Molecular Breeding **2**: 11–24, 1996.

Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population

Hong-Bin Zhang, Sangdun Choi, Sung-Sick Woo, Zhikang Li, and Rod A. Wing*
*Department of Soil and Crop Sciences, Crop Biotechnology Center, Texas A&M University, College Station, TX 77843-2123, USA (*author for correspondence)*

Received 22 February 1995; accepted in revised form 28 August 1995

Key words: bacterial artificial chromosome (BAC), cereal, genome analysis, map-based cloning, monocotyledon, rice

Abstract

Rice is a leading grain crop and the staple food for over half of the "world" population. Rice is also an ideal species for genetic and biological studies of cereal crops and other monocotyledonous plants because of its small genome and well developed genetic system. To facilitate rice genome analysis leading to physical mapping, the identification of molecular markers closely linked to economic traits, and map-based cloning, we have constructed two rice bacterial artificial chromosome (BAC) libraries from the parents of a permanent mapping population (Lemont and Teqing) consisting of 400 F9 recombinant inbred lines (RILs). Lemont (*japonica*) and Teqing (*indica*) represent the two major genomes of cultivated rice, both are leading commercial varieties and widely used germplasm in rice breeding programs. The Lemont library contains 7296 clones with an average insert size of 150 kb, which represents 2.6 rice haploid genome equivalents. The Teqing library contains 14208 clones with an average insert size of 130 kb, which represents 4.4 rice haploid genome equivalents. Three single-copy DNA probes were used to screen the libraries and at least two overlapping BAC clones were isolated with each probe from each library, ranging from 45 to 260 kb in insert size. Hybridization of BAC clones with chloroplast DNA probes and fluorescent *in situ* hybridization using BAC DNA as probes demonstrated that both libraries contain very few clones of chloroplast DNA origin and are likely free of chimeric clones. These data indicate that both BAC libraries should be suitable for map-based cloning of rice genes and physical mapping of the rice genome.

Introduction

Arabidopsis thaliana has been used as a model species for genetic and biological studies of dicotyledonous plants because of its well developed genetic system and small genome (C = 145 Mb) [5]. Such a model species is needed for monocotyledonous plants not only because of their biological differences with dicotyledonous plants but also their economic importance. Monocotyledonous plants include many important grain, forage, and other economic crops. The best known grain crops are wheat, rice, maize, barley, sorghum, pearl millet, oats, and rye, which provide almost

all of the grain food for man and domestic animals. Rice, *Oryza sativa* L., is an ideal species for genetic and biological studies in monocotyledonous plant species [33, 34]. It is diploid ($2n = 2x = 24$,) and has the smallest genome among all grain crop species (C = 420 Mb) [5]. Rice has a genome size about 3-fold larger than *A. thaliana*, but over 5-fold smaller than maize, and over 37-fold smaller than bread wheat. Rice has a well established classical genetic map [24], two high-density molecular linkage maps [12, 26], a yeast artificial chromosome library [41], and a well developed genetic transformation system [6, 11, 21, 22]. The rice map of Causse *et al.* [12] contains 720 DNA markers, the average physical distance of which is 290 kb/cM and 580 kb between adjacent DNA markers. The rice map of Kurata *et al.* [26] contains 1383 DNA markers including 883 expressed sequences, with an average physical distance of 270 kb/cM and an interval of 300 kb between adjacent DNA markers. These physical distances are comparable to that of *A. thaliana*. Recent comparative genome mapping has revealed that the rice genome has many homeologous and conservative regions with other cereal crop genomes [2, 3]. Therefore, rice genome analysis, the construction of overlap contigs, and map-based gene cloning will directly contribute to the genetic and biological studies of rice and other monocotyledonous plants.

The *Arabidopsis* genome project has led to a rapid progress in plant genetics and biology. An essential element for such research is a large insert DNA library. Two vector systems have been developed for cloning large DNA fragments (> 100 kb): yeast artificial chromosome (YAC) [10] and bacterial artificial chromosome (BAC) [39] cloning vectors. In the YAC cloning system, large fragments of exogenous DNA are cloned into a linearized YAC cloning vector and maintained in yeast. In the BAC cloning system, large fragments of exogenous DNA are cloned into an *Escherichia coli* F factor-based plasmid vector and are maintained in *E. coli*. In comparison, BAC libraries are relatively easy to construct, have lower levels of chimeric clones [36, 39, 45], and a large amount of pure BAC DNA is easier to isolate than YAC DNA. These attributes are very important when the library is used for map-based cloning, genome analysis, and the construction of overlap contigs.

YAC DNA libraries have been constructed for a number of plant species including *A. thaliana* [14, 15, 20, 43], maize [16], sugarbeet [17], tomato [31], rice [41], and barley [25]. Some of these YAC libraries have been successfully used for map-based cloning of genes of economic importance [4, 7, 13, 28, 32, 35], the isolation of molecular markers closely linked to or cosegregating with economic traits [25, 31, 37, 47], and the analysis of large genes or specific chromosome regions [37, 47]. Recently, a BAC library has been constructed for sorghum, which has an average insert size equivalent to those of a majority of the plant YAC libraries [45].

To facilitate rice genome research, we have constructed two rice BAC libraries from the parents of a permanent recombinant inbred line (RIL) mapping population [29, 30]. A RIL mapping population is extremely important for genetic studies since it can be easily reproduced and distributed without significant genetic changes. The mapping population consists of 400 F9 RILs and has been developed from a cross between two leading commercial rice varieties, Lemont and Teqing. Lemont is a US variety which has desirable productivity and high quality while Teqing is a Chinese variety which has high productivity. Lemont belongs to ssp. *japonica* and Teqing belongs to ssp. *indica*, which represent the two major genomes of cultivated rice. These two varieties show a high level of polymorphism at the RFLP level [29] and the RIL mapping population segregates for many traits of agronomic importance, including heading date, disease resistance (bacterial blight and blast), fertility, and many component traits for plant type and yield [29, 30, and Z. Li, unpublished]. A molecular linkage map has been developed for this population [29] and a saturated molecular linkage map is under development (Z. Li, unpublished). In this paper, we report the construction and characterization of these two rice BAC libraries and their applications for rice genome research.

Materials and methods

Plant materials

Two rice varieties, Lemont and Teqing, were grown in a greenhouse. Plants 4 to 6 weeks old were transferred into a dark condition and kept for 24–48 h to reduce the carbohydrate content. Whole plants were harvested, cut into small pieces, frozen in liquid nitrogen, and stored at −80 °C.

