Theor Appl Genet (2002) 105:1058–1066
DOI 10.1007/s00122-002-1057-3

Q. Tao · A. Wang · H.-B. Zhang

One large-insert plant-transformation-competent BIBAC library and three BAC libraries of Japonica rice for genome research in rice and other grasses

Received: 27 August 2001 / Accepted: 18 February 2002 / Published online: 22 August 2002

Abstract We report one large-insert BIBAC library and three BAC libraries for japonica rice cv Nipponbare. The BIBAC library was constructed in the *Hin*dIII site of a plant-transformation-competent binary vector (pCLD04541) and the three BAC libraries were constructed in the *Bam*HI, *Hin*dIII and *Eco*RI sites of a BAC vector (pECBAC1), respectively. Each library contains 23,040 clones, has an average insert size of 130 kb, 170 kb, 150 kb and 156 kb, and covers 6.7×, 8.7×, 7.7× and 8.0 × rice haploid genomes, respectively. The combined libraries contain 92,160 clones in total, covering 31.1 × rice haploid genomes. To demonstrate their utility, we screened the libraries with 55 DNA markers mapped to chromosome 8 of the rice genetic maps and analyzed a number of clones by the restriction fingerprinting and contig assembly method. The results indicate that the libraries completely cover the rice genome and, thus, are well-suited for genome research in rice and other gramineous crops. The BIBAC library represents the first plant-transformation-competent large-insert DNA library for rice, which will streamline map-based cloning, functional analysis of the rice genome sequence and molecular breeding in rice and other grass species. These libraries are being used in the development of a whole-genome, BAC/BIBAC-based, integrated physical, genetic and sequence map of rice and in the research of genome-wide comparative genomics of grass species.

Keywords BAC and BIBAC library · Plant transformation · Genomics · Positional cloning · Rice

Communicated by P. Langridge

Q. Tao · A. Wang · H.-B. Zhang (✉)
Department of Soil and Crop Sciences and
Institute for Plant Genomics and Biotechnology, 2123 TAMUS,
Texas A&M University, College Station, TX 77843-2132, USA
e-mail: hbz7049@tamu.edu
Tel.: +979-862-2244, Fax: +979-862-4790

Introduction

Rice, *Oryza sativa* L., is the staple food of about half of the world's population and a model species for genome research of cereal crops due to its small genome (430 Mb/1C, Arumuganathan and Earle 1991), established transformation systems (Hiei et al. 1994; Dong et al. 1996) and wealth of genomic and genetic resources (Causse et al. 1994; Harushima et al. 1998; Khush and Kinoshita 1991). The gene content and order are highly conserved among the grass genomes (Ahn and Tanksley 1993; Ahn et al. 1993; Moore et al. 1995; Chen et al. 1997; Devos and Gale 1997), allowing isolation of genes in large-genome cereal species such as maize and wheat, using rice as the reference species. To discover and decode the genes of rice, an international collaborative effort is being made to sequence the entire genome of rice using *O. sativa* ssp. *japonica* cv Nipponbare as the reference genotype. To accomplish this goal, several large-insert DNA libraries that are of high-quality, of high-genome coverage and complementary to each other are required (Zhang and Wu 2001). This is argued by the recent development of the genome-wide physical and sequence maps of *Arabidopsis thaliana* (Marra et al. 1999; Mozo et al. 1999; Arabidopsis Genome Initiative 2000; Chang et al. 2001), human (International Human Genome Mapping Consortium 2001; International Human Genome Sequencing Consortium 2001; Venter et al. 2001) and indica rice (Tao et al. 2001), in each of which several large-insert, complementary DNA libraries were used. For instance, to develop the physical and sequence maps of the human genome, clones from eight genomic DNA libraries were employed. These libraries were constructed with four different restriction enzymes and the average insert sizes of the eight libraries range from 116 to 196 kb.

To facilitate rice genome research, several rice BAC (Zhang et al. 1996a; Nakamura et al. 1997) and YAC (Umehara et al. 1995) libraries were constructed in the past few years. However, no large-insert BAC library for japonica rice cv Nipponbare has been reported yet, al-

though two BAC libraries and one PAC library are developed for the genotype (http://www.genome.clemson.edu; http://rgp.dna.affrc.go.jp). The PAC library was constructed with *Sau*3AI having 50%-A/T in restriction sites, and the two BAC libraries were constructed with *Hin*dIII and *Eco*RI, respectively, both of which are A/T-rich in their restriction sites, but not in agreement with the fact that the genome of rice is G/C-rich (Zhang et al. 1996a; see Fig. 1). Additionally, the average insert sizes of the three libraries range from 112 to 130 kb. Although these libraries are used for large-scale genome sequencing, their insert sizes may not be sufficiently large to span the regions that contain large-segment repeats. The regions containing large-segment repeats, spanning from 50 to 500 kb, have been discovered in the human genome (International Human Genome Sequencing Consortium 2001). The large-segment repeats have been shown to significantly influence the assembly of correct physical and sequence maps in the Arabidopsis genome sequencing effort (Stupar et al. 2001). Therefore, additional BAC libraries that are generated with restriction enzymes that are G/C-rich in restriction sites and have larger inserts are needed to sequence the entire rice genome. Such restriction enzymes are expected to clone the genomic regions that could not be cloned by *Hin*dIII, *Eco*RI or *Sau*3AI, and thus are expected to complement the BAC and PAC libraries being used in genome sequencing. The complementarity will be further enhanced by using a vector system differing from those employed for the existing BAC and PAC libraries. The much larger inserts of the new libraries are expected to span the genomic regions containing large-segment repeats and help assemble the genome sequence in the regions.

