Theor Appl Genet (2005) 111: 630–639
DOI 10.1007/s00122-005-1992-x

ORIGINAL PAPER

B. Yüksel · A. H. Paterson

Construction and characterization of a peanut *Hin*dIII BAC library

Received: 15 December 2004 / Accepted: 7 March 2005 / Published online: 28 July 2005

Abstract Bacterial artificial chromosome (BAC) libraries have been an essential tool for physical analyses of genomes of many crops. We constructed and characterized the first large-insert DNA library for *Arachis hypogaea L.* The *Hin*dIII BAC library contains 182,784 clones; only 5,484 (3%) had no inserts; and the average insert size is 104.05 kb. Chloroplast DNA contamination was very low, only nine clones, and r-DNA content was 1,208, 0.66% of clones. The depth of coverage is estimated to be 6.5 genome-equivalents, allowing the isolation of virtually any single-copy locus. This rate of coverage was confirmed with the application of 20 overgos, which identified 305 positive clones from the library. The identification of multiple loci by most probes in polyploids complicates anchoring of physical and genetic maps. We explored the practicality of a hybridization-based approach for determination of map locations of BAC clones in peanut by analyzing 94 clones detected by seven different overgos. The banding patterns on Southern blots were good predictors of contig composition; that is, the clones that shared the same size bands and ascribed to the same overgos usually also located in the same contigs. This BAC library has great potential to advance future research about the peanut genome.

Introduction

Peanut (*Arachis hypogaea* L.) is one of the most economically important food and oil-seed crops with 37,057,652 Gg produced world-wide (Food and Agriculture Organization 2003), with China and India the leading producers. The section Arachis of genus *Arachis* comprises about 69 species, most of which are A- and B-genome diploids with the exception of the D-genome diploid *A. glandulifera* (Stalker 1991; Stalker et al. 1991). The cultivated peanut *A. hypogaea* L. and its wild relative *A. monticola* (2 $n=4$, $x=40$) are the only known AABB tetraploids (Fernandez and Krapovickas 1994).

Like other monophyletic amphidiploid species, the genetic variation within and between tetraploid peanut species is very limited, consequently hindering genome analysis by genetic mapping. A large-insert DNA clone library will be valuable for detailed analysis of the peanut genome by permitting the use of physical mapping approaches that are not dependent upon genetic polymorphism.

Bacterial artificial chromosome libraries (BAC) have emerged as the preferred method for constructing large-insert size clone libraries, supplanting yeast (YACs) (Burke et al. 1987) because the former is easier to manipulate (Anderson 1993), has lower levels of instabilities, and has less rearrangement in highly repetitive DNA (Dunford et al. 1993; Neil et al. 1990). An additional disadvantage of YACs is their abundance of chimeric clones (Larionov et al. 1994; Schmidt et al. 1996). The first F-factor-based vector in *Escherichia coli* for the propagation of large inserts has been described by O'Connor et al. (1989). The F-factor-based genes parA and parB ensure the single-copy maintenance of the plasmid, thereby preventing co-transformation and recombinant clones that are common in YACs (Larionov et al. 1994; Willets and Skurray 1987). Overall, bacteria grow much faster than yeast, and bacterial clones are much easier to work with. Although BAC clones with insert sizes up to 300 kb have been reported, the average size of BAC libraries is usually between 100 kb and 150 kb.

BAC libraries for many important crops from very diverse plant families already exist: *Sorghum bicolor* (sorghum; Woo et al. 1994), *Arabidopsis thaliana* (Choi

Requests for the BAC library (or subsets) should be directed to Dr. A. Paterson (paterson@uga.edu).

Communicated by F.J. Muehlbauer

B. Yüksel · A. H. Paterson (✉)
Plant Genome Mapping Lab, The University of Georgia,
111 Riverbend Rd, Athens, GA 30605, USA
E-mail: paterson@uga.edu
Tel.: +1-706-5830162
Fax: +1-706-5830160

et al. 1995), *Sorghum propinquum* (Lin et al. 1999), *Carica papaya* (papaya; Ming et al. 2001), *Helianthus annuus* (sunflower; Gentzbittel et al. 2002), *Lactuca sativa* (lettuce; Frijters et al. 1997), *Prunus persica* (peach; Wang et al. 2001a), *Triticum aestivum* (wheat; Cenci et al. 2003), *Musa acuminata* (banana; Vilarinhos et al. 2003), and *Brassica oleracea* var. *botrytis* (broccoli; Li et al. 2003). In the Leguminosae, multiple libraries have been made for *Glycine max* (soybean), *Medicago truncatula* (Danesh et al. 1998; Meksem et al. 2000; Nam et al. 1999; Salimath and Bhattacharyya 1999; Tomkins et al. 1999), and *Phaseolus vulgaris* (common bean; Vanhouten and MacKenzie 1999). However, no BAC libraries have been previously constructed for *Arachis* spp.

The main approach for the screening of arrayed BAC libraries has been the application of radioactively labeled probes to high-density BAC filters. While individual probes such as cDNA or genomic clones can be used, oligonucleotide-based probes, called ''overgo'', improve screening due to more efficient radiolabeling and low cost. Overgos are two 24-bp oligonucleotides with an 8-bp overlapping region at the 3′ end, thereby allowing the synthesis of complementary strands with radioactive nucleotides (Ross et al. 1999). In addition to the single probe-based approaches, the overgos also enable multiplexing designs, significantly improving the efficiency of library screening. For example, 10,642 overgos designed from expressed sequence tags (ESTs) were applied to 165,888 maize BACs in a 24×24 multiplex with a 88% success rate (Gardiner et al. 2004).

The primary objective of this study was the construction and characterization of an *A. hypogaea* BAC library with sufficient genome coverage for genomic analyses.

Materials and methods

Plant material

Arachis hypogaea L. cv. Florunner UF-439-16-1003-2 was grown in the greenhouse, and leaf material was collected at the seedling stage. The collected leaf tissue was washed in ddH_2O and frozen in liquid nitrogen in packages of 3–4 g, then stored at −80°C until needed.

