Bacterial Artificial Chromosome-Based Physical Map of the Rice Genome Constructed by Restriction Fingerprint Analysis

Quanzhou Tao,[*,1] **Yueh-Long Chang,**[*,1] **Jingzhao Wang,**[*,†] **Huaming Chen,**[*]
M. Nurul Islam-Faridi,[*] **Chantel Scheuring,**[*] **Bin Wang,**[†] **David M. Stelly**[*]
and Hong-Bin Zhang[*]

*[*]Department of Soil and Crop Sciences and Crop Biotechnology Center, Texas A&M University, College Station, TX 77843-2123 and
[†]Institute of Genetics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China*

ABSTRACT

Genome-wide physical mapping with bacteria-based large-insert clones (*e.g.*, BACs, PACs, and PBCs) promises to revolutionize genomics of large, complex genomes. To accelerate rice and other grass species genome research, we developed a genome-wide BAC-based map of the rice genome. The map consists of 298 BAC contigs and covers 419 Mb of the 430-Mb rice genome. Subsequent analysis indicated that the contigs constituting the map are accurate and reliable. Particularly important to proficiency were (1) a high-resolution, high-throughput DNA sequencing gel-based electrophoretic method for BAC fingerprinting, (2) the use of several complementary large-insert BAC libraries, and (3) computer-aided contig assembly. It has been demonstrated that the fingerprinting method is not significantly influenced by repeated sequences, genome size, and genome complexity. Use of several complementary libraries developed with different restriction enzymes minimized the "gaps" in the physical map. In contrast to previous estimates, a clonal coverage of 6.0–8.0 genome equivalents seems to be sufficient for development of a genome-wide physical map of ~95% genome coverage. This study indicates that genome-wide BAC-based physical maps can be developed quickly and economically for a variety of plant and animal species by restriction fingerprint analysis via DNA sequencing gel-based electrophoresis.

G ENOME-WIDE physical mapping using large-insert DNA clones is becoming the centerpiece of current genomics research of virtually all plant and animal species. Genome-wide physical maps provide essential platforms for large-scale genome sequencing, effective positional cloning, high-throughput expressed sequence tag (EST) physical mapping, and target DNA marker development. Bacteria-based large-insert clones, including bacterial artificial chromosomes (BACs; Shizuya *et al.* 1992), bacteriophage P1-derived artificial chromosomes (Iaonnou *et al.* 1994), and large-insert conventional plasmid-based clones (Tao and Zhang 1998), have provided desirable resources for genomics research because of their high stability, low chimerism, and facility for large-scale DNA purification (Zhang and Wing 1997). To develop physical maps from bacteria-based large-insert clones, several approaches have been developed and used (for review, see Zhang and Wu 2001). These include hybridization-based methods such as iterative hybridization (*e.g.*, Mozo *et al.* 1998, 1999; Zhu *et al.* 1999), restriction-based fingerprinting meth-

ods (Coulson *et al.* 1986; Gregory *et al.* 1997; Marra *et al.* 1997, 1999; Zhang and Wing 1997; Ding *et al.* 1999; Zhu *et al.* 1999; Hoskins *et al.* 2000; Y.-L. Chang, Q. Tao, C. Scheuring, K. Meksem and H.-B. Zhang, unpublished results), and integrated BAC end sequencing, fingerprinting, and genome sequencing methods (Venter *et al.* 1996; Mahairas *et al.* 1999). Since the restriction-based fingerprinting method is not significantly affected by repeated sequences as is the iterative hybridization method and is much more rapid and economical than the integrated sequencing and fingerprinting method, it promises to provide a powerful means for rapid development of genome-wide physical maps from bacteria-based large-insert random clones.

In the restriction fingerprinting approach, the restricted fragments of clonal DNA were fractionated on either agarose gels (Marra *et al.* 1997) or denaturing polyacrylamide DNA sequencing gels (Coulson *et al.* 1986; Gregory *et al.* 1997; Tait *et al.* 1997; Zhang and Wing 1997; Tao and Zhang 1998; Ding *et al.* 1999; Zhang and Wu 2001). In the DNA sequence electrophoresis-based restriction fingerprinting method, the restricted fragments of clones are end labeled with either a radioactive nucleotide (Coulson *et al.* 1986; Zhang and Wing 1997; Tao and Zhang 1998) or a fluorescent dideoxynucleotide (Gregory *et al.* 1997; Tait *et al.* 1997; Ding *et al.* 1999).

Corresponding author: Hong-Bin Zhang, Department of Soil and Crop Sciences and Crop Biotechnology Center, 2123 TAMUS, Texas A&M University, College Station, TX 77843-2123.
E-mail: hbz7049@pop.tamu.edu

[1] These authors contributed equally to this work.

Validity of the restriction fingerprinting approach was first demonstrated by the development of genome physical maps of *Saccharomyces cerevisiae* (Olson *et al.* 1986; Riles *et al.* 1993) and *Caenorhabditis elegans* (Coulson *et al.* 1986; Hodgkin *et al.* 1995) with cosmid or λ clones. Recently, BAC-based physical maps were developed for small genome species, *Arabidopsis thaliana* (130 Mb; Marra *et al.* 1999; Mozo *et al.* 1999), chromosome 7 of *Magnaporthe grisea* (4.2 Mb; Zhu *et al.* 1999), and the major autosomes (120 Mb) of *Drosophila melanogaster* (Hoskins *et al.* 2000) using integrated iterative or sequence-tagged site-based hybridization and agarose gel-based fingerprinting (Marra *et al.* 1997) methods. However, the use of the restriction fingerprinting approach for development of genome-wide physical maps of large, complex genomes remains to be investigated. Unlike physical mapping of the small genome species, the development of global physical maps of large, complex genomes must fingerprint and analyze a large number of clones. Therefore, a high-resolution, high-throughput restriction fingerprinting method is needed to generate physical maps of large, complex genomes from large-insert random clones. The DNA sequence electrophoresis-based fingerprinting method (Coulson *et al.* 1986; Gregory *et al.* 1997; Zhang and Wing 1997; Tao and Zhang 1998; Ding *et al.* 1999; Zhang and Wu 2001) is not only high in resolution (one nucleotide), which is several hundredfold higher than that of the agarose gel-based method (10–1000 bp; for review, see Zhang and Wu 2001), but also highly amenable to automation on automated DNA sequencers (Gregory *et al.* 1997; Ding *et al.* 1999) and to high throughput (Zhang and Wu 2001). Therefore, it should be suitable for genome-wide physical mapping of large, complex genomes from bacteria-based large-insert random clones. However, no genome-wide, BAC-based physical maps have been developed to date using the DNA sequence electrophoresis-based fingerprinting method. Demonstration of the feasibility and development of strategies for genome-wide physical mapping with BACs by this method will greatly enhance research of large, complex genomes. This result will also provide a basis of incorporating the newly developed capillary DNA automated sequencing technology into the fingerprinting method for genome-wide physical mapping of large, complex genomes with bacteria-based large-insert random clones.

Rice, *Oryza sativa* L., is considered to be a model species for genome research of monocotyledonous plant species because of its relative small genome size. It has a wealth of genetic and genomic resources and is well established in genetic transformation. Rice has a genome size of 430 Mb/1C (where 1C is the haploid genome; Arumuganathan and Earle 1991) in which about 70% of the DNA is repetitive. The genome of rice is >3.5-fold larger than those of *A. thaliana* (Lin *et al.* 1999) and the major autosomes of *D. melanogaster* (Hoskins *et al.* 2000) in size. Although a yeast artificial chromosome (YAC)-based physical map has been developed for rice by the Japan Rice Genome Program to facilitate rice genome research (Saji *et al.* 2001), it covers only 63% of the rice genome. In addition, YACs are limited in applications for extensive genome research because they are relatively unstable and high in chimerism and their DNA is difficult to purify. Efforts are also being made to develop BAC-based physical maps for rice (http://www.genome.clemson.edu; http://rgp.dna.affrc.go.jp); however, no genome-wide, BAC-based physical maps of the rice genome have been reported to date. Furthermore, all these efforts are working with japonica rice (cv. Nipponbare), which accounts for <10% of world rice production. In this study, we developed a genome-wide BAC-based physical map of indica rice, which accounts for >90% of world rice production, from three complementary large-insert BAC libraries, and demonstrated the feasibility of and developed strategies for genome-wide physical mapping with bacteria-based large-insert random clones using the DNA sequence electrophoresis-based fingerprinting method. Contig reliability of the physical map was verified using different approaches and the results indicate that the physical map is reliable and provides a readily used framework for genomics research of monocotyledonous plants. The results of this study have provided a paradigm for rapid development of genome-wide physical maps of plant and animal genomes from bacterial clone-based, large-insert random clones.

MATERIALS AND METHODS

BAC libraries and DNA markers: Three *O. sativa* ssp. indica cultivar Teqing BAC libraries were used to develop the BAC-based physical map of the rice genome because >90% of the world rice production is indica rice. The libraries were constructed in the *Hin*dIII site of pBeloBAC11 (Kim *et al.* 1996; Zhang *et al.* 1996), the *Bam*HI and *Eco*RI sites of pECBAC1 (Frijters *et al.* 1997; H.-B. Zhang, unpublished results), respectively, and have average insert sizes of 130, 150, and 147 kb, respectively. The vector pECBAC1 was derived from pBeloBAC11 by knocking out the *Eco*RI site in its chloramphenicol resistance gene, thus making the *Eco*RI site in the multiple cloning sites suitable for cloning. These BAC libraries are permanently maintained in 384-well microplates and publicly available at the GENE*finder* Genomic Resources (formerly, the Texas A&M BAC Center) (http://hbz.tamu.edu-BAC Library-Library List).

