Theor Appl Genet (1999) 98: 712–720

S. S. Salimath · M. K. Bhattacharyya

Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the *Rps1*-k disease resistance gene

Received: 12 May 1998 / Accepted: 24 August 1998

Abstract A soybean bacterial artificial chromosome (BAC) library, comprising approximately 45 000 clones, was constructed from high-molecular-weight nuclear DNA of cultivar Williams 82, which carries the *Rps1*-k gene for resistance against *Phytophthora sojae*. The library is stored in 130 pools with about 350 clones per pool. Completeness of the library was evaluated for 21 random sequences including four markers linked to the *Rps1* locus and 16 cDNAs. We identified pools containing BACs for all sequences except for one cDNA. Additionally, when screened for possible contaminating BAC clones carrying chloroplast genes, no sequences homologous to two barley chloroplast genes were found. The estimated average insert size of the BAC clones was about 105 kb. The library comprises about four genome equivalents of soybean DNA. Therefore, this gives a probability of 0.98 of finding a specific sequence from this library. This library should be a useful resource for the positional cloning of *Rps1*-k, and other soybean genes. We have also evaluated the feasibility of an RFLP-based screening procedure for the isolation of BAC clones specific for markers that are members of repetitive sequence families, and are linked to the *Rps1*-k gene. We show that BAC clones isolated for two genetically linked marker loci, Tgm*r* and TC1-2, are physically linked. Application of this method in expediting the map-based cloning of a gene, especially from an organism, such as soybean, maize and wheat, with a complex genome is discussed.

Key words DNA · Gene cloning · Genome analysis · *Glycine max* · Legume

Communicated by G. E. Hart

S. S. Salimath · M. K. Bhattacharyya (✉)
Plant Biology Division, The Samuel Roberts Noble Foundation,
P.O. Box 2180, 2510 Sam Noble Parkway,
Ardmore, OK 73402, USA
Fax: + 1-580-221-7380
E-mail: mkbhattach@noble.org

Introduction

One of the essential requirements for plant genome analysis, especially for physical mapping, gene isolation, and gene structure and function analysis, is the availability of clone libraries with large DNA inserts. During the last two decades two cloning vectors, i.e., cosmid (Collins and Hohn 1978) and yeast artificial chromosomes (YACs) (Burke et al. 1987), have been extensively used in generating genomic libraries with large DNA inserts. The advantages and disadvantages associated with these two large DNA cloning vectors have been eloquently discussed by Ioannou et al. (1994) and Monaco and Larin (1994). The YAC library generated for tomato has proven to be highly useful, especially in map-based gene cloning (Martin et al. 1992, 1993). Recently, other large DNA cloning vectors, such as bacteriophage P1 (Sternberg 1990), the bacterial artificial chromosome (BAC) (Shizuya et al. 1992) and the P1-derived artificial chromosome (PAC) (Ioannou et al. 1994), have been developed. Using the bacteriophage P1 cloning system a plant genomic library for *Arabidopsis*, with an average insert size 80 kb, was developed by Liu et al. (1995). The BAC vector, developed from an *Escherichia coli* 'F'-factor plasmid, was shown to maintain human genomic DNA fragments of > 300 kb quite stably (Shizuya et al. 1992). Recently, this vector was also used for generating genomic DNA libraries in economically important crop plants such as rice (Wang et al. 1995; Zhang et al. 1996; Yang et al. 1997), sorghum (Woo et al. 1994), lettuce (Frijters et al. 1997) and soybean (Danesh et al. 1998; Marek and Shoemaker 1997), as well as in *Arabidopsis* (Wang et al. 1996).

Soybean [*Glycine max* (L.) Merr.] is an important legume cultivated throughout the world as an oilseed crop. As soybean is a rich source of oil (20%) and protein (40%), it is of great significance in human consumption, livestock feed and for industrial purposes (Fehr 1980). In North America, soybean suffers severe

yield losses from the fungal disease, phytophthora root and stem rot, caused by *Phytophthora sojae* M. J. Kaufmann & J. W. Gredemann. The resistance conferred by a series of 14 dominant single genes (*Rps*) present at seven loci has been proven to be successful in controlling 37 recorded races of this fungal pathogen (see Kasuga et al. 1997 for references). Among these *Rps* genes, the *Rps1* locus alone carries six alleles. One of these, *Rps1*-k, is known to confer resistance to at least 21 races of *P. sojae*, and has been widely used in commercial soybean cultivars for nearly two decades (Schmitthenner et al. 1994). We are interested in the positional cloning of this agronomically important disease resistance gene. High-resolution genetic and physical mapping data around the *Rps1*-k region indicated that two AFLP markers flanking the *Rps1*-k gene are located within 125 kb of each other (Kasuga et al. 1997). The main objectives of the present investigation were: (1) to generate a large insert genomic BAC library that is essential for the map-based cloning of *Rps1*-k and other important soybean genes; and (2) to evaluate the feasibility of an RFLP-based screening procedure in the construction of a chromosome contig for molecular markers that are members of repetitive sequence families in soybean.

Materials and methods

Plant materials

Soybean cv Williams 82, which carries the *Phytophthora* disease resistance gene *Rps1*-k, is a backcross derivative of the cross between cv Williams and cv Kingwa, and was developed at the University of Illinois, Urbana. Williams was crossed with Kingwa, the source of *Rps1*-k, followed by six backcrosses and selection for resistance. Four $BC_6\ F_{2:3}$ lines homozygous for *Rps1*-k were bulked to produce Williams 82 (see Kasuga et al. 1997 for references).