Enzymes and chemicals

T4 DNA ligase was purchased from US Biochemicals (USA), β-agarase from New England Biolabs (USA), and proteinase K from Boehringer Mannheim (Germany). Restriction enzymes, calf intestine alkaline phosphatase (CIAP), ultra pure agarose, low-melting-point (LMP) agarose, Luria broth (LB) base, and ElectroMAX DH10B *E. coli* cells were purchased from Gibco BRL (USA). Spermine, spermidine, sucrose (molecular biology grade), and other chemicals were obtained from Sigma (USA).

DNA probes

Nuclear DNA probes RG348 and TGMS1.2 are single-copy RFLP markers from rice, which were kindly provided by Dr S. McCouch, Cornell University and Dr H. Nguyen, Texas Tech University, USA, respectively. pcSh2-1a is a portion of a cDNA of the *Shrunken-2* gene from maize [9], kindly provided by Dr J. Bennetzen, Purdue University, USA. The three barley chloroplast DNA probes, pBHP20, pBPH134, and pBHE319, were kindly provided by Dr J. Mullet, Texas A&M University, USA. pBHP20 is a clone for the *ndhA* gene, pBHP134 for the *rbcL* gene, and pBHE319 for the *psbA* gene. The three probes space almost equally apart in the barley chloroplast genome, ca. 50 kb apart. In all cases, the inserts of the clones were purified on an agarose gel, radioactively labeled [18], and used as probes.

Preparation of BAC vector

pBeloBAC 11 was used as the cloning vector for the construction of the BAC libraries, which was kindly provided by Drs M. Simon and H. Shizuya (unpublished; for address, see [39]). The vector DNA was isolated from a 2 liter overnight culture by the alkaline lysis method as described by Sambrook *et al.* [38]. RNA was removed by incubating with 20 μg/ml DNase free RNase for 45 min followed by extraction with an equal volume of equilibrated phenol once and phenol/chloroform/isoamyl alcohol (25:24:1) once. The DNA was precipitated with ethanol, dried, and dissolved in H_2O.

The pBeloBAC 11 DNA was digested with 3 units of *Hin*d III per microgram DNA at 37 °C. After 2 h of incubation, an additional 2 units of *Hin*d III per microgram DNA was added and the incubation continued for additional 2 h. The digestion was tested on an agarose gel and by transformation to ensure that the digestion was indeed complete (circular plasmid DNA molecules were <2%). The digest was extracted with phenol/chloroform (1:1), precipitated, dried, and dissolved in H_2O. To dephosphorylate the ends of linearized vector DNA, a preliminary experiment was performed to optimize the dephosphorylation conditions to obtain over 95% recombinant clones when the vector DNA was ligated with a *Hin*d III digest of lambda DNA and transformed into *E. coli*. Using the optimal conditions, a large scale of linearized vector DNA was dephosphorylated at 0.1 units of CIAP per pmol of DNA at 37 °C for 30 min. The reaction was stopped by adding 5 mM EDTA, pH 8.0, 0.5% SDS (sodium dodecyl sulfate), and 0.1 mg/ml proteinase K and incubating at 56 °C for 30 min, extracted twice with phenol, once with phenol/chloroform/iso-amy alcohol (25:24:1), precipitated with ethanol, dried, and dissolved in H_2O at 10 ng/μl. The dephosphorylation was tested as above and the vector DNA was stored in 10 μl aliquots at −80 °C.

Isolation of high-molecular-weight (HMW) DNA from rice nuclei

Rice HMW DNA from Lemont and Teqing was isolated from nuclei according to Zhang *et al.* [48]. Nuclei prepared from 240 g (fresh weight) of the tissue of each variety were embedded in 8 ml LMP agarose microbeads.

Preparation of Hin*d III-partially digested rice HMW DNA*

The partial digestion of rice HMW DNA embedded in LMP agarose microbeads was conducted and analyzed according to Zhang *et al.* [48] with modifications. Previous studies showed that it is necessary to perform a preliminary experiment for each preparation of HMW DNA to determine the optimal conditions for partial digestion. The digestion reaction of HMW DNA in microbeads was prepared as that for conventional aqueous phase DNA digestion. Before adding the *Hin*d III enzyme, the mixture, including DNA in microbeads, 1 × reaction buffer, and 2 mM spermidine, was incubated on ice for one hour and then aliquoted into different tubes. A series of dilutions of *Hin*d III was added to different tubes at 1/50 of the total reaction volume, except for one tube which was used as a control (no enzyme was added). The reactions were incubated on ice for additional 30 min, then transferred into a 37 °C water bath, and incubated for one hour. The reactions were transferred onto ice first and then stopped by adding 1/10 volume of 0.5 M EDTA, pH 8.0. The partial digests were analyzed by pulsed-field gel electrophoresis (PFGE). The partial digestion conditions that showed a majority of restricted DNA fragments from 120 to 400 kb in size on the pulsed-field gel were selected for large-scale partial digestion of HMW DNA used for the construction of the BAC libraries. The large scale *Hin*d III partial digest was tested on a pulsed-field gel to confirm the partial digestion and the digest was stored at 4 °C before size selection.

Size selection of partially digested HMW DNA by PFGE

Size selection of partial digests of rice HMW DNA was performed on a pulsed-field gel using the CHEF Mapper (BioRad, USA). The gel was prepared with 1% LMP agarose in 1 × TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) with a sample trough ca. 4 mm wide at the top of the gel. The microbeads containing the partially digested DNA fell to the bottom of the tube during storage. The supernatant fluid of digestion mixture was gently removed with a pipette, and the microbeads were loaded into the trough and sealed in position with the molten LMP agarose. The lambda DNA concatemer ladder (ICN, USA) was used as the molecular weight marker and loaded in the wells adjacent to each side of the sample trough. The gel was subjected to PFGE in 1 × TAE buffer and at 5 V/cm, 12.5 °C, and 120° angle with a 90 s pulse for 15 h (for Lemont BAC library) or a 110 s pulse for 22 h (for Teqing BAC library). After electrophoresis, the lanes containing the lambda DNA ladder were removed and stained with ethidium bromide. The regions of the lanes containing lambda DNA ladders from 120 to 400 kb in size were marked and used for locating the region of the gel containing rice restricted DNA fragments of the same size range. The region containing the rice DNA fragments from 120 to 400 kb was excised from the gel using a glass coverslip. For the construction of the Teqing BAC library, this gel region was sliced into pieces of ca. 80 μl in volume and stored in 50 mM EDTA, pH 8.0, 30 mM NaCl at 4 °C. For the construction of the Lemont BAC library, this gel region was replaced in the sample trough of a new 1% LMP agarose gel prepared as above for the second size selection. The gel was subjected to PFGE at 4.0 V/cm, 12.5 °C, and a 5 s pulse for 9 h. As described above, the compressed portion of DNA in the gel was localized, excised from the gel, sliced into pieces of ca. 80 μl, and stored at 4 °C.

