DOI: 10.1534/genetics.104.034165

Fine Mapping of a Grain-Weight Quantitative Trait Locus in the Pericentromeric Region of Rice Chromosome 3

Jiming Li, Michael Thomson[1] and Susan R. McCouch[2]

Department of Plant Breeding, Cornell University, Ithaca, New York 14853

Manuscript received July 30, 2004
Accepted for publication August 30, 2004

ABSTRACT

As the basis for fine mapping of a grain-weight QTL, *gw*3.1, a set of near isogenic lines (NILs), was developed from an *Oryza sativa,* cv. Jefferson × *O. rufipogon* (IRGC105491) population based on five generations of backcrossing and seven generations of selfing. Despite the use of an interspecific cross for mapping and the pericentromeric location of the QTL, we observed no suppression of recombination and have been able to narrow down the location of the gene underlying this QTL to a 93.8-kb region. The locus was associated with transgressive variation for grain size and grain weight in this population and features prominently in many other inter- and intraspecific crosses of rice. The phenotype was difficult to evaluate due to the large amount of variance in size and weight among grains on a panicle and between grains on primary and secondary panicles, underscoring the value of using multiple approaches to phenotyping, including extreme sampling and NIL group-mean comparisons. The fact that a QTL for kernel size has also been identified in a homeologous region of maize chromosome 1 suggests that this locus, in which the dominant *O. rufipogon* allele confers small seed size, may be associated with domestication in cereals.

RICE grain length and shape are important to consumers because they determine the physical appearance and affect the cooking quality of the grain. Seed and grain weight are important to farmers because they are among the most stable components of yield in rice. Furthermore, seed size or weight is important in the evolution of cereal crops because humans tended to select for large seed size during the early domestication process, as evidenced by the fact that most cultivated species have larger seeds than their wild relatives (Harlan 1992; Doganlar *et al.* 2000). Small seeds may be associated with reduced seedling vigor and difficult mechanical harvesting, which are problems for crop cultivation (Takeda 1991), but small seed is often favored under natural selection because it is frequently associated with a large number of seeds per plant, more rapid maturity, and wider geographic distribution.

Botanically, a rice *seed* consists of the brown rice grain (or kernel) plus the hull, while a *grain* (kernel) is a dehulled seed. However, the use of the terms "seed" and "grain" is often ambiguous in cereals and the terms are frequently used interchangeably. "Grain weight" and "seed weight" are highly correlated in rice and QTL associated with these traits are often identified in the same location along the chromosomes. Both grain and seed weight and grain and seed length are also highly heritable, making them useful characters for genetic analysis (Chauhan 1998).

Several independent studies in rice have identified QTL associated with grain weight, shape (length/width), or kernel elongation after cooking. For example, a QTL associated with grain weight or length has been reported in the centromere region of rice chromosome 3 in at least 10 different inter- and intraspecific populations: a Lemont (tropical *japonica*) × Teqing (*indica*) cross (Li *et al.* 1997); a Zhenshan 97 (*indica*) × Minghui 63 (*indica*) cross (Yu *et al.* 1997); a V20 (*indica*) × *Oryza rufipogon* cross (Xiao *et al.* 1998); a Labelle (*tropical japonica*) × Black Gora (*indica*) cross (Redoña and Mackill 1998); an Asominori (temperate *japonica*) × IR24 (*indica*) cross (Kubo T. 2001); a Zhenshan 97 (*indica*) × Minghui 63 (*indica*) cross (Xing *et al.* 2001, 2002); a Caiapo (tropical *japonica*) × *O. rufipogon* cross (Moncada *et al.* 2001); a BG90-2 (*indica*) × RS-16 (*Oryza glumaepatula*) cross (Brondani *et al.* 2002); a Jefferson (tropical *japonica*) × *O. rufipogon* cross (Thomson *et al.* 2003); and a V20 (*indica*) × *O. glaberrima* cross (Li 2004). These results suggest that the same QTL is expressed in different genetic backgrounds and environments, making it a valuable target for genetic analysis and also for further applications in rice breeding. Moreover, comparative mapping of this seed-weight QTL in maize suggests that a homologous gene determining seed weight or size may be associated with domestication and subsequent selection in different species (Doebley *et al.* 1994; Pat-

[1]*Present address:* Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development, Jl. Tentara Pelajar 3A, Bogor 16111, Indonesia.

[2]*Corresponding author:* Department of Plant Breeding and Genetics, 162 Emerson Hall, Cornell University, Ithaca, NY 14853. E-mail: srm4@cornell.edu

erson *et al.* 1995). While the positional convergence of these QTL is suggestive, proof of structural and functional conservation of orthologs across the grasses awaits the cloning and characterization of the genes underlying these QTL.

Map-based or positional cloning has been successful in isolating genes underlying QTL in several plant species, including rice (Yano *et al.* 2000; Takahashi *et al.* 2001; Kojima *et al.* 2002), barley (Buschges *et al.* 1997), wheat (Yan *et al.* 2002, 2003, 2004), and tomato (Frary *et al.* 2000; Fridman *et al.* 2000; Liu *et al.* 2002). Of fundamental importance to the positional approach is the development of large, segregating populations with informative recombinants in the target region. This study was undertaken to refine the position of a grain-weight QTL, *gw*3.1, mapped by Thomson *et al.* (2003) to an interval of 31.8 cM in the pericentromeric region of rice chromosome 3, and to develop a set of near-isogenic lines (NILs) that would provide the foundation for isolation of the gene underlying this QTL. We aimed to use the NILs to characterize the magnitude and behavior of the *O. rufipogon*-derived allele in a domesticated tropical *japonica* background. Further, because *gw*3.1 is adjacent to the centromere on rice chromosome 3, it was of interest to estimate the genetic:physical distance in the region and to determine whether a positional approach to gene isolation was likely to be successful.

MATERIALS AND METHODS

Plant materials: A BC_2F_2 population was constructed for QTL mapping as described by Thomson *et al.* (2003) using an *O. sativa* ssp. tropical *japonica* cultivar, Jefferson, as the recurrent parent and a wild accession of *O. rufipogon* (IRGC-105491 from Malaysia) as the donor parent. From this population, one BC_2F_2 family, C126-3, was selected as the starting material for fine mapping and NIL development of the QTL *gw*3.1 (Figure 1). This family was selected because it contained an *O. rufipogon* introgression in the target region and had significantly smaller grain weight than other BC_2F_2 individuals and relatively few nontarget background introgressions. NILs were developed by backcrossing to the Jefferson parent followed by selfing to eliminate nontarget genomic regions.