Preparation and partial restriction digestion of high-molecular- weight (HMW) DNA

For DNA, Williams 82 plants were grown in growth chambers under standard conditions (Bhattacharyya and Ward 1986). HMW DNA, from young leaves of Williams 82, was extracted following the protocol of Liu and Whittier (1994). The final steps of proteinase K treatment, PMSF wash and dialysis of plugs in TE buffer were as described by Wang et al. (1995). Approximately 8–10 μg of HMW DNA was present in each agarose plug (76-μl vol). For partial restriction digestion of HMW DNA, plugs were cut into quarters. To each quarter plug, 100 μl of restriction cocktail (*Mbo*I enzyme + 1 × buffer) was added and the tubes were incubated on ice for 24 h. Later, the tubes were transferred to 37°C for 5 min. Restriction digestion was stopped by adding 50 μl ice-cold 0.5 M EDTA (pH 8), and the plugs were equilibrated with 0.5 × TBE.

Clamped homogeneous electric field (CHEF) gel-electrophoresis and isolation of HMW DNA

An agarose gel (1% Sea Plaque GTG agarose, FMC, USA, in 0.5 × TBE) with partially digested HMW DNA and megabase DNA markers (Bio-Rad, USA) was run on a BIO-RAD CHEF mapper using a two-state program. Initially, the gel was run for 6 h with switch times of 42 s (initial), 8 min and 0.01 s (final), an angle of 120° and a 6 V gradient. This helped to release low-molecular-weight DNA trapped in the HMW DNA. The same gel was run for a further 16 h with switch times 6.03 s (initial), 1 min 5.59 s (final), an angle of 120° and a 6 V gradient. After electrophoresis, isolation of HMW DNA (approximately 200–350 kb size) from the CHEF gel was according to Wang et al. (1995). Three units of Gelase enzyme (*Epicentre* Technologies, USA), per 600 mg of gel, were used to completely digest the gel at 45°C. This HMW DNA solution was used directly in ligation reactions.

Preparation of vector DNA, ligation and transformation

The pECBAC1 vector used in this study contains a unique *Eco*RI cloning site (Frijters et al. 1997), and was derived from the pBeloBAC11 vector (Kim et al. 1996). Large-scale preparation and purification of plasmid DNA was as described by Heilig et al. (1995). Purified plasmid DNA was digested completely with the restriction enzyme *Bam*HI, and de-phosphorylated using calf intestine alkaline phosphatase (Boehringer Mannheim, USA). About 50 ng of HMW DNA (approximately 200–350 kb size) was ligated to *Bam*HI-digested pECBAC1 vector DNA (10 M excess) at 16°C overnight in a final volume of 100 μl, using 20 units of T4 DNA ligase enzyme (NEB, USA). Following electroporation, *Escherichia coli* transformants were grown on LB plates containing 12.5 μg/ml of chloramphenicol (Wang et al. 1995).

Division and storage of BAC library

For efficient handling and thorough screening of the library, 45 000 BAC clones were divided into 130 primary pools. Each primary pool comprised about 350 individual clones collected directly from 1 to 3 plates in 5 ml of LB containing 12.5 μg/ml of chloramphenicol. Plasmid DNA was extracted using 3 ml of this bacterial suspension (Sambrook et al. 1989). The remaining 2 ml of the bacterial suspension of each pool were stored (in triplicate) at −80°C as glycerol stocks.

Restriction fragment length polymorphism (RFLP) analysis of BAC "superpools"

Equal amounts of recombinant plasmid DNA from two primary pools were combined to form a "superpool". About 2 μg of DNA from each of 65 superpools were separately digested with *Bgl*II, *Hin*dIII and *Taq*I enzymes (GIBCO-BRL, USA), and were blotted onto nylon membranes. Southern analyses were carried out using one AFLP marker, TC1 (Kasuga et al. 1997); three RFLP markers, Tgm*r*, pA-71 and pA-280, all linked to *Rps1*-k, (Diers et al. 1992; Bhattacharyya et al. 1997); one soybean cDNA, *PLC1* (Shi et al. 1995 a), and two barley photosynthetic (chloroplast) genes (courtesy of V. Rajasekhar, UC Irvine, USA) (Table 1).

Screening of BAC "giantpools" by the polymerase chain reaction (PCR)

PCR-based screening of the BAC library was according to Green and Olson (1990). For use in PCR, an equal amount of DNA from three superpools was combined to form a "giantpool". Therefore, 65 superpools were combined to form 22 "giantpools". Genomic DNA of Williams 82 was used as a positive control. Primers used in PCR amplifications were derived from the sequence information of 16

Table 1 RFLP-based screening of the BAC library

Probe[a]	Restriction[b] Enzyme	No. of superpools detected
Tgm*r*	*Hin*dIII	4
pA-71	*Taq*I	1
pA-280	*Hin*dIII	4
TC1–2	*Taq*I	2
PLC1	*Bgl*II	1
Chloroplast-I[c]	*Hin*dIII	0
Chloroplast-II	*Taq*I	0

[a] Probes are described in the Materials and methods, and in the text
[b] Restriction enzymes used to prepare the DNA blots of BAC superpools
[c] A single monomorphic band for each of chloroplast genes was detected in the genomic DNA of soybean cvs Williams and Williams 82

Table 2 PCR screening of the BAC library

Primer pair no[a]	No. of giant pools detected
1	1
2	0
3	6
4	1
5	1
6	3
7	3
8	6
9	4
10	2
11	1
12	3
13	2
14	10
15	4
16	4

[a] Primer pair nos. 1–13 were based on the sequences of 13 soybean cDNA clones with unknown functions (Bhattacharyya, unpublished). Primer pair nos. 14, 15 and 16 were based on the sequences of Annexin, VCP and *Erg6* cDNAs, respectively (Shi et al. 1995 b)

soybean cDNA clones (Table 2). PCR reactions, following Cai et al. (1995), were carried out in a thermocycler (MJ Research Inc., USA). PCR-amplified products were analyzed by Southern-blot analysis.