Ligation of rice DNA fragments into the BAC vector

The LMP agarose gel slices containing the size-selected rice DNA fragments were dialyzed with ice-cold TE (10 mM Tris, 1 mM EDTA pH 8.0) plus 30 mM NaCl three times, one hour each time. After the last wash, the solution was completely removed. The gel slice was melted at 68 °C for 10–15 min, transferred to a 40 °C water bath, and digested with β-agarase for 2 h as specified by the manufacturer. After one hour of digestion, the DNA concentration of the reaction was determined by electrophoresing an aliquot of 20 μl on an 1% agarose gel using lambda DNA of known concentration as a standard. When the gel was completely digested, the vector DNA was added at 5:1 molar ratio of vector DNA/rice DNA and incubated at 56 °C for 10 min. The mixture was cooled to room temperature, and 1/10 volume of 10 × ligation buffer supplied by the manufacturer and 2 units of T4 DNA ligase were added per 100 μl of reaction containing approximately 40 ng rice DNA. To test the ligation, 20 μl of the reaction was gently removed with a cut-off pipette tip and mixed with 1 μl 0.5 μg/μl *Hin*d III-digested lambda DNA. The reaction was incubated at 16 °C overnight. The ligation was confirmed by analyzing the test ligation on an agarose gel using the unligated *Hin*d III-digested lambda DNA as a control.

Transformation of the recombinant DNA molecules (BAC) into E. coli *by electroporation*

Escherichia coli strain DH10B was used as the host [39]. The ligated DNA was transformed into the ElectroMAX DH10B cells by electroporation using a Cell Porator and Voltage Booster system (BRL, USA) as specified by the cell provider (BRL, USA): 1 μl of ligation was added to 24 μl of the cells for a single electroporation. The Cell Porator settings are 400 V, 330 μF capacitance, low ohms impedance, and fast charge rate, and the Voltage Booster setting is 4 kΩ resistance. After electroporation, the cells were resuspended in SOC medium (2% bacto-tryptone, 0.5% bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose pH 7.0) and incubated at 37 °C for one hour with shaking at 200 rpm. The cells were then plated on LB medium (1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl pH 7.5, 1.5% agar) containing 12.5 μg/ml chloramphenicol, 7 μl 200 μg/ml isopropyithio-β-D-galactoside (IPTG) and 70 μl 20 μg/ml 5-bromo-4-chloro-3-indolyl-β-galactoside (X-gal) per 100 mm diameter plate (28 μl 200 μg/ml IPTG and 280 μl 20 μg/ml X-gal for each 150 mm diameter plate), and grown at 37 °C for 35–48 h. At this time, the recombinant clones (BACs) could be clearly identified by a blue (nonrecombinant) or white (recombinant) phenotype.

Individual BAC analysis

Individual BAC clones were inoculated into 5 ml of LB broth containing 12.5 μg/ml chloramphenicol and grown at 37 °C with shaking at 250 rpm overnight. The circular BAC DNA was isolated by the alkaline lysis method used for the preparation of plasmid DNA [38]. One quarter of the BAC DNA from the 5 ml overnight culture was digested with *Not* I to release the rice DNA insert from the cloning vector for at least 3 h. The digested BAC DNA was analyzed by PFGE with the CHEF Maper (BioRad, USA) or the CHEF (CBS Scientific HEX CHEF 6000, BioRad, USA) using lambda concatemer DNA ladder as the molecular weight marker. The conditions of PFGE for the CHEF Maper were: 1% agarose gel, 0.5 × TBE (1 × TBE is 89 mM Tris-borate, 2 mM EDTA pH 8.0), 12.5 °C, 5 V/cm, 5 s initial pulse time, 15 s final pulse time, linear ramp, and 120° angle for 13 h. The PFGE conditions for the CHEF were: 1% agarose, 0.5 × TBE, 12.5 °C, 175 V, and 15 s pulse for 16 h. The gel was stained with ethidium bromide and photographed. The insert size of each BAC clone was determined as described in the figure legends.

For Southern blotting, the BAC DNA in the gel was nicked with 60 mJ of UV light (254 nm)

using the GS Gene Linker (BioRad, USA) and blotted onto Hybond-N+ membrane (Amersham, USA) with 1.5 NaCl, 0.4 M NaOH for 24–40 h. The Southern blot was hybridized with Lemont (for clones from the Lemont library) or Teqing (for clones from the Teqing library) total DNA. The probe DNA was labeled with ^{32}P-dCTP according to Feinberg and Vogelstein [18] and Southern blot hybridization was performed according to Bernatzky and Tanksley [8].

Storage of BAC clones into ordered BAC libraries

Since over 90% of clones on the medium were recombinant (white), the white colonies as individual clones were directly transferred into a microtiter plate, each well containing 50 μl (for 384-well plates) or 200 μl (for 96-well plates) of LB broth with 35 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM sodium citrate, 0.4 mM $MgSO_4$, 5.8 mM $(NH_4)_2SO_4$, and 4.4% (v/v) glycerol. The cells in the plates were incubated at 37 °C until the medium became cloudy (from 20 to 30 hours), and then stored at −80 °C. To prevent the ordered Lemont and Teqing BAC libraries from contamination and loss and to facilitate library screening by robotics, each library was stored in triplicate, two in 384-well microtiter plates and one in 96-well microtiter plates.

BAC library screening

The ordered Lemont and Teqing BAC libraries stored in 384-well microtiter plates were replicated using a 384-pin replicating device (Genetix, UK) onto Hybond N+ membrane of 7.5 cm × 11.0 cm laid on the LB medium containing 12.5 μg/ml chloramphenicol in a Nunc bio-assay