Segregating populations used for recombinant screening were grown for 3–4 weeks in 2-inch-wide × 7-inch-deep plastic pots in the Guterman greenhouse at Cornell University, Ithaca, New York. Once genotyping analysis identified informative recombinants, selected segregants along with parental controls were transferred to 5-inch clay pots to permit tillering, and seeds were harvested from a single primary panicle of mature plants for phenotyping.

Phenotypic evaluation: Seeds collected from primary panicles were dried for 3 days at 50°. Fifteen to 30 dry seeds were evaluated per line, depending on seed availability. Seed weight and length were evaluated after the removal of awns using an electronic scale (Mettler AE50 from RPBCal) and an electronic digital caliper (Traceable, Model 3415, Control, Houston), respectively. Seeds were dehulled using a manual dehuller and grains were evaluated as described for seeds. The weight of seeds (or grains) was converted to 1000-seed (or grain) weight for easy comparisons with previous studies. Seed or grain lengths were measured as the distance between opposing tips of a seed (after removing awns) or brown rice grain. Therefore, the following four phenotypic traits were evaluated for each NIL and control line: *seed weight* (SW, converted to 1000-seed weight in grams); *grain weight* (GW, converted to weight in grams for 1000 brown-rice grains); *seed length* (SL, measured in millimeters); and *grain length* (GL, measured in millimeters).

DNA extraction: DNA was extracted using a micro-isolation method as described by Cho *et al.* (1996) for samples that required multiple marker assays over time. For high-throughput recombinant screens where DNA was to be discarded following one or two rounds of marker evaluation (within 2–3 weeks of extraction), a Matrix Mill extraction method was used to obtain 1000–2000 ng of DNA or enough for 100–200 PCR reactions for SSR analysis (http://home.twcny.rr.com/htihome/protocol.htm), with the following modifications: Approximately 2 cm^2 of rice leaf were harvested into the 96-well plate with flat-bottom wells, which was kept on liquid nitrogen, followed by grinding these leaf tissues using a HyPure Seed Crusher (HSC-200 from Perkin-Elmer, Norwalk, CT). An alloy dowel pin (1/8 × 5/16 inches) was placed into each well of the plate before adding 120 μl of 0.5 M sodium hydroxide and homogenizing on the Matrix Mill machine for 3 min or cycles (with a change of the plate orientation each cycle). Then 10 μl of the sample extract was pipetted into 140–200 μl of 0.05 M Tris-HCl (pH 7.0, with 1 mM EDTA) for PCR analysis before cold storage in a −80° freezer.

Molecular marker analysis: The required density of molecular markers in the target region was achieved by using previously published SSRs (McCouch *et al.* 2002) as well as SSR and indel markers developed as part of this study. New markers were designed from publicly available rice genome sequence (http://rgp.dna.affrc.go.jp/, http://www.usricegenome.org/) and the likelihood of detecting polymorphism between the Jefferson and *O. rufipogon* parents was predicted by comparing sequence from the *japonica* cultivar, cv. Nipponbare (sequenced by the International Rice Genome Sequencing Project; http://rgp.dna.affrc.go.jp/), and the *indica* cultivar, cv. 93-11 (sequenced by the Beijing Genomics Institute; http://rise.genomics.org.cn/index.jsp). Indel markers were developed within annotated gene sequences, with primers anchored in conserved exons flanking the indel, which was most often detected in an intron. Primer sequences, map position, and amplified length of 14 newly developed SSRs and four indel markers used in this study are listed in Table 1.

Fine-mapping strategies: Two complementary strategies were used to identify informative recombinants in the F_2–F_4 generations of BC_4 and BC_5 populations. Using forward genetics, the extremes of the phenotypic distribution were genotyped at marker loci in the target region and haplotypes of the two groups were compared to identify informative recombination breakpoints. The use of only the phenotypic extremes allowed us to avoid problems associated with environmental variation and genetic background effects in early generation analysis. Using a reverse genetics approach, populations were genotyped with markers in the target region and each marker was then used to divide the population into three genotypic groups (Jefferson/Jefferson, Jefferson/*O. rufipogon*, and *O. rufipogon*/*O. rufipogon*) across a sliding window. The phenotypic means of these genotypic groups were then compared in F_3 and F_4 generations of BC_4–BC_5 NILs to determine the critical recombination breakpoints.

Data analysis: Phenotypic means were compared using both ANOVA and Tukey's multiple comparison using the software package Minitab (Release 13.1). Qgene was used for interval analysis (Nelson 1997) in the F_2 and F_3 generations of BC_4 and BC_5 NILs.