One major goal of rice genome research is to identify every gene and decipher their functions in rice. A plant-transformation-competent large-insert binary BAC (BIBAC) library promises to accelerate this process. Such a large-insert DNA library can not only be used as a BAC library in genome research, but also facilitate gene discovery and functional studies by direct transformation of BIBACs carrying the genes, or gene clusters of interest, into plants via *Agrobacterium*. This is especially important for large-scale gene discovery and cloning, and functional analysis of the genome sequence if the BIBAC library could be incorporated into the rice genome physical and sequence maps under development. Furthermore, such a large-insert BIBAC library will streamline the map-based cloning of genes in rice and other cereal crops. Techniques for BIBAC transformation have been established in several plant species, including tobacco (Hamilton et al. 1996), tomato (Hamilton et al. 1999), rice (Choi et al. 2000, Plant and Animal Genome VIII, Abstract #P609) and *Arabidopsis* (Liu et al. 1999). In contrast, the BACs cloned in BAC vectors, such as pBeloBAC11 (Kim et al. 1996) and pBACe3.6 (Frengen et al. 1999) that are used for rice genome sequencing, must be subcloned into a plant transformation binary vector (e.g., BIBAC2, Hamilton et al. 1996) for plant transformation. However, it is difficult to subclone the entire insert of a BAC of monocotyledonous plants into the BIBAC2 vector because there are often several *Not*I sites in a 100-kb BAC insert (Zhang et al. 1996a; see Fig. 1). The *Not*I is often used in subcloning of a BAC insert into the BIBAC2 vector because it is a rare (8 bp)-cutting enzyme and two of its restriction sites flank the cloning sites of BIBAC2 and pBeloBAC11. Therefore, development of a BIBAC library for Nipponbare and incorporation of it into rice global genome physical mapping and sequencing will significantly accelerate discovery, cloning and engineering of cereal genes and functional analysis of the rice genome sequence. For similar purposes, a whole-genome, BAC/BIBAC-based, integrated physical, genetic and sequence map of *A. thaliana* has been developed (Chang et al. 2001). Although BIBAC libraries for several plant species (Hamilton et al. 1999; Moullet et al. 1999; Meksem et al. 2000; Wu et al. 2000; H.-B. Zhang, unpublished) have been developed, no BIBAC library for rice has been reported.

We present here the generation and characterization of one large-insert, high quality plant-transformation-competent BIBAC library and three BAC libraries for japonica rice cv Nipponbare. The new BAC libraries not only have much larger average insert sizes (130–170 kb) and extensive genome coverage (31.1×), but also complement each other and the BAC and PAC libraries being used in rice genome sequencing. The transformability of the BIBAC library in plants will streamline gene cloning and functional genomics research in rice and other cereal crops. We have also demonstrated the utilities of these libraries for rice genome research using 55 RFLP markers selected from chromosome 8 of the rice genetic maps (Causse et al. 1994; Harushima et al. 1998). In addition, these libraries are being used in a whole-genome, BAC/BIBAC-based, integrated physical, genetic and sequence map of rice, the physical mapping and characterization of chromosome centromeres, and genome-wide analysis of disease resistance genes in cereal species (H.-B. Zhang, unpublished).

Materials and methods

Vector preparation

The vector DNAs of pECBAC1 and pCLD04541 were isolated by the alkaline-lysis method, followed by two rounds of cesium chloride gradient centrifugation (Sambrook et al. 1989). The vector DNA was completely digested and dephosphorylated according to Zhang et al. (1996a).

High-molecular-weight (HMW) DNA preparation

The seeds of japonica rice cv Nipponbare were a gift of Henry T. Nguyen, Texas Tech University, Texas, USA, which were kindly provided by the Japan Rice Genome Research Program. The seedlings of Nipponbare were grown in a greenhouse for 3–4 weeks, harvested, frozen immediately in liquid nitrogen, and stored at –80 °C. HMW nuclear DNA was prepared according to Zhang et

al. (1995, 1996b). The DNA embedded in low-melting-point agarose (FMC, USA) plugs (100 μl/plug) was subjected to pulsed-field gel electrophoresis (PFGE) using the CHEF DRIII apparatus (BioRad, USA) to remove the small DNA fragments and the substances in the plugs that might inhibit subsequent restriction enzyme digestion. The conditions of the PFGE were 1.0% agarose gel, 4 V/cm, 11 °C, 5-s switch time for 8 h in 0.5 × TBE (Sambrook et al. 1989). The plugs containing large DNA fragments (>400 kb) were collected from the pulsed-field gel, dialyzed against cold TE (10 mM Trizma base, pH 8.0, 1.0 mM EDTA, pH 8.0) on ice for at least three-times, with one change of TE per hour and then stored at 4 °C.

BAC and BIBAC library construction

Each HMW DNA plug (about 100 μl in volume) was cut into nine pieces (mini-plugs) of approximately equal size, and transferred into a microfuge tube containing 1 ml of pre-reaction buffer [1 × restriction buffer (GIBCO BRL, USA), 2 mM spermidine (Sigma, USA), 0.1 mg/ml BSA (New England Biolabs, USA) and 1 mM dithiothreitol (DTT) (Sigma, USA)]. The mini-plugs were incubated on ice for 80 min with one change of the pre-reaction buffer during the incubation. Three mini-plugs were transferred to a fresh tube containing 100 μl of reaction buffer (the same as the pre-reaction buffer except that 0.2 mg/ml of BSA was used) and various amounts of enzyme ranging from 0.2 to 5 units per microtube. The reaction was incubated on ice for 60 min, then transferred into a 37 °C water bath and incubated at 37 °C for 10 min. The reaction was immediately stopped by adding 1/10 vol of 0.5 M EDTA, pH 8.0, and subjected to PFGE. The PFGE condition was 1% agarose gel, 6 V/cm, 11 °C, at a 50-s switch time in 0.5 × TBE for 18 h. The partial digestion condition resulting in a majority of restricted fragments in a range from 100 to 350 kb was selected for large-scale partial digestion.

Large-scale partial digestion was carried out using six 100-μl plugs according to the optimal condition predetermined as above. The partially digested DNA was fractionated on a 1% agarose gel at 6 V/cm, 11 °C, at a 90-s switch time in 0.5 × TBE for 18 h, and then 4 V/cm, with a 5-s switch time for 6 h. The DNA fractions ranging from 100 to 150 kb, 150 to 300 kb and 300 to 350 kb were excised. The DNA was recovered from the agarose-gel slices by electroelution in dialysis tubing (molecular-weight exclusion limit = 12,000–14,000 Daltons) (GIBCO BRL, USA) using the PFGE apparatus (BioRad, USA) at 6 V/cm, 11 °C, at a 30-s switch time in 0.5 × TBE for 4 h, followed by reversing the polarity of the currency for 60 s. The DNA was carefully transferred from the dialysis tubing to a microfuge tube with a wide-bore pipette tip, mixed with gel-loading dye (Sambrook et al. 1989), by inverting the tube and subject it to a second round of size selection on a 1% agarose gel at 4 V/cm, 11 °C, at a 5-s switch time in 0.5 × TBE for 8 h. The gel-zone containing the compressed DNA fragments (that appeared as a single band) was excised, and the DNA was recovered by electroelution as above and dialyzed in the same tube against 0.5 × TE for 3 h at 4 °C, with one change of 0.5 × TE per h.