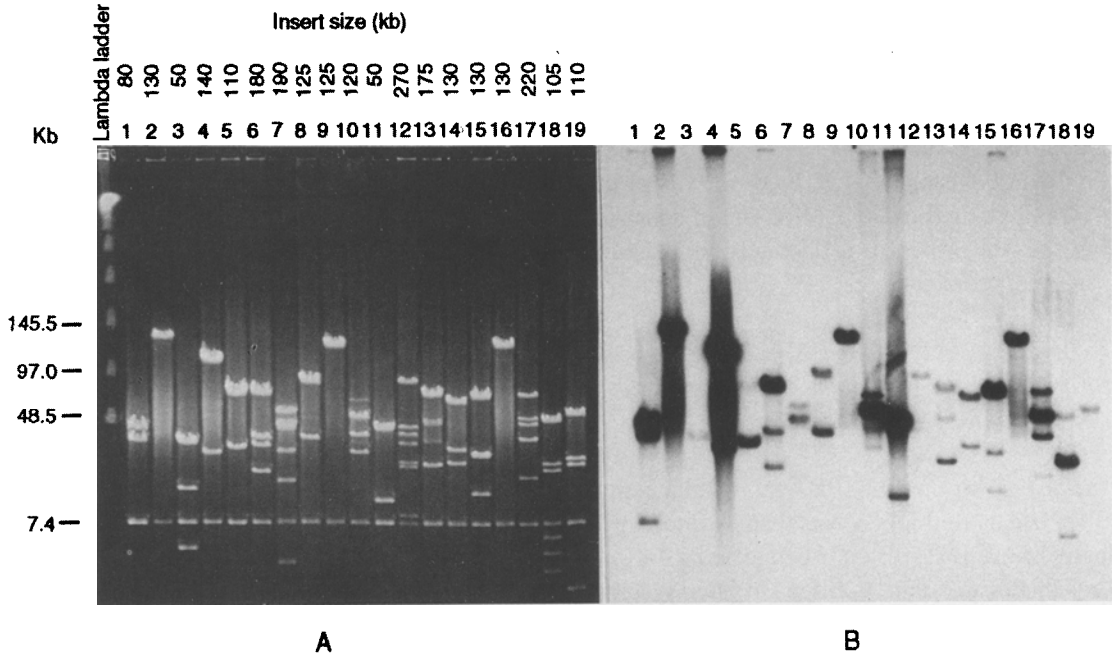

Fig. 1. Random BAC clones from the Teqing BAC library. A. BAC clones analyzed by pulsed-field gel electrophoresis (PFGE). The BAC DNA was isolated, digested with *Not* I to release the insert DNA from the cloning vector, and fractionated on a pulsed-field gel. The 7.4 kb band crossing all lanes of the gel is from the cloning vector pBeloBAC 11 and the insert size of each BAC clone is sum of sizes of all the remaining bands in a lane. B. The BACs blotted from the gel shown in A and hybridized with Teqing total DNA. Note that the fragments only containing repeated sequence elements are able to show hybridization signals when hybridized with total DNA.

dish (24.5 cm × 24.5 cm × 1.2 cm) (NUNC 140835, Denmark), 384 clones on each filter and 6 filters in each dish. The inoculate was incubated at 37 °C until the colonies were ca. 2 mm in diameter (14–16 h). The colony filter was prepared as specified by the membrane manufacturer with modification. The filter was placed colony side up onto two layers of Whatman 3 MM paper saturated with the following solutions, incubated at room temperature, and transferred in order: (1) 10% SDS, 4 min; (2) 1.5 M NaCl, 0.5 M NaOH, 7 min; (3) 1.5 M NaCl, 0.5 M Tris pH 7.2, 1 mM EDTA pH 8.0, 2 × 3 minutes; (4) 2 × SSC, 1 min; (5) on a stack of dried paper towels, 4 min; (6) 0.4 M NaOH, 20 min. The filter was then washed twice in 1000 ml of 5 × SSC, 5 min each time, and twice in 1000 ml of 1 × SSC, 0.1% SDS, 5 min each time. Afterwards, the filter was either stored wet at 4 °C or directly used for Southern blot hybridization as described above.

Results

Construction of Lemont and Teqing BAC libraries

Two rice BAC libraries were constructed, one from Lemont and the other from Teqing, both of which are the parents of a permanent RIL mapping population. The Lemont BAC library contains 7296 clones, which were constructed from the *Hin*d III partially digested fragments that were size-selected twice on pulsed-field gels. The Teqing BAC library contains 14208 clones, which were constructed from the *Hin*d III partially digested fragments that were size-selected once on a pulsed-field gel. Approximately 0.5 ng/μl of size-selected DNA fragments was obtained from the first size selection. For construction of the Teqing library, when the ligation reaction was transformed into *E. coli*, 94% of the transformants were recombinant. Approximately 200 recombinant clones were obtained from each electroporation and 400 recombinant clones from each nanogram of ligated DNA. For construction of the Lemont library, the transformation efficiency of the recombinant DNA was only approximately one quarter of that for the Teqing library. Both the Lemont and Teqing BAC libraries were constructed from single ligations.

Characterization of the BAC libraries

To test the utility of the BAC libraries for map-based cloning, genome analysis, and construction of overlap contigs, both BAC libraries were analyzed for insert size distribution, clone representation, and level of chloroplast DNA content.

Insert size distribution of the clones from the Lemont and Teqing BAC libraries

To determine the distribution of BAC insert sizes and the average insert size of the two BAC li-

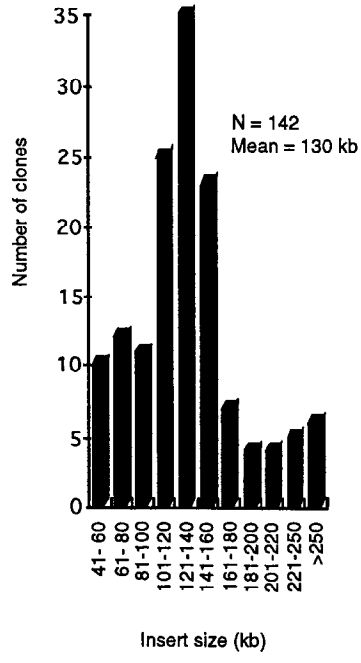

Fig. 2. Insert size distribution of BAC clones randomly taken from the Teqing BAC library. To estimate the insert size of Teqing BAC library, BAC DNA was isolated from 142 random clones and analyzed as shown in Fig. 1A. The clones were grouped by insert size and the insert size of each clone was plotted versus its frequency present in the library. The result shows that the clones in the Teqing library have an insert size range from 40 kb to 315 kb and over 60% of the clones have inserts of larger than 120 kb. The average insert size of the Teqing BAC library is 130 kb.

braries, DNA was isolated from 142 random BAC clones from the Teqing library and 130 random clones from the Lemont library. The BAC DNA was digested with *Not* I to release the rice DNA insert from the cloning vector and analyzed by PFGE (e.g., see Figs. 1A and 3A). The insert size of each clone was determined by adding up the sizes of all fragments in each lane except for the 7.4 kb BAC vector band. The clones were grouped by insert size and the insert size of each clone was plotted versus the frequency of each group of clones present in each library (Figs. 2 and 4). The Teqing BAC library has an average insert size of 130 kb, which is equivalent to 4.4 haploid genomes of rice. The insert sizes of clones in this library range from 40 to 315 kb, and over 60% of the clones have inserts larger than 120 kb (Fig. 2). Theoretically, the probability of obtaining a particular clone from this library is greater than 98%. The Lemont BAC library has an average insert size of 150 kb, which is equivalent to 2.6 haploid genomes of rice. The insert sizes of clones in this library range from 45 to 310 kb, and over 80% of the clones have inserts larger than 120 kb (Fig. 4). The probability of obtaining a particular BAC clone from the Lemont library is greater than 97%.