The DNA markers were selected from the Cornell University (Causse *et al.* 1994) and Japan Rice Genome Research Program (Harushima *et al.* 1998) rice genetic maps and kindly provided by S. McCouch and the Japan MAFF DNA Bank at the National Institute of Agrobiological Resources (http://bank.dna.affrc.go.jp). The random rice EST clones were kindly provided by Dupont Company (G.-H. Miao).

BAC fingerprinting and contig assembly: BAC clones maintained in a 384-well microplate were inoculated in four 96-deep well plates containing 1 ml LB medium plus 12.5 μg/ml chloramphenicol and grown at 37° with shaking at 250 rpm overnight. BAC DNA was isolated and purified in the 96-deep well plates and then in 8- or 12-microtube strips using a modified

alkaline lysis method (Q. Tao, Y.-L. Chang, B. Vinatzer and H.-B. Zhang, unpublished results). The DNA was double-digested with *Hin*dIII and *Hae*III, end labeled with [^{32}P]dATP using reverse transcriptase at 37° for 2 hr, and then subjected to 4.0% (w/v) polyacrylamide DNA sequencing gel electrophoresis at 85 W for ~100 min. The gel was dried and autoradiographed.

The fingerprints on the autoradiographs were scanned into image files using a UMAX Mirage D-16L scanner. The image of the fingerprints was size adjusted to 1.1 MB, transferred to a computer workstation (SUN Microsystems, Utra10), and edited using the Image 3.8 of the FPC (FingerPrinted Contig) package (Sulston *et al.* 1988; Soderlund *et al.* 1997). The fragments ranging from 58 to 673 bases were used in contig assembly, on average, 22 bands per BAC fingerprint. The bands derived from the BAC vectors (pBeloBAC11 and pEC-BAC1) were manually deleted from the image files, and the clones without inserts were excluded.

The BAC contigs of the rice genome were assembled from the fingerprint database using the FPC 3.4 of the FPC package (Soderlund *et al.* 1997) in two steps. We first assembled automated BAC contigs under highly stringent criteria (see below) to ensure that they are accurate. Then we joined automated contigs into larger contigs, using a less stringent criterion for the number of consensus bands (fewer common bands). When the fingerprints on the autoradiograph were scanned into image files, the original image size of each autoradiograph (35 × 43 cm) was 7.8 MB. To facilitate fingerprint analysis, we reduced the image size of each autoradiograph to 1.1 MB before transferring the image to the Image 3.8 of the FPC package at the computer workstation for data analysis. Soderlund *et al.* (1997) recommended that tolerance 7 be suitable to build contigs from the fingerprints fractionated on polyacrylamide DNA sequencing gels. In our case, tolerance 3 was selected for contig assembly, which was equivalent to tolerance 7 for the original size of the autoradiograph image.

To select the cutoff values suitable for contig assembly, we used three DNA probes, *adhA*, *psbA*, and *rbcL*, that are approximately 50 kb apart on the barley chloroplast genome to screen the source rice BAC libraries and obtained 615 positive clones. We supposed that all positive clones should be assembled into a single contig if the tolerance values and cutoff scores were properly selected for contig assembly. After a series of tests according to this criterion, tolerance = 3 and cutoff = 10^{-10}–10^{-18} were selected and used for the BAC physical map contig assembly. The other software parameters used were Diff = 0.3, MinBands = 5, Diffbury = 0.10, and Minends = 8. To achieve the best overlap, each contig was subjected to analysis at cutoff = 10^{-4} and then by running "Calculation," and "Again" until the best result was obtained.

Library screening: The rice BAC libraries or the BACs of the map contigs were double-spotted on Hybond N + membrane (Amersham, Piscataway, NJ) in a 3 × 3 format using the Biomek 2000 robotic workstation (Beckman, Fullerton, CA). The membranes were prepared following a published procedure (Zhang *et al.* 1996). To estimate the realized genome coverage of the rice BAC libraries, the filters of the rice cv. Teqing and Lemont *Hin*dIII BAC libraries (Zhang *et al.* 1996) were probed with 93 DNA markers selected from the rice genetic map (Causse *et al.* 1994). To identify the BACs derived from chloroplast DNA, the filters prepared from the rice physical map BACs were hybridized with three chloroplast DNA probes (see above). The colony hybridization was performed as described at http://hbz.tamu.edu. In the post-hybridization, the filters were washed for three times in 0.1% SDS, 0.5× SSC at 65°, 30 min each wash.

To test the reliability of the rice map BAC contigs, the filters of the rice physical map BACs were probed with 77 markers selected from linkages 8, 11, and 12 of the existing rice genetic maps (Causse *et al.* 1994; Harushima *et al.* 1998) and six random rice EST clones. Clone DNA was prepared by the conventional alkaline lysis method. The insert of each clone was released from its cloning vector by restriction enzyme digestion or PCR amplification using the DNA sequences immediately flanking the cloning site as primers. The insert DNA was purified with the GENECLEAN Kit according to its manufacturer (BIO 101, Vista, CA) and labeled with the Dig high primer labeling kit (Roche Molecular Biochemicals). The BACs on the filters were screened with row and column probe groups of the DNA markers, respectively, with nine DNA markers per probe group. The positive clones of each probe were identified by cross-hybridization between the column and row probe groups to the filters. The BAC clone filters were transferred into an appropriate amount of prewarmed Dig prehybridization buffer (5× SSC, 0.1% *N*-lauroylsarcosine, 0.02% SDS, and 1.0% blocking reagent) and incubated at 65° for 1 hr with gentle agitation. Then the hybridization was conducted by adding denatured Dig-labeled probes to the prewarmed hybridization buffer, mixing well, transferring the filters from the prehybridization buffer into the probe/hybridization buffer mixture, and incubating at 65° with gentle agitation overnight. The filters were washed in 2× SSC, 0.1% SDS for two times, 5 min each time, at room temperature, followed by two washes in 0.1× SSC, 0.1% SDS, 15 min each wash, at 65°. The hybridization signals were detected with the Detection Starter Kit II according to the manufacturer (Roche Molecular Biochemicals).

RESULTS

Development of a genome-wide BAC-based physical map of the rice genome: Bacteria-based large-insert clone libraries of truly high-genome coverage are of significance for genome-wide physical mapping by restriction fingerprint analysis. To develop a BAC-based physical map of the rice genome, we previously developed two large-insert rice BAC libraries, the Teqing *Hin*dIII and Lemont *Hin*dIII BAC libraries (Zhang *et al.* 1996). To test the true genome coverage of the libraries, we screened the Teqing *Hin*dIII BAC library with 97 mapped DNA markers. The Teqing *Hin*dIII BAC library has a theoretical genome coverage of 98% (4.4 × genome coverage; Zhang *et al.* 1996). Surprisingly, the result showed that only 83% of the DNA markers gave one or more positive BACs—there was a 15% difference between the theoretical and realized genome coverage. To further test the relationship between the number of clones in a BAC library and its true genome coverage, we screened the Lemont *Hin*dIII BAC library with the same set of the DNA markers. The Lemont *Hin*dIII BAC library has a theoretical genome coverage of 97% (2.6 × genome coverage; Zhang *et al.* 1996). The result was also surprising in that ~85% of the DNA markers gave one or more positive clones in at least one of these two rice *Hin*dIII BAC libraries. This result indicates that it is necessary to develop several individual source BAC libraries with different enzymes in order to develop a genome-wide physical map of a high-genome coverage. Therefore, we constructed two additional Teqing BAC
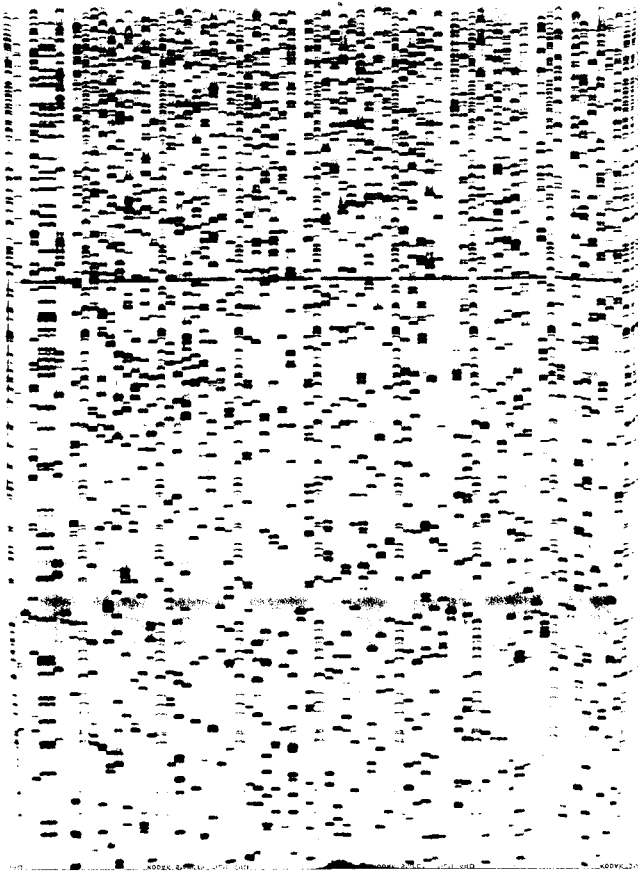
FIGURE 1.—Example of the autoradiographs of BAC fingerprints that were used for contig assembly of the rice BAC-based physical map. DNA markers (λ DNA/*Sau*3AI) were used in the first lane and every ninth lane thereafter. The fragments of BAC DNA were labeled with [^{32}P]dATP and the fragments of marker DNA were labeled with [^{33}P]dATP. The fingerprints were fractionated on a 4% (w/v) denaturing polyacrylamide DNA sequencing gel. The band appearing in all BAC lanes was derived from the BAC cloning vector pBeloBAC11 (Kim *et al.* 1996), which was manually deleted during fingerprint image editing.

libraries with *Bam*HI and *Eco*RI (H.-B. Zhang, unpublished results), respectively, to develop the genome-wide BAC-based physical map of the rice genome. The three rice cv. Teqing BAC libraries have average insert sizes of 130, 150, and 147 kb, respectively (see http://hbz. tamu.edu-BAC Library-Library List).

