Theor Appl Genet (2004) 109: 1041–1050
DOI 10.1007/s00122-004-1712-y

ORIGINAL PAPER

C.-C. Wu · P. Nimmakayala · F. A. Santos · R. Springman · C. Scheuring · K. Meksem · D. A. Lightfoot · H.-B. Zhang

Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping

Received: 6 June 2003 / Accepted: 22 April 2004 / Published online: 26 May 2004

Abstract Two plant-transformation-competent large-insert binary clone bacterial artificial chromosome (hereafter BIBAC) libraries were previously constructed for soybean cv. Forrest, using *Bam*HI or *Hin*dIII. However, they are not well suited for clone-based genomic sequencing due to their larger ratio of vector to insert size (27.6 kbp:125 kbp). Therefore, we developed a larger-insert bacterial artificial chromosome (BAC) library for the genotype in a smaller vector (pECBAC1), using *Eco*RI. The BAC library contains 38,400 clones; about 99.1% of the clones have inserts; the average insert size is 157 kbp; and the ratio of vector to insert size is much smaller (7.5 kbp:157 kbp). Colony hybridization with probes derived from several chloroplast and mitochondrial genes showed that 0.89% and 0.45% of the clones were derived from the chloroplast and mitochondrial genomes, respectively. Considering these data, the library represents 5.4 haploid genomes of soybean. The library was hybridized with six RFLP marker probes, 5S rDNA and 18S-5.8S-25S rDNA, respectively. Each RFLP marker hybridized to about six clones, and the 5S and 18S-5.8S-25S rDNA probes collectively hybridized to 402 BACs—about 1.05% of the clones in the library. The BAC library complements the existing soybean Forrest BIBAC libraries by using different restriction enzymes and vector systems. Together, the BAC and BIBAC libraries encompass 13.2 haploid genomes, providing the most comprehensive clone resource for a single soybean genotype for public genome research. We show that the BAC library has enhanced the development of the soybean whole-genome physical map and use of three complementary BAC libraries improves genome physical mapping by fingerprint analysis of most of the clones of the library. The rDNA-containing clones were also fingerprinted to evaluate the feasibility of constructing contig maps of the rDNA regions. It was found that physical maps for the rDNA regions could not be readily constructed by fingerprint analysis, using one or two restriction enzymes. Additional data to fingerprints and/or different fingerprinting methods are needed to build contig maps for such highly tandem repetitive regions and thus, the physical map of the entire soybean genome.

Communicated by P.J. Heslop-Harrison

C.-C. Wu · P. Nimmakayala · F. A. Santos · R. Springman · C. Scheuring · H.-B. Zhang (✉)
Department of Soil and Crop Sciences and Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, TX 77843-2123, USA
e-mail: hbz7049@tamu.edu
Tel.: +1-979-8622244
Fax: +1-979-8624790

K. Meksem · D. A. Lightfoot
Department of Plant Soil and General Agriculture, Biotechnology and Genomics Core Facility, Center for Excellence in Soybean Research, Teaching and Outreach, Southern Illinois University, Carbondale, IL 62901-4415, USA

Introduction

Large-insert bacterial artificial chromosome (BAC) libraries are crucial to physical map construction, genome analysis, clone-based sequencing, and genomic sequence finishing of large, complex genomes (Zhang et al. 1996b; Tao et al. 2001; The International Human Genome Mapping Consortium 2001; The International Human Genome Sequencing Consortium 2001; Venter et al. 2001; Zhang and Wu 2001; Chen et al. 2002; Gregory et al. 2002; Xu et al. 2003; Wu et al. 2004a). As the insert size of the source library increases, the number of clones needed for construction of a genome-wide physical map is reduced proportionally (Zhang et al. 1996b). The physical maps constructed from larger-insert libraries have increased contig sizes and reduced gaps (Zhang et al. 1996b; Gregory et al. 2002).

Clone-based physical maps have been proven to provide essential platforms for advanced genome research (Chang et al. 2001; Tao et al. 2001; The International Human Genome Mapping Consortium 2001; Zhang and Wu 2001; Gregory et al. 2002). Use of multiple complementary BAC libraries constructed with different restriction enzymes can

enhance physical map construction. The beneficial effects may derive from the different GC contents of the recognition sites of different restriction enzymes that increase genome representation (Sun et al. 2003). The complete genome representation of the source library is key to development of whole-genome physical maps of high genome coverage. The distribution of recognition sites for a restriction enzyme is uneven among chromosomes and genomes (The Arabidopsis Genome Initiative 2000), so that the clones from some genomic regions cannot be found in a BAC library constructed with a single restriction enzyme. A physical map constructed from such a library would have more and larger gaps.

The use of different vector systems for library construction can further enhance the complementarity of libraries because different origins of replication, cloning sites, selectable markers, and insertion sites of different cloning systems can affect cloning ability and insert stability. The BAC libraries that are constructed in transformation-competent binary vectors promise to streamline use of the derived physical maps and genome sequences (Hamilton et al. 1996; Liu et al. 1999). Direct transformation with gene clusters will be important in positional cloning of complex loci, determination of the functions of clustered genes and related sequences, and molecular breeding following genetic transformation.

Soybean, *Glycine max* (L.) Merr., is the world's major legume crop. For genome research of this important crop, BAC or binary bacterial artificial chromosome (BIBAC) libraries encompassing more than 35 haploid genomes have been developed for several genotypes, including Williams 82 (Marek and Shoemaker 1997), Faribault (Danesh et al. 1998), PI 437654 (Tomkins et al. 1999), Forrest (Meksem et al. 2000), and A3244 (Tomkins et al. 2000). However, no comprehensive libraries have been constructed for any of the genotypes; all of them, except for the Forrest libraries that were constructed with *Hin*dII or *Bam*HI in the bacterial P1-based vector pCLD04541, were constructed from the partial digests of DNA with a single restriction enzyme (either *Hin*dIII or *Eco*RI) in the bacterial F factor-based vector pBeloBAC11 or its derivative pESBAC4. The average insert sizes of the libraries, except for that of the Williams 82 library, all are smaller than 150 kbp. Although these libraries have been proven very useful for soybean genome research (Marek and Shoemaker 1997; Danesh et al. 1998; Tomkins et al. 1999; Meksem et al. 2000), they are not well suited for comprehensive genome analysis of the species.

Here we describe a BAC library for soybean cv. Forrest. The library has a larger insert size than any publicly available soybean BAC libraries and complements the existing Forrest BIBAC libraries. Therefore, the Forrest BAC library, along with our existing Forrest BIBAC libraries (Meksem et al. 2000), provides the most comprehensive, large-insert clone resource of a single genotype for extensive analysis of the soybean genome. The library has been made available to the public (http://hbz.tamu.edu) and used in the construction of a genome-wide physical map of the soybean genome (reviewed in Lightfoot et al. 2003a; Wu et al. 2004a).

