5 cm), decreases when panicles are between 7 and 10 cm, and falls to below detection at flowering. In rice, panicles at the 3- to 5-cm stage are undergoing the early stages of inflorescence and ovule development, accompanied by the differentiation of glumes and floral organs. This period spans ovule primordial differentiation, integument primordial differentiation, division of integument primordium and meiosis of meristem mother cells (MMC), and early integument elongation (Itoh *et al.* 2005). Integument elongation begins in panicles that are 5–7 cm in length and finishes when panicles reach 7–10 cm (Itoh *et al.* 2005), which corresponds precisely to the time when expression of *GS3* starts to decline. This tissue- and stage-specific expression of *GS3* suggests that it regulates seed size through control of ovule development.

In rice, the A allele introduces a premature stop codon in the *GS3* gene prior to the VWFC domain (Fan *et al.* 2006). Our RNA analysis showed that the expression pattern of the A allele is similar to that of the wild allele, suggesting that the effect on phenotype is likely due to the truncation of the protein product itself,

rather than to any difference in gene expression. We hypothesize that the wild-type C allele at *GS3* functions as a dominant negative regulator of cell division and/or elongation in the integument.

The VWFC domain of the GS3 functional protein is reported to be important for protein–protein interaction and signaling (van Vlijmen *et al.* 2004; Zhang *et al.* 2007). Our genetic data demonstrate that the A allele of *GS3* masks the effects of other seed length genes in diverse accessions of *O. sativa* and that it interacts differentially with seed width genes in the different subpopulations, supporting the hypothesis that it affects seed morphology via interaction with other genes. Identifying the interacting partners of GS3 will allow identification of how this gene affects seed size differently in diverse genetic backgrounds.

Interaction between *GW2* and *GS3* does not explain subpopulation differences in seed size: Recently, a gene governing grain weight in rice, *GW2*, was cloned and found to encode a RING-type E3 ubiquitin ligase (Song *et al.* 2007). A 1-bp deletion resulting in a premature stop codon in the *GW2* gene was responsible for increasing seed size (Song *et al.* 2007). The deletion increased the number of spikelet hull cells, which increased hull size and enlarged the endosperm cell size in mature rice grains. To investigate whether there might be an interaction between *GS3* and *GW2* that would help explain why *GS3* affected grain size differently in different genetic backgrounds, we screened our germplasm panel to identify accessions that contained the 1-bp deletion in *GW2*. Results of this survey demonstrated that none of the varieties of either *O. sativa* or *O. rufipogon* contained the reported functional mutation in *GW2*. Thus, the subpopulation differences in grain size observed in *O. sativa* are not the result of interaction between *GS3* and *GW2*, but rather of that between *GS3* and other, as yet unidentified, genes. The absence of the *GW2* allele for large seed size in our panel suggests that either it is of very recent origin or it has been selected against by plant breeders and agriculturalists due to undesirable pleiotropic effects on grain quality. This is in direct contrast to the C165A mutation in *GS3* that appears to have played a significant role in rice domestication and is found widely distributed throughout the rice-growing world.

We thank Jiming Li for valuable advice and material used to fine map the *gw3.1* QTL, Han Nguyen for assistance with genotyping and phenotyping early in the project, and Lisa Polewsczak and Anna McClung at the U.S. Department of Agriculture–Agriculture Research Service Rice Research Unit in Beaumont, Texas for generating the Winseedle data on seed size. We thank Yukihiro Ito for providing the GUS vector and Rod Wing and Jose Luis Goicoechea from the Arizona Genomics Institute for providing BAC clones and helping sequence through the *GS3* region in *O. rufipogon*. We thank Lois Swales for help with formatting and administrative support and Michael Kovach for critical review, discussion, and substantial editing of the manuscript. This work was supported in part by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan to A.Y. (Integrated Research Project for Plant, Insect and Animal using Genome Technology MP1125) and grants from the Plant Genome Program of the National Science Foundation (awards 0606461 and 0110004 to S.M.).