We used the DNA sequencing gel-based, radioactive nucleotide labeling method to generate BAC fingerprints (*e.g.*, see Figure 1). A total of 21,087 BACs, covering 6.9 × rice haploid genomes, were fingerprinted on 380 autoradiographs. Of these clones, 3.1 × genome BACs were randomly selected from the *Hin*dIII library, 1.7 × genome BACs from the *Eco*RI library, and 2.1 × genome BACs from the *Bam*HI library. The BAC fingerprints were scanned into image files, edited, and created into FPC database. The overlapping clones were assembled into contigs using the FPC program (Soderlund *et al.* 1997). From the BAC fingerprint database,

the FPC assembled 585 contigs, designated hereon as "automated contigs" (Table 1). With the FPC program, it was established that these 585 contigs encompassed 70,009 unique bands and each band, on average, represented a 6.3-kb fragment of a BAC clone. Therefore, the 585 contigs collectively cover 441 Mb in length. This collective physical length of the contigs is larger than the 430-Mb genome size of rice because most of the contigs are overlapped despite not being detected under the conditions used in the study. Of these automated contigs, the largest one (ctg13) contains 128 clones, encompassing 579 unique bands and spanning 3648 kb in length; 291 contigs contain 26 or more clones; 226 contain 10–25 clones; and 68 contain 5–9 clones. The contigs containing 4 or fewer clones were dismissed, and 1942 clones remained as singletons. We then manually analyzed every contig, extended the automated contigs with the End Extension program, and added the singletons to the contigs with the Singles Hit program of the FPC (Soderlund *et al.* 1997). We assumed that if two contig end clones between contigs had 10 or more bands in common, they were claimed as overlapped. Only after careful comparison of the contig end clones, were suspected overlapping contigs merged to form "extended contigs." As a result, the number of contigs was reduced to 298 contigs (Table 2), encompassing 66,589 unique bands and collectively covering 419 Mb in length. The largest contig (ctg3) contains 257 clones, encompassing 972 unique bands and spanning 6.1 Mb in length. Eight hundred ninety-six clones remained as singletons, each of which consisted of four or fewer bands that were insufficient to be included in contig assembly. Both the automated contigs and extended contigs are posted at http://hbz.tamu.edu-Physical Mapping-Indica Rice Map. Figure 2 shows an example of the automated BAC contigs of the map and the distribution of the BACs from three complementary BAC libraries in the contig.

The reliability of the rice BAC-based physical map: We conducted the following experiments to test the reliability of the automated contigs of the map.

Chloroplast DNA BAC contig analysis: The chloroplast genome of rice is ~140 kb in size. Therefore, all of the chloroplast DNA-derived BACs should be assembled into a single contig if the map contigs were assembled properly. We identified 615 chloroplast DNA-derived BACs from the entire database of the BACs using three chloroplast DNA probes (see materials and methods) and checked their positions in the contigs. The result showed that 588 of them were in a single contig (data not shown) and 27 were as singletons. The 27 singleton BACs were excluded from their assembly into the contig because the fingerprint of each of them consisted of four or fewer bands that were insufficient to be included in the contig assembly. These 615 chloroplast DNA-derived BAC clones were from three BAC libraries, and the fingerprint data were collected from 380 autoradiographs generated by three scientists in different experi-

TABLE 1

The automated (fundamental) BAC contigs of the rice physical map

Contig	C*	M**	B***	Contig	C*	M**	B***	Contig	C*	M**	B***	Contig	C*	M**	B***	Contig	C*	M**	B***	Contig	C*	M**	B***
ctg1	105	-	573	ctg71	46	-	236	ctg141	81	2	327	ctg211	22	1	78	ctg281	31	-	138	ctg351	43	1	154
ctg2	95	-	340	ctg72	104	-	406	ctg142	38	-	139	ctg212	30	-	126	ctg282	28	-	110	ctg352	36	-	147
ctg3	80	-	411	ctg73	39	-	144	ctg143	24	-	109	ctg213	29	1	97	ctg283	38	-	164	ctg353	20	-	87
ctg4	42	-	150	ctg74	61	-	285	ctg144	83	-	262	ctg214	25	1	83	ctg284	51	-	132	ctg354	44	-	141
ctg5	588	3	86	ctg75	60	-	297	ctg145	20	-	51	ctg215	41	2	132	ctg285	42	-	128	ctg355	26	-	128
ctg6	45	-	214	ctg76	42	-	172	ctg146	40	-	156	ctg216	48	-	215	ctg286	33	-	132	ctg356	52	3	166
ctg7	53	2	175	ctg77	52	-	212	ctg147	72	1	239	ctg217	39	-	180	ctg287	27	-	75	ctg357	13	-	47
ctg8	53	-	286	ctg78	51	1	126	ctg148	101	1	412	ctg218	34	-	142	ctg288	21	-	83	ctg358	29	-	90
ctg9	56	-	17	ctg79	53	-	194	ctg149	40	1	148	ctg219	29	-	125	ctg289	25	-	108	ctg359	35	-	93
ctg10	75	1	356	ctg80	117	-	440	ctg150	84	-	329	ctg220	31	2	117	ctg290	22	-	88	ctg360	30	-	108
ctg11	56	-	226	ctg81	52	2	158	ctg151	16	1	86	ctg221	30	-	141	ctg291	22	-	75	ctg361	20	-	71
ctg12	69	-	259	ctg82	57	-	212	ctg152	25	-	97	ctg222	28	-	98	ctg292	22	-	90	ctg362	49	2	161
ctg13	128	-	579	ctg83	60	1	193	ctg153	32	-	136	ctg223	64	-	252	ctg293	23	-	79	ctg363	46	-	133
ctg14	90	1	253	ctg84	88	1	528	ctg154	33	-	157	ctg224	31	-	91	ctg294	24	-	92	ctg364	18	-	74
ctg15	70	-	257	ctg85	42	-	150	ctg155	32	-	92	ctg225	31	1	132	ctg295	23	-	87	ctg365	25	1	80
ctg16	85	-	255	ctg86	41	1	159	ctg156	20	-	64	ctg226	26	-	114	ctg296	30	-	102	ctg366	21	-	89
ctg17	137	2	484	ctg87	42	-	168	ctg157	28	-	85	ctg227	28	-	94	ctg297	21	-	68	ctg367	43	-	148
ctg18	55	-	236	ctg88	63	1	242	ctg158	27	1	109	ctg228	35	-	117	ctg298	38	-	120	ctg368	37	-	117
ctg19	80	1	268	ctg89	66	-	232	ctg159	43	1	146	ctg229	23	-	88	ctg299	39	-	132	ctg369	26	3	95
ctg20	91	-	343	ctg90	60	-	229	ctg160	31	-	144	ctg230	27	-	105	ctg300	20	1	85	ctg370	18	-	80
ctg21	67	-	234	ctg91	94	1	305	ctg161	22	-	96	ctg231	50	-	150	ctg301	23	-	97	ctg371	24	-	56
ctg22	51	-	227	ctg92	116	1	357	ctg162	36	-	119	ctg232	17	-	56	ctg302	22	-	102	ctg372	15	-	77
ctg23	42	-	95	ctg93	68	2	264	ctg163	31	-	141	ctg233	26	-	103	ctg303	20	-	70	ctg373	26	-	86
ctg24	85	-	291	ctg94	42	-	166	ctg164	33	-	125	ctg234	36	1	149	ctg304	20	-	69	ctg374	23	-	81
ctg25	99	1	462	ctg95	20	-	81	ctg165	40	1	169	ctg235	37	-	149	ctg305	23	-	90	ctg375	15	-	45
ctg26	56	-	195	ctg96	22	-	70	ctg166	64	2	208	ctg236	35	-	127	ctg306	25	-	94	ctg376	23	-	68
ctg27	33	-	163	ctg97	54	2	188	ctg167	30	-	121	ctg237	24	-	89	ctg307	20	-	61	ctg377	37	-	67
ctg28	124	-	382	ctg98	58	-	243	ctg168	28	-	103	ctg238	26	-	55	ctg308	20	-	34	ctg378	40	-	151
ctg29	51	-	243	ctg99	65	1	243	ctg169	54	-	168	ctg239	16	-	77	ctg309	43	1	166	ctg379	29	-	83
ctg30	41	1	199	ctg100	75	1	313	ctg170	27	-	109	ctg240	33	-	132	ctg310	67	-	230	ctg380	34	-	109
ctg31	41	-	191	ctg101	86	-	293	ctg171	22	-	127	ctg241	86	1	308	ctg311	28	-	102	ctg381	28	-	74
ctg32	33	1	161	ctg102	94	1	354	ctg172	35	-	128	ctg242	18	-	96	ctg312	22	-	93	ctg382	22	-	83
ctg33	31	1	107	ctg103	52	-	256	ctg173	50	1	129	ctg243	31	1	118	ctg313	22	-	77	ctg383	22	-	74
ctg34	83	-	363	ctg104	70	1	262	ctg174	15	-	63	ctg244	29	-	83	ctg314	26	-	104	ctg384	21	-	79
ctg35	51	-	227	ctg105	72	2	247	ctg175	16	-	73	ctg245	30	-	79	ctg315	36	-	124	ctg385	20	-	99
ctg36	83	1	302	ctg106	86	-	344	ctg176	111	2	400	ctg246	20	1	76	ctg316	46	1	176	ctg386	25	-	66
ctg37	90	-	372	ctg107	6	1	47	ctg177	38	1	137	ctg247	23	-	112	ctg317	36	1	125	ctg387	21	-	57
ctg38	30	-	117	ctg108	23	-	102	ctg178	27	-	89	ctg248	24	-	94	ctg318	27	-	94	ctg388	23	-	72
ctg39	60	-	235	ctg109	56	1	188	ctg179	48	-	174	ctg249	21	-	62	ctg319	20	1	90	ctg389	17	-	72
ctg40	41	-	196	ctg110	51	2	195	ctg180	48	-	256	ctg250	25	-	102	ctg320	25	-	95	ctg390	18	-	82
ctg41	92	1	321	ctg111	30	2	105	ctg181	29	-	117	ctg251	22	-	101	ctg321	20	-	62	ctg391	17	1	66
ctg42	50	-	154	ctg112	46	-	203	ctg182	31	-	152	ctg252	31	1	75	ctg322	21	-	66	ctg392	19	-	70
ctg43	46	-	230	ctg113	74	1	302	ctg183	31	2	139	ctg253	21	-	94	ctg323	21	-	86	ctg393	17	-	38
ctg44	69	-	260	ctg114	51	2	168	ctg184	36	-	125	ctg254	20	1	84	ctg324	26	-	110	ctg394	27	-	95
ctg45	46	-	148	ctg115	31	-	122	ctg185	33	-	123	ctg255	23	-	87	ctg325	26	-	110	ctg395	72	-	193
ctg46	42	-	193	ctg116	25	-	84	ctg186	46	-	179	ctg256	21	-	122	ctg326	31	-	88	ctg396	30	-	119
ctg47	30	-	152	ctg117	33	-	122	ctg187	36	-	165	ctg257	20	-	81	ctg327	34	1	109	ctg397	48	-	134
ctg48	44	-	203	ctg118	30	-	28	ctg188	41	-	185	ctg258	23	-	87	ctg328	22	-	84	ctg398	32	-	131
ctg49	51	1	161	ctg119	39	-	168	ctg189	74	2	271	ctg259	23	-	99	ctg329	21	-	80	ctg399	43	-	139
ctg50	40	1	156	ctg120	46	-	206	ctg190	64	-	300	ctg260	24	-	77	ctg330	22	5	64	ctg400	24	-	55
ctg51	31	-	108	ctg121	42	-	195	ctg191	60	1	226	ctg261	21	-	74	ctg331	30	1	135	ctg401	27	-	86
ctg52	44	-	110	ctg122	27	-	128	ctg192	37	-	175	ctg262	20	-	75	ctg332	30	-	114	ctg402	14	1	61
ctg53	40	-	138	ctg123	76	-	285	ctg193	15	-	72	ctg263	21	-	66	ctg333	22	-	75	ctg403	16	-	46
ctg54	31	1	113	ctg124	15	-	54	ctg194	49	-	184	ctg264	22	-	98	ctg334	20	-	17	ctg404	22	-	80
ctg55	69	1	264	ctg125	37	-	186	ctg195	35	-	132	ctg265	22	-	81	ctg335	20	-	100	ctg405	19	-	49
ctg56	51	-	192	ctg126	37	-	171	ctg196	63	-	244	ctg266	21	-	87	ctg336	15	-	48	ctg406	15	-	59
ctg57	81	1	238	ctg127	28	-	108	ctg197	43	-	148	ctg267	30	-	126	ctg337	37	-	79	ctg407	22	1	109
ctg58	44	-	153	ctg128	34	1	152	ctg198	26	-	139	ctg268	24	1	115	ctg338	29	1	123	ctg408	16	-	42
ctg59	124	2	514	ctg129	65	-	236	ctg199	29	-	116	ctg269	22	-	99	ctg339	23	-	109	ctg409	30	-	105
ctg60	51	1	194	ctg130	28	-	85	ctg200	28	-	119	ctg270	21	-	105	ctg340	22	1	83	ctg410	66	3	239
ctg61	45	-	196	ctg131	27	-	73	ctg201	26	-	91	ctg271	20	-	101	ctg341	21	-	78	ctg411	20	-	67
ctg62	95	-	368	ctg132	32	-	103	ctg202	31	1	84	ctg272	20	-	93	ctg342	47	-	129	ctg412	12	-	35
ctg63	95	2	332	ctg133	35	-	146	ctg203	26	-	81	ctg273	20	-	84	ctg343	36	-	134	ctg413	17	2	52
ctg64	44	-	184	ctg134	22	1	89	ctg204	22	-	107	ctg274	20	-	50	ctg344	18	-	95	ctg414	21	-	83
ctg65	55	-	183	ctg135	33	-	112	ctg205	23	-	98	ctg275	16	1	66	ctg345	54	-	178	ctg415	10	-	24
ctg66	47	-	200	ctg136	37	-	135	ctg206	45	1	155	ctg276	35	-	150	ctg346	17	1	53	ctg416	10	-	30
ctg67	44	-	186	ctg137	31	-	118	ctg207	65	4	285	ctg277	32	-	128	ctg347	27	-	102	ctg417	10	-	35
ctg68	52	4	183	ctg138	43	-	202	ctg208	60	-	243	ctg278	31	-	119	ctg348	26	-	91	ctg418	16	-	46
ctg69	97	1	415	ctg139	30	-	140	ctg209	24	-	74	ctg279	25	-	101	ctg349	35	3	130	ctg419	10	-	35
ctg70	41	-	133	ctg140	40	-	140	ctg210	38	-	145	ctg280	28	-	117	ctg350	27	-	118	ctg420	10	-	39