Isolation of BAC clones from BAC pools

Approximately 60 000 cfu from each primary pool carrying a candidate clone were plated on LB plates containing 12.5 μg/ml of chloramphenicol. Colony lifts on nylon membranes were prepared (Sambrook et al. 1989) and baked at 80°C under vacuum for 2 h. They were washed and hybridized to a DNA probe(s) following standard protocols (Sambrook et al. 1989). The colonies that hybridized to the probe were isolated from the original plates and streaked onto new LB plates containing chloramphenicol, and then re-screened. In most cases the candidate BAC clones were purified at this step.

BAC end-sequencing and generation of BAC end specific probes

For BAC end-sequencing BAC clones were digested separately with *Hin*dIII and *Eco*RI, resulting in a linear vector along with small insert DNA. The linearized DNA was re-ligated using T4 DNA ligase. The re-ligated DNA was electrotransformed and bacteria were spread on LB medium containing chloramphenicol (12.5 μg/ml). Purified individual BAC clones were sequenced using T7 and M13 (reverse) primers. In one case, however, a BAC clone of about 34 kb was sequenced directly using T7 and M13 (reverse) primers without going through this subcloning procedure. From the sequence information obtained for each end of the BAC a pair of primers was designed. PCR products obtained from the respective BAC template and primer-pair combination were used as DNA probes in Southern hybridization.

Probe radiolabeling and hybridization

Gel-purified DNA fragments were radiolabeled by random primer extension (Feinberg and Vogelstein 1984). Pre-hybridization, hybridization and membrane washes were at 65°C according to Church and Gilbert (1984).

Results

Construction of a soybean BAC library

*Mbo*I partially digested HMW DNA of soybean was cloned into the *Bam*HI site of the pECBAC1 vector (Frijters et al. 1997). One microliter of ligation mixture yielded about 150 recombinant clones. Therefore, the efficiency of cloning was 3×10^5 cfu per μg of insert DNA. To measure the average insert size of recombinant BAC clones, 150 individual colonies were randomly picked from all 130 pools, each carrying about 350 cfu. Miniprep DNA of each clone was digested with *Not*I and analyzed on a CHEF mapper. The insert size ranged from 30 kb to over 300 kb with an average insert size of approximately 105 kb. Among 150 clones analyzed, 139 (about 93%) contained inserts and over 43% of these inserts were larger than 100 kb, whereas 61% of the clones were over 75 kb. Only about 11% of the clones contained inserts in the range of 30–50 kb (Fig. 1). Sixteen randomly selected clones were grown and subcultured serially for 5 days. At the end of the 5th day of subculturing the bacteria had undergone 100 cycles of growth. Recombinant plasmid DNA isolated from the day 1 and day 5 cultures of bacteria were digested with *Hin*dIII. Comparison of DNA fingerprints from day 1 and day 5 cultures did not reveal any differences, indicating the stability of soybean DNA in the BAC clones.

RFLP-based screening of the BAC pools

The feasibility of identifying BAC clones carrying either a specific molecular marker, out of several copies, or a specific gene of a multigene family from the soybean

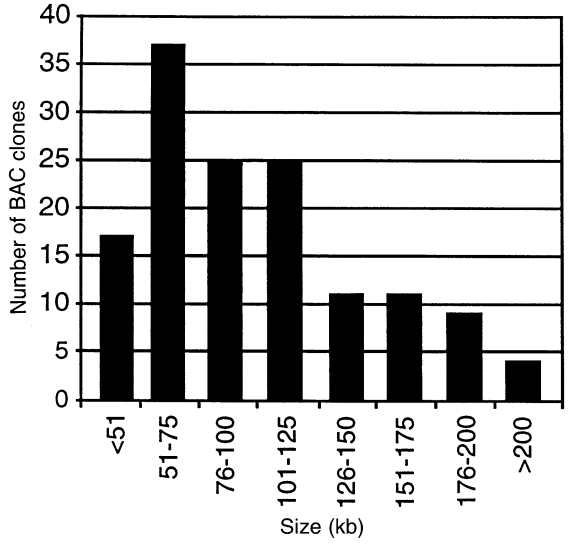

Fig. 1 Size-range distribution in 139 BAC clones. One hundred and fifty BAC clones were randomly selected from 130 BAC pools and the BAC miniprep DNA digested with *Not*I was sized in a CHEF mapper. Eleven clones had no inserts