LITERATURE CITED

Brondani, C., P. H. N. Rangel, R. P. V. Brondani and M. E. Ferreira, 2002 QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. Theor. Appl. Genet. **104:** 1192–1203.

Caicedo, A. L., S. H. Williamson, R. D. Hernandez, A. Boyko, A. Fledel-Alon *et al.*, 2007 Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet. **3:** e163.

Champagne, E. T., K. L. Bett, B. T. Vinyard, A. M. McClung, F. E. Barton *et al.*, 1999 Correlation between cooked rice texture and Rapid Visco Analyses measurements. Cereal Chem. **76:** 764–771.

Chaudhury, A. M., A. Koltunow, T. Payne, M. Luo, M. R. Tucker *et al.*, 2001 Control of early seed development. Annu. Rev. Cell Dev. Biol. **17:** 677.

Chen, W.-B., I. Nakamura, Y.-I. Sato and H. Nakai, 1993 Distribution of deletion type in cpDNA of cultivated and wild rice. Jpn. J. Genet. **68:** 597–603.

Dally, A. M., and G. Second, 1990 Chloroplast DNA diversity in wild and cultivated species of rice (genus Oryza, section Oryza). Cladistic-mutation and genetic-distance analysis. Theor. Appl. Genet. **80:** 209–222.

Dellaporta, S., T. Wood and T. Hicks, 1983 A plant DNA mini preparation: version II. Plant Mol. Biol. Rep. **1:** 19–21.

Doebley, J., A. Bacigalupo and A. Stec, 1994 Inheritance of kernel weight in 2 maize-teosinte hybrid populations—Implications for crop evolution. J. Hered. **85:** 191–195.

Dubcovsky, J., and J. Dvorak, 2007 Genome plasticity a key factor in the success of polyploid wheat under domestication. Science **316:** 1862–1866.

Fan, C. H., Y. Z. Xing, H. L. Mao, T. T. Lu, B. Han *et al.*, 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.

Fitzgerald, M., S. McCouch and R. D. Hall, 2009 Not just a grain of rice: the quest for quality. Trends Plant Sci. **14:** 133–139.

Fuse, T., T. Sasaki and M. Yano, 2001 Ti-plasmid vectors useful for functional analysis of rice genes. Plant Biotech. J. **18:** 219–222.

Garcia, D., V. Saingery, P. Chambrier, U. Mayer, G. Jurgens *et al.*, 2003 *Arabidopsis haiku* mutants reveal new controls of seed size by endosperm. Plant Physiol. **131:** 1661–1670.

Garris, A. J., T. H. Tai, J. Coburn, S. Kresovich and S. McCouch, 2005 Genetic structure and diversity in *Oryza sativa* L. Genetics **169:** 1631–1638.

Glaszmann, J. C., 1987 Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. **74:** 21–30.

Harlan, J. R., J. M. J. De Wet and E. G. Price, 1972 Comparative evolution of cereals. Evolution **27:** 311–325.

Ishii, T., T. Terachi and K. Tsunewaki, 1988 Restriction endonuclease analysis of chloroplast DNA from A-genome diploid species of rice. Jpn. J. Genet. **63:** 523–536.

Itoh, J., K. Nonomura, K. Ikeda, S. Yamaki, Y. Inukai *et al.*, 2005 Rice plant development: from zygote to spikelet. Plant Cell Physiol. **46:** 23–47.

Jofuku, K. D., P. K. Omidyar, Z. Gee and J. K. Okamuro, 2005 Control of seed mass and seed yield by the floral homeotic gene *APETALA2*. Proc. Natl. Acad. Sci. USA **102:** 3117–3122.

Juliano, B. O., 2003 *Rice Chemistry and Quality*. PhilRice, Munoz, Philippines.

Juliano, B. O., and C. P. Villareal, 1993 *Grain Quality Evaluation of World Rices*. International Rice Research Institute, Manila, Philippines.