C, the number of clones in the contig; M, the number of markers landed in the contig; and B, the number of unique bands for the length of the contig, one band, on average, being equivalent to ~6.3 kb as shown by the FPC program (SODERLUND *et al.* 1997).

TABLE 2

The extended BAC contigs of the rice physical map

Contig	Clones	Bands	Contig	Clones	Bands	Contig	Clones	Bands	Contig	Clones	Bands	Contig	Clones	Bands
ctg1	108	497	ctg61	43	129	ctg121	37	67	ctg181	49	136	ctg241	88	308
ctg2	156	487	ctg62	132	490	ctg122	56	216	ctg182	91	300	ctg242	30	73
ctg3	257	972	ctg63	146	441	ctg123	162	579	ctg183	43	138	ctg243	68	239
ctg4	204	704	ctg64	47	184	ctg124	64	215	ctg184	47	186	ctg244	27	108
ctg5	11	44	ctg65	27	53	ctg125	80	339	ctg185	58	234	ctg245	50	158
ctg6	45	188	ctg66	160	553	ctg126	22	74	ctg186	46	102	ctg246	48	176
ctg7	180	713	ctg67	105	354	ctg127	94	304	ctg187	59	183	ctg247	54	189
ctg8	164	521	ctg68	88	298	ctg128	52	164	ctg188	56	188	ctg248	96	278
ctg9	74	34	ctg69	89	276	ctg129	27	86	ctg189	113	368	ctg249	76	233
ctg10	135	451	ctg70	36	132	ctg130	127	398	ctg190	83	324	ctg250	58	204
ctg11	26	75	ctg71	59	250	ctg131	66	213	ctg191	137	445	ctg251	20	67
ctg12	40	178	ctg72	114	406	ctg132	38	117	ctg192	53	182	ctg252	19	24
ctg13	134	489	ctg73	67	198	ctg133	98	368	ctg193	34	122	ctg253	24	97
ctg14	109	469	ctg74	26	89	ctg134	37	139	ctg194	69	210	ctg254	21	84
ctg15	200	672	ctg75	198	644	ctg135	250	815	ctg195	40	132	ctg255	51	169
ctg16	87	283	ctg76	46	172	ctg136	98	292	ctg196	27	118	ctg256	52	164
ctg17	93	320	ctg77	55	212	ctg137	64	166	ctg197	43	148	ctg257	91	266
ctg18	71	257	ctg78	102	356	ctg138	27	108	ctg198	57	190	ctg258	45	125
ctg19	104	333	ctg79	35	93	ctg139	36	152	ctg199	54	178	ctg259	64	217
ctg20	117	415	ctg80	28	126	ctg140	88	328	ctg200	91	346	ctg260	53	195
ctg21	112	424	ctg81	55	173	ctg141	216	752	ctg201	27	91	ctg261	16	35
ctg22	73	310	ctg82	58	207	ctg142	42	139	ctg202	27	102	ctg262	52	158
ctg23	73	224	ctg83	74	266	ctg143	48	149	ctg203	11	38	ctg263	85	288
ctg24	145	645	ctg84	101	294	ctg144	95	262	ctg204	11	38	ctg264	24	99
ctg25	111	462	ctg85	47	136	ctg145	48	179	ctg205	14	48	ctg265	23	81
ctg26	115	300	ctg86	46	159	ctg146	62	110	ctg206	45	155	ctg266	39	132
ctg27	140	520	ctg87	81	296	ctg147	40	123	ctg207	86	295	ctg267	21	40
ctg28	68	218	ctg88	105	384	ctg148	43	108	ctg208	56	164	ctg268	66	107
ctg29	72	223	ctg89	37	135	ctg149	92	233	ctg209	48	206	ctg269	11	35
ctg30	73	262	ctg90	68	208	ctg150	87	331	ctg210	65	164	ctg270	18	57
ctg31	87	222	ctg91	127	428	ctg151	22	91	ctg211	29	123	ctg271	21	57
ctg32	73	244	ctg92	52	138	ctg152	85	290	ctg212	14	38	ctg272	18	66
ctg33	217	722	ctg93	88	334	ctg153	59	231	ctg213	32	99	ctg273	11	53

(*continued*)

TABLE 2

(Continued)