TABLE 1

SSR and indel markers developed to fine map the *gw*3.1-containing region

Marker reagent	Locus name	Marker type	Motif and length	BAC location	Product size in Nipponbare (bp)	Forward primer (5′–3′)	Reverse primer (5′–3′)	Annealing temperature
JL1	RM630	SSR	$(TTC)_{22}$	OSJNBb0058G04	160	GTTCTGAACAACCACCACCA	CTACCCGCCTGCAAGAATTA	55°
JL4	RM631	SSR	$(TA)_{32}$	OSJNBa0017N12	186	CAACAAACTTTGGAGGCACA	CTCCTCAACGCTGGGAATTA	55°
JL8	RM632	SSR	$(GA)_{16}$	OSJNBa0087G11	208	TCTTCAAGGCCTCTTCTACCA	CCCCCTTTTTGTTCACTTCC	55°
JL14	RM633	SSR	$(CT)_{13}$	OSJNBb0108E08	214	GGGGAACAAATTTGGAAGTG	AGGGGAGGTGAGGATCAGAG	55°
JL106	RM634	SSR	$(CGG)_8$	OSJNBa0090P23	379	GAGGATGTGCTCACCAGGAT	AGCAACATTTTGGGATGGAG	55°
JL107	RM635	SSR	$(TC)_{15}$	OSJNBb0041J20	203	ATTTGTGCAAGTGGCAGATG	GCTTGGCCTCAGAGGTTAGA	55°
JL109	RM636	SSR	$(AG)_{16}$	OSJNBa0002D18	202	TGGAGCTGTGGACTACTGGA	TCCCTGAGCCTACCTGTCAT	55°
JL114	RM637	SSR	$(TC)_{21}$	OSJNBb0056B16	223	GGCGTGTGCTCTTCGTGTAG	TTGGCAGCAGCAGTACTTTC	55°
JL118	RM638	SSR	$(GA)_{29}$	OSJNBa0083F15	191	CTTGCATCTCCCACCCTGT	AGTCCTCGCCTCTTTCTCCT	55°
JL120	RM639	SSR	$(TC)_7$	OSJNBa0002D18	189	AACCGACGCCTTTTCAAAC	GCCACCGTGCAAGATAGAAG	55°
JL123	RM640	SSR	$(GCG)_6$	OSJNBa0002D18	217	GGGTACAACGAGAGGGATGA	CAGCCCAGCATAATTTAGCC	55°
JL127	RM641	SSR	$(AG)_7$	OSJNBb0074M06	193	CACAATCGCAGCAGGAACTA	ACACTATACGCTCCCGGATG	55°
JL138	RM642	SSR	$(AC)_{10}$	OSJNBb0041J20	234	GGAACGGAGGGAGTACGTAAA	CATGTTCCATTGCTCGGTAT	55°
JL141	RM643	SSR	$(CGC)_6$	OSJNBb0041J20	206	CAGGCCCACCTTATGTTTTG	AGGGTGTACCGTTCCATGTC	55°
IND17	RID1	Indel		OSJNBb0074M06	233	CACTGTCGTCCCTTCAGCTT	ACGAACCTTCTCCACCTCCT	55°
IND31	RID2	Indel		OSJNBb0041J20	224	ACGGAGCTTACTCCCTTCGT	CGTCGAGCTTCCTCTCATCT	55°
IND42	RID3	Indel		OSJNBb0041J20	211	TGTCACCGTAGGTGTTGCTT	TCCACCCGTGTGAGTATCTTC	55°
IND51	RID4	Indel		OSJNBb0074M06	206	TTCCCGATAGTACCACAGACG	GGAAGGAAAGCACGGAGATA	55°

FIGURE 1.—Procedures of NIL development and fine mapping. P, phenotyping; G, genotyping; A, generation advance without genotyping or phenotyping.

RESULTS

Fine mapping of *gw*3.1: C126-3, the BC_2F_2 source line for NIL development, had small seeds and contained an *O. rufipogon* introgression across the entire 31.8-cM interval defining the original grain-weight QTL in the centromere region of chromosome 3. C126-3 also contained eight additional background introgressions located on 6 of the 12 chromosomes (chromosome 1, 2, 8, 9, 10, and 11). To rapidly eliminate the nontarget introgressions, C126-3 was backcrossed to Jefferson for one or two generations to generate BC_3 and BC_4 lines. In each case the lines were genotyped using markers flanking *gw*3.1 to ensure retention of the desired *O. rufipogon* introgression (Figure 1).

In BC_3F_2 and BC_4F_2 families derived from C126-3, the difference in grain weight between plants having an *O. rufipogon* introgression in the 31.8-cM target region and those having *O. sativa*, cv. Jefferson DNA was as large as 6.68 g (a 32.4% difference) in greenhouse experiments at Cornell (Figure 2). These results confirmed primary QTL observations based on field evaluations in Texas and Arkansas (THOMSON *et al.* 2003) and demonstrated that transgressive segregants could be identified in subsequent generations. The range of phenotypic variation observed in BC_3F_2/BC_4F_2 families was signi-

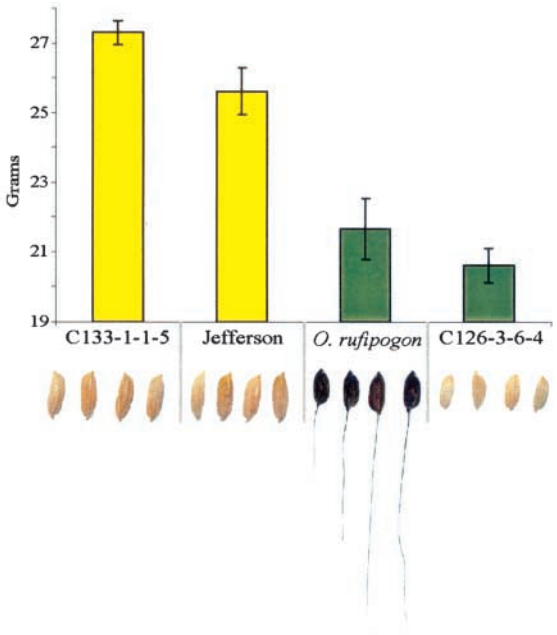

FIGURE 2.—Comparison of grain weights between two rice lines derived from the same BC_3 backcross generation, with parental controls. C133-1-1-5 is the BC_3F_2 individual with Jefferson DNA across the target region (between RM473D and RM135 as shown in Figure 3); C126-3-6-4 (NIL69) is the BC_3F_2 individual with an *O. rufipogon* introgression in the target region (between RM473D and RM135); *O. rufipogon* is the wild donor parent; Jefferson is the cultivated recurrent parent.

cantly greater than that observed between the *O. rufipogon* and Jefferson parental lines. This is illustrated in Figure 2 for BC_2F_2 segregants C133-1-1-5 (containing Jefferson DNA in the target region) and C126-3-6-4 (containing *O. rufipogon* DNA in the target region).

To identify useful recombinants for fine mapping, populations of 113 BC_3F_2 and 92 BC_4F_2 individuals were evaluated both phenotypically and genotypically using five markers in the region flanked by RM473D and RM135 (Figure 3). One BC_3F_2 individual, NIL69 (C126-3-6-4), and two BC_4F_2 individuals, NIL96 (C126-3-6-18-2) and NIL99 (C126-3-6-42-1), were selected for further population development because they each had an *O. rufipogon* introgression in the target region and low seed weight. Each line was selfed to promote recombination in the target region, backrossed to Jefferson to further reduce the number and size of unwanted background introgressions, and selfed again to generate segregating BC_4F_2 and BC_5F_2 populations (Figure 1).