Preparation of high-molecular weight (HMW) DNA

The protocol for isolating peanut nuclei is a modification of the DNA extraction method originally developed by Burow et al. (2001). The frozen plant material was ground in liquid nitrogen with a mortar and pestle until very fine powder was obtained; the powdered sample was then dissolved in extraction buffer [0.005 *M* citric acid, 0.5 *M* glucose, 0.01 *M* Na_2EDTA, 2.0% (w/v) polyvinylpyrrolidone-40 (PVP-40), 5% (v/v) Triton X-100, 0.25% (w/v) spermidine, titrated to pH 6.5 with NaOH] to which fresh antioxidants [0.1% (w/v) ascorbic acid, 0.2% (v/v) 2-mercaptoethanol, 0.1% (w/v) disodium diethylthiocarbamate (Na_2Et_2dtc), and 0.4% (w/v) $NaHSO_3$] were added. The samples were dissolved on ice for about 10 min and filtered twice through eight layers of cheese cloth and two layers of miracloth (instead of filtering once as DNA extraction). The filtrates were centrifuged for 20 min at 160 *g*, which is a much lower speed than normal DNA extraction (500 *g*), to prevent the precipitation of organelles but to precipitate the nuclei. The precipitated nuclei were dissolved in extraction buffer and centrifuged at the same speed three to four times until the nuclei looked clean, The final wash of the nuclei was in the same extraction buffer without Triton X-100. The washing steps were necessary to ensure the purity of the final product, which is required for a successful digestion of the embedded HMW DNA with a minimal amount of enzyme. The quality of the DNA used by Burow et al. (2001) was not required to be as good as that required for BAC library construction; thus, they were able to omit extra washing steps. Instead of being lysed directly in the liquid medium, the nuclei were pre-warmed at 42°C for 5 min before being mixed with an equal volume of 1% LMP agarose. The plugs were formed in order to provide a solid supporting matrix to keep DNA intact after lysing. The following steps were only followed for HMW DNA extraction: the plugs were incubated at 50°C for 24 h in lysis buffer [0.005 *M* citric acid, 0.14 *M* NaCl, 0.05 *M* Na_2EDTA, 2% (w/v) PVP-40, 1% (w/v) sodium dodecyl sulfate (SDS), 1% sodium lauryl sarcosine titrated to pH 6.5 with NaOH, and autoclaved] and the same antioxidants at similar proportions as in the extraction buffer, and 0.06 mg/ml Proteinase K (added to ensure the intactness of the final product) was added. The buffer was replaced, and plugs were incubated at the same temperature for another 24 h. The plugs were incubated at room temperature for at least 4 h in 70% ethanol then stored at −20°C until use.

We increased the pH of the extraction and lysis buffers from 5.0 to 6.5 following our observation that the low pH damaged the DNA and resulted in low ligation efficiency. HMW DNA larger than 1,000 kb is necessary.

Test digestion, size selection, and elution

The protocols developed by Ming et al. (2001) and Peterson et al. (2000) were followed. Briefly, the plugs were incubated in 0.05 *M* EDTA on ice until ethanol effused, following which they were incubated for 1 h at 50°C in 0.5 *M* EDTA, pH 9.3, for 1 h on ice in 0.05 *M* EDTA, for 30 min on ice in $T_{10}E_{10}$, and finally for 1 h in a 0.1 μ*M* PMSF, $T_{10}E_{10}$solution. The plugs were then washed with $T_{10}E_{10}$ for 2 h on ice. The washed plugs were minced and incubated in a 1-ml aliquot of buffer cocktail (1× NEB buffer 2, 1 m *M* BSA, 1 m *M* DTT, and 4 m *M* spermidine) for 30 min, following which the buffer was refreshed and incubated another 30 min on

ice. The buffer cocktail was removed a second time, and 250 µl fresh buffer cocktail was added to each sample. Serial dilutions of *Hin*dIII (NEB, Beverly, Mass.) (0, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1, and 2 U per milligram of plug) were added to the samples and the samples incubated at 4°C for 4 h. Partial digestion was carried out at 37°C for 7 min; 0.5 *M* EDTA was added to the tubes to stop the reactions. The partially digested samples were resolved on 1% agarose gels run in 0.5× TBE buffer by pulsed field gel electrophoresis (PFGE) at 6 V/cm, with 1- to 40-s switch times and a linear ramp, for 18 h. Optimum enzyme concentration was determined by visualizing maximum fragment concentration in the 150- to 350-kb range.

The plugs were treated as described for the test digestions, partially digested in pre-determined enzyme concentrations and the DNA electrophoresed under the same conditions. Markers on the gel were stained, and the gel retaining the 125- to 350-kb fraction was excised. This gel fraction was divided into three different equal sub-fractions, X, Y, and Z. A second round of size selection was applied to the X, Y, and Z fractions in order to augment the desired size DNA fragments and purge small fragments that had been stuck behind large fragments after the first size selection. Electrophoresis parameters were set to 3- to 5-s switch times, 6 V/cm, a linear ramp, and a 18-h run time. In the second size selection gel, the gel containing the size fraction larger than 125 kb was excised and the gel fragments saturated in 70% ethanol at room temperature for 4 h before being stored at −20°C until use. Excised gel pieces were washed in 1× TAE buffer on ice until ethanol diffused out of them. The DNA was eluted with an Electro-Eluter (Bio-Rad, Hercules, Calif.) by applying a constant 10 mA of current for each cuvette for 2 h. The concentration of the eluted DNA was checked and concentrated with 0.025-µm nitrocellulose membranes (Millipore, Billerica, Mass.) on 10% PEG if needed.

BAC vector

Three types of BAC vector were used in the construction of the library: pBeloBAC (prepared at our laboratory), pCUGIBAC1 (kindly provided by Dr. J. Tomkins, Clemson University Genomics Institute, S.C.), a pIndigo536 cloned into pGEM-4Z vector (Luo et al. 2001) and pIndigoBAC-5 (Epicentre Technologies, Madison, Wis.). The vector used in each ligation and number of colonies picked are denoted in Table 1.

The pBeloBAC plasmid was isolated by an alkaline-lysis method (Sambrook and Russell 2001) from 10 l of LB CM culture. Isolated plasmid was dissolved in TE overnight at 4°, and closed-circle plasmid was purified by two passages through a CsCl-ethidium bromide gradient (Sambrook and Russell 2001). The purified plasmid samples were dialyzed twice in TE for 24 h at 4°, and plasmid was precipitated with 95% ethanol, 1/10 (v/v) of 7.5 *M* $NH_4CH_3CO_2$, and washed once in 70% ethanol. Ten micrograms of the plasmid was digested with *Hin*dIII and dephosphorylated with HK phosphatase (Epicentre Technologies). The dephosphorylated vector was self-ligated and run on a 0.8%, 1× TAE gel, following which the open-circle plasmid band was excised and eluted. λHIII fragments were ligated into the vector to test the efficiency of dephosphorylation.