Kovach, M., M. Sweeney and S. McCouch, 2007 New insights into the history of rice domestication. Trends Genet. **23:** 579–587.

Kubo, T., N. Takano-Kai and A. Yoshimura, 2001 RFLP mapping of genes for long kernel and awn on chromosome 3 in rice. Rice Genet. Newsl. **18:** 26–28.

Kubo, T., Y. Aida, K. Nakamura, H. Tsunematsu, K. Doi *et al.*, 2002 Reciprocal chromosome segment substitution series derived from Japonica and Indica cross of rice (*Oryza sativa* L.). Breed. Sci. **52:** 319–325.

Li, C., A. Zhou and T. Sang, 2006 Rice domestication by reducing shattering. Science **311:** 1936–1939.

Li, J. M., W. Thomson and S. R. McCouch, 2004 Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. Genetics **168:** 2187–2195.

Li, Y. H., L. Y. Zheng, F. Corke, C. Smith and M. W. Bevan, 2008 Control of final seed and organ size by the DA1 gene family in *Arabidopsis thaliana*. Genes Dev. **22:** 1331–1336.

Li, Z. K., S. R. M. Pinson, W. D. Park, A. H. Paterson and J. W. Stansel, 1997 Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics **145:** 453–465.

Ma, J., and J. Bennetzen, 2004 Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. USA **101:** 12404–12410.

Moncada, P., C. P. Martinez, J. Borrero, M. Chatel, H. Gauch *et al.*, 2001 Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. **102:** 41–52.

Morishima, H., Y. Sano and H. I. Oka, 1992 Evolutionary studies in cultivated rice and its wild relatives. Oxf. Surv. Evol. Biol. **8:** 135–184.

Nakano, M., A. Yoshimura and N. Iwata, 1992 Phylogeneticstudy of cultivated rice and its wild relatives by RFLP. Rice Genet. Newsl. **9:** 132–134.

Ni, J., A. Pujar, K. Youens-Clark, I. Yap, P. Jaiswal *et al.*, 2009 Gramene QTL database: development, content and applications. Database: J. Biol. Databases Curation **2009:** bap005.

Ohto, M., R. L. Fischer, R. B. Goldberg, K. Nakamura and J. J. Harada, 2005 Control of seed mass by *APETALA2*. Proc. Natl. Acad. Sci. USA **102:** 3123–3128.

Oka, H. I., 1988 *Origin of Cultivated Rice*. Elsevier Science/Japan Scientific Societies Press, Tokyo.

Orsi, C. H., and S. Tanksley, 2009 Natural variation in an ABC transporter gene associated with seed size evolution in tomato series. PLoS Genet. **5:** e1000347.

Paterson, A. H., Y. R. Lin, Z. K. Li, K. F. Schertz, J. F. Doebley *et al.*, 1995 Convergent domestication of cereal crops by independent mutations at corresponding genetic-loci. Science **269:** 1714–1718.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics **155:** 945–959.

Redona, E. D., and D. J. Mackill, 1998 Quantitative trait locus analysis for rice panicle and grain characteristics. Theor. Appl. Genet. **96:** 957–963.

Roxrud, I., S. E. Lid, J. C. Fletcher, E. D. L. Schmidt and H.-G. Opsahl-Sorteberg, 2007 GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development. Plant Cell Physiol. **48:** 471–483.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Scheet, P., and M. Stephens, 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. **78:** 629–644.

Schruff, M. C., M. Spielman, S. Tiwari, S. Adams, N. Fenby *et al.*, 2006 The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. Development **133:** 251–261.

Second, G., 1985 Evolutionary relationships in the Sativa group of Oryza based on isozyme data. Genet. Sel. Evol. **17:** 89–114.

Semon, M., R. Nielsen, M. P. Jones and S. R. McCouch, 2005 The population structure of African cultivated rice *Oryza glaberrima* (Steud.): evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. Genetics **169:** 1639–1647.