The concentrations of both insert and vector DNAs were measured by running the samples with standard lambda DNA of known concentration on a 0.8% agarose gel in 0.5 × TBE. The concentration of insert DNA was adjusted to 1 to 2 ng/μl. Ligation was carried out in a tube containing 100 μl of insert DNA and an appropriate vector DNA (the molar ratio was 4:1 with vector excess) with five units of T4 DNA ligase (GIBCO BRL, USA). The ligation reaction was incubated at 16 °C for 10 h.

The ligation mixture was used directly for the transformation of bacterial cells by electroporation using the Cell-Porator system (GIBCO BRL, USA). For single transformation, 1 μl of ligation was mixed with 20 μl of ElectroMAX DH10B competent cells (GIBCO BRL, USA). The electroporation conditions were: 350 V, 330 μF, low Ohms and 4 KV with a fast charge. Transformed cells were transferred into 1 ml of SOC medium (Sambrook et al. 1989) and incubated at 37 °C with gentle shaking for 1 h. The cells were plated on LB agar medium with either 12.5 μg/ml of chloramphenicol for the pECBAC1 vector or 15 μg/ml of tetracycline for the pCLD04541 vector, and incubated at 37 °C for 20 h for the pCLD04541clones or 30 h for the pECBAC1 clones to allow the colony color to completely develop.

BAC and BIBAC analysis and library assembly

White colonies on the selective medium were randomly picked from different plates, inoculated in 5 ml of LB medium with appropriate antibiotics (see above) and incubated at 37 °C with shaking at 250 rpm overnight. DNA was prepared, digested with *Not*I and electrophoresed on a pulsed-field gel (Zhang et al. 1996a). The PFGE condition was 6 V/cm, 11 °C, at a 5-s initial switch time and a 15-s final switch time with a linear ramp, 1% agarose gel in 0.5 × TBE for 16 h. The DNA fragments of BACs or BIBACs were visualized by staining the gel with ethidium bromide and then photographed. The insert size of each clone was estimated by adding the sizes of all DNA fragments derived from the clone insert. The ligation that gave the largest insert size and high transformation efficiency was selected for library assembly. The BAC and BIBAC colonies were arrayed in 384-well microtiter plates containing 50 μl of LB plus freezer medium (Zhang et al. 1996a) per well with appropriate antibiotics, grown at 37 °C overnight and stored at –80 °C.

Library screening

High-density colony filters were made from the arrayed libraries using the Biomek 2000 robot (Beckman, USA). Clones were double-spotted onto Hybond N+ membrane (Amersham, USA) in a 3 × 3 format. The high-density clone filters were processed according to Zhang et al. (1996a), followed by baking at 80 °C for 2 h. The DNA markers selected from rice chromosome 8 (Causse et al. 1994; Harushima et al. 1998), and the clones of chloroplast genes, *ndh*A, *rbc*L and *psb*A, were used as probes to screen the libraries. The DNA markers are well distributed along the chromosome of the rice genetic linkage maps. Probe preparation, hybridization and signal detection were performed according to Tao et al. (2001) using a non-radioactive labeling and signal-detection kit (Roche Molecular Biochemicals, USA).

BAC fingerprinting and contig assembly

BAC DNA was prepared from 1 ml of overnight culture and fingerprinted according to a published procedure that was used for whole-genome physical mapping with BACs and BIBACs on DNA sequencing gels (Chang et al. 2001; Tao et al. 2001). The BAC fingerprint autoradiographs were scanned into image files using the Mirage D-16L scanner (UMAX, USA) and edited by using the Image 3.9a of the FPC package downloaded from the Sanger Center, UK (Soderlund et al. 1997). BAC and BIBAC contigs were assembled using the FPC3.4 of the FPC package (Soderlund et al. 1997) at Tolerance = 7 and Cutoff = 1e-10, Diff = 0.30, Diffbury = 0.10, MinEnds = 8 and MinBands = 5.

Results

We have constructed one large-insert BIBAC library and three large-insert, complementary BAC libraries for japonica rice cv Nipponbare (Table 1). The three BAC libraries were constructed from the partial digests of the Nipponbare nuclear DNA with *Bam*HI, *Hin*dIII and *Eco*RI, respectively, and the BIBAC library was constructed from the partial digest of the Nipponbare nuclear DNA with *Hin*dIII. Nipponbare HMW DNA was par-