Materials and methods

Plant materials and BAC cloning vector

The seeds of soybean cv. Forrest were provided and are maintained in perpetuity by Southern Illinois University at Carbondale, Ill., USA (Hnetkovsky et al. 1996; Lightfoot et al. 2003b). Plants were grown in a greenhouse at 16 h light, 8 h dark, and 25°C. Leaves were harvested from 2-week-old seedlings, frozen immediately in liquid nitrogen, and stored at −80°C before use. High-molecular weight (HMW) DNA was prepared from nuclei as described by Zhang et al. (1995) with the following modifications. Chloroplasts and mitochondria were minimized by four washes of the nuclei pellet. Nuclei were prepared from 50 g of frozen leaves and embedded in 6 ml of 0.5% (w/v) low-melting-point (LMP) agarose plugs. After nuclei lysis, the DNA concentration of the plugs was about 10 μg/100-μl plug.

The pECBAC1 vector (Frijters et al. 1997) was used in the library construction. Vector DNA was isolated by the alkaline lysis method, purified by cesium chloride gradient centrifugation (Sambrook et al. 1989), digested completely with *Eco*RI, and dephosphorylated with calf intestinal alkaline phosphatase (Zhang 2000; Wu et al. 2004b; http://hbz.tamu.edu). The digested vector DNA was precipitated, dissolved in TE (10 mM Tris.HCl, 1 mM EDTA, pH 8.0), adjusted to 10 ng/μl, and stored at −20°C before use.

BAC library construction

The BAC library was constructed according to a procedure developed in our laboratory (Zhang 2000; Wu et al. 2004b; http://hbz.tamu.edu). Before partial digestion, HMW DNA embedded in LMP agarose plugs was subjected to presize selection by PFGE (pulsed-field gel electrophoresis) on a CHEF DRIII (Bio-Rad, USA) to remove smaller fragments and substances that might inhibit restriction enzyme activities. The presize selection was conducted on a 1.0% (w/v) agarose gel at 4 V/cm, 11°C, 5-s pulse time for 8 h in 0.5× TBE (Sambrook et al. 1989). The plugs containing HMW DNA were then collected, dialyzed against cold TE on ice for 3 h, one TE change per hour, and used for partial digestion.

To determine the desirable condition for partial digestion, each plug was cut into 12 slices approximately equal in size and incubated in 1 ml reaction buffer [1× restriction buffer (Gibco BRL, USA), 2 mM spermidine and 1 mM DTT (Sigma, USA), and 0.2 mg/ml BSA (New England Biolabs, USA)] on ice for 2 h, with one buffer change after 1 h. Four slices of the plug were transferred into a 1.5-ml tube containing 100 μl fresh reaction buffer plus 0.2–10 U *Eco*RI. The reaction mixture was incubated on ice for another hour, then transferred into a 37°C water bath and

incubated for 8 min. The reaction was stopped immediately by adding 1/10 (v/v) of 0.5 M EDTA (pH 8.0) and analyzed on a 1.0% (w/v) agarose gel at 6 V/cm, 11°C, and 50-s pulse time for 18 h in 0.5× TBE. The enzyme concentration that generated fragments with a majority ranging from 100 kbp to 350 kbp was selected for large-scale partial digestion.

Large-scale partial digestion of HMW DNA for BAC library construction was carried out using five plugs and the *Eco*RI concentration determined above. The plug slices containing partially digested DNA were subjected to first size selection on a 1.0% (w/v) agarose gel at 6 V/cm, 11°C, and 90-s pulse time for 18 h, followed by 4 V/cm, 11°C, and 5-s pulse time for 6 h in 0.5× TBE. The gel fractions containing DNA fragments from 100 kbp to 150 kbp, 150 kbp to 300 kbp, and 300 kbp to 350 kbp were excised. The DNA fragments were recovered from the gel by electroelution in dialysis tubing (12,000–14,000 Daltons molecular-weight exclusion, Gibco BRL), using the CHEF DRIII at 6 V/cm, 11°C, and 30-s pulse time for 4 h, followed by reversing the polarity of currency for 90 s in 0.5× TBE. The DNA was collected gently with a wide-bore pipette tip and then subjected to second size selection on a 1% (w/v) agarose gel at 4 V/cm, 11°C, and 5-s pulse time for 8 h in 0.5× TBE. The DNA from each fraction was excised and recovered by electroelution as above. The eluted DNA fragments were dialyzed in the same tubing against 0.5× TE for 3 h at 4°C, with one TE change per hour.

The DNA concentration was measured by electrophoresis on a 0.8% (w/v) agarose gel and adjusted to 1–2 ng/μl. Two hundred microliters of the DNA fragments from each fraction were ligated into the cloning vector at a molar ratio of 4 vector:1 insert DNA, using 10 U T4 DNA ligase (Gibco BRL) at 16°C for 12 h.

Ligated DNA was transformed into *Escherichia coli* strain ElectroMAX DH10B competent cells (Gibco BRL) by electroporation, using a Cell Porator System (Gibco BRL). The setting conditions were 350 V, 330 μF, low ohms, and 4 KΩ with fast charge. Transformed cells were transferred into 1 ml SOC medium and recovered at 37°C for 1 h with gente shaking. Recombinant transformants were selected on an LB agar (Gibco BRL) plate containing 12.5 mg/l chloramphenicol, 0.5 mM IPTG, and 50 μg/ml X-gal. After a 32-h incubation at 37°C, white colonies were randomly selected, and BAC DNA was isolated, digested with *Not*I, and subjected to size analysis by PFGE (Zhang et al. 1996a; Zhang 2000). The ligation that had a transformation efficiency of 200 or more white colonies/μl ligation and that generated clones with the largest inserts was selected for library construction. White colonies were manually arrayed as individual clones in 384-well microtiter plates containing 50 μl LB plus freezing broth with 12.5 mg/l chloramphenicol (Zhang et al. 1996a; Zhang 2000). After incubation at 37°C for 14 h, the microtiter plates were stored at −80°C.