A total of 1920 F_2 individuals derived from BC_4 and BC_5 plants were genotyped using 12 SSR markers in the target region and 173 recombinants were detected between RM1164 and RM6266 (Figure 3). These informative recombinants were evaluated for seed weight. Fine mapping confirmed a QTL peak between markers JL8 and RM3180 with a LOD score of 14.89 and an R^2 of 37.9% (Figure 3). This peak overlaps with the original BC_2F_2 mapping results reported by THOMSON *et al.* (2003).

To further refine the position of *gw*3.1, 65 of the 173 informative F_2 individuals were selfed to produce BC_4F_3 and BC_5F_3 seed. These F_3 materials were initially used to explore the dominance relationships among alleles at *gw*3.1. To do this, phenotypic means were compared among the three genotypic classes defined by the allele constitution of markers in the interval JL8–RM16 (Figure 4). The mean seed weight of the homozygous *O. rufipogon* class (17.06 g) was not significantly different from that of the heterozygous class (16.33 g), but both statistically differed from that of the large-seeded homozygous Jefferson class (20.11 g). These results demonstrated that the *O. rufipogon* allele was completely dominant over the Jefferson allele. The implication for this analysis was that the heterozygotes and the homozygous *O. rufipogon* class could be combined for phenotypic analysis during fine mapping, providing larger sample sizes and greater statistical power.

From the 65 F_3 families, a total of 532 BC_4F_3 and BC_5F_3 plants were evaluated for grain and seed weight and for grain and seed length. The same individuals were also evaluated for the 12 SSR markers between RM1164 and RM6266 used for fine mapping in the previous generation. QTL analysis in this generation confirmed the highly significant peak between markers JL8 and RM3180, with JL14 identified as the peak marker. LOD scores associated with the analysis of seed weight and grain length in the BC_4F_3 and BC_5F_3 generations were 21.63 and 34.87, respectively, with corresponding R^2 values of 41.4 and 58.5% for the two phenotypic characters, SW and GL (Figure 3).

In addition to the QTL analysis, an extreme sampling strategy was employed to clarify on which side of JL14 the *gw*3.1 locus was located. Of the 532 BC_4F_3 and BC_5F_3 plants, 159 that had either a homozygous *O. rufipogon* (73 individuals) or a heterozygous introgression (86 individuals) between JL8 and RM3180 were identified. Of these 159 plants, 39 individuals with low-extreme phenotypic values (25% of the population) and 39 individuals with high-extreme phenotypic values (25% of the population) were compared genotypically using nine of the markers in the target region. This comparison identified 13 individuals that fell into three different genotypic classes and were classified into the low-seed-weight group (genotypic classes 1–3, Figure 5A) and 9 individuals falling into the high-seed-weight group (genotypic classes 4–6) with 10 Jefferson parental genotypes used as a control (genotypic class 7; Figure 5A). Of greatest interest was the comparison between groups 3 and 4, where the genotypes were identical between markers JL14 and RM16, but the phenotypes were significantly different ($P < 0.01$). From these results we concluded that the location of *gw*3.1 had to lie in the

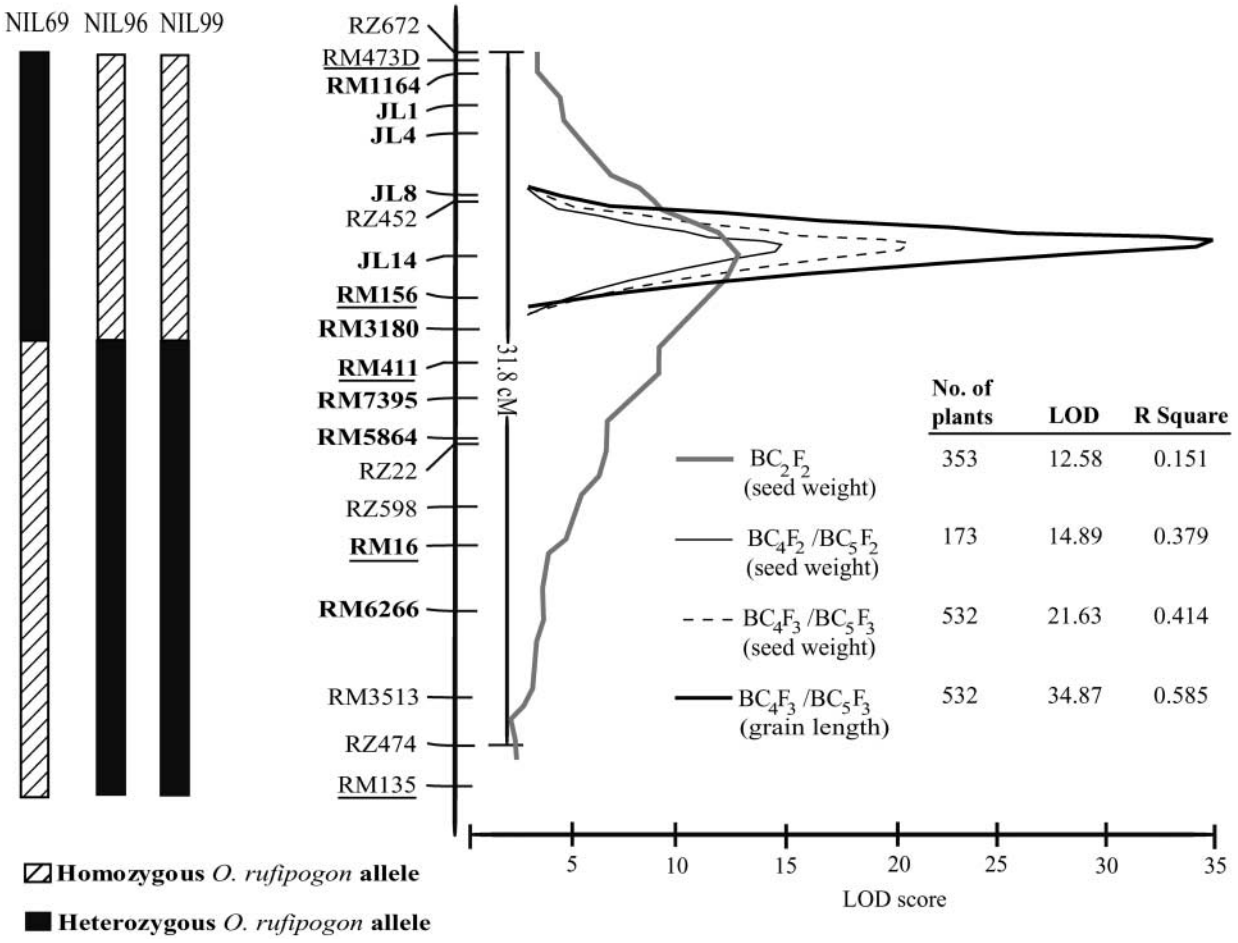

FIGURE 3.—Genotypes of three source NILs in the generation of BC_3F_2 and BC_4F_2 and QTL analysis of *gw*3.1 in the F_2 and F_3 generations of BC_4 and BC_5 populations. Underlined markers indicate the markers used in genotyping of three source NILs (NIL69, NIL96, and NIL99). Markers in boldface type indicate the markers used in QTL analysis in the F_2 and F_3 generations of BC_4 and BC_5 populations.