To further confirm that the insert DNA in the BACs was from rice and to estimate the repeated nucleotide element contents in the BAC clones, the BAC DNA on the gels shown in Figs. 1A and 3A was Southern-blotted and hybridized with total genomic DNA of Teqing for Fig. 1A (Fig. 1B) or Lemont for Fig. 3A (Fig. 3B). All BAC clones gave hybridization signals, which suggests that the insert DNA of all the BAC clones is from rice. The hybridization of the BACs with rice total DNA also suggests that repeated nucleotide elements are contained in the insert DNA of all BACs.

Representation of the Lemont and Teqing BAC libraries for rice genomes

The above analysis of insert size of the two BAC libraries indicated that the Lemont library is equivalent to 2.6 rice haploid genomes and the

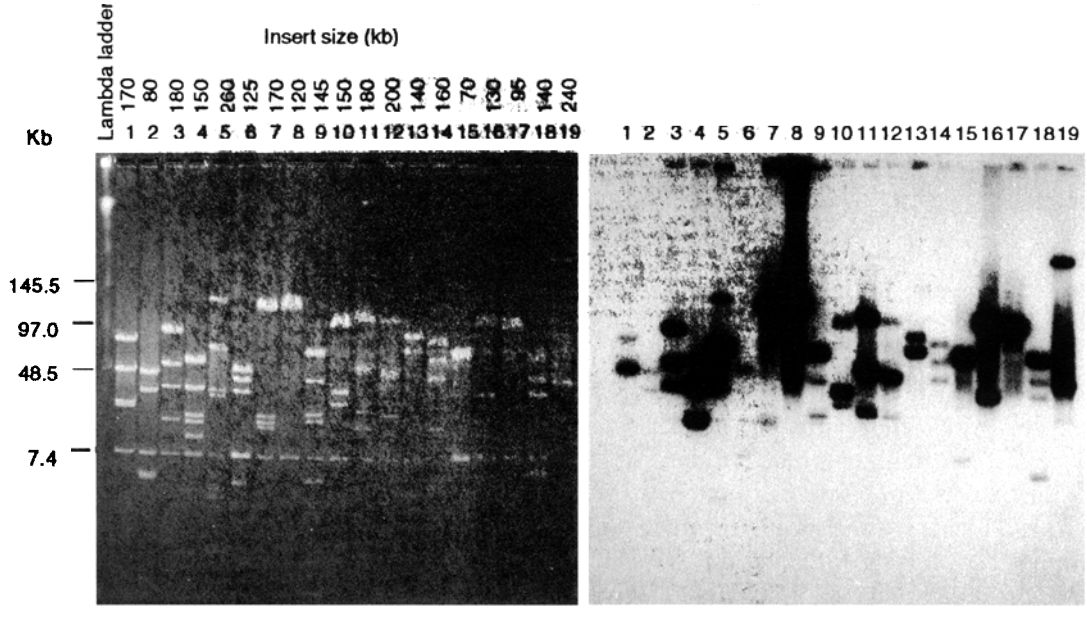

Fig. 3. Random BAC clones of the Lemont library analyzed with PFGE (A) and hybridized with Lemont total DNA (B). The insert size of each clone was estimated as that of Teqing BAC library. Note that some bands of each BAC hybridized with total Lemont DNA and others did not, indicating the distribution of repeated elements and single or low copy sequences in the DNA inserts of the clones.

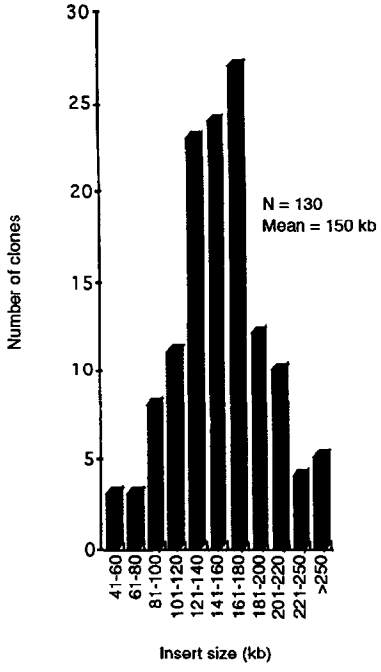

Fig. 4. Insert size distribution of BAC clones randomly taken from the Lemont BAC library. To estimate the insert size of Lemont BAC library, BAC DNA was isolated from 130 random clones and analyzed as shown in Fig. 2. The Lemont library has an insert size range from 45 kb to 310 kb and over 80% of the clones have inserts of larger than 120 kb. The average insert size of this library is 150 kb.

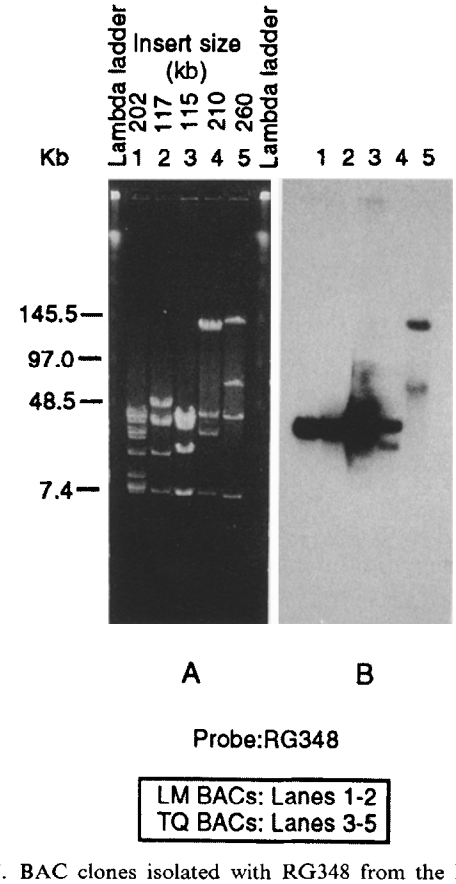

Fig. 5. BAC clones isolated with RG348 from the Lemont (lanes 1 and 2) and Teqing (lanes 3–5) BAC libraries. A. BAC clones analyzed by pulsed-field gel electrophoresis. BAC DNA was isolated, digested with *Not* I to release the insert DNA from the BAC vector and fractionated on a pulsed-field gel. The insert size of each BAC was estimated as Fig. 1A. Panel B. The BACs blotted from the gel shown in A and hybridized with the purified insert of RG348.