BAC library screening

A Biomek 2000 robotic workstation (Beckman, USA) was used to double spot the BAC library onto 8 × 12-cm Hybond N^+ filters (Amersham-Pharmacia, USA) in 3 × 3 format, so that the high-density clone filters contained two spots of each clone from four 384-well microtiter plates (1,536 × 2 spots). The filters were processed according to Zhang et al. (1996a) and Zhang (2000). Chloroplast and mitochondrial DNA probes were generated from soybean cv. Forrest total genomic DNA by PCR, using primers specific for five chloroplast genes, *ndhA*, *psbA*, *rbcL*, *petA*, and *rpoB*, and seven mitochondrial genes, *atp6*, *atp9*, *atpA*, *cox1*, *cox2*, *cox3*, and *nad4*. The 5S rDNA used as a probe was isolated from wheat DNA (Gerlach and Dyer 1980). The 18S-5.8S-25S rDNA probe was generated from an *Arabidopsis* 18S-5.8S-25S rDNA-containing BAC clone F1A18 (Mozo et al. 1998). Six soybean RFLP markers, pA186, pA504, pA708, pA757, pB212, and pK644, were purchased from Biogenetic Services Inc. (Brookings, South Dakota, USA). All probes were labeled with [^{32}P]-dCTP by random priming. Colony hybridization was carried out using standard procedures with a washing condition of 1× SSC, 0.1% (w/v) SDS at 65°C twice, 10 min each time, followed by 0.5× SSC, 0.1% SDS at 65°C twice, 10 min each time (Sambrook et al. 1989).

BAC fingerprinting and contig assembly

BAC DNA was isolated and fingerprinted following Chang et al. (2001) and Tao et al. (2001). The BAC DNA was digested with *Hin*dIII/*Hae*III and end labeled at the *Hin*dIII sites with [^{33}P]-dATP and reverse transcriptase. To test the feasibility of constructing contigs for the rDNA regions, DNA of the rDNA-containing clones was also digested with *Bam*HI/*Hae*III, or *Sau*3AI alone and end labeled with [^{33}P]-dATP and reverse transcriptase. The reactions were subjected to electrophoresis on 3.5% (w/v) denaturing polyacrylamide gels. The gels were dried onto 3-mm chromatography papers and used to expose X-ray BioMax MR films (Kodak, USA).

The fingerprints on the films were scanned into image files, using a UMAX Mirage D-16L scanner and edited using Image 4.0 (Sulston et al. 1988; Soderlund et al. 1997). To reduce the influence of the relative lower resolution (>1 base) of larger fragments (>773 bases) on the sequencing gels, only the fragments ranging from 58 bases to 773 bases were used. The single band derived from the BAC vector was deleted from the data files. The clones that failed in fingerprinting, had no inserts, or produced five or fewer bands were excluded during fingerprint editing because they could not provide sufficient information for contig assembly. Overlapping clones were assembled into contigs, using the computer program Fingerprinted Contig (FPC), version 4.7 (Soderlund et al. 2000). The automatic contigs were assembled using a fixed tolerance of 2 and cutoffs between 1e-28 and

1e-12, and the contigs were merged by using cutoffs ranging from 1e-26 to 1e-15.

Results and discussion

BAC library construction and characterization

We constructed a BAC library for soybean cv. Forrest with *Eco*RI-derived inserts in the F factor-based BAC vector pECBAC1. This library complements our two BIBAC libraries of the same genotype by using different restriction enzymes and vector systems (Meksem et al. 2000). The BAC library contains 38,400 clones arrayed in 100 384-well microtiter plates. A random sampling of 112 BACs from the library showed that the clones have an average insert size of 157 kbp, with a range from 85 kbp to 320 kbp. Only one of the clones had no insert (0.9%, Figs. 1, 2). The large-insert size and small ratio of vector: insert size (7.5:157) of the BAC library makes it well suited for clone-based genomic sequencing. Of the 112 BACs analyzed, 103 (92.0%) had no *Not*I site in their inserts (one insert band on the gel), whereas eight (7.1%) had one or more *Not*I sites (two or more insert bands). Dicotyledonous plant genomes have a lower G/C content and hence, fewer *Not*I sites than the monocotyledonous plant genomes (Zhang et al. 1996a; Meksem et al. 2000; Tao et al. 2002; Chang et al. 2003).

BAC libraries that were constructed with different enzymes and have larger insert sizes have been used in genome research (Zhang et al. 1996b; Tao et al. 2001; The International Human Genome Mapping Consortium 2001; The International Human Genome Sequencing Consortium 2001; Venter et al. 2001; Zhang and Wu 2001; Chen et al. 2002; Gregory et al. 2002). The Forrest *Eco*RI BAC library constructed in this study, combined with the existing *Hin*dIII and *Bam*HI BIBAC libraries (Meksem et al. 2000), collectively provides a total of 115,200 clones, equivalent to 13.2 haploid genomes, for a single soybean genotype. These BAC and BIBAC libraries should be sufficient for extensive analysis of the soybean genome, including genome physical mapping (Chang et al. 2001; Tao et al. 2001; Wu et al. 2004a), BAC end sequencing, clone-based genomic sequencing and functional analysis of the genome sequences.

Genome origin of the BAC clones

Plants have three genomes: nuclear, chloroplast, and mitochondrial. To estimate genome origin of the clones, we screened the library with probes derived from five chloroplast genes, *ndhA*, *psbA*, *rbcL*, *petA*, and *rpoB*, and seven mitochondrial genes, *atp6*, *atp9*, *atpA*, *cox1*, *cox2*, *cox3*, and *nad4*, respectively. A total of 341 and 174 positive clones were identified to contain chloroplast and mitochondrial genomic sequences, respectively, suggesting that approximately 0.89% and 0.45% of the clones in the library originated from the chloroplast and mitochondrial genomes, respectively, and greater than 98% were from the nuclear genome (Table 1). The low percentage of organelle genome clones would facilitate use of the library for nuclear genome analysis. Based on the soybean haploid genome size estimate of 1,115 Mb (Arumuganathan and Earle 1991), the BAC library represents 5.4-fold haploid genome equivalents, theoretically providing an over 99% probability of obtaining at least one clone,

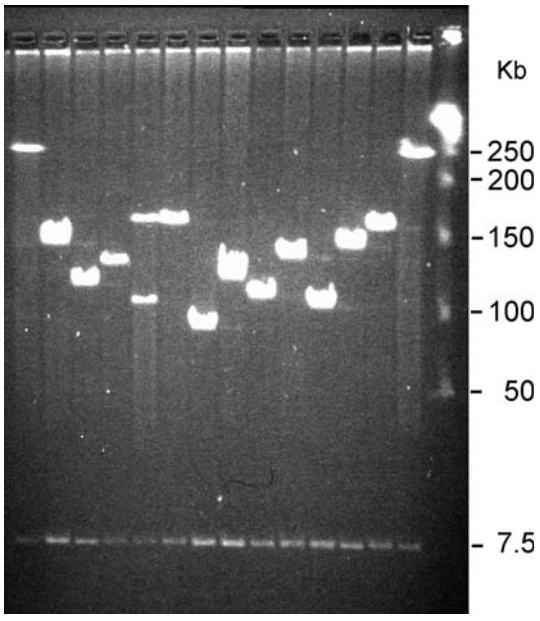

Fig. 1 Bacterial artificial chromosomes (BACs) randomly selected from the soybean cv. Forest BAC library constructed in the *Eco*RI cloning site of the BAC vector pCEBAC1. White clones were selected randomly and grown in LB medium with chloramphenicol (12.5 mg/l) for 20 h at 37°C, and BAC DNA was isolated according to Zhang et al. (1996a) and Zhang (2000). Inserts of the clones were excised by digestion with *Not*I and resolved on 1.0% (w/v) agarose CHEF gels. The *last lane from left* is the λ ladder for molecular weight marker and the *remaining lanes* are BACs randomly selected from the library. The band appearing in all BAC lanes was from the vector pECBAC1 and the remaining bands were from inserts. The insert size of each clone was estimated by adding all insert band sizes in each lane