Shomura, A., T. Izawa, K. Ebana, T. Ebitani, H. Kanegae *et al.*, 2008 Deletion in a gene associated with grain size increased yields during rice domestication. Nat. Genet. **40:** 1023–1028.

Song, X., W. Huang, M. Shi, M. Zhu and H. Lin, 2007 A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat. Genet. **39:** 623–630.

Sun, C., X. Wang, A. Yoshimura and K. Doi, 2002 Genetic differentiation for nuclear mitochondrial and chloroplast genomes in common wild rice (*O. rufipogon* Griff) and cultivated rice (*Oryza sativa* L.) Theor. Appl. Genet. **104:** 1335–1345.

Sundaresan, V., 2005 Control of seed size in plants. Proc. Natl. Acad. Sci. USA **102:** 17887–17888.

Sweeney, M., M. Thomson, Y.-G. Cho, Y.-J. Park, S. Williamson *et al.*, 2007 Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet. **3**(8): e133.

Takeda, T., Y. Suwa, M. Suzuki, H. Kitano, M. Ueguchi-Tanaka *et al.*, 2003 The *OsTB1* gene negatively regulates lateral branching in rice. Plant J. **33:** 513–520.

Temnykh, S., W. D. Park, N. Ayres, S. Cartinhour, N. Hauck *et al.*, 2000 Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100:** 697–712.

Thomson, M. J., T. H. Tai, A. M. McClung, X. H. Lai, M. E. Hinga *et al.*, 2003 Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor. Appl. Genet. **107:** 479–493.

Toki, S., 1997 Rapid and efficient *Agrobacterium*-mediated transformation in rice. Plant Mol. Biol. Rep. **15:** 16–21.

van Vlijmen, H. W. T., A. Gupta, L. S. Narasimhan and J. Singh, 2004 A novel database of disulfide patterns and its application to the discovery of distantly related homologs. J. Mol. Biol. **335:** 1083–1092.

Vitte, C., T. Ishii, F. Lamy, D. Brar and O. Panaud, 2004 Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol. Genet. Genomics **272:** 504–511.

Wang, E., J. Wang, X. Zhu, W. Hao, L. Wang *et al.*, 2008 Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat. Genet. **40:** 1370–1374.

Wang, Z., and S. Tanksley, 1989 Restriction fragment length polymorphism in *Oryza sativa* L. Genome **32:** 1113–1118.

Weng, J., S. Gu, X. Wan, H. Gao, T. Guo *et al.*, 2008 Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. Cell Res. **18:** 1199–1209.

Xiao, H., N. Jiang, E. Schaffner, E. J. Stockinger and E. van der Knaap, 2008 A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science **319:** 1527–1530.

Xiao, J. H., J. M. Li, S. Grandillo, S. N. Ahn, L. P. Yuan *et al.*, 1998 Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics **150:** 899–909.

Xing, Y. Z., Y. F. Tan, J. P. Hua, X. L. Sun, C. G. Xu *et al.*, 2002 Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor. Appl. Genet. **105:** 248–257.

Yamanaka, S., I. Nakamura, K. Watanabe and Y.-I. Sato, 2004 Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theor. Appl. Genet. **108:** 1200–1204.

Yoine, M., T. Nishii and K. Nakamura, 2006 Arabidopsis UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. Plant Cell Physiol. **47:** 572–580.

Yu, S. B., J. X. Li, C. G. Xu, Y. F. Tan, Y. J. Gao *et al.*, 1997 Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. USA **94:** 9226–9231.

Zhang, J. L., Y. Huang, L. Y. Qiu, J. Nickel and W. Sebald, 2007 von Willebrand factor type C domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms. J. Biol. Chem. **282:** 20002–20014.

Zhou, Y., X. Zhang, X. Kang, X. Zhao, X. Zhang *et al.*, 2009 SHORT HYPOCOTYL UNDER BLUE1 associates with MINISEED3 and HAIKU2 promoters in vivo to regulate Arabidopsis seed development. Plant Cell **21:** 106–117.