Teqing library is equivalent to 4.4 rice haploid genomes. Theoretically, the probability of obtaining a particular clone from each library is over 97%. To further confirm these inferences, 11520 clones (30 × 384) of the Teqing BAC library and 5912 (18 × 384) clones of the Lemont BAC library were inoculated, grown, and colony-blotted. The filters were hybridized with a RFLP marker RG348 mapped on rice chromosome 3 [29] and a probe set from the combined molecular markers TGMS1.2 and pcSh2-1a. TGMS1.2 is a single copy RFLP marker from rice and pcSh2-1a is a cDNA clone for the maize *Shrunken-2* gene [9]. The probe set was prepared by mixing TGMS1.2 and pcSh2-1a at a DNA weight ratio of 1:1 labeled with ^{32}P-dCTP. Five positive clones were isolated with RG348, ranging from 115 to 260 kb in insert size (Fig. 5). Two of the BAC clones are from the Lemont library and 3 from the Teqing library. Fourteen positive clones were isolated with the probe set of TGMS1.2 and pcSh2-1a. To confirm this result and classify these clones based on their homology with the probe, DNA was isolated from these clones, digested with *Not* I, fractionated on a pulsed-field gel, Southern-blotted, and hybridized with TGMS1.2 and pcSh2-1a, respectively. The result showed that seven clones hybridized with TGMS1.2, ranging from 45 to 210 kb in insert size, 3 of which are from the Lemont library and 4 from the Teqing library. Six clones hybridized with pcSh2-1a, ranging from 70 to 250 kb in insert size, three of which are from each library (not shown). These results confirm the above inferences and suggest

that both BAC libraries are suitable for map-based cloning of rice genes, physical mapping of the rice genome, and construction of overlap contigs. The results also show that multiple RFLP markers, even from different species, can be combined together and used for screening the BAC libraries.

Ratio of chloroplast DNA clones in the rice BAC libraries

The two BAC libraries were investigated in terms of contamination with clones originating from chloroplast DNA. The degree of contamination with chloroplast DNA clones is an important characteristic for a large insert DNA library, especially when the library is used for map-based cloning. This is because chloroplast DNA clones may mislead a chromosome walk toward a target gene when a seriously contaminated library is used for map-based cloning due to homology between some organellar and nuclear DNA sequences [40]. Additionally, if a large insert DNA library has a higher percentage of clones containing chloroplast DNA, more clones need to be generated, screened, and maintained to have an equal chance of isolating a particular nuclear DNA clone. Therefore, little or no contamination from chloroplast DNA clones is preferred for a large insert DNA library. Both Lemont and Teqing BAC libraries were constructed from DNA isolated from rice nuclei. It is expected that both libraries contain few clones generated from chloroplast DNA. To test this prediction, 1920 clones from the Lemont BAC library and 3072 clones from the Teqing BAC library were hybridized with 3 barley chloroplast DNA probes, *adh* A, *rbc* L, *psb*A, simultaneously. These three probes space approximately equally apart in the chloroplast genome such that any BAC insert larger than 50 kb should hybridize to one of the probes, indicating that the insert is probably of chloroplast DNA origin. The results showed that 2.89% (55 clones) of the Lemont clones and 0.94% (29 clones) of the Teqing clones hybridized with this probe set, indicating that 2.89% clones of the Lemont library and 0.94% clones of the Teqing library originated from chloroplast DNA. These percentages of chloroplast DNA clones are 4–15 fold lower than those of the available plant YAC and BAC libraries which were constructed from the HMW DNA isolated from protoplasts [31, 44, 45, 46]. This suggests that both libraries are well suited for map-based cloning of rice genes.

Discussion

The Lemont and Teqing BAC libraries

Large-insert DNA libraries are essential for genome analysis, construction of overlapping contigs, and map-based gene cloning. Critical components of such libraries are large insert size, genome representation and clone stability. Low levels of organellar DNA and chimeric clones also enhance the quality of the library by reducing the number of clones required for library screening. The Lemont and Tequing BAC libraries constructed in this study meet most of the above criteria and thus should be useful for many aspects of genome research.

The average DNA insert size of the Lemont and Tequing BAC libraries are 150 kb and 130 kb, respectively. These insert sizes are comparable to the majority of plant YAC and BAC libraries [16, 17, 20, 25, 31, 42, 43, 45] except for the recently reported rice and two *A. thaliana* YAC libraries with insert sizes of 350 [41] and 250 [15] and 420 kb [14], respectively. The combined Lemont and Tequing BAC libraries represent about seven haploid equivalents of the rice genome which statistically means that one has a 99% chance of obtaining at least one specific BAC clone during library screening with a specific sequence. Although the insert size of the Lemont and Tequing BAC libraries are about 2.5-fold smaller than the Japanese rice BAC library, the BAC libraries should be very useful for many aspects of rice genome research either alone or in conjunction with the rice YAC library [41]. The relatively small insert sizes of BACs to YACs might be compensated for by constructing several BAC libraries with different restriction enzymes. Such

libraries may be more suitable for assembling long continuous contigs than a single library constructed with one enzyme.

To demonstrate the estimated clone representation of both BAC libraries, each library was screened with three unlinked RFLP markers. In each case and at least two BAC clones was obtained from each library which suggests that our numerical estimates for clone representation are correct. Although this study did not specifically test the stability of large rice DNA inserts in *E. coli*, data from other BAC libraries with human [39] and sorghum [45] DNA have shown that DNA inserts larger than 300 kb can be stably maintained over 100 generations without detectable rearrangements. Such results suggest that the BAC cloning vector can faithfully maintain most large foreign DNA in *E. coli*. We have not detected any changes in insert size or restriction patterns of over 300 rice BAC clones so far analyzed by CHEF electrophoresis.

The Lemont and Teqing BAC libraries were constructed from DNA isolated from the nuclei of healthy growing plants. This is in contrast to the material used for the construction of the rice YAC library reported by Umehara *et al.* [41], for which the DNA was isolated from cultured callus cells of *O. sativa* cv. Nipponbare. DNA isolated from nuclei decreases the amount of chloroplast DNA sequences in a library, 2.9% and 0.94% for the Lemont and Teqing BAC libraries, respectively, and thus reduces the number of clones required to have a representative genomic library. Additionally, constructing a library with DNA from healthy plant tissue versus callus may help to prevent potential problems associated with the use of DNA from a tissue that is prone to DNA rearrangements and mutations [27].