critical area of recombination between JL8 and JL14. This conclusion was further supported by the comparison between group 6 (large-seeded group) and group 2 (small-seeded group). Thus, we concluded that a gene(s) located in the interval between JL8 and JL14 controlled the performance of all four phenotypic traits.

To more precisely determine the location of the critical breakpoints, three new markers were added to further subdivide the JL8–JL14 interval (Figure 5B). All BC_4F_3 and BC_5F_3 plants with recombination breakpoints between JL8 and JL14 were genotyped with the new markers (JL106, JL107, and JL109) and, on the basis of this data, 24 recombinants were identified within the JL106–JL14 interval. By comparing the mean phenotypic performance of individuals with and without *O. rufipogon* DNA across each of the three marker intervals, we observed that group 3 (underlined in Figure 5B) was characterized by intermediate seed/grain size, compared to groups 1–2 (small seeds) and groups 4–5 (large seeds). We reasoned that group 3, with its intermediate mean seed/grain size, was likely to consist of individuals that had both large and small grains due to the fact that the critical recombination breakpoint(s) fell between JL107 and JL109. Therefore we narrowed down the location of *gw*3.1 to a 337-kb window between markers JL107 and JL109 (Figure 5B).

To further fine map *gw*3.1, 26 BC_4F_3 and BC_5F_3 individuals that had recombination breakpoints between JL107 and JL109 were selfed to produce F_4 populations. A total of 1700 BC_4F_4 and BC_5F_4 plants were genotyped using eight new markers (four SSRs and four indels) that were developed to further subdivide the 337-kb interval (Figure 5C). Seventy-six recombinants were identified and they were grouped into three genotypic classes consisting of 41, 23, and 12 individuals each. When phenotypic means for each class were compared to both *O. rufipogon* and Jefferson controls, the most informative comparison was between genotypic class 4 and Jefferson, both of which had large seed. The presence of *O. rufipogon* DNA in the critical area confers small seed size, so the fact that class 4 had an *O. rufipogon* introgression extending from marker JL107 to JL123 indicated that

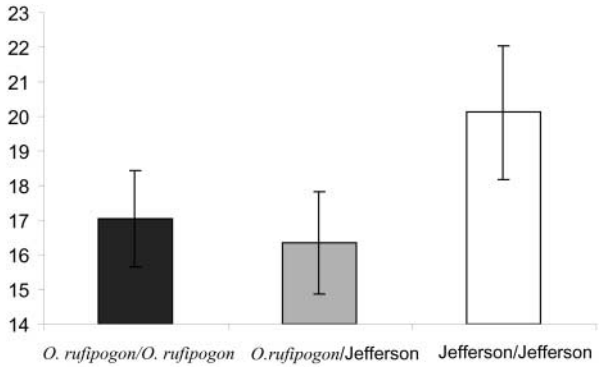

FIGURE 4.—Dominance of seed weight in the generations of BC_4F_2 and BC_5F_2. *O.rufipogon/ O.rufipogon* indicates the homozygous *O.rufipogon* genotype between JL8 and RM16; *O. rufipogon/*Jefferson indicates the heterozygous *O.rufipogon* genotype between JL8 and RM16; and Jefferson/Jefferson indicates the homozygous Jefferson genotype between JL8 and RM16.

*gw*3.1 must reside in the 93.8-kb region between JL123 and JL109. This is consistent with a higher R^2 value for seed/grain weight and length associated with JL109 (40–56%) compared to JL107 (20–25%) in the F_3 generation of BC_4 and BC_5 populations.

Candidate genes in the 93.8-kb region: On the basis of available sequence annotation (http://rgp.dna.affrc.go.jp/; http://www.usricegenome.org/), there are 14 predicted genes (5303.t0005–5303.t00018) in the 93.8-kb target region. Of these genes, 12 had unknown functions, and the 2 genes with functional annotation included (1) 5302.t00012 (TIGR_TU ID), a gene with a transcript length of 2061 bp and only 1 exon, which was classified as a putative transposase family *tnp2* gene; and (2) 5302.t00018, a gene of 3405 bp consisting of 8 exons and having a transcript length of 822 bp, which contained a putative EF-1 guanine nucleotide exchange domain. Gene 5302.t00008 consisted of 15 exons and had a transcript length of 2154 bp; it had no functional annotation but was of interest because many cDNAs corresponding to this hypothetical protein were identified in rice, maize, and sorghum (http://www.gramene.org). There were no cDNAs corresponding to any of the other genes in the region.

No recombination suppression observed in pericentromeric region of rice chromosome 3: To determine the position of *gw*3.1 in relation to the genetically defined centromere on chromosome 3, we compared the positions of molecular markers used to map the centromere on chromosome 3 and those used to map *gw*3.1. From this analysis, it was determined that *gw*3.1 lies in the interval flanked by S14055-RZ394 on the short arm and S2274-RZ576 on the long arm of chromsome 3, which positions it within the genetically defined centromere (SINGH *et al.* 1996; HARUSHIMA *et al.* 1998).

To determine how close *gw*3.1 is to the physically defined centromere, we analyzed ~6 Mb of rice genome sequence in the target region to identify the location(s) of CentO sequences (156 bp) that are considered diagnostic of centromeres (CHENG *et al.* 2002). This region spanned a genetic distance of 18 cM and contained three gaps in the genomic sequence. It was determined that *gw*3.1 was located in the interval 16,221,231–16,366,892 bp on The Institute for Genome Research pseudomolecule (76.6–83.0 cM on the genetic map, http://www.gramene.org). The nearest CentO sequence was ~2.46 Mb away. The gaps may harbor additional centromere repeats and until they are closed, the exact physical location of the centromere on rice chromosome 3 cannot be confirmed.