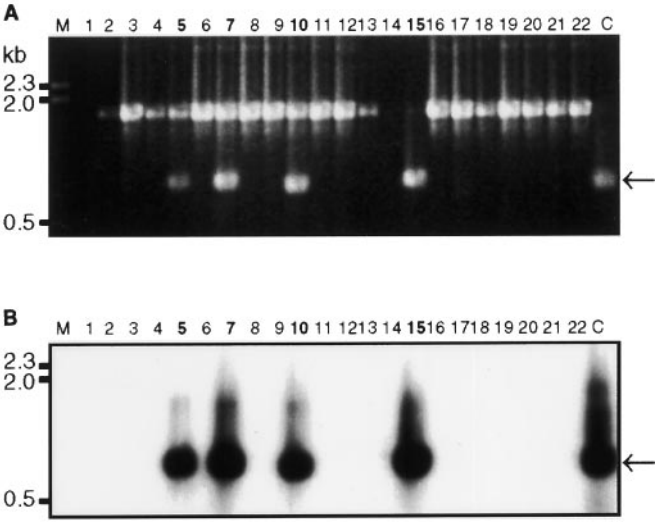

Fig. 2A, B PCR-based screening of the BAC library. **A** Ethidium bromide-stained gel showing PCR-amplified bands. The *Arrow* indicates the amplified band from positive control (*C*) – Williams 82, and giantpools 5, 7, 10 and 15. **B** PCR products from **A** were transferred to a nylon membrane and hybridized with ^{32}P-labeled soybean *Erg6* cDNA. The *arrow* indicates PCR products strongly hybridizing with their specific cDNA

BAC library, was tested by a method based upon RFLPs, or DNA fingerprints in Southern blots of BAC DNA pools. This method of BAC library screening involves three steps: (1) pooling of BAC clones, (2) restriction digestion and Southern blotting of recombinant plasmid DNA from BAC pools, and (3) probing the Southern blots with DNA markers. This approach can directly lead to the identification of pools carrying the BAC clones with sequences identical to that of the marker. Finally, the BAC clones can be purified from a specific primary pool, carrying only a limited number of clones, through colony hybridization.

Southern-hybridization experiments with BAC superpools revealed that BAC clones containing sequences homologous to Tgm*r* and pA-280 were found in four superpools, whereas BAC clones containing pA-71 and *PLC1*-related sequences were found in one superpool (see Table 1). The AFLP marker TC1, upon conversion into an RFLP marker(s), mapped to two independent loci. One copy of TC1 was tightly linked to *Rps1*-k, and the other was tightly linked to the Tgm*r* locus. From the Southern analysis of *Taq*I-digested genomic DNA we only distinguished alleles of the TC1 locus linked to Tgm*r*. This locus was termed TC1-2. Screening of the library revealed two TC1-2-specific BAC superpools (see Table 1).

The library was screened for possible contaminating BAC clones carrying photosynthetic (chloroplast) genes. Southern blots carrying BAC pool DNA were hybridized to two barley photosynthetic (chloroplast) genes. The chloroplast genes were not detected in any of the BAC superpools. However, a single monomorphic band for each of the chloroplast genes was detected in the genomic DNA of soybean cvs Williams and Williams 82 (see Table 1).

PCR-based screening of the BAC pools

PCR-based screening of the BAC library was carried out according to Green and Olson (1990). For PCR, DNA samples from three superpools were consolidated to form one giantpool, representing 0.18 genome equivalents. PCR products of identical size, originating from both Williams 82 and BAC pools, and the occurrence of one band per giantpool were considered to represent the presence of at least one cDNA – specific genomic clone in the BAC library (Fig. 2). The number of giantpools that yielded PCR amplified products varied from 0 to 10 among the 16 pairs of primers (see Table 2).

Isolation of candidate BAC clones from BAC pools

The feasibility of isolating marker-specific BAC clones from the BAC pools was tested through colony hybridization for two tightly linked molecular markers, TC1-2 and Tgm*r*. DNA samples isolated from BAC superpools, and genomic DNA of cvs Williams and Williams 82 were digested with *Taq*I and *Hin*dIII, and DNA filters were prepared for Southern hybridization with TC1- and Tgm*r*-specific probes, respectively. Two BAC superpools, (47 and 48) showed the TC1-2-specific band (Fig. 3A), while superpools 38, 42, 48 and 61