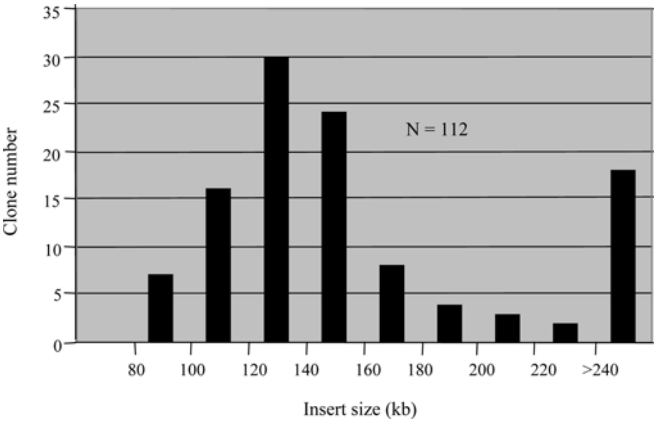

Fig. 2 Insert size distribution of 112 clones randomly selected from the soybean cv. Forrest *Eco*RI BAC library. The average insert size of the library was estimated to be 157 kbp as in Fig. 1

Table 1 Positive clones of the soybean *Eco*RI bacterial artificial chromosome (BAC) library identified by library screening with probes derived from chloroplast genes, mitochondrial genes, RFLP markers, and rDNA

Chloroplast gene probes											
01G20	19M12	33H06	52B04	64J11	72P13	77O14	81K05	84N05	89F04	94B10	98J07
03J13	19M17	33K02	52H05	64L23	73D11	77P15	81K08	85F23	89L16	94D18	98K22
04A20	20M09	33K19	52H08	65E21	73G02	77P18	81N09	85F24	89M14	94E07	98O01
04D14	20M10	33P12	52K07	66L21	73I02	78D13	81N12	85G16	89N04	94F12	98O10
04I17	21L04	36B21	52N24	68M15	73J03	78E13	82A18	85H01	89N18	94F16	98O13
04O21	22E18	37F04	53E08	69A17	73M13	78G23	82C01	85I22	90D01	94K01	98O20
05E08	23L02	37G07	53I19	69C02	74A03	78J12	82C06	85M15	91A13	94L08	99C05
05J04	24C12	38N16	53L12	69E02	74B13	78K07	82D17	85O14	91C24	94N13	99D13
05M02	24C18	39M19	53L24	69J22	74C23	78L14	82F17	85P03	91F13	95A07	99I13
06E11	24J24	41E24	53M19	69O04	74E17	78M04	82G14	86B03	91H06	95A21	99E20
06E14	24P23	46H10	55F11	70B01	74H16	79E20	82G16	86D18	91 M09	95A22	99G19
06F03	25B21	47D11	55N13	70E11	74I18	79G12	82I22	86G09	91P05	95A24	99G20
06L12	25E05	47H01	56C17	70G10	74J07	79I14	82N04	86O23	91P16	95E15	99K02
07C05	25G16	47P22	56C22	70H08	74K09	79O05	82O14	87E22	92A23	95G08	99M22
07I01	26E13	48C12	56D10	70I21	74 K12	80A05	82P17	87F23	92D16	95I02	99O14
07P02	26O13	49C08	56D20	70L19	74 K21	80B13	83A18	87 J06	92E22	95I16	99P05
08H05	27F16	49 J16	57B19	71H03	74 M08	80D10	83A21	87 N04	92G03	95 J10	100C24
09D02	27P20	50A13	58J21	71H07	75D12	80J20	83D12	87P07	92I13	95K01	100E17
10C08	28B06	50C07	58N13	71K07	75E06	80K11	83D22	87P16	92L20	95K04	100G18
10M04	28C15	50E14	58P12	71K14	75M06	80L19	83G14	88A02	92O21	96B14	100J10
14F20	28C17	50E15	60D22	71M05	75N19	80M14	83I04	88A09	92P08	96D08	100J21
15H05	28J14	50I23	60H20	71N13	76D13	80M21	83I21	88A14	93B01	96D11	100K10
16K20	29K21	50J22	60J24	71O01	76D17	80O13	84C24	88E22	93B23	96F22	
17D09	29P24	50L23	62B18	72A15	76E17	80O15	84E09	88G12	93D16	96I17	
17E02	30F17	50O22	63D04	72B19	76M14	80P21	84I18	88H16	93F03	97G10	
17I08	30M04	51B01	63D22	72I16	77A15	81A10	84I19	88H24	93F22	98B12	
18L19	31A09	51C04	63K02	72K23	77A17	81E21	84I20	88K13	93G12	98E04	
19F10	31D06	51O01	64A11	72L22	77G16	81H14	84L04	88N06	93H05	98I16	
19G07	33G10	51O24	64C20	72M06	77I23	81I07	84L15	89C08	93K16	98J06	
Mitochondrial gene probes											
01H12	09B02	17E08	22D03	30 N15	35C06	45A04	63B14	73P21	77O16	90 J19	98F22
01N03	09B03	17N04	23K06	31A01	35M01	45C07	63I04	74F13	78I15	91I01	98L01
01P11	09B04	17O04	23 N23	31B23	37G22	48H06	63K09	75G10	78J16	92F20	99K05
02G03	09B05	18H15	24B03	31B24	37M04	48L17	64A23	75I01	80F13	92 J11	100C23
02I03	09B06	18I04	25G16	31D04	37O16	50F10	65F13	75I15	81K08	92K12	100G10
03D04	09B07	18J14	26H22	31H06	38N23	50I05	65I04	75O06	83E20	92O07	100G11
03F16	09B08	18K04	26J21	31J21	39H07	52F02	66O22	75P13	83J08	93A18	100J20
04B05	09B09	20C18	26 M08	32D02	39M04	52F09	68D24	76 J15	83L07	93A21	100N23
04I05	09B10	20C19	26M21	32E18	42A03	52F14	68K15	76M14	84C06	93F22	100O16
05A08	09B11	20D19	27I06	33A24	42H19	54H13	69I23	77A07	85E11	93M19	
05C23	09B12	20F10	29A17	33D21	42I19	55F04	71B14	77A08	86M10	94I06	
06C03	09B13	20M17	29A18	33H08	42L04	55I12	72L22	77C09	86P04	94J17	
07E22	09B14	20N24	29B24	34A13	43A23	56D16	73B13	77D23	87D15	96D13	
07O21	09B15	21A18	30A06	35A08	43K09	57O06	73J20	77J07	88A02	96F03	
08P18	16C20	21H03	30N14	35C03	44P19	61C12	73N17	77M11	89F04	97E04	
RFLP marker probes											
pA186	12C14	24A12	59J08	71M02							
pA504	42F17	48A22	63H03	49N17	60E23	64H12	75I15				
pA708	22N05	41N07	44J07	54G20	56B13	63G21	64A17	76A16	77E03	78B05	
pA757	11E16	18 K20	19G06	65K13	80C11						
pB212	57B10	66G10	66G10	71L20							
pK644	01H14	34L17	41M09	58017	58A20	71L20					