Contig	Clones	Bands	Contig	Clones	Bands	Contig	Clones	Bands	Contig	Clones	Bands	Contig	Clones	Bands
ctg34	45	142	ctg94	134	442	ctg154	45	156	ctg214	12	18	ctg274	20	62
ctg35	92	305	ctg95	32	105	ctg155	32	92	ctg215	52	147	ctg275	86	283
ctg36	113	408	ctg96	27	70	ctg156	54	160	ctg216	11	40	ctg276	99	309
ctg37	209	761	ctg97	55	188	ctg157	71	244	ctg217	42	185	ctg277	38	129
ctg38	55	170	ctg98	7	32	ctg158	61	221	ctg218	53	191	ctg278	18	77
ctg39	51	173	ctg99	70	243	ctg159	72	218	ctg219	73	190	ctg279	22	89
ctg40	138	502	ctg100	75	269	ctg160	43	158	ctg220	64	237	ctg280	28	117
ctg41	115	407	ctg101	43	139	ctg161	88	302	ctg221	42	109	ctg281	24	76
ctg42	105	368	ctg102	102	354	ctg162	42	102	ctg222	52	170	ctg282	16	46
ctg43	99	397	ctg103	58	126	ctg163	39	133	ctg223	70	255	ctg283	19	70
ctg44	84	281	ctg104	123	395	ctg164	16	70	ctg224	68	206	ctg284	10	22
ctg45	225	751	ctg105	56	192	ctg165	57	202	ctg225	31	88	ctg285	11	13
ctg46	56	247	ctg106	48	134	ctg166	78	245	ctg226	61	238	ctg286	38	132
ctg47	54	144	ctg107	90	288	ctg167	82	262	ctg227	29	94	ctg287	21	78
ctg48	114	444	ctg108	54	197	ctg168	56	188	ctg228	22	87	ctg288	13	38
ctg49	125	438	ctg109	132	362	ctg169	39	163	ctg229	23	61	ctg289	5	15
ctg50	56	285	ctg110	104	379	ctg170	17	78	ctg230	80	285	ctg290	24	109
ctg51	26	53	ctg111	23	72	ctg171	56	219	ctg231	29	110	ctg291	16	77
ctg52	89	381	ctg112	52	150	ctg172	57	181	ctg232	19	56	ctg292	24	73
ctg53	119	469	ctg113	81	307	ctg173	87	316	ctg233	28	103	ctg293	12	37
ctg54	27	91	ctg114	106	392	ctg174	60	182	ctg234	38	149	ctg294	21	66
ctg55	26	89	ctg115	53	160	ctg175	123	388	ctg235	72	266	ctg295	5	24
ctg56	26	110	ctg116	40	151	ctg176	66	236	ctg236	133	408	ctg296	6	15
ctg57	54	146	ctg117	71	191	ctg177	55	190	ctg237	29	115	ctg297	8	14
ctg58	48	153	ctg118	102	323	ctg178	95	331	ctg238	15	16	ctg298	615	85
ctg59	194	727	ctg119	40	131	ctg179	54	174	ctg239	86	309			
ctg60	77	206	ctg120	49	206	ctg180	38	109	ctg240	84	301			

Clones, the number of clones in the contig; bands, the number of unique bands for the length of the contig, one band, on average, being equivalent to ~6.3 kb as shown by the FPC program (SODERLUND et al. 1997).
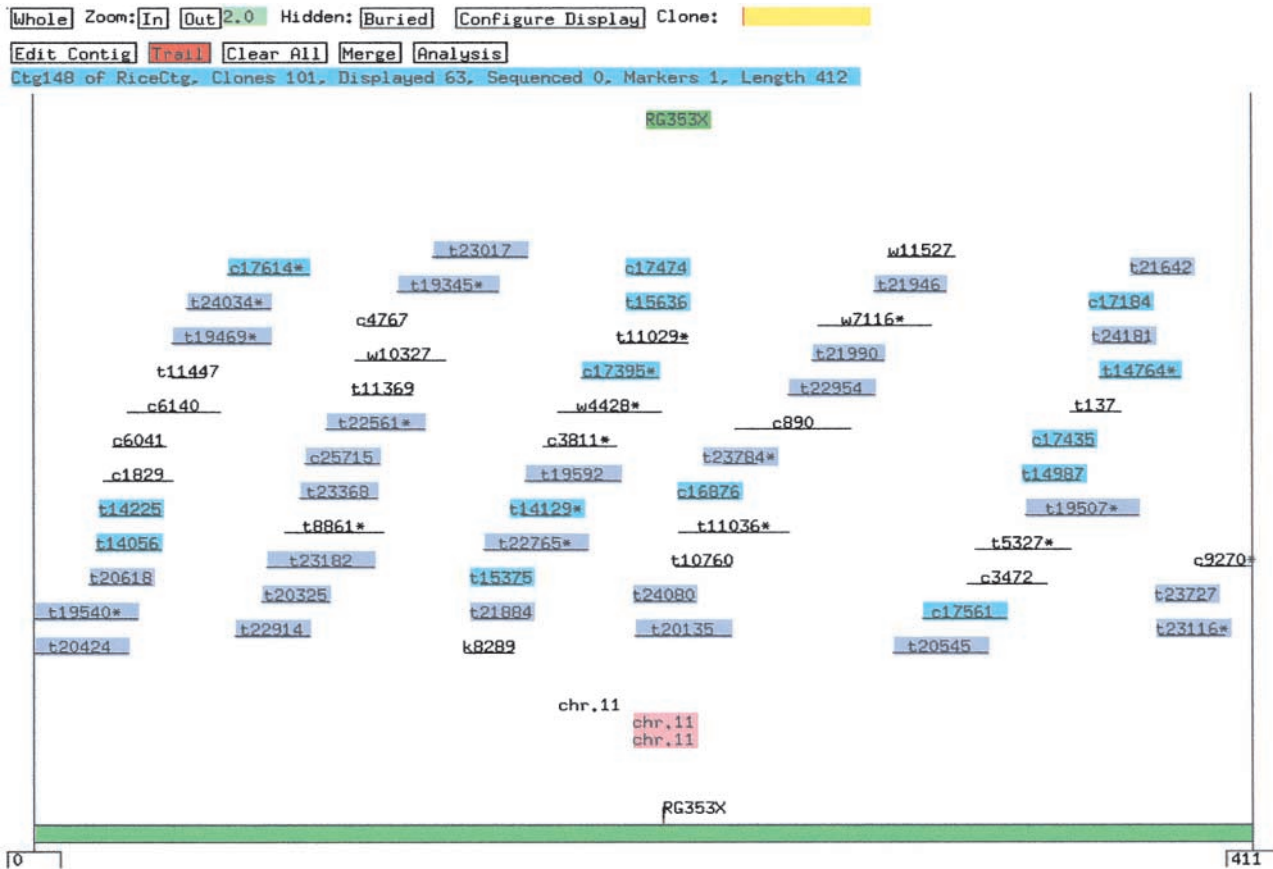
FIGURE 2.—Example of the BAC contigs of the rice physical map showing the distribution of the BACs from the three complementary libraries (ctg148 in Table 1). The contig includes 101 clones and has a length of 412 unique bands, being equivalent to 2595 kb. The highlighted clones in blue color were from the rice cv. Teqing *Eco*RI BAC library, the highlighted clones in green color from the rice cv. Teqing *Bam*HI BAC library, and the remaining clones from the rice cv. Teqing *Hin*dIII BAC library (see http://hbz.tamu.edu). Asterisk indicates a parent clone that covers one or more clones.

ments. The assembly of all 588 chloroplast DNA derived BACs having five or more bands in each of their fingerprints into a single contig indicated that the tolerance and cutoff values were properly selected and the map contigs were properly assembled.

Screening the contig BACs with mapped DNA markers: We hypothesized that if the map contigs are "reliable," the BACs selected with a single-copy DNA marker should all be located to a single contig. To test this hypothesis, we screened the BACs of the contigs with 77 mapped DNA markers and six random EST clones. The result is shown in Table 3 and summarized in Table 4. Library screening showed that 61 of the 83 DNA markers and ESTs gave two or more positive clones, 18 gave one positive clone, and 4 gave no positive clone (Tables 3 and 4). Note that of the 6.9 × genome coverage clones analyzed, 1.7 ×, 2.1 × and 3.1 × genome coverage clones were selected from each library, respectively. The uneven numbers of clones from each library might result in 18 of the 83 markers identifying one positive clone. Overall, 79 of the 83 markers (95%) gave one or more positive clones, which is consistent with the

estimate of the map contig genome coverage (97%) based on the total length of the contigs.

We then checked the positions of the BACs selected with each of the 61 markers that hybridized to two or more BACs in the 585 automated contigs. For 45 of the 61 markers, all of the clones selected with each marker were found to be members of a single contig (Tables 3 and 4), indicating that the contigs containing these DNA markers were properly assembled. Furthermore, we investigated the clones selected by 2 or more closely linked DNA markers and found that they were located at a single contig in 28 cases (Table 1 and Figure 3). These results also agreed with the genetic maps (CAUSSE *et al.* 1994; HARUSHIMA *et al.* 1998) from which the DNA markers were selected and thus further verified the reliability of these contigs.