To determine the ratio of genetic to physical distance in the target region, we evaluated a population of 150 F_2 individuals derived from an O. *rufipogon* × *O. sativa*, cv. Jefferson cross. This population had a map distance of 5.3 cM between markers JL14 and RM3180. In addition, 750 randomly selected BC_4F_2/BC_5F_2 individuals were genotyped for the same markers, and 45 (or 6%) were recombinant between JL14 and RM3180. Results from both mapping populations were in agreement, with the genetic distance of 5–6 cM between these two markers. On the basis of the most recent sequencing information for BAC/PAC clones on rice chromosome 3 (http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/irgspMini.pl?chr=3), the physical distance between these two markers was estimated to be 1440 kb. Using the estimate of genetic distance as 6 cM, the ratio of genetic-to-physical distance in the region of interest is calculated as 1 cM:240 kb. This is consistent with the genome-wide average for genetic-to-physical distance in rice as reported by ARUMUGANATHAN and EARLE (1991), NAKAMURA *et al.* (1997), and HARUSHIMA *et al.* (1998), but it is unexpected on the basis of reports of suppression of recombination in interspecific crosses (CAUSSE *et al.* 1994) and, more specifically, in centromere regions (COPENHAVER *et al.* 1999; WU *et al.* 2002).

DISCUSSION

A QTL associated with grain weight located in the pericentromeric region of rice chromosome 3 has been reported in numerous interspecific crosses in rice (XIAO *et al.* 1998; MONCADA *et al.* 2001; BRONDANI *et al.* 2002; THOMSON *et al.* 2003; LI 2004). In all cases, the wild parent was associated with a dominant allele for small grain weight. The effect was consistent across diverse *indica* and *japonica* cultivars and these results suggest that a recessive mutation leading to larger grain size may have been part of the domestication syndrome in rice. While this is only one of the many loci contributing to this trait, it consistently explains the largest proportion of phenotypic variance for grain weight in diverse

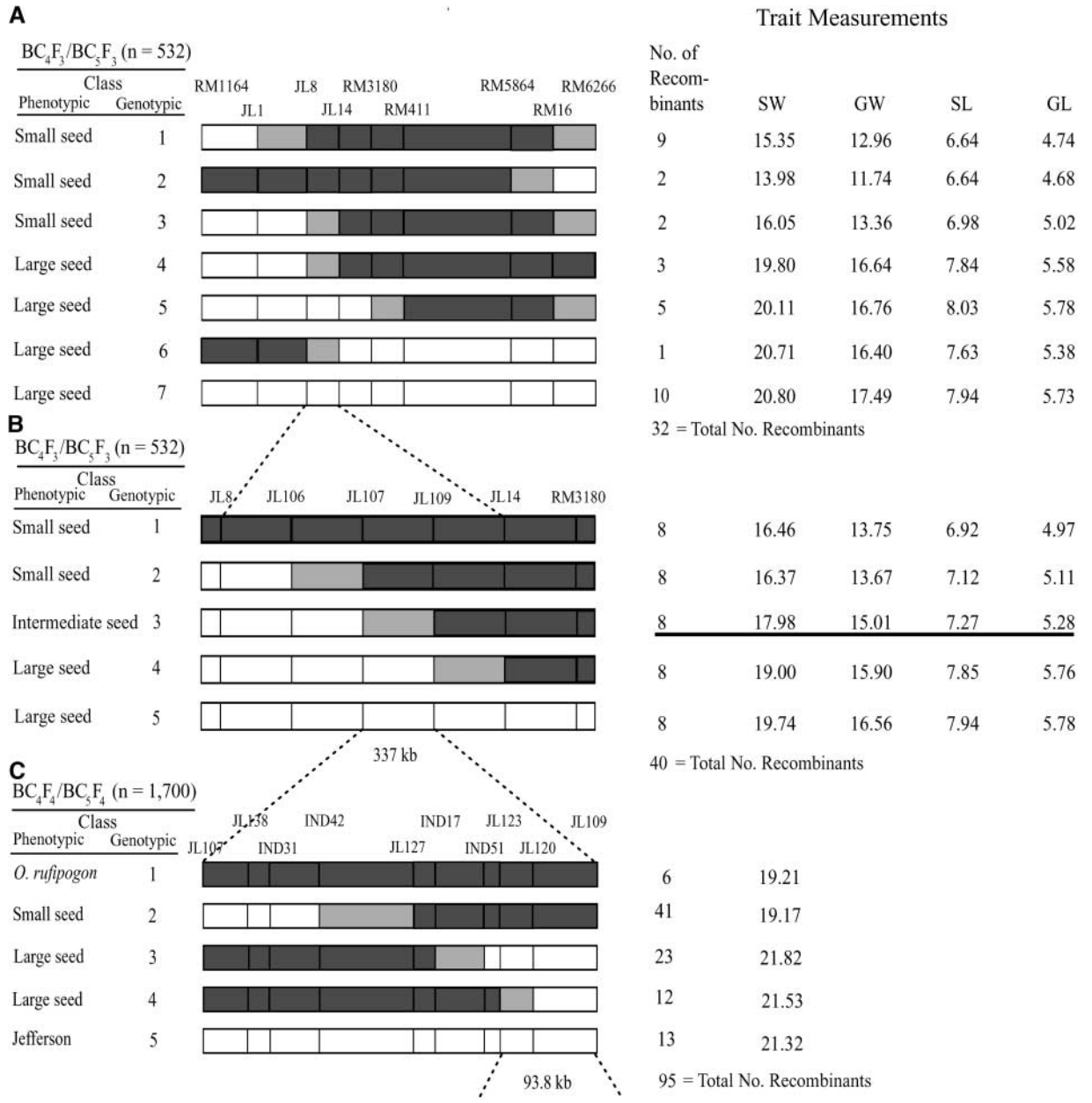

FIGURE 5.—Fine mapping of *gw3.1* in F_3 and F_4 generations of BC_4 and BC_5 populations. Darkly shaded rectangles, the homozygous or heterozygous *O. rufipogon* genotype; open rectangles, the homozygous Jefferson genotype; lightly shaded rectangles, the marker intervals containing recombination breakpoints. SW, seed weight (in grams); GW, grain weight (in grams); SL, seed length (in millimeters); GL, grain length (in millimeters). (A) BC_4F_3/BC_5F_3 generation: population size, $n = 532$; dotted lines indicate fine-mapped *gw3.1* location between JL8 and JL14. (B) BC_4F_3/BC_5F_3 generation: population size, $n = 532$; trait measurements of the intermediate seeded group are underlined; dotted lines indicate fine-mapped *gw3.1* location equivalent to a 337-kb region between JL107 and JL109. (C) BC_4F_4/BC_5F_4 generation: population size, $n = 1700$; dotted lines indicate fine-mapped *gw3.1* location equivalent to a 93.8 kb between JL123 and JL109.