Table 1 (continued)

5S and 18S-5.8S-25S rDNA probes											
01A05	07C14	15D03	22E25	28M09	35F13	43N17	53G09	60B24	66B24	75P12	89H17
01A11	07N24	15K11	23B02	28M23	35K24	43N18	53K23	60C03	66C03	76A11	90P03
01A19	07O01	16E04	23D02	28 N04	35L14	43O04	54A22	60C21	66C04	77A06	91B10
01H02	08C03	16G14	23M01	29A07	36C12	44E07	54H22	60G03	66C16	77H23	91C10
01H12	08F23	16L13	24A01	29B07	36G11	44 J24	54I06	60L16	66J19	78A24	91K15
01I10	08O17	17A11	24D17	29E06	36H10	44M10	54N16	61D05	669K23	78B06	91N24
01J06	09B13	17C06	24E25	29E15	36I12	45E08	54O01	61G06	66L21	78L04	92A12
02C10	09C18	17 K22	24G24	29F12	36 K23	45F17	55C20	61H18	66N01	78P23	92B15
02D04	09C24	17L17	24 J10	29H13	36L15	45N14	55H09	61J19	66P09	79M02	92C14
02D23	09J02	17M06	24J16	29J24	37L09	45O06	55J06	61L10	67D08	80A02	92K13
02E09	09K08	17M17	24M17	29K15	37M20	46E21	55M13	62A16	67E13	80F14	93I04
02G16	09M12	17N01	24P02	29N06	37N17	46L11	55O13	62E07	67F20	80G03	93M08
02I08	09N06	18B04	25A01	29N23	38A11	46P01	55P06	62J07	67G06	80I20	96A04
02L24	09N15	18F04	25H10	30I04	38E06	47A21	56E10	62 J08	68A04	81H07	96G02
02N23	09P16	18M22	25J14	30M08	38G07	47C16	56H02	62J16	68J08	82E24	96P06
02P08	10B23	18P13	25P13	30P21	38I16	47G03	56P11	62M15	68O06	82F11	97F05
03E15	10H15	19E17	26B21	31C01	38K06	47M04	57A18	62P10	68O13	82O12	97J05
03H18	11B04	19J05	26C22	31E23	39B11	47N18	57C22	63C12	68P01	83 N19	97L23
03K20	11D22	20A20	26 M17	31G11	39C06	48I01	57H22	63D10	69B22	84B03	98C11
03M11	11H09	20C23	27A23	31K12	39G10	48O11	57J04	63E19	69F13	84C03	98P13
03N18	11L18	20E02	27C23	31N10	39N18	48P24	57K01	63I22	69P23	84L05	99A01
04B07	11O15	20G13	27H21	31N14	40J17	49A07	57O11	63J22	70C14	85P17	99D17
04F07	12C14	20H12	27I01	32C10	40L01	49O05	58A03	63 N08	70L15	86G02	99H24
04F22	12H23	20I22	27L18	32H01	41D09	50A20	58A05	63O11	70P02	86K05	99N11
04O19	12K05	20J09	27M17	32I21	41E10	50M20	58D05	64D16	71A18	87E15	99P24
05C23	12K11	20K14	27M18	33E22	41H12	51F20	58P09	64I11	71G06	87F05	100F12
05D01	12L01	20L14	27O03	33 M10	41I18	51N05	58P22	64I17	71G21	87L23	100P01
05K17	12O08	20N06	27O19	34B16	41J22	51P09	59D12	64O09	73F23	88A01	100P03
05O01	13A17	20P03	28A18	34 J16	41P11	52D06	59D15	64P07	73H06	88F20	
06B09	13C06	21C06	28A19	34J21	42M18	52D13	59E20	65D04	74A05	88G13	
06C17	13F21	21F01	28E05	34P04	43B15	52G24	59G11	65H01	74G04	88H05	
07A12	14F07	21K10	28J06	35C05	43B16	53A04	59N10	65H13	75A10	88I04	
07C03	14I03	21L11	28K22	35D15	43J16	53E04	60B12	65 J16	75G03	88I22	
07C13	14O23	21L15	28M08	35D21	43K20	53F05	60B22	65P16	75H16	89C01	

using a single-copy probe from the library (Zhang et al. 1996b; Wu et al. 2004b).

BAC library screening with RFLPs and rDNAs

To further characterize the library and facilitate its applications for soybean genome research, we screened the library with six RFLP markers selected from the soybean molecular genetic linkage map (Cregan et al. 1999), 5S rDNA and 18S-5.8S-25S rDNA, respectively. Screening of the library with the DNA markers identified an average of six positive clones per marker, with a range from 4 to 10 BACs (Table 1). Screening of the library with the 5S and 18S-5.8S-25S rDNAs identified a total of 402 clones (Table 1), representing 1.05% of the library. The percentage (1.05%) of rDNA clones was close to that observed in the BIBAC library of *Arabidopsis* ecotype Landsberg (1.86%, Chang et al. 2003), but dramatically less than that observed in the large-insert BAC (12.4%, Mozo et al. 1998) and YAC (15.1%, Creusot et al. 1995) libraries of *Arabidopsis* ecotype Columbia. Cytological analysis by fluorescent in situ hybridization showed that rRNA genes in different plant genera (and even in different species of a genus) exhibited variations in the position and copy number of their repeat units (Shishido et al. 2000). These variations are in sharp contrast to the sequence homology of rDNA throughout plant species (Eckenrode et al. 1985). In contrast to ploidy expectations, the diploidized tetraploid soybean had only one 5S rDNA locus and one 45S rDNA locus, whereas the common bean (*Phaseolis vulgaris* L.), a diploid related member of the same tribe (*Phaseoleae*), had two or more 5S rDNA loci and two or more 45S rDNA loci (Shi et al. 1996). In tobacco, the allotetraploid *Nicotiana tabacum* (SSTT) had a reduced proportion of rRNA gene units per genome (3.0×10^3 copies, 0.5% rDNA/1C), compared with its diploid progenitors *N. sylvestris* [(SS) 8.1×10^3 copies, 2.6% rDNA/1C], and *N. tomentosiformis* [(TT) 2.8×10^3 copies, 1.1% rDNA/1C; Lim et al. 2000]. Moreover,