BAC screening with the DNA markers showed that BACs identified by each of the remaining 16 markers were members of two or more contigs. For these 16 markers, it was possible that some of them actually detected two adjacent contigs that could not be linked by fingerprint analysis although further investigation is

TABLE 3

BACs selected with DNA markers and their positions in the physical map

Number	Probe	Positive BAC clones/contig	Contig
		Rice chromosome 8 markers	
1	CDO464	t21274/207, c16542/207	207
2	CDO595	w10299/225	225
3	RG1	w2161/81, t2620/81	81
4	RG28	t11404/207, c1858/207 c1859/207, c25397/207, c17053/S	207, S
5	RG29X	c1679/340, c3731/340	340
6	RG885	w6220/410, c9495/59, w12071/59	410, 59
7	RG978	c16181/7	7
8	RZ323	t21245/365, t22395/365	365
9	RZ562	w10063/211	211
10	RZ617	t3217/17, w4397/17, c5601/17, t21363/17	17
11	RZ952	c9901/330	330
12	C277[a]	w4436/59, w7077/117, t21171/207, c16542/207	207, 59, 117
13	C390	c1858/207, c1859/207, k8122/207, c25397/207	207
14	C626	k8411/128	128
15	C905	c9901/330, t23216/330, c24791/330	330
16	C922B	t20948/275, t22178/275	275
17	C929	t20506/349, t23675/349	349
18	G104	t5140/32, t14377/32, t14807/32	32
19	G1073	t8963/349, t9076/349, c9655/349 c9655/349	349
20	G2132		
21	G278	c1030/107, w6329/S, t22883/107, w3049/107	107, S
22	G56	w10325/215, c13123/215, w16245/215	215
23	R1813[a]	c4817/S, t9128/410, c9248/5, c24908/5, c24656/410	410, 5, S
24	R1010	c9901/330, t23216/330	330
25	R1943[b]	c4965/356, c9815/356, t20282/59, t20381/356, t14359/356	356, 59
26	R1963[b]	w12149/5, w12071/5, w12260/9	5, 9
27	R2007	w10652/434, c17015/434	434
28	R2367	t67/474	474
29	R2662[a]	t3195/33, t9119/33, c9656/33, c13460/215, t15632/S, w16245/215	33, 215, S
30	R2676[a]	t20134/97, t20604/298	97, 298
31	R2976	c16189/105	105
32	R622	c5471/S, w6220/410, t24130/410, c25592/410	410, S
33	R727[a]	c1551/565, w4552/105, w7021/105, t21043/105	105, 565
34	R902	c3601/166, t14855/166	166
35	S10324B	c3528/437, k8118/437	437
36	S10622	t11298/220, t20060/220	220
37	S10631	w10251/189, t11175/189, w11704/189, t21452/189, t21538/189, t14834/189	189
38	S1633A	t8963/349, t9076/349, c9655/S, w11597/143	349, 143, S
39	S1850A	t22771/338, t21470/338, t21210/338, c16875/338	338
40	S2014	c1792/434, t11019/434	434
		Rice chromosome 11 markers	
41	BCD808	t21195/113, c24695/113	113
42	RG1022X	c209/369, w7930/584, w10307/369, c17216/584, c17491/584, t22223/369, t22167/584	369, 584
43	RG103X	w7517/202, k8236/214, t18905/176, t23332/191, w13033/69	202, 214, 176 191, 69
44	RG1094	t23042/362, c24693/362	362
45	RG1109	w6716/92	92

(*continued*)

needed to establish this. The localization of the clones selected with each of the 16 DNA markers at two or more contigs could also be due to the multiple copies of the DNA markers in the rice genome, contig assembly errors or both. To answer this latter question, we investigated the copy number of the 16 markers in the rice genome by Southern hybridization. At the Japan Rice Genome Program website (http://www.dna.affrc.go.jp:84/

TABLE 3

(Continued)

Number	Probe	Positive BAC clones/contig	Contig
46	RG118	c1052/104, w8042/104, t22970/104, t23195/104, t15006/104	104
47	RG131	c16849/331	331
48	RG16	w10328/309	309
49	RG167	w15765/447, c17484/447	447
50	RG2	t19821/446, t20064/446, t21883/446	446
51	RG303	t20859/526, w16090/25, c17254/9, c13377/9	526, 25, 9
52	RG304	c5837/407, t11451/57, w11779/165, c5406/S	407, 57, 165, S
53	RG98	t11154/173, c17079/S	173, S
54	RZ141	t23054/234	234
55	RG353	c306/S, w4428/148, t10812/252, t11264/252, t19147/S, t22126/379, t21947/357, t15636/148, c17474/148	148, 252, 357 379, S
56	RZ536	c3541/78, w6625/78, w6772/78, t21357/78	78
57	RZ525		
58	RZ537	t1268/110, w11787/110, t19885/110, t21607/110, t14038/110	110
59	RZ638	t11014/141	141
60	RZ722		
61	RZ797	t11241/243, t19420/243, t22503/243, w25253/243, c13515/243	243
		Rice chromosome 12 markers	
62	CDO459	w7312/585	585
63	RG181	t3094/356, t3298/5, w7793/5, w7993/5, t20214/356, t21464/5, t21482/356, t21690/356, t23864/356, c25575/356, c13090/5, c13377/5, t15495/5, t17254/5	5, 356
64	RG235	w4308/68, t5221/68, w6548/68, w6994/68, c9469/68, w10533/68	68
65	RG241	c5604/213, w6953/S, k8291/S, w1041/213, t11101/63, t11104/63, t19921/63, t20770/93, t21374/59, t23761/93, t12914/213	213, 63, 59, 93, S
66	RG341X	t1299/68, w4308/68, w6548/68, t11153/68	68
67	RG457	t21284/545	545
68	RG463	c1660/413, t19919/413, t21156/413	413
69	RG574	w6459/102, c9234/183, w12752/S	102, 183
70	RG81	t2417/14, w3039/14, c4643/14, c9678/14, w11541/14, w11845/14	14
71	RG869	t129/68, c9469/68, t1299/68	68
72	RG9	t10094/S, t19532/99, t20360/99	99, S
73	RG901X	t129/68, t1299/68, c1660/413, w6548/68, c17564/68	68, 413
74	RG958	c17057/S, c9292/S, c13688/S	S
75	RZ397	w10626/88	88
76	RZ76		
77	RZ816	w6642/351	351
		Random EST clones	
78	1A2	t23168/111, t22806/111, t14891/111, t14891/111, c17084/111	111
79	1A9	t19664/49, t20335/49, t15232/49, w16404/49, w4205/49	49
80	1F9	w7907/5, w7993/5, w12134/5, t21902/5	5
81	1G10	t3190/369, t19170/369	369
82	4H10	c1747/176	176
83	4H11	t20216/481	481

S, singleton.

[a] The DNA marker is likely to be present in multiple copies in the haploid rice genome.

[b] The DNA marker is likely to be present in single copy in the haploid rice genome. The copy numbers of the remaining markers in the rice genome were not investigated in this study.

publicdata/naturegenetics/ricegmap.html), we were able to find the restriction patterns of 7 of the 16 DNA markers. Southern hybridization patterns indicate that 5 of the 7 DNA markers are multiple copy and 2 are single copy in the rice haploid genome. It is estimated from these 7 DNA markers that ∼71% (5/7) of the 16 DNA markers (5/7 × 16 = 11.4) are multiple copy in the rice genome. Therefore, it was possible to explain

that those clones selected with such DNA markers were located on multiple contigs. If the 11.4 marker contigs were properly assembled, ∼92.5% [(45 + 11.4)/61] of the automated contigs of the rice physical map were properly assembled. Furthermore, we assumed that the association of the remaining 7.5% DNA markers with two or more contigs resulted from "misassembly" of some of the BACs selected with the markers although

Distribution of the BACs selected with mapped DNA markers in the rice BAC-based physical map

Hybridization results	No. of DNA markers	% of markers	Characteristics of the map
No positive clone	4	4.82	4.8% (gap)
Single positive clone	18	21.69	
2 or more positive clones	61	73.49	
The positive clones located in a single contig	45	73.77 (45/61)	95.2% (coverage)
The positive clones distributed in 2 or more contigs	16[a]	26.23 (16/61)	92.5% (reliability) [45 + (16 × 71.43%)]/61
Total	83		

[a] Southern hybridizations were used to investigate the 16 DNA markers associated with two or more contigs. The hybridization patterns of 7 of the 16 markers were found at the web site (www.dna.affrc.go.jp:84/publicdata/naturegenetics/ricegmap.html). Five of the 7 markers were found to be multiple copy and two were found to be single copy in the rice genome (see Table 3). It was estimated on the basis of this result that 71.43% (5/7) of the 16 markers were likely to be multiple copy in the genome.

it was possible that they actually hybridized two adjacent overlapping contigs (see above). We studied the clones selected by single-copy markers (*e.g.*, R1943) each of which was shown to be associated with BACs in two contigs. We found that most of the selected BACs were located on one of the two contigs and one or two on the other contig. This indicated that for the contigs that might have some errors in contig assembly, most of their BACs were properly assembled.

DISCUSSION

We have successfully developed a genome-wide BAC-based physical map of indica rice from 21,078 BACs randomly selected from three complementary libraries by the DNA sequence electrophoresis-based restriction fingerprinting method. The map consists of 298 BAC contigs, which were merged from 585 automated contigs, and covers ~97% of the rice genome. This may represent a slight overestimate because it is possible that some of the 298 contigs are overlapped even though the overlaps could not be detected by fingerprint analysis under the conditions used in this study. Since the method used in this study is well suited for contig assembly from large-insert random BACs derived from centromeric and rDNA regions (T. UHM, C. WU and H.-B. ZHANG, unpublished results), the contigs for these regions are included in the 298 contigs. Hybridization analysis of the chloroplast DNA BAC contig and screening of the physical map BACs with numerous DNA markers consistently indicate that the BAC contigs constituting the physical map are properly assembled. Consistence was also observed between this BAC-based physical map and the rice genetic maps (CAUSSE *et al.* 1994; HARUSHIMA *et al.* 1998; see Figure 3), which further verifies the reliability of the physical map contigs in a long range. The physical mapping result of the rice genome is strongly supported by that of the Arabidopsis

genome using the approach employed in this study, in which nearly all contigs were tested to be accurate by the international Arabidopsis genome sequencing results (ARABIDOPSIS GENOME INITIATIVE 2000) and numerous mapped DNA markers (Y.-L. CHANG, Q. TAO, C. SCHEURING, K. MEKSEM and H.-B. ZHANG, unpublished results).