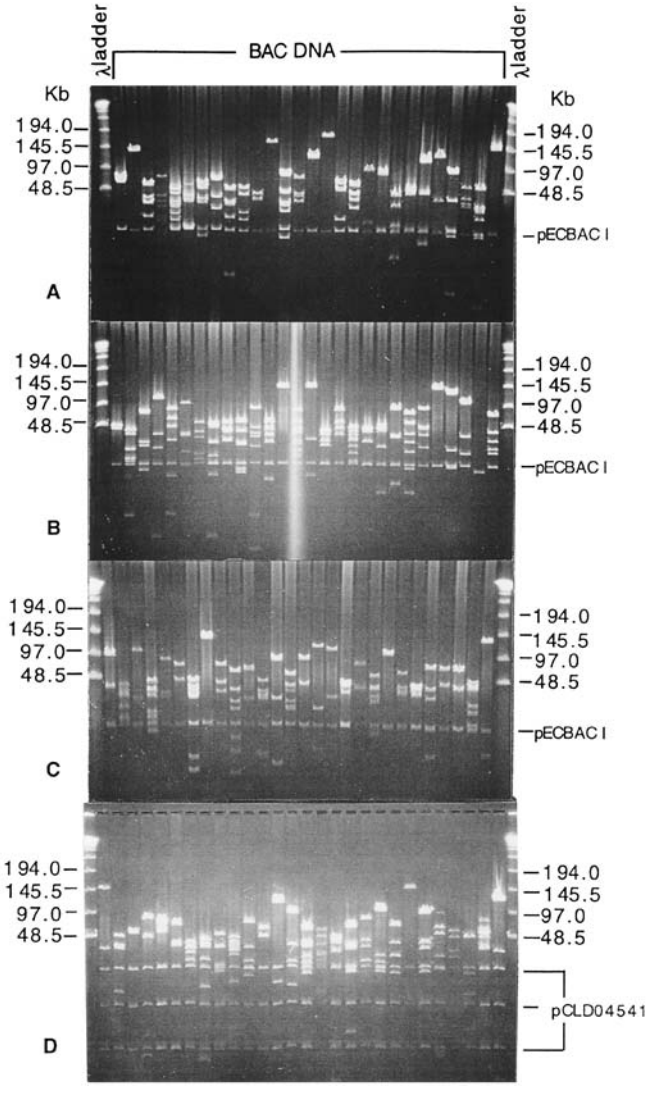

Fig. 1 Random clones from the *Bam*HI (**A**), *Eco*RI (**B**) and *Hin*dIII (**C**) BAC libraries and the *Hin*dIII BIBAC library (**D**) of japonica rice cv Nipponbare. BAC or BIBAC DNA was isolated, digested with *Not*I to release the insert DNA from the cloning vector, pECBAC1 (Panels **A, B and C**) or pCLD04541 (Panel **D**), run on pulsed-field gels, stained with ethidium bromide and photographed. The band(s) appearing in all lanes was from the cloning vectors and the remaining bands in a lane were from the rice DNA insert. Note that several bands were generated from the rice DNA insert of each clone, indicating that the rice genome is rich in the G/C content. The insert size of each clone was estimated by adding the sizes of all insert DNA bands of the clone

tially digested with an appropriate enzyme, size-selected on pulsed-field gels, ligated into a suitable site of the BAC vector pECBAC1 or the binary vector pCLD04541 and transformed into bacterial cells. All four libraries were generated from the selected restricted fragments ranging from 150 to 300 kb, and obtained from single ligations. Transformation efficiencies for the libraries ranged from 3.0×10^6 to 1.5×10^7 BAC clones per microgram of DNA. For each library, 23,040 clones were arrayed in 60 384-well microplates, and three copies of each library were generated and maintained in three –80 °C freezers, respectively, for long-term use in genome research.

To estimate the insert sizes of clones in these libraries, 84 clones were randomly selected and grown in LB medium containing appropriate antibiotics from each library. DNAs of the clones were isolated, digested with *Not*I to release the insert DNAs from the cloning vector and separated on pulsed-field gels. The gels were stained with ethidium bromide and photographed (e.g., see Fig. 1). The insert size of each clone was estimated using the lambda DNA ladder and the cloning vector fragment(s) as the molecular-weight standards. The insert sizes of the clones from the *Eco*RI, *Bam*HI and *Hin*dIII BAC libraries and the BIBAC library ranged from 85 to 236 kb, 98 to 260 kb, 50 to 200 kb and 40 to 210 kb, respectively (Fig. 2). The average insert sizes of the *Eco*RI, *Bam*HI and *Hin*dIII BAC libraries and the BIBAC library are 156 kb, 170 kb, 150 kb and 130 kb, respectively (Table 1; Figs. 1 and 2). From 0 to 3% of clones for each library were found to have no inserts. We also estimated the percentage of clones that were of chloroplast DNA origin. Four filters of 4 × 384 clones (6,144 clones) from each library were screened with three barley chloroplast genome-specific genes, *ndh*A, *rbc*L and *psb*A, and from 93 to 106 positive clones were identified to probably originate from chloroplast DNA. Since all libraries were produced from the same DNA preparation, we combined the screening results from all four libraries. As a result, about 1.6% of the clones in the libraries were estimated to originate from chloroplast DNA. Because each library contains 23,040 clones, 0–3% of the clones have no inserts and 1.6% are of chloroplast DNA origin, each of the *Eco*RI, *Bam*HI and *Hin*dIII BAC libraries and the BIBAC library represents a genome coverage of 8.0×, 8.7×, 7.7× and 6.7×

Table 1 Japonica rice cv Nipponbare BAC and BIBAC libraries

Vectors	Vector size (kb)	Cloning sites	Average insert size (kb)	Genome equivalents	No. of clones stored in a library
pECBAC1[a]	7.4	*Eco*RI	156	8.0	23,040
pECBAC1	7.4	*Bam*HI	170	8.7	23,040
pECBAC1	7.4	*Hin*dIII	150	7.7	23,040
pCLD04541[b]	27.6	*Hin*dIII	130	6.7	23,040
Combined libraries:				31.1	92,160

[a] pECBAC1 (Frijters et al. 1997) is a modified vector of pBeloBAC11 (kim et al. 1996)
[b] pCLD04541 is an *Agrobacterium*-mediated plant transformation binary vector (Jones et al. 1992; Tao and Zhang 1998)

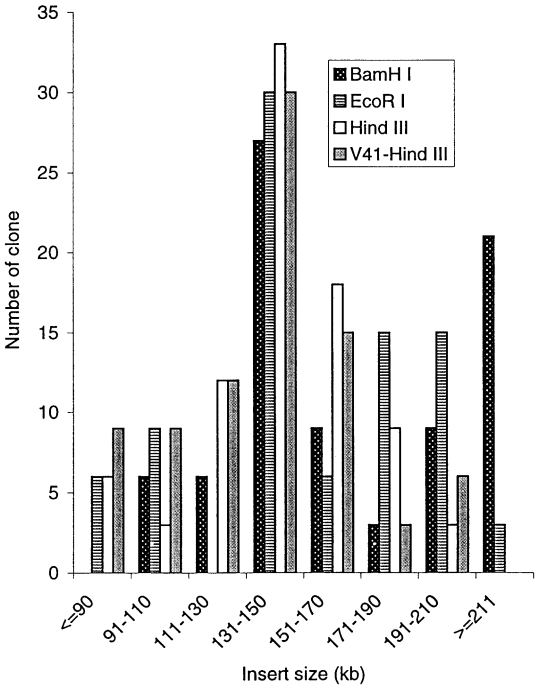

Fig. 2 The insert size distribution of the clones randomly selected from the three BAC libraries and one BIBAC library of Nipponbare. Eighty four clones from each library were analyzed for insert sizes as described in Fig. 1. "*Bam*HI, *Eco*RI, *Hin*dIII and V41-*Hin*dIII" indicate the *Bam*HI, *Eco*RI and *Hin*dIII BAC libraries and the *Hin*dIII BIBAC library, respectively. The V41 represents the vector pCLD04541

equivalents of the rice haploid genome, respectively, considering that rice has a genome size of 430 Mb/1C (Arumuganathan and Earle 1991). Theoretically, the probability of obtaining a specific clone from any of these libraries is greater than 99%. The combined four libraries, therefore, contain 92,160 clones, representing a coverage of 31.1 × haploid genomes (Table 1).