The pCUGIBAC1 vector was isolated in the same manner as pBeloBAC except that a much smaller amount of culture (1 l) was required because of the high-copy nature of the plasmid. The CsCl-ethidium bromide gradient was not applied, but the vector was digested, dephosphorylated, and tested in the same manner as described for the pBeloBAC plasmid.

Ligation, transformation, insert-size characterization, and picking

For ligation, a constant 15 ng of vector was used, and varying amounts of insert ranging from 60 to 120 ng were tested for each size selection. The vector/insert (V/I) ratio, which gave the best efficiency and average insert size, was chosen. Ligation reactions were performed in 60-µl volumes and incubated at 16° for 10 h. After desalting, 2–3 µl of reactions were transformed into *Escherichia coli* DH10B competent cells (DH10B ElectroMAX T1 phage-resistant and ElectroMAX DH10B-competent cells; Invitrogen, Carlsbad, Calif.) by electroporation (BioRad Gene Pulser II). For electroporation, 2.5 kV and 200 Ω of resistance was used. The electroporated cells were immediately mixed with 1 ml of SOC media (Sambrook and Russell 2001) and grown for 1 h at 37°C before separation on selective medium (LB medium) with 12.5 µg chloroamphenicol, 0.55 m *M* IPTG, and 80 µg/ml X-gal. After 18 h of incubation at 37°C, a sampling of 10–20 colonies were picked and tested.

Randomly selected white colonies were inoculated into 1 ml LB CM liquid medium and grown for 16 h at 37°C. The liquid cultures were miniprepped by an alkaline lysis protocol, and the DNA was digested with 10 U of *Not*I (NEB) for 4 h at 37°C. The digested samples were resolved on 1%, 0.5× TBE agarose by PFGE with the following parameters: 3- to 20-s linear ramp, 6 V/cm, and a 16-h run time. The ligation reactions with average insert size of 100 kb or more were mass-transformed, plated, and directly picked with a Q BOT (Genetix, Dorset, UK) into 384-well plates. The clones were stored in FM medium [LB + 36 m *M* K_2HPO_4, 13.2 m *M* KH_2PO_4, 1.7 m *M* sodium citrate, 0.4 m *M* $MgSO_4$, 6.8 m *M* $(NH_4)_2SO_4$, and 4.4% glycerol] at −80°C.

Gridding the peanut BAC library and preparing high-density filters for hybridization

A total of 182,784 *A. hypogaea* BAC clones were gridded on 22.5-cm^2 Hybond N+ membranes (Amersham Life Sciences, Arlington Heights, Ill.) with a Q BOT (Genet-

Table 1 Specificities of the peanut BAC library derived from randomly picked colonies

Ligation number	Vector	Total no. of colonies picked	Number of 0 (kb)[a] inserts/tested (%)	Average insert size[b]	Total Mb cloned
1	PBeloBAC	3,072	0/26 (0%)	95.20	292.45
2	PBeloBAC	4,224	0/8 (0%)	106.25	448.80
3	PBeloBAC	3,456	½0 (5%)	94.20	309.28
4	PBeloBAC	2,688	0/13 (0%)	90.23	242.54
5	PBeloBAC	8,832	0/13 (0%)	105.30	930.01
6	PCUGIBAC	2,688	3/21 (14%)	96.00	221.18
7	PIndigoBAC-5	18,048	1/42 (2%)	108.46	1,910.96
8	PIndigoBAC-5	14,592	0/9 (0%)	106.35	1,551.86
9	PCUGIBAC	8,832	0/10 (0%)	100.00	883.20
10	PIndigoBAC-5	14,208	0/14 (0%)	100.05	1,421.51
11	PIndigoBAC-5	4,992	0/10 (0%)	99.78	498.10
12[c]	PIndigoBAC-5	36,864	0/21 (0%)	111.90	4,125.08
13[c]	PIndigoBAC-5	16,128	0/8 (0%)	114.37	1,844.56
14[e]	PIndigoBAC-5	13,056	0/7 (0%)	90.33	1,179.35
15[c]	PIndigoBAC-5	17,664	1/7 (14%)	90.33	1,367.65
16[c]	PIndigoBAC-5	13,440	1/7 (14%)	106.10	1,222.30
Total		182,784	5,484.07[d]	104.05[e]	18,448.83

[a]False positives (i.e., white colonies without inserts)
[b]Colonies without an insert were included in the calculation of the average insert size
[c]T1 phage-resistant host cells were used for these ligations
[d]he total number of colonies without an insert $= \Sigma_{I=1-16}$ [(%) of 0 (kb) inserts×(number of colonies picked)], where I = ligation number
[e]Grand average = $\{\Sigma_{I=1-16}$ [(average insert size×(total no. of colonies picked)$_I$]/total number of colonies}

ix). Each clone was double-spotted in 4×4 arrays, thereby allowing representation of 18,432 different clones per filter. Thus, the whole peanut library fit onto ten filters. The filters were incubated on 1% LB agar containing 12.5 μg/μl CM at 37°C for 12–18 h until an optimal colony growth-size was obtained. The high-density BAC filters were processed according to a standard alkaline-lysis method (Sambrook and Russell 2001). The filters were dried overnight and stored at −4°C.

Fingerprinting, blotting, and labeling

BAC clones were inoculated into 1.5 ml 2× YT medium (Sambrook and Russell 2001) containing 12.5 μg/μl CM and grown at 37°C for 18 h. Minipreps were performed according to a standard alkaline-lysis method. The samples were then digested with 40 U of *Hin*dIII (NEB) at 37°C for 4.5 h and the digestion products resolved on 1% agarose gels in 1× TAE buffer for 16 h at 95 V. The gel was stained with SYBR Green and imaged. Following the analysis of the gel image with IMAGE software (Sulston et al. 1989), the contigs were built with FPC ver. 4.7 (Soderlund et al. 2000).

The fingerprinting gel was blotted onto Hybond N+ filters (Amersham-Pharmacia, Piscataway, N.J.) in 0.4 *N* NaOH solutions.