The BAC-based physical map of the rice genome is suitable for genomics research of rice and other grass species, including large-scale genome sequencing, effective positional cloning, high-throughput EST physical mapping, and target DNA marker development. First, although there is no published data available for comparison between the reliability of this map and those of the physical maps developed with other methods, it is possible that some errors exist in a genome-wide physical map developed with any or combined existing methods (see ZHANG and WU 2001). The development of the genome-wide physical map of *A. thaliana* using the method employed in this study (Y.-L. CHANG, Q. TAO, C. SCHEURING, K. MEKSEM and H.-B. ZHANG, unpublished results) is an indication of the powerfulness of the method for genome-wide physical mapping from large-insert random BACs. The accuracy of the *A. thaliana* physical map was verified by both the Arabidopsis genome sequencing results (ARABIDOPSIS GENOME INITIATIVE 2000) and numerous mapped DNA markers. Second, the rice map developed in this study has a 7.0× redundancy; *i.e.*, about seven clones could be selected for any region of the map. To build the tiling clone path of the genome for the above research purposes, analysis of the BAC fingerprints in a target contig with an aid of computers (see below) would minimize, if not completely eliminate, the clones that were not assembled properly, if any. Third, for genome sequencing a BAC that is anchored to the region of interest is selected from its contig and sequenced. The 1–3 BACs that overlap with the sequenced BAC at each end are then se-
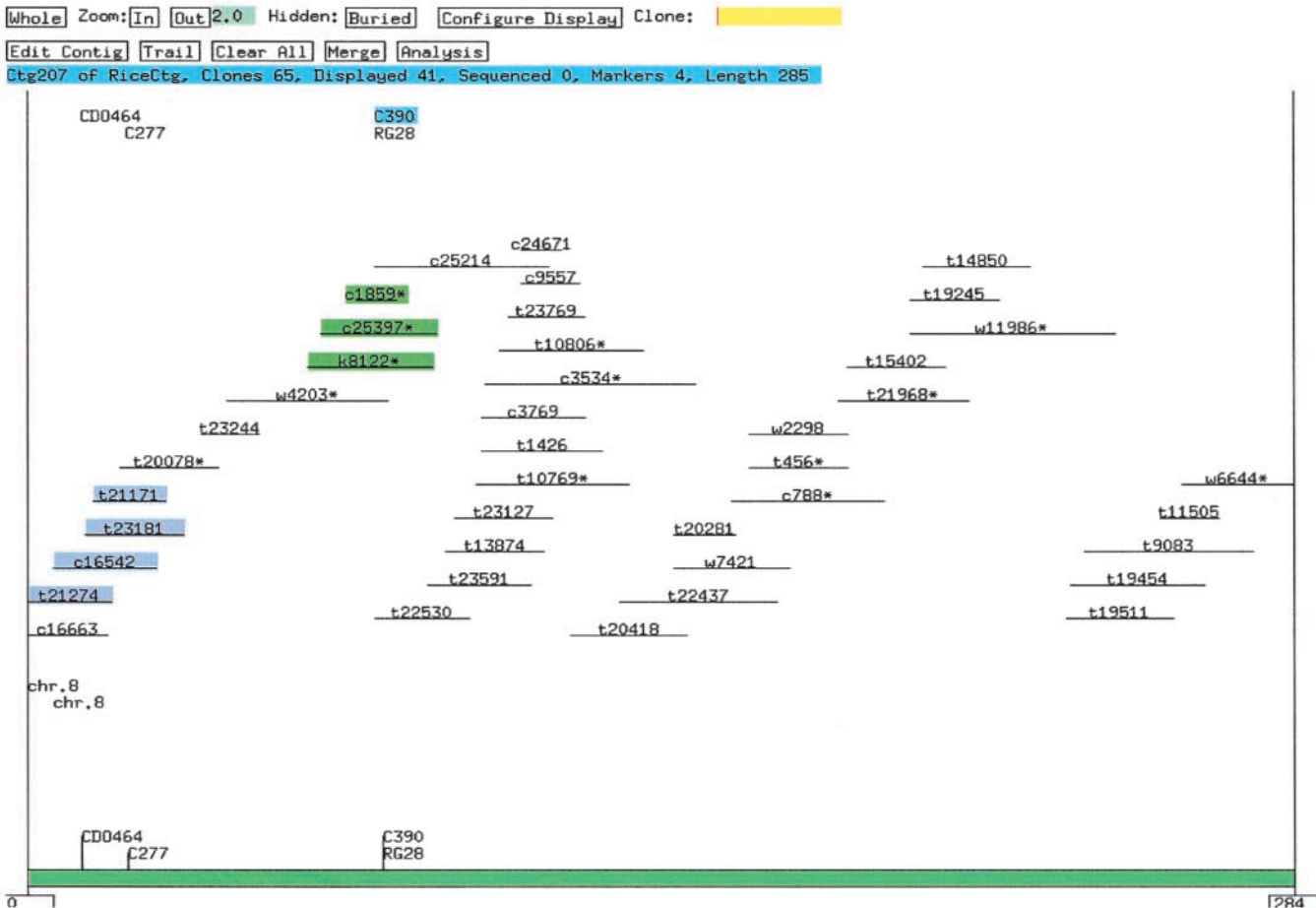
FIGURE 3.—Example of the contigs of the rice BAC-based map containing the positive clones of four DNA markers (ctg207 in Table 1). This contig contains 65 clones and has a length of 285 unique bands, estimated equivalent to 1796 kb. Note that four DNA markers, CDO464, RG28, C277, and C390, were located to this contig, all of which were also located at the same region of linkage group 8 of both rice genetic maps (Causse *et al.* 1994; Harushima *et al.* 1998). The highlighted clones indicate the positive clones of C390 and CDO464. Asterisk indicates a parent clone that covers one or more clones.

lected and end sequenced. The end sequence analysis of the selected BACs against the sequenced BAC will further verify the selection of the BACs for continuous sequencing (Mahairas *et al.* 1999). Fourth, the misassembled BACs, if any, in a contig of interest could be readily eliminated by refingerprinting the BACs of the contig, followed by contig reassembly. Because this experiment includes only the BACs of a target contig, it is much simpler than genome-wide physical mapping. The BACs that were previously assembled into the contig by chance (improperly) will be assembled as singletons and thus excluded, whereas the BACs that were correctly assembled will be reassembled into a single contig. Although this involves some additional work, it is manyfold simpler to develop contigs of interest from the genome-wide physical map contigs than from libraries by chromosome walking. Alternatively, the clones selected could also be verified by using the Clone–Fingerprint Map tool of the Genomic Information System (GIS) developed by this group (see below). Fifth, the BAC fingerprint database generated in this study has pro-

vided a means for chromosome walking and the construction of minimally overlapping clone tiling paths for the above research puproses via web-based tools. This is because the tiling clone path construction and chromosome walking can be directly conducted using the fingerprint database, without need of the assembled contigs by using the FPC Hitting tool (see http://hbz. tamu.edu-Physical Mapping-Indica Rice Map and Marra *et al.* 1999). To facilitate the management and use of integrated physical maps of agricultural genomes, we have created a database, developed the GIS system (H. Chen, Q. Tao, Y.-L. Chang and H.-B. Zhang, unpublished data), and posted the contigs of the indica rice physical map at http://hbz.tamu.edu-Physical Mapping-Indica Rice Map. Using the GIS, users can readily access the rice BAC fingerprint database and the physical map, perform chromosome walking on the rice genome, select clones and contigs of interest, and build contig tiling clone paths via WWW by using not only the FPC Hitting tool as Marra *et al.* (1999), but also four additional tools: Clone–Graphic Contig Map, Clone–

Fingerprint Map, Contig No.–Graphic Contig Map, and Marker/EST–Positive Clones–Contig/PFC Hit/ Fingerprint Matches.

The indica rice BAC-based physical map has provided a readily used platform for genomics research of rice and other monocotyledonous species. Two major subspecies of *O. sativa*, indica rice and japonica rice, are cultivated. Although both are equally good as models for grass genome research and japonica rice cv. Nipponbare is being used in rice genome sequencing by an international rice genome sequencing consortium led by the Japan Rice Genome Program, >90% of the world rice production is indica rice. Therefore, the genome research of indica rice, the staple food of about half of the world population, is far more important than that of japonica rice for the world rice economy. Because of this, sequencing of the indica rice genome is also ongoing in several countries. Additionally, we are developing a genome-wide BAC-BIBAC-based physical map of japonica rice cv. Nipponbare using the method and strategies employed in this study (Y. Li and H.-B. Zhang, unpublished data). The indica rice physical map reported here will provide a framework within which to perform evolutionary genomics research between the two rice subspecies and between rice and other gramineous crop plants. Studies have demonstrated that the gene content and order are highly conserved among the grass genomes (Ahn and Tanksley 1993; Ahn *et al.* 1993; Moore *et al.* 1995; Paterson *et al.* 1995; Bennetzen *et al.* 1996; Chen *et al.* 1997; Devos and Gale 1997, 2000). Therefore, the rice physical map developed in this study could also be used as a reference to expedite DNA marker development, gene identification, and gene cloning in gramineous crops with large genomes such as maize, wheat, and barley.

This rice genome BAC-based physical map represents the first report of the genome-wide physical mapping of large, complex genomes with large-insert, ordered random BACs using the DNA sequence electrophoresis-based restriction fingerprinting method. This method seems to offer a paradigm for genome-wide physical mapping of different plant and animal species of economic importance. The rice BAC-based map was developed in 1.5 scientist years. Similarly, we have developed a genome-wide BAC-BIBAC-based, integrated genetic, physical, and sequence map of the *A. thaliana* genome in 4 scientist months using the method and strategies of this study (Y.-L. Chang, Q. Tao, C. Scheuring, K. Meksem and H.-B. Zhang, unpublished results). In addition, we are developing the genome-wide physical maps of soybean, chicken, wheat, and cotton from BACs and BIBACs using the method and strategies developed in this study. The physical mapping results of rice, *A. thaliana*, and other species have demonstrated that it is feasible to rapidly develop genome-wide physical maps of the genomes of crop plants, farm animals, and humans at a reasonable cost using the method and strategies used in this study.