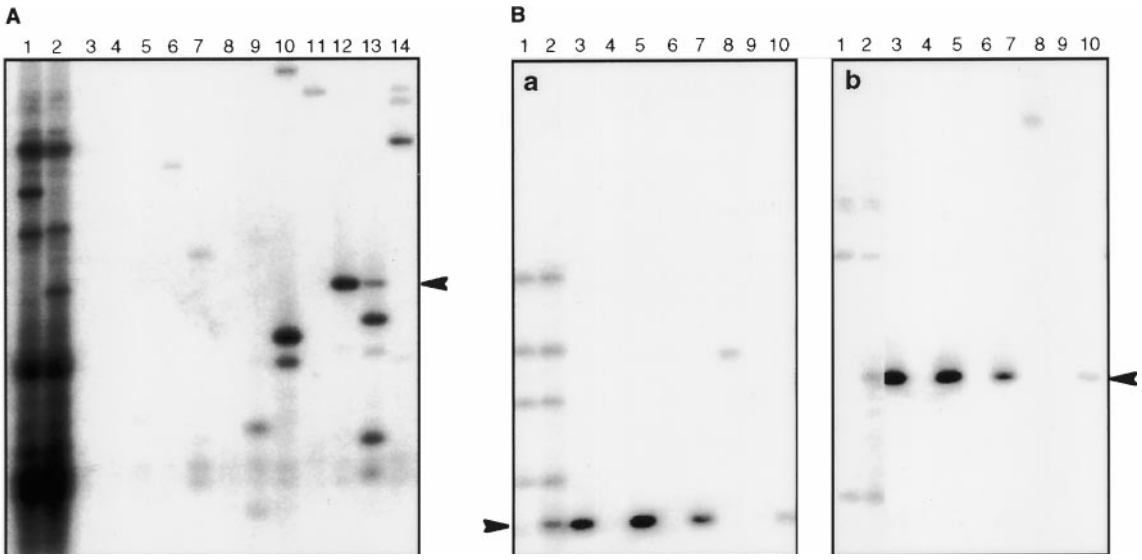

Fig. 3A, B RFLP-based screening of BAC pools. **A** RFLP-based screening of BAC superpools for TC1-2. DNA samples of the super-pools (see Materials and methods) were digested with *Taq*I and electrophoresed along with *Taq*I-digested genomic DNA of cvs Williams (*lane 1*) and Williams 82 (*lane 2*). *Lanes 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, and 14* represent superpools 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, and 49 respectively. Super-pools 47 (*lane 12*) and 48 (*lane 13*) showed the TC1-2-specific band (*arrow*). DNA probe TC1 was PCR-amplified from an AFLP gel (Kasuga et al. 1997). **B** RFLP-based screening of BAC pools for Tgm*r*. Four superpools (nos. 38, 42, 48 and 61) containing Tgm*r*-specific sequences were identified through an RFLP-based analysis of superpools. DNA of individual primary pools from each of the candidate superpools carrying Tgm*r* sequences was digested with either *Hin*dIII (a) or *Eco*RV (b). The Tgm*r*-specific band that is tightly linked to the *Rps*1-k allele was identified in four primary BAC pools, 75 (*lane 3*), 83 (*lane 5*), 95 (*lane 7*) and 122 (*lane 10*), and is shown by *arrows*. *Lane 1*, cv Williams, and lane 2, Williams 82. DNA probe OPRK15 was used in identifying the Tgm*r*-specific band (Bhattacharyya et al. 1997)

showed the Tgm*r*-specific band. Further, RFLP-based Southern analysis of the primary pools 92, 93, 94, and 95 that constituted the superpools 47 and 48 revealed that primary pools 92 and 95 contain TC1-2-specific sequences. In pool 95 many DNA fragments hybridized to TC1 (similar to superpool 48 of Fig. 3 A). Therefore, pool 92 carrying only the TC-2-specific sequences was further characterized. The RFLP-based Southern-blot analysis of the eight primary pools that constituted superpools 38, 42, 48 and 61 revealed that four of the primary BAC pools, 75, 83, 95 and 122, contained the Tgm*r*-specific polymorphic fragments in both *Hin*dIII- and *Eco*RV-digested BAC DNA (Fig. 3B). TC1-2, and Tgm*r*-specific BAC clones BAC92, BAC75, BAC83, BAC95, and BAC122 were purified through colony hybridization (see Materials and methods) from the primary BAC pools 92, 75, 83, 95, and 122, respectively. Based on CHEF gel analysis, the approximate sizes of these five BAC clones were 34, 85, 160, 100, and 60 kb, respectively.

Tgm*r*- and TC1-2-specific BAC clones are physically linked

DNA fingerprinting analysis indicated that all four Tgm*r*-specific BAC clones, i.e., BAC75, BAC83, BAC95 and BAC122, have overlapping sequences, whereas BAC83 and BAC95 carry unique sequences. Therefore, using sequence information, obtained through BAC-end rescue and BAC-end sequencing, primer pairs were synthesized for each end of BAC83 (approximately 160 kb) and BAC95 (approximately100 kb). Two primer pairs, no. 1 and no. 2, were synthesized for BAC83 ends, whereas primer pairs no. 3 and no. 4 were synthesized for BAC95 ends. All four primer pairs were used in the PCR of Tgm*r*-specific BAC clones. Primer pair nos. 1 and 4 respectively amplified sequences from BAC83 and BAC95 only, whereas primer pair nos. 2 and 3 amplified from all four Tgm*r*-specific BAC clones. This indicated that DNA sequences amplified from primer pair nos. 2 and 3 were common to all Tgm*r*-specific BACs. PCR products obtained from each end of BAC83 and BAC95 were used as DNA probes in Southern hybridization. Probe nos. 1 and 2 were generated from a BAC83 template using primer pair nos. 1 and 2, respectively. Probe nos. 3 and 4, were generated from a BAC95 template using primer pair nos. 3 and 4, respectively. Southern-blot analysis of the *Bam*HI-, *Eco*RV-, *Hin*dIII-, *Pvu*II- and *Xba*I-digested DNA of BAC75, 83, 95 and 122 revealed that probe nos. 1 and 4 hybridized to BAC83 and BAC95, respectively. On the other hand, probe nos. 2 and 3 hybridized to all 4 Tgm*r*-specific BACs, thus confirming the PCR results that a common overlapping DNA sequence exists among them. Further, radiolabeled soybean total genomic DNA was hybridized to the Southern blot of all four BAC clones. The DNA fingerprinting pattern and the common restriction fragments that hybridized to the genomic DNA indicated an approximately 55 kb overlap between BAC83 and BAC95 (data not shown).