To test the utilities of the libraries for genome research, we screened the *Eco*RI and *Bam*HI BAC libraries and the *Hin*dIII BIBAC library with 55 DNA markers selected from a single chromosome (chromosome 8) of the rice genetic maps (Causse et al. 1994; Harushima et al. 1998). Since the size of chromosome 8 is about 33 Mb, the 55 markers represent an average distribution of markers on the chromosome of about 600 kb per marker. Our previous studies (Chang et al. 2001; Tao et al. 2001) indicated that the BAC clones of 6–8 × haploid genome equivalents should be sufficient to develop a global physical map with a genome coverage of over 95% if they are generated with several restriction enzymes that are complementary in their restriction sites. Therefore, 7,680 (20,384-well microplates) clones were randomly selected from each of the *Eco*RI and *Bam*HI BAC libraries and the *Hin*dIII BIBAC library, which represent 2.6×, 2.8× and 2.2 × rice haploid genome equivalents, respectively. The combined 23,040 (3 × 7,680) clones from all three libraries represent 7.6 × rice haploid genome equivalents. The clones were robotically double-spotted onto filters and probed with the labeled insert DNA of the 55 marker clones. From 1 to 23 positive clones per probe were identified, each probe giving at least one positive clone (Table 2). This result is in agreement with our previous studies (Tao et al. 2001) and strongly indicates that these libraries are highly amenable for physical mapping and sequencing of the entire rice genome. For each probe, on average, six positive clones were detected, which is slightly less than the predicted 7.6 × haploid genome equivalents of the combined clones. This phenomenon was widely observed in existing plant BAC libraries (e.g., Budiman et al. 2000; Meksem et al. 2000), which may be due to the fact that some clones were not spotted onto the filters and/or the inaccurate estimation of the genome size. The percentage of DNA markers that gave at least one positive clone was further analyzed for the clones from each library. In the 7,680 clones from the *Bam*HI BAC library, 41 of the 55 probes gave at least one positive clone, estimated to be 74.6%; in the 7,680 clones from the *Eco*RI BAC library, 48 of the 55 probes gave at least one positive clone, estimated to be 87.3%; and in the 7,680 clones from the BIBAC library, 40 of the 55 probes gave at least one positive clone, estimated to be 72.7%. According to Clarke and Carbon (1976), the genome coverage of the 7,680 clones of the *Bam*HI BAC library, the *Eco*RI BAC library and the *Hin*dIII BIBAC library should be 95.2%, 93.8% and 90.2%, respectively. The genome coverage estimated by library screening with the 55 DNA markers was smaller than the predicted genome coverage by 20.6% (95.2–74.6%) for the *Bam*HI BAC library, 6.5% (93.8–87.3%) for the *Eco*RI BAC library and 17.5% (90.2–72.7%) for the *Hin*dIII BIBAC library. This difference is most likely caused by the uneven distribution of each of the restriction enzyme sites in the genome, and would lead to a large number of gaps in the physical and sequence maps if a library generated with either of the enzymes is used. This has further demonstrated the importance of developing these complementary libraries for finally physical mapping and sequencing of the entire rice genome.

To confirm the utilities of these libraries for global physical mapping of the rice genome, we fingerprinted the BACs and BIBACs identified with 10 of the 55 probes using the DNA sequencing gel-based restriction fingerprinting method (Chang et al. 2001; Tao et al. 2001; Zhang and Wu 2001). From 20 to 98 bands were generated for each BAC. The BAC and BIBAC fingerprints were scanned into image files, edited, and assembled into contigs using the FPC package (Soderlund et al. 1997). All BACs and BIBACs identified with a single probe were assembled into a single contig, whereas the BACs and BIBACs identified with different probes were assembled into separate contigs (e.g., see Fig. 3). This result has further demonstrated the utilities of these libraries in development of a genome-wide physical map and for sequencing of the whole rice genome.

Table 2 Screening results of the Nipponbare *Bam*HI and *Eco*RI BAC libraries and the Nipponbare *Hin*dIII BIBAC library with the 55 DNA markers selected from chromosome 8 of the rice genetic linkage maps (Causse et al. 1994; Harushima et al. 1998); 7,680 clones randomly selected from each library were used in this experiment and a total of 23,040 (7.6 × genome equivalents) were screened. Letter "B" indicates that the clone was from the *Bam*HI BAC library, "H" indicates that the clone was from the *Hin*dIII BIBAC library, and "R" indicates that the clone was from the *Eco*RI BAC library