rice studies. Previous study identified a seed-size QTL in a putatively homeologous location near the centromere on maize chomosome 1 (DOEBLEY *et al.* 1994). While the resolution of the maize QTL was low, the positional correspondence raises the possibility that this locus may have been a target of selection associated with domestication in both rice and maize.

Comprehensive phenotyping coupled with selective genotyping was first proposed for linkage analysis in 1989 (LANDER and BOTSTEIN 1989) when the costs involved in genotyping far exceeded most phenotypic assessments. The approach was successfully employed for fine mapping of a QTL by DARVASI (1997) and for association analysis by RISCH and ZHANG (1995) and

Van Gestel *et al.* (2000). The strategy does not rely on the assumption of a normal distribution (Barrett 2002) and allows sampling of only 5–10% of the individuals at each phenotypic extreme, as discussed by Darvasi and Soller (1992) or Ayoub and Mather (2002). In this study, the selective sampling approach was modified to suit the needs of our fine-mapping effort such that recombinants in the target region were identified on the basis of comprehensive genotyping (which is now faster and cheaper than the phenotyping for some traits), followed by phenotyping of only the informative recombinants.

During pre-NIL analysis, groups of progeny with extreme phenotypes were compared to determine the location of the gene(s) underlying the *gw*3.1 QTL. This selective phenotyping helped reduce the "noise" associated with random environmental effects that are known to affect grain weight (the weight and length of seeds or grains are known to vary among individuals with the same genotype, between main and secondary tillers of the same plant, or even among seeds on different branches of the same panicle) or unknown genetic effects in the genetic background (*i.e.*, due to "ghost introgressions"), all of which make it difficult to reliably detect differences between genotypic classes. The construction of NILs serves to simplify or eliminate the effect of genetic background on the expression of the QTL, facilitating the fine mapping and isolation of the target gene(s). Sets of NILs with overlapping introgressions in the target region facilitate statistical analysis because each NIL need only be compared to the recurrent parent to determine whether it contains the critical QTL allele or not, allowing use of a simple experimental design that is much less affected by experimentwise error than designs involving large populations.

Thomson *et al.* (2003) reported that *O. rufipogon* introgressions in six QTL on chromosomes 2, 3, 9, 10, and 12 decreased grain weight (more exactly seed weight) whereas those in two QTL on chromosomes 1 and 5 increased seed weight. The fact that multiple QTL with opposing effects originally resided in the same parental line explains why "Mendelizing" the trait using NILs can give rise to lines that show a more extreme phenotype than either parent. Thus, one explanation for transgressive variation in the offspring of a diverse cross is that QTL alleles with similar effects that were dispersed among the parental lines can be combined into one genetic background in the progeny.

Centromeres are important in cell division and stable transmission of genetic information, but they are often associated with a depression of meiotic recombination both within centromeric sequences and within the flanking pericentromeric regions (van Daelen *et al.* 1993; Haupt *et al.* 2001). The observed suppression of recombination is known as the "centromere effect" or "spindle fiber effect" (Beadle 1932; Mather 1938). Recombination suppression was reported to be 6-fold higher in the *Tm-2a* region of the tomato (a centromeric region) compared to noncentromeric regions (Ganal *et al.* 1989), but it can be as high as 10- to 40-fold (Tanksley *et al.* 1992). In rice, the average kilobase/centimorgan ratio was estimated to be 250–300 kb/cM (Arumuganathan and Earle 1991; Nakamura *et al.* 1997; Harushima *et al.* 1998), while Wu *et al.* (2002) indicated that the ratio of physical-to-genetic distance in the centromere regions was 2740 kb/cM, or ~10 times higher than that throughout the rest of the genome.

In our study, the centromeric intervals on rice chromosome 3 were defined by molecular markers flanking the location of *gw*3.1 and yet we see no evidence of recombination suppression. This contradicts the general belief that positional cloning will not be feasible for the isolation of genes in centromeric or pericentromeric regions and is strongly supported by the relative ease with which we have identified highly informative recombinants in this interspecific combination using F_2–F_4 generations of BC_4 and BC_5 populations. On the other hand, the precise boundaries of the centromere on chromosome 3 remain unresolved and until all gaps are closed and the sequencing and annotation in this region is complete, we are unlikely to fully understand the centromere structure of rice chromosome 3. Cheng *et al.* (2002) reported a relatively low signal intensity when pachytene chromosome 3 was hybridized with CentO satellite and centromere-specific retrotransposon probes. These data suggest that chromosome 3 may be an exception among rice chromosomes with relatively little heterochromotin and little or no supression of recombination in the centromere region.

We are grateful to the National Science Foundation (NSF00-151/011-0004) and to RiceTec for financial support; to Ray Wu and Margaret Smith for reviewing this article; to Jiming Jiang, C. Robin Buell, and Shu Ouyang for helpful discussions regarding the position of the centromere; to Sandra Harrington and Fumio Onishi for technical assistance in the lab and greenhouse; and to Lois Swales for help with formatting.

LITERATURE CITED

Arumuganathan, K., and E. D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9:** 211–215.

Ayoub, M., and D. E. Mather, 2002 Effectiveness of selective genotyping for detection of quantitative trait loci: an analysis of grain and malt quality traits in three barley populations. Genome **45:** 1116–1124.

Barrett, J., 2002 Association studies, pp. 3–12 in *Quantitative Trait Loci—Methods and Protocols,* edited by N. J. Camp and A. Cox. Humana Press, Totowa, NJ.

Beadle, G. W., 1932 A possible influence of the spindle fibre on cross-over in Drosophila. Proc. Natl. Acad. Sci. USA **18:** 160–165.

