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

Jiming Li, Michael Thomson\textsuperscript{1} and Susan R. McCouch\textsuperscript{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, \textit{gwa3.1}, a set of near isogenic lines (NILs), was developed from an \textit{Oryza sativa}, cv. Jefferson × \textit{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 homologous region of maize chromosome 1 suggests that this locus, in which the dominant \textit{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 \textit{japonica}) × Teqing (indica) cross (Li et al. 1997); a Zhenshan 97 (indica) × Minghui 63 (indica) cross (Yu et al. 1997); a V20 (indica) × \textit{Oryza rufipogon} cross (Xiao et al. 1998); a Labelle (tropical \textit{japonica}) × Black Gora (indica) cross (Redoña and Mackill 1998); an Asominori (temperate \textit{japonica}) × IR24 (indica) cross (Kubo T. 2001); a Zhenshan 97 (indica) × Minghui 63 (indica) cross (Xing et al. 2001, 2002); a Caiapo (tropical \textit{japonica}) × \textit{O. rufipogon} cross (Moncada et al. 2001); a BG90-2 (indica) × RS-16 (\textit{Oryza glumaepatula}) cross (Brondani et al. 2002); a Jefferson (tropical \textit{japonica}) × \textit{O. rufipogon} cross (Thomson et al. 2003); and a V20 (indica) × \textit{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 (Doebly et al. 1994; Pat-
er son 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, gw3,1, 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 gw3,1 is adjacent to the centromere on rice chromosome 3, it is 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 BC2F1 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 BC2F2 family, C126-3, was selected as the starting material for fine mapping and NIL development of the QTL gw3,1 (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 BC2F2 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°C. 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 5415, 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 2–5 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/ htipro/protocol.htm), with the following modifications: Approximately 2 cm3 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°C 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 F1–F3 generations of BC4 and BC5 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 F2 and F3 generations of BC4–BC5 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 F2 and F3 generations of BC4 and BC5 NILs.
RESULTS

Fine mapping of the grain-weight QTL

The procedures for NIL development and fine mapping are as follows:

1. **NIIL Development**
   - **Step 1:** Identify the grain-weight QTL (gw) using phenotyping.
   - **Step 2:** Select the BC2F2 source line (C126-3) for NIL development.
   - **Step 3:** Cross C126-3 with the recipient line (Jefferson) to generate BC3 and BC4 lines.

2. **Fine Mapping**
   - **Step 1:** Use markers flanking the gw region to identify the desired introgression in BC3F2 and BC4F2 families.
   - **Step 2:** Confirm the presence of the introgression by phenotyping.

3. **Phenotyping**
   - **Step 1:** Evaluate the BC3F2 and BC4F2 families for grain weight.
   - **Step 2:** Compare the mean grain weight of lines with and without the desired introgression.

4. **Summary**
   - The results confirmed the primary QTL observations and demonstrated transgressive segregants could be identified in subsequent generations.