Overgo labeling reactions were set in a total volume of 15 μl containing 0.0067 n *M* forward and reverse oligonucleotide primers that were denatured at 94°C for 5 min and cooled on ice, 1 μg BSA, 2.5 U *Taq* polymerase, 1 μl α-[^{32}]-dATP (6,000 Ci/mmol) (MP Biomedicals, Irvine, Calif.), and 3 μl OLB (Oligo labeling buffer without dATP, dCTP, and random hexamers) (Ross et al. 1999). The reaction mixture was incubated at 37°C for 2 h (Fig. 1). The labeled overgos were filtered through Sephadex columns to remove any unincorporated radioactive nucleotides and then heat denatured at 94°C for 5 min before being added to the hybridization bottle.

A 100-ng aliquot of four cotton chloroplast sequences (300–400 bp) were labeled by the random hexamer protocol using 25 μCi of α-[^{32}]-dCTP(6,000 Ci/mmol) and the Klenow fragment for primer extension at 37°C for 2 h (Chittenden et al. 1994). The unincorporated radioactivity was removed by filtering through a Sephadex column.

The membranes separated by a nylon mesh were placed into the rotisserie bottles with 15 ml (Southern blot) and 85 ml (BAC filters) of hybridization buffer [0.5 *M* sodium phosphate, pH 7.2, 7% (w/v) SDS, 1 m *M* EDTA, and 0.01% (w/v) BSA] and incubated at 58°C for Southern and 55°C for BAC filters before being added to the denatured radioactive overgo probes. Hybridizations were performed at 58°C and 55°C (for BAC filters) with a rotation speed of 4.5 rpm for 18–36 h in a rotisserie oven. Following the hybridization process, the membranes were washed at 55°C for 30 min each treatment with constant shaking in a tray; first at low stringency in wash buffer II [1× SSPE (0.15 *M* NaCl, 10 m *M* $NaH_2PO_4 \cdot H_2O$, and 1 m *M* Na_2EDTA, pH 7.4), 1% (w/v) SDS)], then at high stringency in wash buffer III (same as wash buffer II except for 0.5× SSPE), followed by one last wash in wash buffer II in order to remove nonspecific bindings to reduce background without effecting the specific probe hybridization.

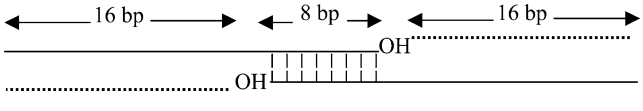

Fig. 1 Schematic representation of an overgo structure and labeling. *Dashed lines* indicate the [^{32}P] labeled portion of the probe

Afterwards, the membranes were blotted dry, wrapped in a sheet protector, and autoradiographed with two intensifying screens (10×12-inch L-Plus, Optonix, N.J.) and X-ray film (Blue, Medical X-ray film, SourceOne, Calif.) for 10 days at 80°C before developing.

For hybridization of the chloroplast clones, BAC membranes were prepared the same way as for the overgo and Southern hybridizations. Before the addition of the radioactively labeled probes, the membranes were pre-hybridized in 85 ml of hybridization buffer [0.5 *M* Na_2HPO_4, 1 m *M* EDTA, and 7% SDS] for 4 h at 62°C, and the hybridization was performed at the same temperature for another 20 h. Following hybridization, the membranes were washed twice at 62°C with washing solution [0.5× SSPE (0.15 *M* NaCl, 10 m *M* $NaH_2PO_4{\cdot}H_2O$, and 1 m *M* Na_2EDTA, pH 7.4), 0.5% SDS] for 15 min each. Subsequently, the membranes were autoradiographed as described above.

Results

Construction and characterization of the peanut BAC library

Size selections

The number of plugs used for the size selection of suitable fragments for ligations was adjusted according to the quality of the DNA and its concentration. Even though three different equal fractions from the first size selection, in the range of 125–350 kb, were subjected to the second size selection for further purging of the smaller fragments, the best results were always obtained from the middle fraction, which corresponds to the fragments between approximately 160 kb and 200 kb. The lowest fraction usually produced an insert size that was too small, while the top fraction yielded too much variation in insert size.

Ligation

For most of the ligation reactions, a 1/4 V/I ratio produced the best average insert size and the best ligation efficiency, but we adjusted this ratio for each size selection.

BAC library of *A. hypogaea*

We used 16 ligation reactions for constructing the peanut BAC library. Efficiency was variable from ligation to ligation (Table 1) due to the intactness and cleanliness of the HMW DNA used. The proportion of colonies without an insert was also significantly different between ligations, ranging from 0 to 14%. Overall, we estimated that 5,484 (3%) of the 182,784 total clones picked had no inserts. This percentage is typical of lacZ-gene based cloning systems (Li et al. 2003)

Three kinds of vector were used, pBeloBAC, pCUGIBAC1, and pIndigoBAC-5 (Table 1). The host cells of 97,152 clones were T1 phage-resistant. The best efficiency was obtained from pIndigoBAC-5 in terms of the blue/white ratio and number of colonies.

Assuming that the genome size of *A. hypogaea* is about 2,813 Mb (Arumuganathan and Earle 1991; Singh et al. 1996; Temsch and Greihulber 2000), the library provides a coverage of about 6.5 genome equivalents, resulting in a 99.88% probability of recovering any single-copy sequence from the genome.

DNA from 206 randomly sampled colonies was digested with *Not*I and resolved by PFGE; a sample of 28 of these clones is shown in Fig. 2. Only 8 out of 206 colonies contained a insert smaller than 20 kb or had no insert at all (Fig. 3). The insert sizes of about 50% of colonies were between 100 kb and 110 kb, and less than 2% of the colonies contained inserts of 160 kb or larger (Fig. 3).