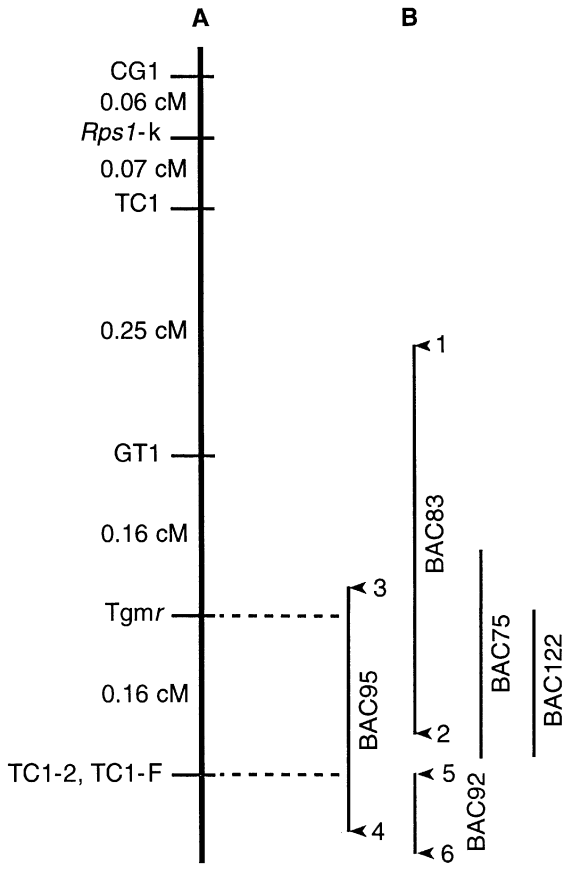

Fig. 4 Chromosome contig near the *Rps1*-k region. Tgm*r*- and TC1-2- specific BAC clones were purified and aligned based on their DNA-fingerprint analysis. The physical map of the BACs is proportional to the size of the BAC clones (BAC95 = 100 kb). *Arrows* indicate the approximate positions of individual PCR primer combinations (see text)

BAC75 and BAC122 overlapped with both BAC83 and BAC95 (Fig. 4).

The ends of TC1-2-specific clone BAC92 (TC1-BAC in Kasuga et al. 1997) were directly sequenced, and primer pairs 5 and 6 were synthesized. One end of BAC92 which is specific to primer pair 5 was converted to a RFLP marker. This marker co-mapped with TC1-2, and was termed TC1-F (Kasuga et al. 1997). TC1-2-, and TC1-F-specific probes were evaluated for hybridization with the Tgm*r*-specific BAC clones. Both probes hybridized to BAC95 indicating overlaps between BAC92 and BAC95. This also supported our initial observation that superpool 48 carrying BAC95 contains a TC1-2-specific band (Fig. 3A). Probe no. 4 obtained from BAC95 also hybridized to BAC92 in a Southern analysis. PCR analysis of BAC92 and BAC95 also showed that primer pairs 4 and 5 can amplify products from BAC92 and BAC95, respectively. DNA fingerprint analysis of BAC92 and BAC95 using TC1-2-, TC1-F-, probe no. 4, and total genomic DNA as probes suggested that there is an overlap of over 20 kb between these two BAC clones (Fig. 4). These results confirmed that the Tgm*r*- and TC1-2-specific BAC clones, independently isolated by the RFLP-based screening approach, are physically linked.

Discussion

Construction of a soybean BAC library

In the present investigation a soybean BAC library was generated from HMW DNA prepared from nuclei of leaf tissues (Liu and Whittier 1994). The library consists of about 45 000 BAC clones with an average insert size of about 105 kb. Therefore, this BAC library carrying 93% recombinant clones represents at least about 44 000 megabases of DNA. Completeness of the soybean BAC library, for random sequences, was evaluated for 21 sequences, including four markers linked to the *Rps1* locus, and 16 cDNAs (Tables 1 and 2). We identified pools carrying BACs for all but one cDNA. Considering the 1115 Mb haploid genome size of soybean (Arumuganathan and Earle 1991), this library represents approximately four genome equivalents. Therefore, this gives us a probability of 0.98 of finding a specific sequence from this library.

In plants, extraction of HMW DNA has been mainly through the nuclear isolation method (Liu and Whittier 1994) or by protoplast isolation (Ganal and Tanksley 1989). To protect from shearing, the HMW DNA extracted by either of these two methods is embedded in low-melting point agarose to form plugs or mixed with low-melting point agarose and mineral oil to form microbeads (Wing et al. 1993; Liu and Whittier 1994; Zhang et al. 1995). One of the major concerns of plant HMW DNA libraries is the occurrence of organellar DNA clones, especially from chloroplasts (Eyers et al. 1992; Martin et al. 1992; Woo et al. 1994; Wang et al. 1995; Umehara et al. 1995). Chloroplast sequence contamination contributes to the overestimation of the genomic content of the library. YAC libraries developed for several plant species contained as much as 27% chloroplast DNA contamination (Schmidt et al. 1994). Recently, the BAC libraries developed for the rice cultivar IR-BB21 (Wang et al. 1995), *indica* and *japonica* rice (Zhang et al. 1996), and *Sorghum* (Woo et al. 1994) were reported to contain 0.3%, 2.9%, 0.94% and 14% chloroplast DNA sequences, respectively. The occurrence of an unusually high per cent of chloroplast DNA in the *Sorghum* BAC library is due mostly to extraction of HMW DNA through the protoplast isolation method (Woo et al. 1994). The use of HMW DNA prepared from nuclei in the construction of BAC libraries for rice greatly reduced the number of contaminating BAC clones carrying chloroplast DNA (Wang et al. 1995; Zhang et al. 1996). We used nuclear DNA in constructing the BAC library and observed no

BAC clones containing sequences similar to two barley chloroplast genes.

Chromosomal rearrangements, deletions and chimerism are frequently associated with YAC clones (Monaco and Larin 1994) and they pose serious problems, especially in map-based gene cloning and physical mapping (Cai et al. 1995; Kim et al. 1996; Zhang et al. 1996). As observed through serial culture for over 100 generations of *E. coli*, large insert-containing soybean BAC clones are highly stable and do not undergo deletions or rearrangements. It is reported that YAC libraries may contain up to 40% chimeric genomic clones (Umehara et al. 1995). However, as evident through chromosomal in situ hybridization experiments, BACs do not, or seldom, exhibit chimerism (Woo et al. 1994; Cai et al. 1995; Kim et al. 1996; Zhang et al. 1996).