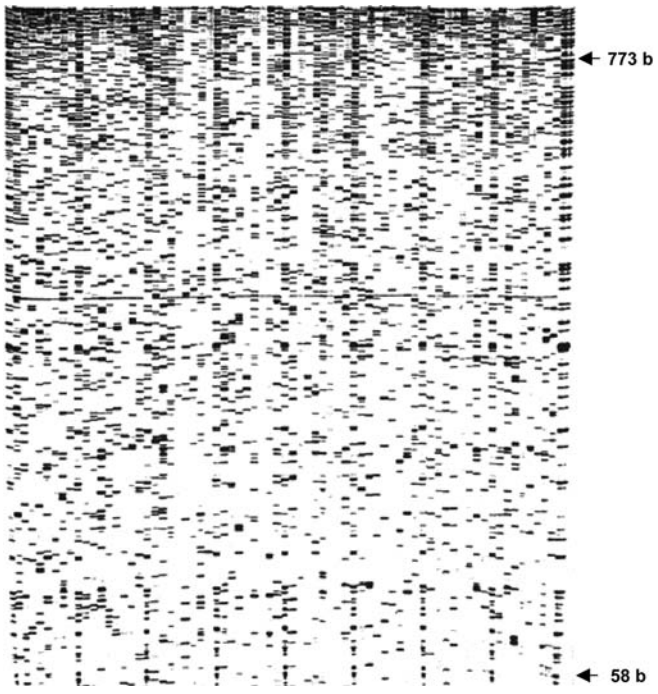

Fig. 3 An autoradiograph of soybean BAC fingerprints. From *left to right*, *lanes 1*, *10*, *19*, *28*, *37*, *46*, *55*, *64*, and *73* are λ/*Sau*3AI DNA markers and the *remaining lanes* are soybean cv. Forrest *Eco*RI BAC clones. BAC DNA was digested with *Hin*dIII and *Hae*III, labeled at the *Hin*dIII fragment ends with ^{33}P-dATP, and fractionated on a 3.5% (w/v) denaturing polyacrylamide gel. The band appearing in all BAC lanes was derived from the pECBAC1 vector and was deleted during fingerprint editing. The bands within the range between 58 bases and 773 bases were used for contig assembly

Shishido et al. (2000) reported that of diploid rice species, *Oryza sativa* ssp. *japonica* (AA) had only one 18S-5.8S-25S rDNA locus on chromosome 9, but *O. sativa* ssp. *indica* (AA), in addition to the locus on chromosome 9, had another 18S-5.8S-25S rDNA locus on chromosome 10, and both *O. eichingeri* (CC) and *O. officinalis* (CC) had three 18S-5.8S-25S rDNA loci, located on chromosomes 4, 9, and 11, respectively. Thus, the significant difference in the percentage of rDNA-containing BAC clones may be mainly due to the different rDNA contents in the nuclear genome rather than the consequence of the restriction enzymes used for BAC library construction (Mozo et al. 1998; Chang et al. 2003). *Eco*RI was used in the construction of *Arabidopsis* ecotype Columbia CIC YAC (Creusot et al. 1995) and IGF BAC (Mozo et al. 1998) and soybean Forrest BAC (this study) libraries, whereas *Bam*HI was used for the construction of *Arabidopsis* ecotype Landsberg BIBAC library (Chang et al. 2003).

Significance of multiple complementary libraries for genome physical mapping

To test the significance of multiple complementary libraries for genome physical mapping and facilitate development of a whole-genome physical map of the soybean genome, we fingerprinted 30,720 clones from the BAC library, using a DNA sequencing gel-based restriction method (Chang et al. 2001; Tao et al. 2001; Zhang and Wu 2001; Fig. 3). After fingerprint editing, fingerprints from 27,944 clones (4.05 × soybean haploid genomes), with an average number of 39.12 bands per BAC, were used for contig assembly (the fingerprint database is available at http://hbz.tamu.edu). Overlapping clones were assembled into contigs, using the FPC program (Soderlund et al. 2000). We assembled the overlapping BACs into contigs by two steps. We first assembled automated contigs, manually edited every automatic contig, and removed all questionable clones and chimeric contigs to ensure that they were accurate. Then we joined automatic contigs into larger contigs, using lower stringency cutoff values. The contig pairs were merged if their terminal clones shared ten or more bands and their overall fingerprint patterns supported joining (see Materials and methods). The results were summarized in Table 2. Of the 27,944 fingerprinted clones, 26,715 were assembled into 3,388 contigs, while 1,279 clones remained as singletons. These contigs collectively consisted of 241,049 unique bands. If each band represented 4.01 kbp (157 kb/39.12), the map spanned 967 Mb in physical length. If they were overlapped at about 30% (Ren et al. 2003; Wu et al. 2004a; http://hbz.tamu.edu), the contigs collectively covered only about 677 Mb (60.7%) of the soybean genome. If the contigs were not overlapped, they collectively covered only 87% of the soybean genome. In comparison, the 27,944 clones theoretically provide 98% coverage of the soybean genome (Zhang et al. 1996b; Wu et al. 2004b). Therefore, the difference between experimental and calculated genome coverage was about 11–39%. A BAC library constructed with a single restriction enzyme, if it is used for genome physical mapping, would result in a large number of true gaps in the derived physical map due to absence of clones in the library for some genomic regions. The fingerprinted clones have been used in our whole-genome physical mapping of the soybean genome (Wu et al. 2004a).