Assessment of the depth of genome coverage by hybridization

A sample of 24 overgos (40-bp oligonucleotides) derived from genetically mapped peanut restriction fragment length polymorphism (RFLP) probe sequences (Burow et al. 2001) were applied to the peanut BAC library to assess the depth of the genome coverage. Overgos that are likely to be present as a single locus in the tetraploid peanut genome were chosen on the basis of number of *Hin*dIII bands (i.e., two or three bands for tetraploid genome) on survey blots (Burow et al. 2001). A total of 655 hits were identified by the 24 overgos (Table 2). The number of hits obtained from overgos POVS1131 (41 hits), POVS1161 (42 hits), POVS1129 (106 hits), and

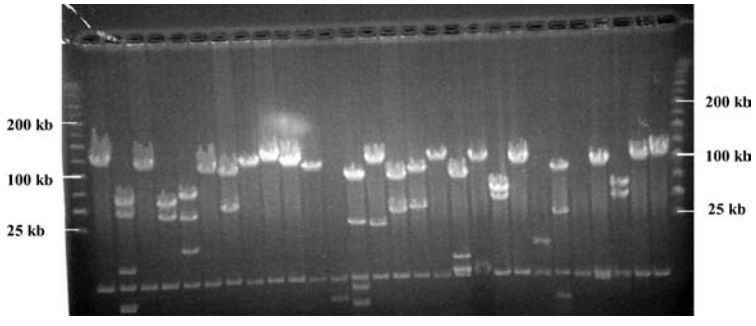

Fig. 2 Characterization of the peanut BAC library. DNA was extracted from randomly selected clones and digested with *Not*I. The *first and last lanes* are MidRange II PFG marker (NEB)

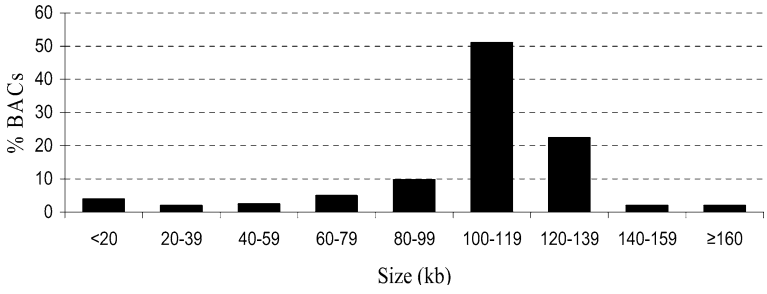

Fig. 3 Randomly selected BAC clones from each ligation were digested and the percentage of colonies in different size ranges calculated

POVS1265 (161 hits) was significantly different than the expectation at the presumed depths of genome coverage ($P = 0.001$), suggesting possible multiple locus recognition by these overgos. Thus, these overgos were not considered for further calculations. The average number of hits from the remaining 20 overgos was 15.25 (± 7.49), which suggests a genome coverage of the peanut genome of 7.6 equivalents. The range of number of hits observed, from 5 to 28, is consistent with a range of library coverage from 6 to 9.5 genome equivalence based on a Poisson distribution and a likelihood threshold of 95%.

Table 2 The results of the peanut BAC library screening with different types of probes

Overgo Id	GenBank Id[a]	Number of hits	Number of *Hin*dIII[b]
POVS1170	CX018126	2	3
POVS1279	CX018168	5	2
POVS1091	CX018167	7	3
POVS1098	CX018165	9	3
POVS1114	CX018163	9	2
POVS1130	CX018162	10	3
POVS1090	CX018160	11	2
POVS1186	CX018154	11	3
POVS1066	CX018153	13	2
POVS1159	CX018151	13	2
POVS1178	CX018150	14	2
POVS1105	CX018148	18	3
POVS1108	CX018146	18	3
POVS1071	CX018145	19	2
POVS1034	CX018144	21	3
POVS1027	CX018143	22	2
POVS1136	CX018142	23	2
POVS1002	CX018123	25	3
POVS1053	CX018138	27	2
POVS1067	CX018152	28	3
POVS1131	CX018136	41	3
POVS1161	CX018134	42	2
POVS1129	CX018129	106	3
POVS1265	CX018127	161	3
SOG6231[c]	–	9	–
SOG2761[c]	–	0	–
Garb6VRC12[d]	–	0	–
Garb11VRA06[d]	–	0	–
Garb19VRC09[d]	–	0	–
Garb29VRG08[d]	–	0	–
SOG6768[e]	–	866	–
SOG6770[e]	–	1,123	–

[a]GenBank accession number for the template sequences from which overgo is designed
[b]The bands were obtained from surveying blots, where *Hin*dIII-digested peanut genomic DNA was hybridized to the probes that were used for derivation of the overgos (Burow et al. 2001).
[c]These overgos are specific to the sorghum chloroplast
[d]These probes were derived from a *G. arboreum* chloroplast sequence
[e]These overgos are specific to sorghum rDNA (45S)

Four clones (300- to 400-bp long sequences) were included from different segments of the *G. arboreum* chloroplast (Garb6VRC12, Garb11VRA06, Garb19VRC09, and Garb29VRG08) having high homology to *Fabaceae* chloroplast sequences to ensure the completeness of the screening (Table 2). In addition, two overgos specific to the sorghum chloroplast (SOG6231 and SOG2761) were used to screen the peanut high-density BAC filters to elucidate the level of chloroplast DNA contamination in the library. Hybridization of neither cotton chloroplast-based probes nor the sorghum chloroplast overgo, SOG2761, yielded any hits, but one overgo, SOG6231, had nine hits. This could mean that the percentage of chloroplast clones is very low. To rule out possible problems during the labeling stage of the cotton probes, we counted the total incorporated α-[P^{32}]; this was on average 2.3×10^{-6} dpm which, based on past experience, is more than enough to produce a signal on BAC filters.

Two sorghum overgos with good homology to *Fabaceae* 45S rDNA sequences were employed to assess the copy number and organization of rDNA clusters in peanut. One of the overgos, SOG6768, identified 866 BAC clones and the other one, SOG6770, had 1,123 hits (Table 2). Since both of these targeted the same cluster, the hits overlapped, and the total number of unique hits was 1,208. SOG6768 detected fewer loci, possibly due to its homology to the ITS (internal transcribed sequence), which is comparatively less conserved than transcribed rDNA sequences (Nickrent and Doyle 1995). Nevertheless, about 0.66% of the BAC clones contain rDNA clusters, which means that at least 0.1% of the peanut genome is composed of rDNA clusters, assuming a 6.5 genome equivalence of the peanut BAC library.