Markers	Chr.	No. of hits	Hit addresses in the libraries
C1009	8	5	B5N6, H7D16, H6E17, H3J22, R8J22
C1017	8	6	B13I15, H8G11, H5D6, H1H20, R4I16, R13G23
C1107	8	1	B9G23
C1121	8	2	B10N9, B17M19
C166	8	13	B20O10, B12C17, H9E4, H14H10, H3J8, R16F3, R3J10, R12G20, R8E22, R14G15, R14F10, R5A6, R10J22
C181	8	9	B13J14, B4A15, B5N6, B4A15, B13J14, H7D16, H3J22, H8J22, R9M24
C277	8	5	B5J9, B19I18, R4B17, R12G9, R4B14
C347	8	6	B9E12, B9J22, H8N22, H17C19, R8N13, R5B17
C390	8	5	H15H1, H8I3, H9C12, R10D11, R8I3
C626	8	6	B8L15, B3I1, H15A24, H10K10, H18K17, R19K4
C770	8	5	B13J14, B4A15, B5N6, H3J22, H8J12
C825	8	7	B13J14, B6P12, B16P17, H8J12, H14L17, R12D10, R9M24
C905	8	4	B5N6, H6E17, H7D16, R8J22
C922B	8	3	B7G4, B2O8, R6I7
C929	8	6	B19C24, R10K21, R14E12, R11K8, R19F15, R2P14
CDO116	8	2	H1D3, R8E19
G104	8	3	B11N4, H19I7, R12L11
G1073	8	4	B16O9, H8F4, H6F19, H5N8
G278	8	7	H18I16, H12N1, H17I22, R19O18, R14I4, R4F11, R5K3
G56	8	23	B6M2, B7E17, B9E6, B9M21, B17D5, B17N12, B18B15, H1F10, H1M20, H8F13, H10F2, H12L20, H12N9, H13C16, R2A5, R4B13, R4E11, R4H11, R5B17, R8B21, R14H3, R17M9, R20J23
R1010	8	8	B4A15, B6P12, B16P17, H8J12, H14L17, R5G9, R12D10, R9M24
R1394A	8	7	B13I11, H17M8, H8I7, H4E21, R12F7, R14J5, R4G18
R1813	8	3	H1L3, R17K3, R19O18
R1943	8	5	B5N21, H9K6, H9I17, H20H6, R6O9
R1963	8	6	R4B13, R4G5, R12J8, R20J23, R8D10, R8G5
R2007	8	6	B19I11, B10E8, B13D11, H2L24, R15D5, R1E1
R202	8	1	B7F13
R2367	8	5	B1D6, R6K17, R20E13, R9G24, R17J2
R2662	8	3	H20I1, R15D10, R15B20
R2676	8	11	B15O1, B10I2, B17N7, B18E12, H4C23, R15B14, R4M24, R18B23, R20P3, R15B14, R14O3
R2976	8	7	B13I11, H17M8, H8I7, H4E21, R12F7, R14J5, R4G18
R622	8	3	H13N6, R8C18, R15M4
R727	8	20	B10H14, B14L10, B14F21, B1F13, B14L6, B10L20, H4E15, H17B6, H11N17, R10E18, R7L6, R3H14, R4N8, R11N12, R8P8, R2I20, R2L5, R8O3, R9E19, R19F18
R902	8	8	B4J24, B6B15, H12K15, H4F23, H18J7, H12N21, R16I15, R1J18
RG1	8	5	B18A17, R10F10, R6L13, R19B20, R1L10
RG136	8	9	B19B13, B6P7, B12E18, B17E12, H14B12, H16F11, R11F19, R8G8, R16M17
RG28	8	9	B14F21, B1F13, B14L6, B10L20, H17B6, H11N17, R16H21, R9E19, R19F18
RG29X	8	1	R4B14
RG333X	8	4	H5C19, H16A7, H12K3, R19C6
RG885	8	4	B14F20, B19M23, H11H4, R19E17
RG978	8	5	B13J14, B4A15, B5N6, H3J22, H8J12
RZ323	8	4	B9E23, B8H1, H5D2, H6K10
RZ572	8	8	B17H21, B8N14, H4K3, H5E3, H6B1, H20N22, H20H18, R4E16
RZ617	8	6	B17C10, B6K19, B7L10, H2P16, R18C16, R8P21
RZ649	8	3	B7G4, B2O8, R6I7
RZ926	8	2	H12M5, R14B3
RZ952	8	7	B13I11, H17M8, H8I7, H4E21, R12F7, R14J5, R4G18
RZ997	8	4	B3J19, B20N24, R8D10, R8G5
S10324B	8	11	B12C17, B20O10, R12G20, R8E22, R14G15, R14F10, R9K11, R3J10, R16F3, R15B18, R5A8
S0622	8	1	R15K6
S10631	8	7	B12B18, H10J7, H11D11, H1C9, R2A1, R6F13, R6G14
S1633A	8	6	H8F13, H12A16, H12L20, H13C16, R14H3, R8P10
S1850A	8	6	B4J24, H12N21, H18J17, H12K15, R1J18, R16I15
S2014	8	7	H10L24, H6A9, H14H10, H16J8, R12I5, R13J19, R6C6
S815	8	2	R4M6, R13N1

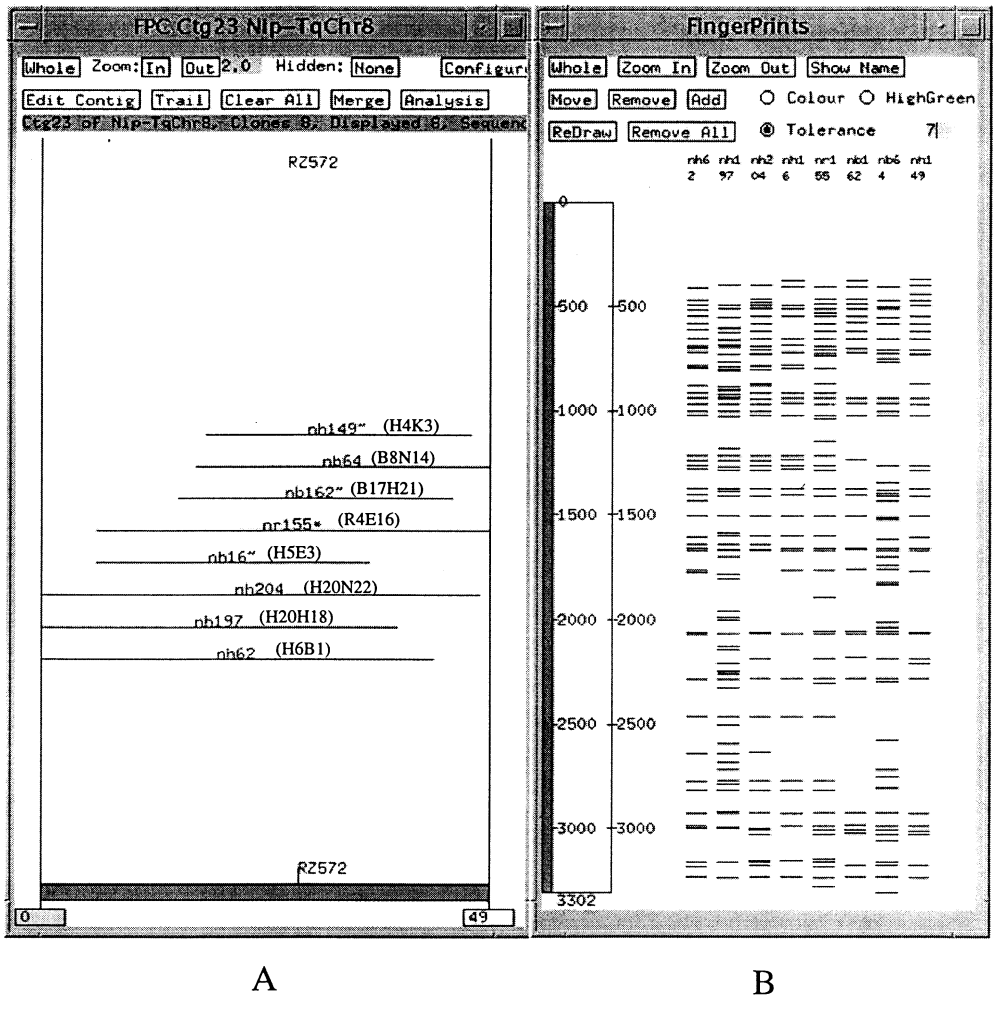

Fig. 3 A contig of the BACs and BIBACs identified with RZ572 (**A**) and the digitized fingerprints of the BACs and BIBACs (**B**). The contig consists of 49 unique bands, approximately equivalent to 300 kb.The clone prefixed with "R'' (in parenthesis) was from the *Eco*RI BAC library, the clones prefixed with "B'' (in parenthesis) were from the *Bam*HI BAC library, and the clones prefixed with "H'' (in parenthesis) were from the *Hin*dIII BIBAC library