Brondani, C., N. Rangel, V. Brondani and 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.

Buschges, R., K. Hollricher, R. Panstruga, G. Simons, M. Wolter *et al.*, 1997 The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell **88:** 695–705.

Causse, M. A., T. M. Fulton, Y. G. Cho, S. N. Ahn, J. Chunwongse *et al.*, 1994 Saturated molecular map of the rice genome based

on an interspecific backcross population. Genetics **138:** 1251–1274.

Chauhan, J.-S., 1998 Inheritance of grain weight, size and shape in rainfed rice (*Oryza sativa*). Indian J. Agric. Sci. **68:** 9–12.

Cheng, Z., F. Dong, T. Langdon, S. Ouyang, C. R. Buell *et al.*, 2002 Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14:** 1691–1704.

Cho, Y.-G., M.-W. Blair, O. Panaud and S.-R. McCouch, 1996 Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. Genome **39:** 373–378.

Copenhaver, G. P., K. Nickel, T. Kuromori, M. I. Benito, S. Kaul *et al.*, 1999 Genetic definition and sequence analysis of Arabidopsis centromeres. Science **286:** 2468–2474.

Darvasi, A., 1997 The effect of selective genotyping on QTL mapping accuracy. Mamm. Genome **8:** 67–68.

Darvasi, A., and M. Soller, 1992 Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor. Appl. Genet. **85:** 353–359.

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

Doganlar, S., A. Frary and S.-D. Tanksley, 2000 The genetic basis of seed-weight variation: tomato as a model system. Theor. Appl. Genet. **100:** 1267–1273.

Frary, A., T. C. Nesbitt, S. Grandillo, E. Knaap, B. Cong *et al.*, 2000 fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science **289:** 85–88.

Fridman, E., T. Pleban and D. Zamir, 2000 A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. USA **97:** 4718–4723.

Ganal, M. W., N. D. Young and S. D. Tanksley, 1989 Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the Tm-2a region of chromosome 9 in tomato. Mol. Gen. Genet. **215:** 395–400.

Harlan, J., 1992 *Crops & Man.* American Society of Agronomy/Crop Science Society of America, Madison, WI.

Harushima, Y., M. Yano, A. Shomura, M. Sato, T. Shimano *et al.*, 1998 A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics **148:** 479–494.

Haupt, W., T. C. Fischer, S. Winderl, P. Fransz and R. A. Torres-Ruiz, 2001 The centromere1 (CEN1) region of *Arabidopsis thaliana*: architecture and functional impact of chromatin. Plant J. **27:** 285–296.

Kojima, S., Y. Takahashi, Y. Kobayashi, L. Monna, T. Sasaki *et al.*, 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:** 1096–1105.

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.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Li, J., J. Xiao, S. Grandillo, L. Jiang, Y. Wan *et al.*, 2004 QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. Genome **47:** 697–704.

Li, Z., S. R. 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.

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

Mather, K., 1938 Crossing over and heterochromatin in the X chromosome of *Drosophila melanogaster*. Genetics **24:** 413–435.

McCouch, S. R., L. Teytelman, Y. Xu, K. B. Lobos, K. Clare *et al.*, 2002 Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res. **9:** 199–207.

Moncada, P., C.-P. Martinez, J. Borrero, M. Chatel, H. Gauch, Jr. *et al.*, 2001 Quantitative trait loci for yield and yield components in an Oryza sativa × Oryza rufipogon BC2F2 population evaluated in an upland environment. Theor. Appl. Genet. **102:** 41–52.

Nakamura, S., S. Asakawa, N. Ohmido, K. Fukui, N. Shimizu *et al.*, 1997 Construction of an 800-kb contig in the near-centromeric region of the rice blast resistance gene Pi-ta2 using a highly representative rice BAC library. Mol. Gen. Genet. **254:** 611–620.

Nelson, J.-C., 1997 QGENE: software for marker-based genomic analysis and breeding. Mol. Breed. **3:** 239–245.

Paterson, A.-H., Y.-R. Lin, Z. 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.

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

Risch, N., and H. Zhang, 1995 Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science **268:** 1584–1589.

Singh, K., T. Ishii, A. Parco, N. Huang, D. S. Brar *et al.*, 1996 Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). Proc. Natl. Acad. Sci. USA **93:** 6163–6168.

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

Takeda, K., 1991 *Rice Genetics II*, pp. 181–189. International Rice Research Institute, Manila, The Philippines.

Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. de Vicente, M. W. Bonierbale *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics **132:** 1141–1160.

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.

van Daelen, R. A., F. Gerbens, F. van Ruissen, J. Aarts, J. Hontelez *et al.*, 1993 Long-range physical maps of two loci (Aps-1 and GP79) flanking the root-knot nematode resistance gene (Mi) near the centromere of tomato chromosome 6. Plant Mol. Biol. **23:** 185–192.

Van Gestel, S., J. J. Houwing-Duistermaat, R. Adolfsson, C. M. van Duijn and C. Van Broeckhoven, 2000 Power of selective genotyping in genetic association analyses of quantitative traits. Behav. Genet. **30:** 141–146.

Wu, J., T. Maehara, T. Shimokawa, S. Yamamoto, C. Harada *et al.*, 2002 A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell **14:** 525–535.

Xiao, J., J. Li, S. Grandillo, S. N. Ahn, L. 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, C. G. Xu, J. P. Hua and X. L. Sun, 2001 Mapping quantitative trait loci for grain appearance traits of rice using a recombinant inbred line population. Acta Bot. Sin. **43:** 721–726.

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.

Yan, L., V. Echenique, C. Busso, P. SanMiguel, W. Ramakrishna *et al.*, 2002 Cereal genes similar to Snf2 define a new subfamily that includes human and mouse genes. Mol. Genet. Genomics **268:** 488–499.

Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima *et al.*, 2003 Positional cloning of the wheat vernalization gene VRN1. Proc. Natl. Acad. Sci. USA **100:** 6263–6268.

Yan, L., A. Loukoianov, A. Blechl, G. Tranquilli, W. Ramakrishna *et al.*, 2004 The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science **303:** 1640–1644.

Yano, M., Y. Katayose, M. Ashikari, U. Yamanouchi, L. Monna *et al.*, 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, 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.

Communicating editor: J. A. Birchler