Determination of the subgenomic specificity of individual BACs

To test the practicality of a hybridization-based approach for peanut physical mapping, we selected 94 BAC clones, which were identified by seven different probes. These BAC clones were first subjected to fingerprinting (Fig. 4b), and contig analyses were

performed with FPC V4.7. The results of the contig analysis are shown in Table 3. However, grouping of BACs into contigs does not anchor contigs to genetic maps. To resolve this, the fingerprinting gel was blotted and labeled with the same overgos. The expectation was that the BAC clones belonging to different map locations would produce different banding patterns when labeled with respective overgos, which can be directly compared with RFLP patterns from the genetic map. This would assist in integration of physical and genetic maps, making subgenomic chromosome walking relatively plausible. The results of the experiment are shown in Fig. 4a. For example, positive BAC clones for POVR2609 produced two distinctive banding patterns, suggesting different subgenomic localization, and the contig data from the fingerprinting gel confirmed the dual characteristics of the clones (Table 3). The contig data for overgo POVR2080 was almost 100% in agreement with the observed banding pattern; the only exceptional clone had no fingerprint data. Likewise, another overgo, POVR2100, produced very similar results. However, some of the clones in POVR2100 showed more than one band on the film (at least four), suggesting a possible partial digestion. For overgo POVR2029, only 50% of the banding patterns agreed with the contig information; however; it needs to be

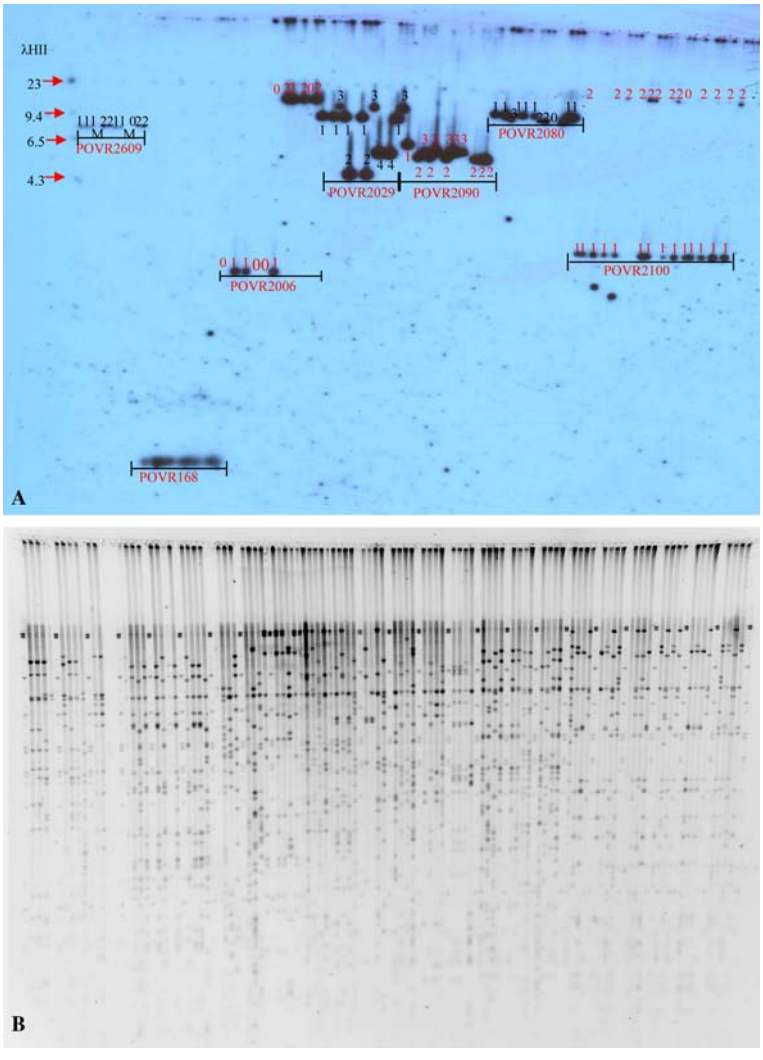

Fig. 4 **a** BAC-RF (Lin et al. 2000) on single BAC clones hybridizing to low-copy probes. Starting from the first lane *every fifth lane* is a marker (*M*). The *second lane* is λHIII, where the sizes of the fragments are indicated (in kilobases) The number of clones for the overgos, POVR2609, POVR168, POVR2006, POVR2029, POVR2090, POVR2080, and POVR2100 are 10, 12, 12, 12, 12, 12, and 24, respectively. The gel was hybridized with corresponding overgo primers. The polymorphic bands on the autoradiograph depict possible contigs belonging to different loci. The approximate positions of positive clones for each probe are indicated. The *numbers* (i.e., 1, 2, 3, and 4) denote possible polymorphic bands between contigs within the same probe, and "*0*" indicates the clones that were recognized as positives at screening the BAC library but failed to produce hybridization in the southern blot. The numbers for the clones specific to the consecutive overgos were depicted in alternating *black* and *red colors*. **b** *Hin*dIII-digested BACs from peanut. Starting from the first lane, *every fifth lane* is a marker. The *first ten samples* are for overgo POVR2609, and *next 60 samples* are, respectively, for POVR168, POVR2006, POVR2029, POVR2090, and POVR2080 overgos, where each probe has 12 positives. The *last 24 sample lanes* are positive clones for POVR210

Table 3 The listing of the contigs per overgo and their relationship with the polymorphic bands on the autoradiograph (Fig. 4a)

Probe	Fragment size[a]	Number of clones	Number of clones that form a contig	False positives[b]	Agreement (%)[c]
POVR2609	1	5	5	1	89
	2	4	3		
POVR168	–[d]	–	–	–	–
POVR2006	1	3	0	5	57
	2	4	4		
POVR2029	1	5	2	0	50
	2	2	2		
	3	3	2		
	4	2	0		
POVR2090	1	1	0	0	50
	2	6	4		
	3	5	2		
POVR2080	1	3	2	0	92
	2	7	7		
	3	2	2		
POVR2100	1	14	14	1	88
	2	9[e]	9		

[a]The polymorphic fragments within the probe are shown on Fig. 4a
[b]Normal clones that were recognized as positives in BAC library screening but failed to show hybridization in the Southern blot
[c]The overall percentage of the clones that were contiged within the overgo
[d]–, Not able to discern the different band sizes where three different contigs were formed
[e]Some fragments were not clearly discernible, perhaps due to partial digestion

noted that the resolution of the film may not be sufficient to discern very closely spaced bands. Some of the other disagreements might have also been caused by an insufficient resolution on the fingerprinting gel (Fig. 4a). Overall, the banding pattern on the Southern blot was usually concordant with the contig association of the clones (71% overall average). We believe that with a better resolution of samples, the technique could be further improved.

It can be concluded that this technique can be employed for large-scale applications. Perhaps the biggest advantage of this approach is that it allows the hybridization of multiple probes in a single hybridization experiment. The subgenomic identity of contigs could be determined with this method rather efficiently.