Organization and RFLP-based screening of the BAC library

The conventional method of screening genomic libraries (e.g. YAC, BAC, PAC and P1 libraries) is by colony hybridization or by arraying individual clones into microtiter plates and then replicating them onto nylon membranes for further screening and identification of positive clones (Ioannou et al. 1994; Liu et al. 1995; Umehara et al. 1995; Zhang et al. 1996). Arraying individual clones of a large – genome library containing few to several-fold genome equivalents into microtiter plates is a time-consuming and laborious procedure. An alternative approach for screening HMW DNA libraries, through the PCR of pooled clones was first described by Green and Olson (1990). High sensitivity and specificity is the major advantage associated with PCR-based screening of pooled DNA libraries (Green and Olson 1990). However, PCR conditions may fail to distinguish specific sequences from a family of multiple copies.

We have described here the screening of a BAC library through Southern-blot analysis of DNA isolated from BAC pools (Fig. 3). Based on RFLPs between two cultivars, one can identify the pools that carry the polymorphic sequence. This approach eliminates the characterization of several hybridizing clones of repetitive sequences, required in a conventional screening approach. This RFLP-based method should aid in the rapid identification of BAC clones specific to a polymorphic band of a multicopy DNA marker. Here, we have isolated four BAC clones that carried the Tgm*r* marker and two BAC clones that carried the TC1-2 marker, tightly linked to the *Rps1*-k gene. This method allowed us to avoid the unnecessary characterization of 17 additional superpools carrying Tgm*r*-related monomorphic sequences (data not shown). Furthermore, BAC clones isolated by this procedure allowed us confidently to land onto *Rps1*-k-linked molecular loci. This approach eliminates the need for mapping the BAC ends to confirm the map positions of BAC clones. In the present study we have shown that BAC clones isolated for two independent, genetically linked loci are physically linked (Fig. 4). Therefore, this technique can expedite the map-based cloning of a gene, especially from an organism, such as soybean, maize and wheat, with a complex genome.

Correlation between genetic and physical distance in the *Rps1*-k region

Molecular markers Tgm*r* and TC1-2 respectively mapped at about a 0.5 cM and a 0.66 cM distance from the *Rps1*-k gene (Fig. 4; Kasuga et al. 1997). Four Tgm*r*-specific and two TC1-2-specific BAC clones were isolated by RFLP-based screening. Southern analysis of BAC clones using TC1-2- and Tgm*r*-specific probes showed that in a single approximately 100-kb clone, BAC95, DNA fragments specific to these two markers are present (Fig. 3). Hybridization with BAC end-specific probes, and PCR analysis using BAC-specific primer pairs also supported the fact that Tgm*r*- and TC1-2-specific BACs formed a contig. This clearly indicated that a single clone, BAC95, of 100 kb size contained both Tgm*r* and TC1-2. The approximate genetic distance between these two molecular markers is 0.16 cM (Kasuga et al. 1997). Therefore, based upon the available evidence, it is likely that the 0.16 cM genetic distance around *Rps1*-k may correspond to $<$ 100 kb of DNA. This relationship between genetic and physical distances is comparable to our earlier observation for the *Rps1*-k region (Kasuga et al. 1997).

Application of the soybean BAC library in positional cloning

The utility of large insert libraries is mainly in the generation of chromosome contigs and the isolation of genes through map-based cloning (Martin et al. 1992, 1993; Wang et al. 1995). Genomic libraries with insert sizes less than 100 kb are considered to be undesirable (Cai et al. 1995). This is especially true when long chromosomal walks are to be performed. However, with the invention of powerful marker-isolation technologies, such as amplified fragment length polymorphisms (AFLP), hundreds of loci can be screened in a single experiment. This provides a greater chance to identify markers in the close vicinity of the target gene. Hence, a highly refined genetic map with markers tightly linked to the target gene can be developed within a short time (Thomas et al. 1995). With the help of such a well-defined molecular marker system, it may be possible to directly land on the target gene, thus eliminating the time-consuming chromosomal walking step

involved in a positional cloning approach (Tanksley et al. 1995). In addition to the two currently available BAC libraries (Marek and Shoemaker 1997; Danesh et al. 1998), this relatively large soybean BAC library with a stable and large insert size DNA should be a useful resource for soybean genome analysis and gene cloning. This library is currently being employed in constructing a contig around the super-nodulation locus *nts-1* of soybean (Gresshoff et al. 1998). Soybean researchers interested in using this BAC library should contact M.K.B.

Acknowledgments We thank Richard Michelmore, UC Davis, USA, for providing the pECBAC1 vector. We also thank Richard Dixon, Robert Gonzales, Maria Harrison and Takao Kasuga for critically reading the manuscript, and Evan Burrows and Cuc Ly for preparing photographs. This research project was supported by The Samuel Roberts Noble Foundation.