Discussion

We have developed one large-insert BIBAC library and three large-insert BAC libraries with three restriction enzymes in two different vector systems (pECBAC1 – bacterial F factor-based, Frijters et al. 1997; pCLD04541 – bacterial P1-based, Jones et al. 1992 and Tao and Zhang 1998) for japonica rice cv Nipponbare (see Table 1 and Fig. 1). These libraries represent the most comprehensive resource for rice genome research to-date. They are complementary to each other, and have the largest insert sizes and the deepest genome coverage among all existing rice BAC and PAC libraries. Of these libraries, the *Bam*HI BAC library is a new library for Nipponbare and the BIBAC library is the first plant-transformation-competent large-insert DNA library generated for rice.

Global physical mapping and sequencing of a complex genome require readily usable DNA libraries of large inserts, high integrity and complete genome coverage (International Human Genome Mapping Consortium 2001; International Human Genome Sequencing Consortium 2001). The four Nipponbare BAC and BIBAC libraries developed in this study have provided such kinds of resources for genome research of rice and other grasses. Screening of the *Bam*HI and *Eco*RI BAC libraries and the *Hin*dIII BIBAC library with the 55 selected DNA markers indicates that these libraries are well-balanced in distribution of clones in the rice genome, complemented to each other, and cover the entire rice genome (Table 2). This indicates that these libraries are well-suited for genome research of rice and other grasses. The positive clones identified with the DNA markers have provided a set of anchor BACs as an initial step for physically mapping and sequencing chromosome 8 for which large-scale physical mapping and sequencing have not been initiated yet. Restriction fingerprint analysis and contig assembly has been proven to be a powerful approach to rapidly develop genome-wide physical maps from bacteria-based large-insert random clones (Hodgkin et al. 1995; Coulson et al. 1986; Gregory et al. 1997; Marra et al. 1997, 1999; Soderlund et al. 1997; Zhang and Wing 1997; Ding et al. 1999; Hoskins et al. 2000; Chang et al. 2001; International Human Genome Mapping Consortium 2001; Tao et al. 2001; Zhang and Wu 2001). The fingerprint analysis and contig assembly of the BAC and BIBAC clones identified with mapped DNA markers in this study has further demonstrated the feasibility of these libraries for eventual physical mapping and sequencing of the entire Nipponbare genome.

The Nipponbare BAC and BIBAC libraries developed in this study have provided useful resources for genome research of rice and other gramineous crops. It is desirable for various goals of genome research to have a DNA library that has large inserts and is suited for direct plant transformation. The direct transformability of large-insert clones into plants will streamline map-based cloning, functional analysis of genome sequence, gene, gene cluster and QTL engineering and plant molecular breeding. Furthermore, if such a binary library could be incorporated into integrative physical and genetic mapping and genome sequencing, the efficacy of the above research would be increased by several fold. The Nipponbare BIBAC library developed in this study has provided such a resource for rice genome research. The binary vector pCLD04541 used for the Nipponbare BIBAC library has routinely been used to transform the DNA fragments of about 30 kb in a variety of plant species (e.g., Jones et al. 1992; Bent et al. 1994). It has also been shown that the large-insert clones of pCLD04541 were stable in *Agrobacterium* (Wu et al. 2000) and could be transformed into plants (Y. Wu, personal communication). It should be possible to transform large-insert (150-kb) pCLD04541 clones via *Agrobacterium* in rice, although further investigation is needed to establish the related techniques. The *Agrobacterium*-mediated transformation techniques have been established in rice (Hiei et al. 1994; Dong et al. 1996). Techniques for large-insert (about 150-kb) clone transformation via *Agrobacterium* have been established in Arabidopsis (Liu et al. 1999), rice (Choi et al. 2000, Plant and Animal Genome VIIII, Abstract #P609), tobacco (Hamilton et al. 1996) and tomato (Hamilton et al. 1999).

The libraries developed in this study have added additional, comprehensive tools to finally sequencing the complete rice genome and to rice genome research in general. This is especially true for the *Bam*HI BAC library and the BIBAC library. First, the *Bam*HI library is the only one that has been developed with *Bam*HI for Nipponbare so far. The G/C content consistence of the *Bam*HI sites with the rice genome makes the *Bam*HI library complementary to the BAC and PAC libraries being used in rice genome sequencing. Second, the BIBAC library was cloned in a bacterial plasmid P1-based vector (pCLD04541), differing from those used for the construction of existing BAC (bacterial plasmid F-based) and PAC (bacteriophage P1-based) libraries. This library is also helpful to clone the genomic regions that could not be cloned in the Nipponbare BAC and PAC libraries being used in genome sequencing. Third, the three new BAC libraries, especially the *Bam*HI library (170 kb), have much larger average insert sizes (150, 156, 170 kb) than any of the existing three Nipponbare BAC and PAC libraries (112, 120, 130 kb). The significantly larger inserts of the new BAC libraries will facilitate not only map-based cloning and the larger contig assembly of the rice physical map (Zhang et al. 1996b), but also long-range sequence assembly of the genomic regions containing large-segment repeats (International Human Genome Sequencing Consortium 2001; Stupar et al. 2001). Additionally, we have developed the physical maps of the entire centromeric regions for all 12 chromosomes of the rice genome from these new libraries (H.-B. Zhang, unpublished) for which it is intractable to develop as indicated by the *Arabidopsis* (Arabidopsis Genome Initiative 2000), *Drosophila* (Adams et al. 2000; Hoskins et al. 2000) and human (International Human Genome Sequencing Consortium 2001; Venter et al. 2001) genome projects. These physical maps will accelerate the process of sequencing the entire rice genome. We are also developing a BAC/BIBAC-based, integrated physical, genetic and sequence map of the rice genome from the existing and new Nipponbare libraries. This integrated map will provide a platform for functional analysis of the rice genome sequence, and for transferring of the sequence information from japonica rice (Nipponbare) to indica rice (Teqing, Tao et al. 2001), the rice subspecies that provides over 90% of the world rice production.