Discussion

We describe the construction and characterization of the first BAC library for any *Arachis* spp. This BAC library will considerably advance genomic studies in peanut. The amount of coverage, 6.5-fold, will enable the isolation of clones containing virtually any gene of interest; consequently, in the near future, map-based cloning may become a reality in peanut. Because of its low level of diversity at the DNA level, the identification and characterization of economically significant genes is urgent for the future of peanut cultivation. This BAC library will also be a very useful tool for comparative evolutionary analysis among different legume species.

Optimization of the enzyme concentration and digestion time was pivotal for the maximization of the numbers of DNA fragments that contain cohesive ends in the desired size range. The cleanliness of the embedded DNA in the plugs was one of the major factors determining the success of a partial digestion reaction. In terms of the efficiency of digestion, we usually obtained more promising results by increasing the amount of enzyme rather than increasing the duration of the digestion

Adjusting the DNA concentration during size fractionation was very critical. Too much DNA caused a wide variation in insert size, probably due to poor resolution during gel electrophoresis. Therefore, the higher the DNA concentration, the higher the number of small fragments 'stuck' in the desired size range. Even a second round of size selection is not enough to completely purge the small fragments, since the small fragments tend to ligate better. At high concentrations of DNA, the uniformity and average size of the inserts dropped dramatically. On the other hand, a low concentration of DNA usually makes it impossible to obtain a sufficient amount of DNA for ligation. Therefore, it is necessary to optimize the amount of DNA used in size selection. Another major problem we observed was that if the initial DNA was too degraded, size selections usually failed to produce DNA fragments of sufficient quality to ligate. We could not obtain any successful ligation from inserts isolated by the β-agarose method, perhaps because large DNA fragments may be damaged during the process (Osoegawa et al. 1998; Strong et al. 1997).

Although the DNA fragments used for ligations were selected to be in the range of 150–190 kb, the final average insert size was much lower, 104 kb. Similar results were observed in the construction of other BAC libraries (Danesh et al. 1998; Meksem et al. 2000). This result is consistent with the suggestions made by Frijters

et al. (1997); that is, the resolution of the agarose gels is not sufficient to remove all the 'trapped' smaller fragments, resulting in a much smaller average insert size than expected. A more effective way of purging the trapped fragments might significantly improve the average insert size.

The efficiency of the ligation reactions was highly variable depending on the quality of the source DNA and efficiency of the size selections. More than 10 h of ligation at 16°C usually produced more false positives (clones with no insert and rearranged vector). The proportion of false positives in BAC libraries has been a problem, with up to 17% being reported (Wang et al. 2001b). In our library, only 3% of the clones were false positives. The V/I ratio had to be adjusted for each size selection, suggesting that the quality (i.e., the ratio of fragments with damaged ends) was variable among the insert DNA samples used for library construction.

The depth of coverage was confirmed by screening of the library with 24 probes (Table 2). The average number of hits was 15.25, which is somewhat more than the expected number of hits per probe for an estimated 6.5 genome equivalence. The inflated number of hits could mean that some of these overgos might have targeted multiple loci. This is not unusual, especially for overgos derived from evolutionarily conserved domains. Nevertheless, the hybridization results approximately confirm the presumed coverage depth. Typically around 1% of most BAC libraries are chloroplast clones (Ming et al. 2001; Wu et al. 2004). The level of chloroplast DNA contamination in our library was very low, which may be a result of the technique utilized for the isolation of HMW DNA in which the nuclei were precipitated at a very low centrifugal force, 160 *g*, possibly preventing precipitation of any organelle.

rDNA sequences are very well conserved across distantly related plant taxa, even though the length and composition of the clusters vary greatly (Shi et al. 1996). By relying on this assumption, we used two sorghum 45S rDNA sequence-derived overgos for screening the peanut BAC library. The total number of BAC clones recognized by the two overgos was 1,208. Raina et al. (1999) has detected two major clusters that are associated with nucleolar organizer regions (NORs; regions of the chromosomes involved in nucleolus organization and rRNA synthesis) of the chromosomes, and six inactive condensed rDNA sites in the *A. hypogaea* genome. In addition to the number of ribosomal RNA cluster sites, the number of *Hin*dIII sites within the repeat would also affect the ratio of BAC clones carrying rDNA repeats. A segment of the 45S rDNA repeat unit in peanut has been cloned and sequenced (Bhagwat et al. 2001), and a single *Hin*dIII restriction site was found within the 12-kb rDNA unit; thus, the frequency of *Hin*dIII sites is less than the expected average of one per 4 kb for a six-cutter. In conclusion, the length of the clusters in peanut could be an explanation for the anomaly in the observed number of hits. Alternatively, one of these overgos, SOG 6770, could have detected other loci as well as the rDNA clusters. Nevertheless, the number of hits roughly verifies previous assessments about the depth of library coverage.

Like many agriculturally important plants, cultivated peanut is polyploid. A complication in the physical mapping of polyploids is the presence of ambiguities associated with anchoring contigs to genetic loci due to presence of multiple loci that hybridize to most probes. Several strategies have been devised to deal with this issue; however, approaches based on hybridization have been the most promising for high-throughput applications. We investigated possible high-throughput techniques for the assignment of subgenomic specificity of the probes with multiple hits in the peanut genome. Although very little genetic variation exists among the *A. hypogaea* cultivars, the A and B subgenomes usually differ with regard to most loci, thereby increasing the chance of success for subgenomic assignment on the basis of polymorphism. The technique tested in this paper was rather easy and would allow the testing of multiple contigs in a single experiment. The overall success rate, 71%, was acceptable, and this rate could be improved with better resolution. Hence, the integration of physical and genetic maps for polyploid species such as *A. hypogaea* becomes more manageable.