Zhu, Q., and S. Ge, 2005 Phylogenetic relationships among A-genome species of the genus Oryza revealed by intron sequences of four nuclear genes. New Phytol. **167:** 249–265.

Communicating editor: M. Kirst

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.109.103002/DC1

Evolutionary History of *GS3*, a Gene Conferring Grain Length in Rice

Noriko Takano-Kai, Hui Jiang, Takahiko Kubo, Megan Sweeney, Takashi Matsumoto, Hiroyuki Kanamori, Badri Padhukasahasram, Carlos Bustamante, Atsushi Yoshimura, Kazuyuki Doi and Susan McCouch

DOI: 10.1534/genetics.109.103002

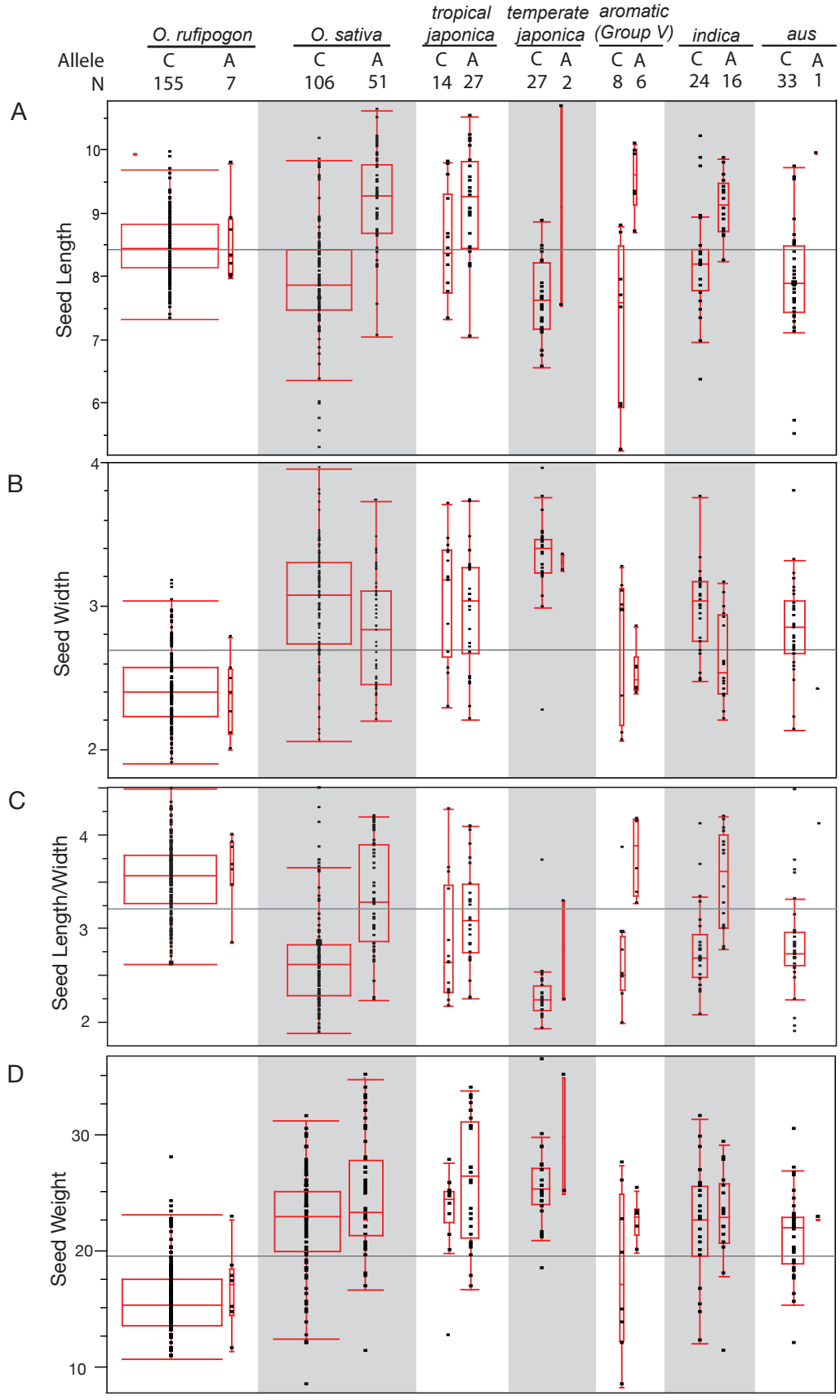

FIGURE S1.—Distribution of seed size and seed weight in *O. rufipogon* and five varietal groups of *O. sativa*.

TABLE S1

List of the primers used in this study

Table S1 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.103002/DC1.

TABLE S2

Germplasm accession information and relationship to phenotype and haplotype

Table S2 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.103002/DC1.

TABLE S3

Comparisons among *O. sativa* subpopulations for seed morphology between accessions carrying the A- versus C-allele at *GS3* (mean±SD) (A) seed width, (B), seed length/width ratio, (C) and seed weight

A

Comparison	C-allele			A-allele			R^2	F ratio	P[a]
	Seed width (mm)	N	Level[c]	Seed width (mm)	N	Level[c]			
tropical japonica	3.08 ±0.42	14	AB	2.98 ±0.41	27	A	0.01	0.52	0.47
temperate. japonica	3.35 ±0.29	27	A	3.30 ±0.08	2	AB	0.00	0.06	0.81