All four Nipponbare BAC and BIBAC libraries have been made available to the public at the GENE*finder* Genomic Resources, Texas A&M University. To request a copy of these libraries, high-density colony filters of the libraries and/or individual clones, please contact the Resources at http://hbz.tamu.edu – BAC Library – Library List.

Acknowledgements This study was partially supported by the Texas Higher Education Coordinating Board (999902-042), Texas Agricultural Experiment Station (8536-203104), and the Rockefeller Foundation (RF97001#555) to H.-B.Z. The authors thank Paul P. Ling, Chantel Scheuring and Teofila S.J. Santos for critically reading the manuscript.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2204

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

Ahn S, Anderson JA, Sorrells ME, Tanksley SD (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, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ (1994) RPS2 of *Arabidopsis thaliana*: a leucine-rich repeated class of plant disease resistance genes. Science 265:1856–1860

Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. Genome Res 10:129–136

Causse MA, Fulton TM, Cho YG, Ahn AN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251–1274

Chang Y-L, Tao Q, Scheuring C, Ding K, Meksem K, Zhang H-B (2001) An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. Genetics 159:1231–1242

Chen M, SanMiguel P, De Oliveira AC, Woo S-S, Zhang H-B, Wing RA, Bennetzen JL (1997) Microlinearity in *sh-2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 94:3431–3435

Clarke L, Carbon J (1976) A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell 9:91–100

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

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3–15

Ding Y, Johnson MD, Colayco R, Chen YJ, Melnyk J, Schmitt H, Shizuya H (1999) Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting. Genomics 56:237–246

Dong J, Teng W, Buchholz WG, Hall TC (1996) *Agrobacterium*-mediated transformation of javanica rice. Mol Breed 2:267–276

Frengen E, Weichenhan D, Zhao B, Osoegawa K, van Geel M, de Jong PJ (1999) A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. Genomics 58:250–253

Frijters ACJ, Zhang Z, van Damme M, Wang G-L, Ronald PC, Michelmore RW (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 SG, Howell GR, Bentley DR (1997) Genome mapping by fluorescent fingerprinting. Genome Res 7:1162–1168

Hamilton CM, Frary A, Lewis C, Tanksley SD (1996) Stable transfer of intact high-molecular-weight DNA into plant chromosomes. Proc Natl Acad Sci USA 93:9975–9979

Hamilton CM, Frary A, Xu Y, Tanksley SD, Zhang H-B (1999) Construction of tomato genomic DNA libraries in a binary-BAC (BIBAC) vector. Plant J 18:223–229

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2,275 markers using a single F2 population. Genetics 148:479–494

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J 6:271–282

Hodgkin J, Plasterk RHA, Waterston RH (1995) The nematode *Caenorhabditis elegans* and its genome. Science 270:410–414

Hoskins RA, Nelson CR, Berman BP, Laverty TR, George RA, et al. (2000) A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. Science 287:2271–2274

International Human Genome Mapping Consortium (2001) A physical map of the human genome. Nature 409:934–941

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Jones JG, Shlumukov L, Carland F, English J, Scofield SR, Bishop GJ, Harrison K (1992) Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. Transgenic Res 1:285–297

Khush GS, Kinoshita T (1991) Rice karyotype, marker genes, and linkage groups. In: Khush GS, Toenniessen GH (eds) Rice biotechnology. CAB International, Wallingford, UK, pp 83–108

Kim U-J, Birren BW, Slepak T, Mancino V, Boysen C, Kang H-L, Simon MI, Shizuya H (1996) Construction and characterization of a human bacterial artificial chromosome library. Genomics 34:213–218

Liu Y-G, Shirano Y, Fukaki H, Yanai Y, Tasaka M, Tabata S, Shibata D (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. Proc Natl Acad Sci USA 96:6535–6540

Marra MA, Kucaba TA, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997) High throughput fingerprint analysis of large-insert clones. Genome Res 7:1072–1084

Marra MA, Kucaba T, Sekhon M, Hiller L, Martienssen R, Chinwalla A, Crockett J, Fedele J, Grover H, Gund C, McCombie WR, McDonald K, McPherson J, Mudd N, Parnell L, Schein J, Seim R, Shelby P, Waterston R, Wilson R (1999) A map for sequence analysis of the *Arabidopsis thaliana* genome. Nature Genet 22:265–270

Meksem K, Ruben E, Zobrist K, Hyten D, Tao Q, Zhang H-B, Lightfoot AD (2000) Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome-wide physical mapping. Theor Appl Genet 101:747–755

Moore G, Devos KM, Wang Z, Gale MD (1995) Gramineouses, line up and form a circle. Curr Biol 5:737–739

Moullet O, Zhang H-B, Lagudah ES (1999) Construction and characterization of a large DNA insert library from the D genome of wheat. Theor Appl Genet 99:305–313

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R, Meier-Ewert S, Altmann T (1999) A complete BAC-based physical map of the *Arabidopsis thaliana* genome. Nature Genet 22:271–275

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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. CABIOS 13:523–535

Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang J (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proc Natl Acad Sci USA 98:5099–5103

Tao Q, Chang Y-L, Wang J, Chen H, Schuering C, Islam-Faridi MN, Wang B, Stelly DM, Zhang H-B (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158:1711–1724

Tao Q-Z, Zhang H-B (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

Umehara Y, Miyazaki A, Tanoue H, Yasukochi Y, Saji S, Otsuki Y, Fujimura T, Kurata N, Minobe Y (1995) Construction and characterization of rice YAC library for physical mapping. Plant Breed 1:79–89

Venter JC, Adams MD, Myers EM, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291:1304–1351

Wu Y, Tulsieram L, Tao Q, Zhang H-B, Rothstein SJ (2000) A binary vector-based large insert library for *Brassica napus* and identification of clones linked to a fertility restorer locus for Ogura cytoplasmic male sterility (CMS). Genome 43:102–109

Zhang H-B, Wing RA (1997) Physical mapping of the rice genome with BACs. Plant Mol Biol 35:115–127

Zhang H-B, Wu C (2000) BAC as tools for genome sequencing. Plant Physiol Biochem 39:195–209

Zhang H-B, Zhao X, Ding X, Paterson AH, Wing RA (1995) Preparation of megabase-size DNA from plant nuclei. Plant J 7:175–184

Zhang H-B, Choi S, Woo S-S, Li Z, Wing RA (1996a) 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

Zhang H-B, Woo S-S, Wing RA (1996b) BAC, YAC and Cosmid Library Construction. In: Foster G, Twell D (eds) Plant gene isolation: principles and practice. John Wiley and Sons, Ltd, England, pp 75–99