Many YAC libraries have been shown to have significant levels of chimeric DNA clones [1]. For example the rice YAC library has reported that ca. 40% of the library is chimeric [41]. Although chimeric clones can be eliminated quickly by DNA fingerprinting, the use of large insert DNA libraries free of chimerism for map-based cloning and physical mapping are highly desirable. The results from all of the available BAC libraries, one human [39] and two plant [23, 45], indicate that BAC libraries have few, if any, chimeric clones as determined by fluorescent *in situ* hybridization (FISH). Preliminary FISH to interphase and metaphase rice nuclei with 15 BAC clones from the Teqing library indicate that 14 BAC clones hybridize to single pairs of unidentified rice chromosomes whereas a single BAC clone, TQ7A4 (135 kb), hybridizes strongly to a single pair of chromosomes and weakly to another pair ([23], B. S. Gill personal communication). It is uncertain whether the FISH pattern for BAC TQ7A4 indicates it is a chimera or hybridizes to a related second sequence in the rice genome. Further analysis will be required to determine between these two possibilities. The FISH analysis suggests that the Tequing BAC library constructed in this study is relatively free of chimeric clones. Additionally, a similar FISH study in progress to determine the level of chimerism in the Lemont BAC library.

Utility of the Lemont and Teqing BAC libraries for rice genomics

As described earlier, rice has a small genome (C = 420 Mb [5], and two high density RFLP maps which have a combined total of 2103 loci [12, 26]. Physically, the rice map developed by Causse *et al.* [12] is ca. 580 kb/DNA marker and 290 kb/cM. The rice map developed by Kurata *et al.* [26] is approximately 300 kb/DNA marker and 270 kb/cM. Therefore, if a target locus is linked to a RFLP marker one centiMorgan away, the locus could be isolated by walking 1–3 steps using the two rice BAC libraries.

The two rice BAC libraries were constructed from the parents, Lemont and Teqing, of a permanent mapping population consisting of 400 F9 RILs which segregates for many traits of agronomic importance [29, 30]. Lemont belongs to ssp. *japonica* and Tequing belongs to ssp. *indica*, which represent the two major genomes of cultivated rice. Both Lemont and Teqing are leading commercial varieties and used as germplasm for the development of many other commercial vari-

eties. The RIL permanent mapping population is maintained as seed and is thus much easier to distribute than a F2 mapping population. Therefore, the RIL mapping population should provide an invaluable framework for rice genome research. Similarly, the libraries constructed from the parents of such a widely available RIL mapping population could provide a common framework for physical mapping of the rice genome. Since the target genes and the RFLP markers isolated from a chromosome walking experiment can be more accurately mapped using this large population, the BAC libraries can be readily used to clone the target genes by map-based cloning and assemble long range overlapping contigs.

Notes from the construction of the Lemont and Teqing BAC libraries

The Lemont and Teqing BAC libraries were constructed from two single ligations, one for each library. The transformation efficiency of recombinant BAC DNA was 4×10^5 cfu/μg ligated DNA, which is ca. 100-fold higher than that of YAC constructs with the similar insert size (ca. 10^3 cfu/mg ligated DNA) [16, 20, 31, 43, 44, 46]. This result indicates that a BAC library can be constructed more rapidly and needs less amount of clonable DNA than a YAC library. It was noted that the size-selected rice DNA fragments from 120 to 400 kb were ligated into the BAC cloning vector and transformed into *E. coli* by electroporation, however, no BAC with a rice DNA insert of up to 400 kb was obtained in this study. The same result was obtained in the construction of a sorghum BAC library [45]. Since *E. coli* can maintain its own genome of several million base pairs, it is theoretically possible to obtain BACs with a Mb size insert. Further investigation is needed to answer the question whether BACs larger than 400 kb can be obtained by using the current method for BAC cloning.

The Teqing and Lemont BAC libraries were constructed with the DNA fragments from single or double-size selections by PFGE, respectively. The average insert size of the Lemont library was increased by 20 kb over that of the Teqing library. The portion of clones larger than 120 kb in insert size in the Lemont BAC library was increased to 80% from 60% in the Teqing BAC library. A similar result was obtained by Woo *et al.* [45] for the construction of a sorghum BAC library. These results suggest that a double size selection is necessary to increase the insert size of a BAC library. However, it was observed that the transformation efficiency of the ligated DNA was decreased by about 4 to 10 fold. This was partially due to elimination of smaller, trapped DNA fragments during the second size selection, the BACs with which might be more efficiently transformed than the BACs with larger inserts.

Acknowledgements

This study was supported in part by the Rockefeller Foundation to H.B.Z. and R.A.W. and the Texas Agricultural Experiment Station and Institute of Biosciences and Technology to R.A.W. The authors appreciate Drs. M. Simon and H. Shizuya for permission to use their BAC cloning vector pBeloBAC 11 prior to publication, Ms J. Svalberg, Linghua Zhu, Chantel Scheuring, and Xiaoling Ding for assistance in storing the Teqing and Lemont BAC libraries, and Drs B.S. Gill and J. Jiang for the Teqing BAC fluorescent *in situ* hybridization data.

References

1. Anderson C: Genome shortcut leads to problems. Science 259: 1684–1587 (1993).
2. Ahn S, Anderson JA, Sorrells ME, Tanksley SD: Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 241: 483–490 (1993).
3. Ahn S, Tanksley SD: Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA, 90: 7980–7984 (1993).
4. Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HN, Somerville CR: Map-based cloning of gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258: 1353–1355 (1992).
5. Arumuganathan K, Earle ED: Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208–218 (1991).