Table 2 Contigs assembled from the clones selected from the *Eco*RI Forest BAC library

BAC clones in FPC database	27,944
Number of singletons	1,279
Number of contigs	3,338
Contigs containing	
>25 clones	44
10–25 clones	821
3–9 clones	2,123
Two clones	400
Unique bands of the contigs	241,049
Physical length of the contigs (in Mb)	967[a]

[a]Each fingerprint band was estimated to represent an average of 4.01 kb

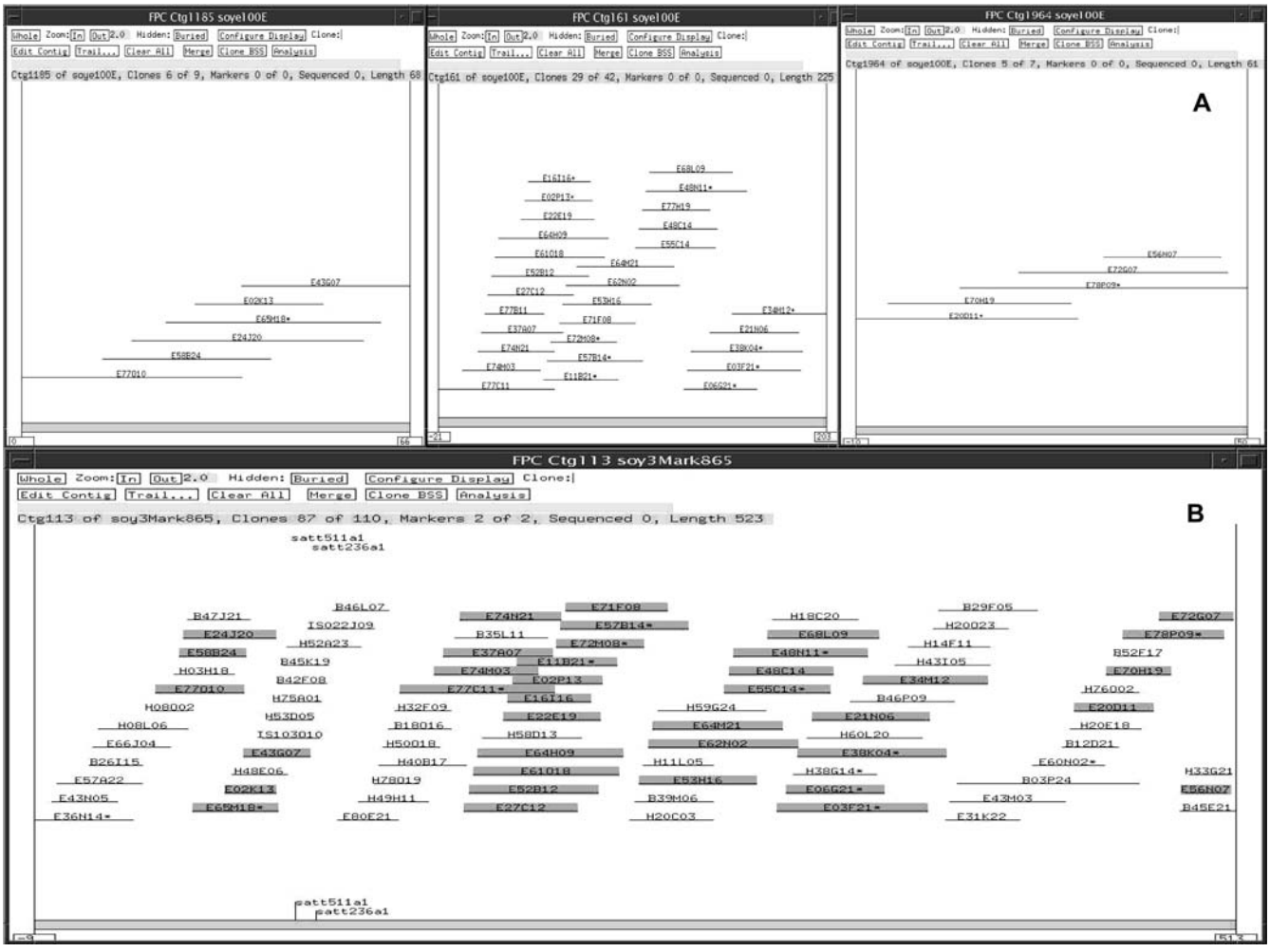

Fig. 4a, b Complementarity of the Forest BAC library to the existing binary clone bacterial artificial chromosome libraries. The three contigs of the Forest *Eco*RI BACs constructed in this study (**a**) were merged into a single 2,123-kb contig in the soybean whole-genome physical map, using all three libraries (Wu et al. 2004a; **b**). *The highlighted clones* in **b** correspond to the clones of the contigs shown in **a**. *Asterisk*s indicate a parent clone that covers one or more clones

Figure 4 shows the complementarity of the BAC library to the existing two Forrest BIBAC libraries to form a supercontig from three contigs. These results illustrate the essentiality of the BAC library constructed in this study for comprehensive research of the soybean genome and the significance of multiple complementary BAC/BIBAC libraries constructed with different restriction enzymes for development of a genome-wide physical map of high genome coverage. Other planned uses of the BAC library include refinement of the whole-genome physical map of the soybean genome (Wu et al. 2004a) for large-scale clone-based genome sequencing (unpublished), physical map gap closure and genomic sequence finishing.

Construction of contig maps for the genomic regions of rDNAs

Construction of contig maps for the genomic regions abundant in tandem repeats such as rDNA has become a headachy challenge, but it is essential for genome physical map and sequence finishing. To test the feasibility of constructing physical map contigs for the rDNA regions, we fingerprinted all 402 positive BAC clones identified with the 5S and 18S-5.8S-25S rDNA probes with *Hin*dIII/*Hae*III, *Bam*HI/*Hae*III, or *Sau*3AI alone, and then assembled them into contigs, separately, by restriction enzyme combinations used (Chang et al. 2001; Tao et al. 2001). Of the 402 rDNA positive BACs, 333 produced only 2–4 bands in their fingerprints in spite of their large inserts, providing insufficient information for contig assembly. Only the remaining 69 clones could be assembled into contigs, probably due to the rDNA locus-flanking sequences of larger-insert clones and/or higher sequence variation. The result indicates that construction of physical map contigs for regions of the soybean genome containing rDNA and other highly tandem repeats might not be possible by restriction fingerprint analysis, using *Hin*dIII/*Hae*III, *Bam*HI/*Hae*III or *Sau*3AI alone. Different techniques, different restriction enzymes and/or their different combinations may be needed to construct contig maps for these genomic regions by fingerprint analysis.

Acknowledgements This study was supported in part by the National Science Foundation Plant Genome Program (award no. 9872635), the Texas Agricultural Experiment Station (account no. 8536-203104), and the Illinois Soybean Program Operating Board (award no. 98-24-128-3).