---

**TABLE 1**

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

<table>
<thead>
<tr>
<th>Marker reagent name</th>
<th>Locus name</th>
<th>Marker type</th>
<th>Motif and length</th>
<th>BAC location</th>
<th>Product size in Nipponbare (bp)</th>
<th>Forward primer (5'-3')</th>
<th>Reverse primer (5'-3')</th>
<th>Annealing temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>JL1 RM630 SSR</td>
<td>(TTG)12</td>
<td>OSJNBb0058G04</td>
<td>160</td>
<td>GTTCTGAAACAGCCACACCA</td>
<td>CTACCCCGCCCTGCAAGAATT</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL4 RM631 SSR</td>
<td>(TA)_{12}</td>
<td>OSJNBa0017N12</td>
<td>186</td>
<td>CAACAACTTTGAGGACCA</td>
<td>CTCTCAACCGCTGGGAACATT</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL8 RM632 SSR</td>
<td>(GA)_{16}</td>
<td>OSJNBa0087G11</td>
<td>192</td>
<td>TCTTGAAGCCTCTGCTACCA</td>
<td>CCCCCGTTTGGTTCTACCTCC</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL14 RM633 SSR</td>
<td>(CT)_{13}</td>
<td>OSJNBa0105E08</td>
<td>214</td>
<td>GGGGAAACATTGTTAGGAT</td>
<td>AGGGGAGTGATGAGTTGAGA</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL106 RM634 SSR</td>
<td>(CGG)_{8}</td>
<td>OSJNBa0090P29</td>
<td>395</td>
<td>GAGGATGTCCTCACAGGAT</td>
<td>AGCCACATTCTGGGAGATTG</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL107 RM635 SSR</td>
<td>(TC)_{13}</td>
<td>OSJNBb0004J20</td>
<td>203</td>
<td>ATTTGCAAGTGAGGACAGT</td>
<td>GTTGGCCTGAGGTTGAGA</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL109 RM636 SSR</td>
<td>(AG)_{10}</td>
<td>OSJNBb0020D18</td>
<td>202</td>
<td>GGGGTTGTTTGGTTTATTG</td>
<td>TGGGACGAGGAGGACTTTC</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL114 RM637 SSR</td>
<td>(TC)_{11}</td>
<td>OSJNBb0056B16</td>
<td>223</td>
<td>GGCCGCGCCCTTTCAGAC</td>
<td>GACCGACGCCAGATGAGA</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL118 RM638 SSR</td>
<td>(GA)_{20}</td>
<td>OSJNBa0083F15</td>
<td>191</td>
<td>AATCCGATACCCAGCTCCTG</td>
<td>GCAGCCGACATTTTACCAC</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL119 RM639 SSR</td>
<td>(TC)_{12}</td>
<td>OSJNBb0004J20</td>
<td>189</td>
<td>GAGGAGGAGGAGGAGGAT</td>
<td>GCCACCGGCTGCAAGATGAG</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL123 RM640 SSR</td>
<td>(CGG)_{10}</td>
<td>OSJNBb0087G11</td>
<td>217</td>
<td>GGGTACGCGCTCTGCTTTT</td>
<td>AGGCGGTACGCGTTCTACATC</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL127 RM641 SSR</td>
<td>(AG)_{17}</td>
<td>OSJNBb0074M06</td>
<td>193</td>
<td>GACATCAGGAGGAGGAGG</td>
<td>GAGCGGACCTTATTTACCAC</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL138 RM642 SSR</td>
<td>(AC)_{10}</td>
<td>OSJNBb0004J20</td>
<td>234</td>
<td>GGGACACACTTATGTTTTT</td>
<td>AAGGCGGTACGCGTTCTACATC</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JL141 RM643 SSR</td>
<td>(CGG)_{10}</td>
<td>OSJNBb0004J20</td>
<td>206</td>
<td>GAGGCGGACCTTATTTACC</td>
<td>AGGCGGTACGCGTTCTACATC</td>
<td>55°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND17 RID1 Indel</td>
<td>OSJNBb0074M06</td>
<td>233</td>
<td>CACTGTCGCTTACGTG</td>
<td>TGGGACGAGGAGGACTTTC</td>
<td>55°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND31 RID2 Indel</td>
<td>OSJNBb0004J20</td>
<td>224</td>
<td>ACGGAGCTAGTTACGCTTGGT</td>
<td>TCGGAGGCTGCTCCTGATC</td>
<td>55°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND32 RID3 Indel</td>
<td>OSJNBb0004J20</td>
<td>211</td>
<td>GCGGACAGCGCTTACGCTTGGT</td>
<td>TCGGAGGCTGCTCCTGATC</td>
<td>55°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND51 RID4 Indel</td>
<td>OSJNBb0074M06</td>
<td>206</td>
<td>TTCCGAGATGCTACACAGACG</td>
<td>GAGAAGGAAGCACCGGAGATA</td>
<td>55°</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

*Figure 1.—Procedures of NIL development and fine mapping.*
O. rufipogon (containing Jefferson DNA in the target region) and C126-3-6-4, SW and GL (Figure 3). Of these 159 plants, 39 individuals with low-extreme values of 41.4 and 58.5% for the two phenotypic characters among the three genotypic classes de
gned both phenotypically and genotypically using Figure 2 for BC 2F2 segregants C133-1-1-5 (containing markers in the region RM473D and RM135); O. rufipogon (C126-3-6-4) is the BC3F2 individual with Jefferson parental lines. This is illustrated in Figure 3). One BC3F2 individual, NIL69 (C126-3-6-4), was selected for further population phenotypic values (25% of the population) and 39 indi-
cions, and selfed again to generate segregating BC 4F2 and BC 5F3 populations (Figure 1). A total of 1920 F 2 individuals derived from BC 4 and BC 5F3 seed. These F3 materials were initially used to explore the dominance relationships among alleles at gw3.1. 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 homo-
zgygs. The same individuals were also evaluated for grain and seed length in the BC 4F3 and BC 5F3 generations were 21.63 and 34.87, respectively, with corresponding $R^2$ values of 41.4 and 58.5% for the two phenotypic character,
cantly greater than that observed between the O. rufipogon and Jefferson parental lines. This is illustrated in Figure 2 for BC3F2 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, pop-
ulations of 113 BC3F2 and 92 BC4F2 individuals were evaluated both phenotypically and genotypically using five markers in the region flanked by RM473D and RM135 (Figure 3). One BC3F2 individual, NIL69 (C126-3-6-4), and two BC4F2 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, backcrossed to Jefferson to further reduce the number and size of unwanted background introgressions, and selfed again to generate segregating BC3F2 and BC4F2 populations (Figure 1).