References

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

Bhattacharyya MK, Ward EWB (1986) Expression of gene-specific and age-related resistance and the accumulation of glyceollin in soybean leaves infected with *Phytophthora megasperma* f. sp. *glycinea*. Physiol Mol Plant Pathol 29:105–113

Bhattacharyya MK, Gonzales RA, Kraft M, Buzzell RI (1997) A *copia*-like retrotransposon Tgm*r* closely linked to the *Rps1*-k allele that confers race – specific resistance of soybean to *Phytophthora sojae*. Plant Mol Biol 34:255–264

Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806–812

Cai L, Taylor JF, Wing RA, Gallagher DS, Woo SS, Davis SK (1995) Construction and characterization of a Bovine Artificial Chromosome library. Genomics 29:413–425

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Collins J, Hohn B (1978) Cosmids: A type of plasmid gene – cloning vector that is packageable in vitro in bacteriophage-l heads. Proc. Natl. Acad. Sci. USA 75:4242–4246

Danesh D, Penuela S, Mudge J, Denny R, Nordstrom H, Martinez JP, Young ND (1998) Construction and characterization of a soybean bacterial artificial chromosome library. Theor Appl Genet 96:196–202

Diers BW, Mansur L, Imsande J, Shoemaker RC (1992) Mapping *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci 32:377–383

Eyers M, Edwards K, Schuch W (1992) Construction and characterization of a yeast artificial chromosome library containing two haploid *Beta vulgaris* L. genome equivalents. Gene 121: 195–201

Fehr WR (1980) Soybean. In: Fehr WR, Hadley HH (eds) Hybridization of crop plants. American Society of Agronomy and Crop Science Society of America Publishers, Madison, pp 589–599

Feinberg AP, Vogelstein S (1984) A technique for radio-labeling DNA fragments to high specific activity. Anal Biochem 137:266–267

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 *Hind*III genomic fragments of lettuce. Theor Appl Genet 94:390–399

Ganal MW, Tanksley SD (1989) Analysis of tomato DNA by pulsed-field gel electrophoresis. Plant Mol Biol 7:17–27

Green ED, Olson MV (1990) Systemic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 87:1213–1217

Gresshoff PM, Stiller J, Men A, Radutoiu S, Pillai S, Landau-Ellis D, Chian RJ, Ghassemi F, Bhattacharyya M, Hussain A, Lohar D, Jiang Q (1998) Functional genomics of legumes: map-based cloning and gene trapping advances in soybeans and *Lotus japonicus*. Plant and Soil (in press)

Heilig JS, Lech K, Brent R (1995) Large-scale preparation of plasmid DNA. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) Current protocols in molecular biology. John Wiley and Sons, New York, pp. 1.7.1–1.7.15

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nature Genet 6:84–89

Kasuga T, Salimath SS, Shi J, Gijzen M, Buzzell RI, Bhattacharyya MK (1997) High-resolution genetic and physical mapping of molecular markers linked to the *Phytophthora* resistance gene *Rps1*-k in soybean. Mol Plant-Microbe Interact 10: 1035–1044

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

Liu YG, Whittier RF (1994) Rapid preparation of megabase plant DNA from nuclei in agarose plugs and microbeads. Nucleic Acids Res 22:2168–2169

Liu YG, Mitsukawa N, Vazquez-Tello A, Whittier RF (1995) Generation of a high-quality P1 library of *Arabidopsis* suitable for chromosome walking. Plant J 7:351–358

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

Martin GB, Ganal MW, Tanksley SD (1992) Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. Mol Gen Genet 233:25–32

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

Monaco AP, Larin Z (1994) YACs, BACs, PACs and MACs: artificial chromosomes as research tools. Trends Biotechol 12:280–286

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

Schmidt R, Putterill J, West J, Cnops G, Robson F, Coupland G, Dean C (1994) Analysis of clones carrying repeated DNA sequences in two YAC libraries of *Arabidopsis thaliana* DNA. Plant J 5:735–744

Schmitthenner AF, Hobe M, Bhat RG (1994) *Phytophthora sojae* races in Ohio over a 10-year interval. Plant Dis 78:269–276

Shi J, Gonzales RA, Bhattacharyya MK (1995 a) Characterization of a plasma membrane-associated phosphoinositide-specific phospholipase C from soybean. Plant J 8:381–390

Shi J, Dixon RA, Gonzales, RA, Kjellbom P, Bhattacharyya MK (1995 b) Identification of cDNA clones encoding valosin-containing protein and other plant plasma membrane-associated proteins by a general immunoscreening strategy. Proc Natl Acad Sci USA 92:4457–4461

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

Sternberg N (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc Natl Acad Sci USA 87:103–107

Tanksley SD, Ganal MW, Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 11:63–68

Thomas CM, Vos P, Zabeau M, Jones DA, Norcott KA, Chadwick BP, Jones JDG (1995) Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato *Cf-9* gene for resistance to *Cladosporium fulvum*. Plant J 8:785–794

Umehara Y, Inagaki A, Tanoue, Yasukochi Y, Nagamura Y, Saji S, Otsuki Y, Fujimura T, Kurata N, Minobe Y (1995) Construction and characterization of a rice YAC library for physical mapping. Mol Breed 1:79–85

Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC (1995) 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

Wang GL, Warren R, Innes R, Osborne B, Baker B, Ronald PC (1996) Construction of an *Arabidopsis* BAC library and isolation of clones hybridizing with disease resistance, gene-like sequences. Plant Mol Biol Rep 14:107–114

Wing RA, Rastogi VK, Zhang HB, Paterson AH, Tanksley SD (1993) An improved method of plant megabase DNA isolation in agarose microbeads suitable for physical mapping and YAC cloning. Plant J 4:893–898

Woo SS, Jiang J, Gill, BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res 22:4922–4931

Yang D, Parco A, Nandi S, Subudhi P, Zhu Y, Wang G, Huang N (1997) Construction of a bacterial artificial chromosome (BAC) library and identification of overlapping BAC clones with chromosome 4-specific RFLP markers in rice. Theor Appl Genet 95:1147–1154

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

Zhang HB, Choi S, Woo SS, Li Z, Wing RA (1996) Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 2:1–14