References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

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

Chang Y-L, Henriquez X, Preuss D, Copenhaver GP, Zhang H-B (2003) A plant transformation-competent binary BAC library of the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomics. Theor Appl Genet 106:269–276

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DJ, Chung J, Specht JE (1999) The integrated map of the soybean genome. Crop Sci 39:1464–1490

Creusot F, Fouilloux E, Dron M, Lafleuriel J, Picard G, Billault A, Paslier DL, Cohen D, Chaboute M-E, Durr A, Fleck J, Gigot C, Camilleri C, Bellini C, Caboche M, Bouchez D (1995) The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*. Plant J 8:763–770

Danesh D, Penuela S, Mudge J, Denny RL, Nordstorm H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor Appl Genet 96:196–202

Eckenrode VK, Arnold J, Meagher RB (1985) Comparison of the nucleotide sequence of soybean 18S rRNA with the sequences of other small-subunit rRNAs. J Mol Evol 21:259–269

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

Gerlach WL, Dyer TA (1980) Sequence organization of the repeating units in the nucleus of wheat which contains 5S rRNA genes. Nucleic Acids Res 8:4851–4865

Gregory SG, Sekhon M, Schein J, Zhao S, Osoegawa K, Scott CE, Evans RS, Burridge PW, Cox TV, Fox CA, Hutton RD, Mullenger IR, Phillips KJ, Smith J, Stalker J, Threadgold GJ, Birney E, Wylie K, Chinwalla A, Wallis J, Hillier L, Carter J, Gaige T, Jaeger S, Kremitzki C, Layman D, Maas J, McGrane R, Mead K, Walker R, Jones S, Smith M, Asano J, Bosdet I, Chan S, Chittaranjan S, Chiu R, Fjell C, Fuhrmann D, Girn N, Gray C, Guin R, Hsiao L, Krzywinski M, Kutsche R, Lee SS, Mathewson C, McLeavy C, Messervier S, Ness S, Pandoh P, Prabhu AL, Saeedi P, Smailus D, Spence L, Stott J, Taylor S, Terpstra W, Tsai M, Vardy J, Wye N, Yang G, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Shvartsbeyn A, Gebregeorgis E, Krol M, Russell D, Overton L, Malek JA, Holmes M, Heaney M, Shetty J, Feldblyum T, Nierman WC, Catanese JJ, Hubbard T, Waterston RH, Rogers J, de Jong PJ, Fraser CM, Marra M, McPherson JD, Bentley DR (2002) A physical map of the mouse genome. Nature 418:743–750

Hamilton C, Frarey 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

Hnetkovsky N, Chang SJC, Doubler TW, Gibson PT, Lightfoot DA (1996) Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). Crop Sci 36:392–400

Lightfoot DA, Meksem K, Zhang H-B (2003a) Integrative physical mapping of the soybean genome. In: VandenBosch KA, Stacey G (eds) Summaries of legume genomics projects from around the globe. Community resources for crops and models. Plant Physiol 131:19–20

Lightfoot DA, Njiti VN, Gibson PT, Schmidt ME, Myers O Jr, Iqbal JM, Meksem K (2003b) Registration of Essex × Forrest recombinant inbred line (RIL) mapping population. Crop Sci 43:2422–2423

Lim KY, Kovarik A, Matyasek R, Bezdek M, Lichtenstein CP, Leitch AR (2000) Gene conversion of ribosomal DNA in *Nicotiana tabacum* is associated with undermethylated, decondensed and probably active gene units. Chromosoma 109:161–172

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

Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. Genome 40:420–427

Meksem K, Zobrist K, Ruben E, Tao Q, Zhang H-B, Lightfoot DA (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

Mozo T, Fischer S, Shizuya H, Altmann T (1998) Construction and characterization of the IGF *Arabidopsis* BAC library. Mol Gen Genet 258:562–570

Ren C, Lee M-K, Yan B, Ding K, Cox B, Romanov MN, Price JA, Dodgson JB, Zhang H-B (2003) A BAC-based physical map of the chicken genome. Genome Res 13:2754–2758

Sambrook J, Maniatis T, Fritsch EF (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Shi L, Zhu T, Keim P (1996) Ribosomal RNA genes in soybean and common bean: chromosomal organization, expression, and evolution. Theor Appl Genet 93:136–141

Shishido R, Sano Y, Fukui K (2000) Ribosomal DNAs: an exception to the conservation of gene order in rice genomes. Mol Gen Genet 263:586–591

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

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 10:1772–1787

Sulston J, Mallett F, Staden R, Durbin R, Horsnell T, Coulson A (1988) Software for genome mapping by fingerprinting techniques. CABIOS 4:125–132

Sun S, Xu Z, Wu C, Ding K, Zhang H-B (2003) Genome properties and their influences on library construction and physical mapping. In: Proceedings of the plant and animal genomes XI conference, P77 (abstract). San Diego, California

Tao Q, Chang YL, Wang J, Chen H, Islam-Faridi MN, Scheuring C, Wang B, Stelly DM, Zhang H-B (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158:1711–24

Tao Q, Wang A, Zhang H-B (2002) One large-insert plant-transformation-competent BIBAC library and three BAC libraries of japonica rice for genome research in rice and other grasses. Theor Appl Genet 105:1058–1066

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

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

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

Tomkins JP, Mahalingham R, Miller-Smith H, Goicoechea JL, Knapp HT, Wing RA (1999) A soybean bacterial artificial chromosome library for PI 437654 and the identification of clones associated with cyst nematode resistance. Plant Mol Biol 41:25–32

Tomkins JP, Frisch DA, Byrum JR, Wing RA (2000) Construction and characterization of a BAC library for the soybean cultivar A3244. Soybean Genetics Newsletter 27. http://www.soygenetics.org/articles/sgn2000-002.htm

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science 291:1304–1351

Wu C, Sun S, Padmavathi N, Santos FA, Springman R, Meksem K, Lightfoot D, Zhang H-B (2004a) A BAC and BIBAC-based physical map of the soybean genome. Genome Res 14:319–326

Wu C, Xu Z, Zhang H-B (2004b) DNA Libraries. In: Meyers RA (ed) Encyclopedia of molecular cell biology and molecular medicine, vol 3, 2nd edn. Wiley, Weinheim, pp 385–425

Xu Z, Sun S, Covaleda L, Ding K, Zhang A, Zhang H-B (2003) Whole-genome physical mapping with BACs: fingerprinting methods, source clone genome coverage, and map contig accuracy and size. In: Proceedings of the plant and animal genomes XI conference, W224 (abstract). San Diego, California

Zhang H-B (2000) Construction and manipulation of large-insert bacterial clone libraries—manual. Texas A&M University, Texas. http://hbz.tamu.edu

Zhang H-B, Wu C-C (2001) BAC as tools for genome sequencing. Plant Physiol Biochem 39:1–15

Zhang H-B, Zhao XP, Ding XD, Paterson AH, Wing RA (1995) Preparation of megabase-sized 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. Wiley, England, pp 75–99