References

Anderson C (1993) Genome shortcut leads to problems. Science 259:1684–1687

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

Bhagwat AS, Krishna TG, Jawali N, Mitra RM (2001) Cloning and characterization of a ribosomal RNA gene repeat unit from groundnut. Plant Cell Rep 20:193–197

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

Burow MD, Simpson CE, Starr JL, Paterson AH (2001) Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. Genetics 159:823–837

Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. durum). Theor Appl Genet 107:931–939

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) RFLP mapping of a cross between *Sorghum bicolor* and *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of Sorghum chromosomes. Theor Appl Genet 87:925–933

Choi S, Creelman RA, Mullet JE, Wing RA (1995) Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis-thaliana*. Plant Mol Biol Rep 13:124–128

Danesh D, Penuela S, Mudge J, Denny RL, Nordstrom 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

Dunford R, Vilageliu L, Moore G (1993) Stabilization of a yeast artificial chromosome containing plant DNA using recombination-deficient host. Plant Mol Biol 21:1187–1189

Fernandez A, Krapovickas A (1994) Cromasomas y evolucion en *Arachis* (Leguminosae). Bonplandia 8:187–220

Food and Agriculture Organization (2003) FAOSTAT Agriculture. Available at: http://apps.fao.org/page/collections?subset = agriculture; verified 10 February 2004

Frijters ACJ, Zhang Z, van Damme M, Wang GL, 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

Gardiner J, Schroeder S, Polacco ML, Sanchez-Villeda H, Fang Z, Morgante M, Landewe T, Fengler K, Useche F, Hanafey M, Tingey S, Chou H, Wing R, Soderlund C, Coe EH Jr (2004) Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol 134:1317–1326

Gentzbittel L, Abbott A, Galaud JP, Georgi L, Fabre F, Liboz T, Alibert G (2002) A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. Mol Gen Genet 266:979–987

Larionov V, Kouprina N, Nikolaishvili N, Resnick M (1994) Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). Nucleic Acids Res 22:4154–4162

Li L, Lu S, O'Halloran DM, Garvin DF, Vrebalov J (2003) High-resolution genetics and physical mapping of the cauliflower high-beta-carotene gene *Or Orange*). Mol Gen Genet 270:132–138

Lin Y-R, Zhu L, Ren S, Yang J, Schertz KF, Paterson AH (1999) A Sorghum propinquum BAC library, suitable for cloning genes associated with loss-of-function mutations during crop domestication. Mol Breed 5:511–520

Lin Y, Draye X, Qian X, Ren S, Zhu L, Tomkins J, Wing R, Li Z, Paterson A (2000) Locus-specific contig assembly in highly-duplicated genomes using the BAC-RF method. Nucleic Acids Res 28:e23

Luo MZ, Wang YH, Frisch D, Joobeur T, Wing RA, Dean RA (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon Fusarium wilt (*Fom-2*). Genome 44:154–162

Meksem K, Zobrist K, Ruben E, Hyten D, Quanzhou T, Zhang H-B, Lightfoot D (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

Ming R, Moore PH, Zee F, Abbey CA, Ma H, Paterson AH (2001) Construction and characterization of a papaya BAC library as a foundation for molecular for molecular dissection of a tree-fruit genome. Theor Appl Genet 102:892–899

Nam YW, Penmetsa RV, Endre G, Uribe P, Kim D, Cook DR (1999) Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. Theor Appl Genet 98:638–646

Neil DL, Villasante A, Fisher RB, Vetrie C, Cox B, Tyler-Smith C (1990) Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nucleic Acids Res 18:1421–1428

Nickrent DL, Doyle JJ (1995) A molecular phylogeny of diploid *Glycine* (Fabaceae) based upon nuclear ribosomal ITS sequences. Am J Bot 82:153

O'Connor M, Peifer M, Bender W (1989) Construction of large DNA segmets in *Escherichia coli*. Science 244:1307–1312

Osoegawa K, Woon PY, Zhao B, Eirik F, Tateno M, Catanese JJ, de Jong PJ (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52:37994

Peterson DG, Tomkins JP, Frisch DA, Wing RA, Paterson AH (2000) Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. J Agric Genomics 5 (http://www.ncgr.org/research/jag/index.html)

Raina SN, Mukai Y (1999) Detection of a variable number of 18S-5.8S-26S and 5S ribosomal DNA loci by fluorescent in situ hybridization in diploid and tetraploid Arachis species. Genome 42:52–59

Ross MT, LaBrie T, McPherson J, Stanton VM (1999) Screening large-insert libraries by hybridization. Wiley, New York, N.Y.

Salimath SS, Bhattacharyya MK (1999) Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the *Rps-1* disease resistance gene. Theor Appl Genet 98:712–720

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Cold Spring Harbor, New York

Schmidt R, Putterill J, West J, Cnops G, Robson F, Coupland G, Dean C (1996) Detailed description of four YAC contigs representing 17 Mb of chromosome 4 of *Arabidopsis thaliana* ecotype Columbia. Plant J 9:755–760

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

Singh KP, Raina SN, Singh AK (1996) Variation in chromosomal DNA associated with the evolution of *Arachis* species. Genome 39:890–897

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

Stalker H (1991) A New Species in section *Arachis* of peanuts with a D genome. Am J Bot 78:630–637

Stalker H, Dhesi J, Parry D, Hahn J (1991) Cytological and interfertility relationships of *Arachis* Section Arachis. Am J Bot 78:238–246

Strong SJ, Ohta Y, Litman GW, Amemiya C (1997) Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. Nucleic Acids Res 25:3959–3961

Sulston J, Mallett F, Durvin R, Horsnell T (1989) Image analysis of restriction enzyme fingerprint audioradiograms. Comput Appl Biosci 5:101–106

Temsch EM, Greihulber J (2000) Genome size variation in *Arachis hypogaea* and *A-monticola* re-evaluated. Genome 43:449–451

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

Vanhouten W, MacKenzie S (1999) Construction and characterization of a common bean bacterial artificial chromosome library. Plant Mol Biol 40:977–983

Vilarinhos AD, Priffanelli P, Lagoda P, Thibivilliers S, Sabau X, Carreel F, D'Hont AD (2003) Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Colla). Theor Appl Genet 106:1102–1106

Wang Q, Zhang K, Qu X, Jia J, Shi J, Jin D (2001a) Construction and characterization of a bacterial artificial chromosome library of peach. Theor Appl Genet 103:1174–1179

Wang W, Zhai W, Luo M, Jiang G, Chen X, Li X, Wing RA, Zhu L (2001b) Chromosome landing at the bacterial blight resistance gene *Xa4* locus using a deep coverage rice BAC library. Mol Gen Genet 265:118–125

Willets N, Skurray R (1987) Structure and function of F-factor and mechanism of conjugation. In: *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol 2. American Society of Microbiology, Washington D.C., pp 1110–1137

Woo S-S, 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

Wu C-C, Nimmakayala P, Santos FA, Springman R, Scheuring C, Meksem K, Lightfoot DA, Zhang HB (2004) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet 109:1041–1050