This study indicates that genome-wide physical mapping by restriction fingerprint analysis is not significantly influenced by genome size, genome complexity, and/or abundance of repeated sequences. This result was further confirmed by fingerprint analysis of BACs of 14 different plant and animal species with genome sizes ranging from 120 to 23,000 Mb/1C and repetitive sequences from 10 to 95% of the genomes (our unpublished results). Use of several complementary, bacteria-based large-insert clone libraries developed with different restriction enzymes, respectively, is an efficient strategy for minimizing "gaps" in the physical map because such libraries are balanced in distribution of clones in the genome and thus are equivalent to physically sheared shotgun libraries. A similar strategy has been or is being used for the physical mapping of Arabidopsis (Marra *et al.* 1999; Y.-L. Chang, Q. Tao, C. Scheuring, K. Meksem and H.-B. Zhang, unpublished results), Drosophila (Hoskin *et al.* 2000), and *Neurospora crassa* (Kelkar *et al.* 2001). The number of clones covering 6.0–8.0 haploid genomes seems to be sufficient for development of a genome-wide physical map of 95% genome coverage if they are truly random clones from the genome. This genome coverage of clones has been widely used for genome-wide shotgun genome sequencing (Fleischmann *et al.* 1995; Lin *et al.* 1999) and confirmed by this and our Arabidopsis (Y.-L. Chang, Q. Tao, C. Scheuring, K. Meksem and H.-B. Zhang, unpublished results) physical mapping results. A high-resolution electrophoresis system for fingerprint generation is crucial for ensuring the accuracy and reliability of contig assembly. This is especially true for genome-wide physical mapping of large, complex genomes because the data from tens of thousands of BAC fingerprints are needed to assemble the target physical map. The DNA sequence electrophoresis-based fingerprinting method has been proven to be reliable, high throughput, and economical for rapid genome-wide physical mapping of large, complex genomes with bacteria-based large-insert random clones. Furthermore, the physical mapping process could be further accelerated by a few fold by incorporating the newly developed capillary DNA automated sequencing technology into the fingerprinting approach.

LITERATURE CITED

Ahn, S., and S. D. Tanksley, 1993 Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA **90:** 7980–7984.

Ahn, S., J. A. Anderson, M. E. Sorrells and S. D. Tanksley, 1993 Homoeologous relationships of rice, wheat and maize chromosomes. Mol. Gen. Genet. **241:** 483–490.

Arabidopsis Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature **408:** 796–815.

Arumuganathan, K., and E. D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Reporter **9:** 208–218.

Bennetzen, J. L., P. Sanmiguel, C.-N. Liu, M. Chen, A. Tikhonov *et al.*, 1996 Microcolinearity and segmental duplication in the evolution of grass genomes, pp. 1–3 in *Unifying Plant Genomes* edited by J. S. Heslop-Harrison. Company of Biologists, Ltd., Cambridge.

Causse, M. A., T. M. Fulton, Y. G. Cho, A. N. Ahn, J. Chunwongse *et al.*, 1994 Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics **138:** 1251–1274.

Chen, M., P. Sanmiguel, A. C. De Oliveira, S.-S. Woo, H.-B. Zhang *et al.*, 1997 Microlinearity in *sh-2*-homologous regions of the maize, rice, and sorghum genomes. Proc. Natl. Acad. Sci. USA **94:** 3431–3435.

Coulson, A., J. Sulston, S. Brenner and J. Karn, 1986 Toward a physical map of the genome of the nematode *Caenorhabditis elegans.* Proc. Natl. Acad. Sci. USA **83:** 7821–7825.

Devos, K. M., and M. D. Gale, 1997 Comparative genetics in the grasses. Plant Mol. Biol. **35:** 3–15.

Devos, K. M., and M. D. Gale, 2000 Genome relationships: the grass model in current research. Plant Cell **12:** 637–646.

Ding, Y., M. D. Johnson, R. Colayco, Y. J. Chen, J. Melnyk *et al.*, 1999 Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting. Genomics **56:** 237–246.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirtness *et al.*, 1995 Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:** 496–511.

Frijters, A. C. J., Z. Zhang, M. Van Damme, G.-L. Wang, P. C. Ronald *et al.*, 1997 Construction of a bacterial artificial chromosome library containing large *Eco*RI and *Hin*dIII genomic fragments of lettuce. Theor. Appl. Genet. **94:** 390–399.

Gregory, S. G., G. R. Howell and D. R. Bentley, 1997 Genome mapping by fluorescent fingerprinting. Genome Res. **7:** 1162–1168.

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 F2 population. Genetics **148:** 479–494.

Hodgkin, J., R. H. A. Plasterk and R. H. Waterston, 1995 The nematode *Caenorhabditis elegans* and its genome. Science **270:** 410–414.

Hoskins, R. A., C. R. Nelson, B. P. Berman, T. R. Laverty, R. A. George *et al.*, 2000 A BAC-based physical map of the major autosomes of *Drosophila melanogaster.* Science **287:** 2271–2274.

Iaonnou, I., C. T. Amemiya, J. Garnes, P. M. Kroisel, H. Shizuya *et al.*, 1994 A new bacteriophage P1-derived vector for propagation of large human DNA fragments. Nat. Genet. **6:** 84–89.

Kelkar, H. S., J. Griffith, M. E. Case, S. F. Covert, R. D. Hall *et al.*, 2001 The *Neurospora crassa* genome: cosmid libraries sorted by chromosome. Genetics **157:** 979–990.

Kim, U.-J., B. W. Birren, T. Slepak, V. Mancino, C. Boysen *et al.*, 1996 Construction and characterization of a human bacterial artificial chromosome library. Genomics **34:** 213–218.

Lin, X., S. Kaul, S. Rounsley, T. R. Shea, M.-I. Benito *et al.*, 1999 Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana.* Nature **402:** 761–768.

Mahairas, G. G., J. C. Wallance, K. Smith, S. Swartzell, T. Holzman *et al.*, 1999 Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. Proc. Natl. Acad. Sci. USA **96:** 9739–9744.

Marra, M. A., T. A. Kucaba, E. D. Green, B. Brownstein, R. K. Wilson *et al.*, 1997 High throughput fingerprint analysis of large-insert clones. Genome Res. **7:** 1072-1084.

Marra, M. A., T. Kucaba, M. Sekhon, L. Hiller, R. Martienssen *et al.*, 1999 A map for sequence analysis of the *Arabidopsis thaliana* genome. Nat. Genet. **22:** 265–270.

Moore, G., K. M. Devos, Z. Wang and M. D. Gale, 1995 Gramineouses, line up and form a circle. Curr. Biol. **5:** 737–739.

Mozo, T., S. Fischer, S. Meier-Ewert, H. Lehrach and T. Altmann, 1998 Use of the IGF BAC library for physical mapping of the *Arabidopsis thaliana* genome. Plant J. **16:** 377–384.

Mozo, T., K. Dewar, P. Dunn, J. R. Ecker, S. Fischer *et al.*, 1999 A complete BAC-based physical map of the *Arabidopsis thaliana* genome. Nat. Genet. **22:** 271–275.

Olson, M. V., J. E. Dutchik, M. Y. Graham, G. M. Brodeur, C. Helms *et al.*, 1986 Random-clone strategy for genomic restriction mapping in yeast. Proc. Natl. Acad. Sci. USA **83:** 7826–7830.

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:** 1714–1717.

Riles, L., J. E. Dutchik, A. Baktha, B. K. McCauley, E. C. Thayer *et al.*, 1993 Physical maps of six smallest chromosomes of *Saccharomyces cerevisiae* at a resolution of 2.6 kilobase pairs. Genetics **134:** 81–150.

Saji, S., Y. Umehara, B. A. Antonio, H. Yamane, H. Tanoue *et al.*, 2001 A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. Genome **44:** 32–37.

Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak *et al.*, 1992 Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA **89:** 8794–8797.

Soderlund, C., I. Longden and R. Mott, 1997 FPC: a system for building contigs from restriction fingerprinted clones. Comput. Appl. Biosci. **13:** 523–535.

Sulston, J., F. Mallett, R. Staden, R. Durbin, T. Horsnell *et al.*, 1988 Software for genome mapping by fingerprinting techniques. Comput. Appl. Biosci **4:** 125–132.

Tait, E., M. C. Simon, S. King, A. J. Brown, N. A. R. Gow *et al.*, 1997 A *Candida albicans* genome project: cosmid contigs, physical mapping, and gene isolation. Fungal Genet. Biol. **21:** 308–314.

Tao, Q., and H.-B. Zhang, 1998 Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors. Nucleic Acids Res **26:** 4901–4909.

Venter, J. C., H. O. Smith and L. Hood, 1996 A new strategy for genome sequencing. Nature **381:** 364–366.

Zhang, H.-B., and R. A. Wing, 1997 Physical mapping of the rice genome with BACs. Plant Mol. Biol. **35:** 115–127.

Zhang, H.-B., and C. Wu, 2001 BAC as tools for genome sequencing. Plant Physiol. Biochem. **39:** 195–209.

Zhang, H.-B., S. Choi, S.-S. Woo, Z. Li and R. A. Wing, 1996 Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol. Breed. **2:** 11–24.

Zhu, H., B. P. Blackmon, M. Sasinowski and R. A. Dean, 1999 Physical map and organization of chromosome 7 in the rice blast fungus, *Magnaporthe grisea.* Genome Res. **9:** 739–750.