A total of 1920 F 2 individuals derived from BC 3 and BC 4 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 BC3F2 and BC4F2 generations were 21.63 and 34.87, respectively, with corresponding $R^2$ values of 41.4 and 58.5% for the two phenotypic character.

In addition to the QTL analysis, an extreme sampling strategy was employed to clarify on which side of JL14 the gw3.1 locus was located. Of the 532 BC3F2 and BC4F2 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 compari-
son 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 signifi-
cantly different ($P < 0.01$). From these results we concluded that the location of gw3.1 had to lie in the
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$_4$F$_3$ and BC$_5$F$_3$ 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 _gw3.1_ to a 337-kb window between markers JL106, JL107, and JL109 (Figure 5B).

To further fine map _gw3.1_, 26 BC$_4$F$_3$ and BC$_5$F$_3$ individuals that had recombination breakpoints between JL107 and JL109 were selfed to produce F$_4$ populations. A total of 1700 BC$_4$F$_4$ and BC$_5$F$_4$ 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 3**—Genotypes of three source NILs in the generation of BC$_4$F$_2$ and BC$_5$F$_2$ and QTL analysis of _gw3.1_ 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.
Corresponding to this hypothetical protein were an annotation but was of interest because many cDNAs had a transcript length of 2154 bp; it had no functional domain. Gene 5302.t00008 consisted of 15 exons and 1994 bp, which positions it within the genetically defined centromere of rice chromosome 3.1 lies in that a recessive mutation leading to larger grain size and S2274-RZ576 on the long arm of chromosome 3, 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 cultivars and these results suggest that a recessive mutation leading to larger grain size may have been part of the domestication syndrome in rice. When 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

discussion

To determine how close gw3.1 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 gw3.1 was located in the interval 16,221,291–16,366,892 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 F2 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 BC4F2/BC5F2 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/irgspMinipl?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).

**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.t0018) 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 gw3.1 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 gw3.1. From this analysis, it was determined that gw3.1 lies in the interval flanked by S14055-RZ394 on the short arm and S2274-RZ576 on the long arm of chromosome 3, which positions it within the genetically defined centromere (Singh et al. 1996; Harushima et al. 1998).

**Figure 4.**—Dominance of seed weight in the generations of BC4F2 and BC5F2. *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.
Figure 5.—Fine mapping of gw3.1 in F$_1$ and F$_2$ 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$_4$F$_3$/BC$_5$F$_3$ generation: population size, $n = 532$; dotted lines indicate fine-mapped gw3.1 location between JL8 and JL14. (B) BC$_4$F$_3$/BC$_5$F$_3$ 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$_4$F$_4$/BC$_5$F$_4$ generation: population size, $n = 1700$; dotted lines indicate fine-mapped gw3.1 location equivalent to a 93.8 kb between JL123 and JL109.

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 gw3.1 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 regressions”), 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 experiment-wise 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 Daalen 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 (Arunuganathan 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 gw3.1 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 F2–F3 generations of BC, and BC2 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 heterochromatin and little or no supression of recombination in the centromere region.

We are grateful to the National Science Foundation (NSF00-515/011-0004) and to RiceTec for financial support; to Ray Wu and Margaret Smith for reviewing this article; to Jimming Jiang, C. Robin Buell, and Shu OuYang for helpful discussions regarding the position of the centromere; to Sandra Harrington and Fumio Oishi for technical assistance in the lab and greenhouse; and to Lois Scales for help with formatting.

LITERATURE CITED


Causse, M. A., T. M. Fuloton, Y. G. Gao, S. N. Arn, J. Chunwongse et al., 1994 Saturated molecular map of the rice genome based...


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) rice. Genome 47: 697–704.