6. Battraw M, Hall TC: Expression of a chimeric neomycin phosphotransferase II gene in first and second generation transgenic rice plants. Plant Sci 85: 191–202 (1992).
7. Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ: *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeated class of plant disease resistance genes. Science 265: 1855–1860 (1994).
8. Bernatzky R, Tanksley SD: Methods for detection of single or low copy sequence in tomato on Southern blots. Plant Mol Biol Rep 4: 37–41 (1986).
9. Bhave MR, Lawrence S, Barton C, Hannah C: Identification and molecular characterization of *Shrunken-2* cDNA clones of maize. Plant Cell 2: 581–588 (1990).
10. Burke DT, Carle GF, Olson MV: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 805–811 (1987).
11. Cao J, Zhang W, McElroy D, Wu R: Assessment of rice genetic transformation techniques. In: Khush GS, Toenniessen GH (eds) Rice Biotechnology, pp. 175–198. CAB International, Wallingford, UK (1991).
12. 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: Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138: 1251–1274 (1994).
13. Chang C, Kwok SF, Bleecker AB, Meyerowitz EM: *Arabidopsis* ethylene-response gene *ETR1*: similarity of product to two-component regulators. Science 252: 539–544 (1993).
14. Creusot F, Fouilloux E, Dron M, Lafleuriel J, Picard G, Billault A, Le Paslier D, Cohen D, Chaboute M, Durr A, Fleck J, Gigot C, Camilleri C, Bellini C, Caboche M, Bouchez D: The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*. Plant J, in press (1995).
15. Ecker JR: PFGE and YAC analysis of the *Arabidopsis* genome. Methods 1: 186–194 (1992).
16. Edwards KJ, Thompson H, Edwards D, de Saizieu A, Sparks C, Thompson JA, Greenland AJ, Eyers M, Schuch W: Construction and characterization of a yeast artificial chromosome library containing three haploid maize genome equivalents. Plant Mol Biol 19: 299–308 (1992).
17. Eyers M, Edwards K, Schuch W: Construction and characterization of a yeast artificial chromosome library containing two haploid *Beta vulgaris* L. genome equivalents. Gene 121: 195–201 (1992).
18. Feinberg AP, Vogelstein S: A technique for radio-labeling DNA fragments to high specific activity. Anal Biochem 137: 266–267 (1984).
19. Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM: Isolation of the *Arabidopsis 3AB1* gene by positional cloning. Plant Cell 4: 1251–1251 (1992).
20. Grill E, Somerville: Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. Mol Gen Genet 226: 484–490 (1991).
21. Hall TC, Xu Y, Huntley CC, Yu H, Seay J, Connell JC, Lepetit M, Dong J, Wallace D, Way MO, Buchholz WG: Transgene expression and agronomic improvement of rice. Phil Trans R Soc London B 341: 351–374 (1993).
22. Hodges TK, Peng J, Lyznik LA, Koetje D: Transformation and regeneration of rice protoplasts. In: Khush, GS, Toenniessen GH (eds) Rice Biotechnology, pp. 157–176. CAB International, Wallingford, UK (1991).
23. Jiang J, Gill B, Wang G, Ronald P, Ward D: Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci USA 92: 4487–4491 (1995).
24. Khush GS, Kinoshita T: Rice karyotype, marker genes, and linkage groups. In: Khush GS, Toenniessen GH (eds), Rice Biotechnology, pp. 83–108. CAB International, Wallingford, UK (1991).
25. Kleine M, Michalek W, Graner A, Herrmann RG, Jung C: Construction of a barley (*Hordeum vulgare* L.) YAC library and isolation of a *Hor1*-specific clone. Mol Gen Genet 240: 255–272 (1993).
26. Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin S-Y, Inoue T, Fukuta A, Shimano T, Kuboki Y, Toyama, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang Z-X, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y: A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genet 8: 365–372 (1994).
27. Larkin J, Scowroft W: Somatoclonal variation: a novel source of variability from cell cultures for plant improvement. Theor Appl Genet 60: 197–214 (1981).
28. Leung J, Bouvier-Durand M, Morris P-C, Guerrier D, Chefdor F, Giraudat: *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulate protein phosphatase. Science 264: 1448–1452 (1994).
29. Li ZK, Pinson SRM, Stansel JW, Park WD: Identification of QTL for heading date and plant height in rice using RFLP markers. Theor Appl Genet, in press (1995).
30. Li ZK, Pinson SRM, Marchetti MA, Stansel JW, Park WD: Characterization of quantitative trait loci contributing to field resistance to sheath blight (*Rhizoctonia solani*) in rice. Theor Appl Genet, in press (1995).
31. Martin GB, Ganal MW, Tanksley SD: Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. Mol Gen Genet 223: 25–32 (1992).
32. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD: Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 252: 1432–1438 (1993).
33. McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD: Molecular mapping of rice chromosomes. Theor Appl Genet 76: 815–829 (1988).

34. McCouch SR, Tanksley SD (1991) Development and use of restriction fragment length polymorphism in rice breeding and genetics. In: Khush, GS, Toenniessen GH (eds) Rice Biotechnology, pp. 109–124. CAB International, Wallingford, UK (1991).
35. Meyer K, Leube MP, Grill E: A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. Science 154: 1452–1455 (1994).
36. Oannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ: A new bacteriophage PI-derived vector for the propagation of large human DNA fragments. Nature Genet 5: 84–89 (1994).
37. Putterill J, Robson F, Lee K, Coupland G: Chromosome walking with YAC clones in *Arabidopsis*: isolation of 1700 kb of contiguous DNA on chromosome 5, including a 300 kb region containing the flowering-time gene *CO*. Mol Gen Genet 239: 145–157 (1993).
38. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
39. Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiiri Y, Simon M: 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 (1992).
40. Timmis JN, Scott NS: Sequence homology between spinach nuclear and chloroplst genomes. Nature 305: 65–67 (1983).
41. Umehara Y, Inagaki A, Tanoue H, Yasukochi Y, Nagamura Y, Saji S, Otsuki Y, Fujimura T, Kurata N, Minobe Y: Construction and characterization of a rice YAC library for physical mapping. Mol Breed 1: 79–89 (1995).
42. Wang G, Holsten T, Wang H, Ronald P: Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa*-21 disease resistance locus. Plant J 7: 525–533 (1995).
43. Ward ER, Jen GC: Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. Plant Mol Biol 14: 561–568 (1990).
44. Wing RA, Rastogi VK, Zhang H-B, Paterson AH, Tanksley SD: An improved method of plant megabase DNA isolation in agarose microbeads suitable for physical mapping and YAC cloning. Plant J 4: 893–898 (1993).
45. Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA: Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucl Acids Res 22: 4922–4931 (1994).
46. Woo SS, Rastogi VK, Zhang H-B, Paterson AH, Schertz KF, Wing RA: Isolation of megabase-size DNA from sorghum and applications for physical mapping and bacterial and yeast artificial chromosome library construction. Plant Mol Biol Rep 13: 38–50 (1995).
47. Zhang H-B, Martin GB, Tanksley SD, Wing RA: Map-based cloning in crop plants: tomato as a model system. II. Isolation and characterization of a set of overlapping YACs encompassing the jointless locus. Mol Gen Genet 224: 513–521 (1994).
48. Zhang H-B, Zhao X, Ding X, Paterson AH, Wing RA: Preparation of megabase-size DNA from plant nuclei. Plant J 7: 175–184 (1995).