aromatic (Group V)	2.75 ±0.49	8	B	2.53 ±0.17	6	AB	0.08	1.03	0.33
indica	2.99 ±0.29	24	B	2.63 ±0.31	15	B	0.26	13.8	<0.001
aus	2.85 ±0.32	33	B	2.41 ±0.21	1	AB	0.05	1.85	0.18

B

Comparison	C-allele			A-allele			R^2	F ratio	P[a]
	SL/SW ratio	N	Level[c]	SL/SW ratio	N	Level[c]			
tropical japonica	2.83 ±0.65	14	A	3.14 ±0.53	27	A	0.07	2.76	0.1
temperate japonica	2.30 ±0.32	27	B	2.77 ±0.74	2	A	0.11	3.4	0.08
aromatic (Group V)	2.67 ±0.56	8	AB	3.79 ±0.41	6	A	0.58	16.76	<0.001
indica	2.78 ±0.45	24	A	3.52 ±0.51	15	A	0.38	23.51	<0.001
aus	2.82 ±0.51	33	A	4.12 ±0.43	1	A	0.17	6.34	0.017

C

Comparison	C-allele			A-allele			R^2	F ratio	P[a]
	Seed weight (g)[b]	N	Level[c]	Seed weight (g)[b]	N	Level[c]			
tropical japonica	23.31 ±3.65	14	AB	25.96 ±5.28	27	A	0.07	2.78	0.1
temperate japonica	25.30 ±3.38	27	A	29.76 ±7.01	2	A	0.1	2.87	0.1
aromatic (Group V)	17.84 ±6.91	8	C	22.53 ±1.70	6	A	0.17	2.6	0.13
indica	22.20 ±4.92	24	ABC	22.67 ±4.35	15	A	0	0.33	0.56
aus	21.29 ±3.64	33	BC	22.70 ±0.25	1	A	0	0.14	0.7

[a]Probability calculated using the F-statistics

[b]Calculated as seed weight/1000 seeds

[c]Multiple means comparisons based on Tukey-Kramer Honestly Significant Difference (HSD) test. Levels not connected by same letter are significantly different.

TABLE S4

Numbers of SNPs and indels within *GS3* based on sequencing of 54 accessions of *O. sativa*

Region	No. of SNPs	No. of indels
Coding region	2	1
Synonymous (*Ks*)	0	
Non-synonymous (*Ka*)	2	
In frame		1
Frame shift		0
Introns	72	21
5' UTR	12	6
3' UTR	0	0
Total	**86**	**28**