Theor Appl Genet (2003) 106:269–276
DOI 10.1007/s00122-002-1074-2

Y.-L. Chang · X. Henriquez · D. Preuss
G. P. Copenhaver · H.-B. Zhang

A plant-transformation-competent BIBAC library from the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomics

Received: 1 October 2001 / Accepted: 26 April 2002 / Published online: 30 August 2002

Abstract The genome of the model plant *Arabidopsis thaliana* has been sequenced to near completion. To facilitate experimental determination of the function of every gene in the species, we constructed a large-insert library from the Landsberg ecotype using a plant-transformation-competent binary BAC vector, BIBAC2. The library contains 11,520 clones with an estimated average insert size of 162 kb. Of a sample of 102 clones, 17.6% had no inserts; further, in the library as a whole, 287 clones contained chloroplast DNA, and 25 contained mitochondrial DNA. Thus it is estimated that 9,295 clones originated from the nuclear genome, representing a 11.5 × coverage. The library was further characterized by screening with probes corresponding to 180-bp repeats, 5S rDNA, 18S–25S rDNA and 23 single-copy RFLP markers. The results showed that 92 clones contained 180-bp centromeric repeats, 78 contained 5S rDNA and 95 contained 18S–25S rDNA, approximately 1%, 0.8% and 1%, respectively, of the nuclear clones in the library. Screening the library with the 23 RFLP markers showed that each one hybridized to an average of seven clones. This library is the first large-insert DNA library for the widely studied Landsberg *erecta* strain. It will greatly facilitate gene identification by complementation screening, and will enhance analysis of the structure, organization and evolution of the *A. thaliana* genome.

Keywords BIBAC library · Plant transformation · Genomics · Positional cloning · *Arabidopsis thaliana* · Landsberg ecotype

Communicated by L. Willmitzer

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

D. Preuss · G.P. Copenhaver
Howard Hughes Medical Institute and Department of
Molecular Genetics and Cell Biology, University of Chicago,
Chicago, Illinois 60637, USA

Present address:
G.P. Copenhaver, The Biology Department,
The University of North Carolina at Chapel Hill, CB 3280,
Coker Hall 305, Chapel Hill, NC 27599, USA

Introduction

Previously, several genomic libraries have been constructed for *Arabidopsis thaliana*, nearly all derived from the ecotype Columbia, the variety chosen for complete genome sequencing (The Arabidopsis Genome Initiative 2000). These libraries include yeast artificial chromosome (Ward and Jen 1990; Grill and Somerville 1991; Ecker 1992; YAC, Creusot et al. 1995), bacterial artificial chromosome (BAC, Choi et al. 1995; Mozo et al. 1998), bacteriophage P1 (Liu et al. 1995) and plant transformation-competent artificial chromosome (TAC, Liu et al. 1999; unpublished) libraries. Several of these libraries were used to develop physical maps corresponding to individual chromosomes or to the entire genome (Hauge et al. 1991; Schmidt et al. 1995; Zachgo et al. 1996; Canilleri et al. 1998; Marra et al. 1999; Mozo et al. 1999; Kotani et al. 1999; Chang et al. 2001). These libraries have been proven useful for positional cloning of a number of genes (e.g., Arondel et al. 1992; Giraudat et al. 1992; Bent et al. 1994; Helliwell et al. 1998). In addition to the Columbia ecotype, the Landsberg *erecta* strain has been used by many laboratories for genetic and molecular studies (Redei 1962). This variety differs from Columbia by 1 to 0.5% at the DNA sequence level (Chang et al. 1988), and polymorphisms between the Landsberg and Columbia genes have been identified by several laboratories (Konieczny and Ausubel 1993; Bell and Ecker 1994; Cho et al. 1999). However, no large-insert BAC or BIBAC library was previously developed for Landsberg.

Following the release of the genome sequence of *A. thaliana* the experimental determination of the func-

tion(s) of every gene has been targeted as a goal for the coming decade (Somerville and Dangl 2000). To this end, several methods have been developed, including T-DNA-(Azpiroz-Leehan and Feldmann 1997; Feldmann 1991; Sussman et al. 2000) or transposon-based (Martienssen et al. 1998) gene tagging, DNA microarray or gene-chip analysis (Schena et al. 1995; Desprez et al. 1998; Ruan et al. 1998) and genetic transformation (Chang et al. 1994; Liu et al. 1999). Transformation of *A. thaliana* via *Agrobacterium* is efficient and can be accomplished without tissue-culture procedures (Feldmann and Marks 1987; Koncz et al. 1989; Bechtold et al. 1993). These methods allow routine complementation studies to be performed with mutant lines identified in natural or mutagenized populations (Redei 1970; Koornneef et al. 1983; Feldmann 1991). Although many complementation studies are performed with relatively small-insert clones, recent investigations demonstrate that very large DNA fragments (from 80 to 150 kb) can be successfully transferred into the genomes of several plants species, including tomato, tobacco and *A. thaliana*, via *Agrobacterium*-mediated transformation (Hamilton et al. 1996, 1999; Liu et al. 1999). As described below, we developed a large-insert, direct plant-transformation-competent BIBAC library from Landsberg *erecta*. This library will facilitate analysis and positional cloning of mutations identified in the Landsberg ecotype, as well as a thorough examination of sequence variations between Landsberg and Columbia (C. Wu et al., submitted). These studies will provide important insight into the structure, organization and evolution of the *Arabidopsis* genome.

Materials and methods

Preparation of high-molecular-weight (HMW) DNA from nuclei

A. thaliana HMW nuclear DNA was prepared according to Zhang et al. (1995). Twenty grams of tissue were collected from the aerial portion of 14 day old Landsberg plants, frozen in liquid nitrogen, and stored at –80 °C. The tissue was ground in liquid nitrogen, nuclei were isolated, and HMW nuclear DNA was prepared and purified in 100-µl of low-melting-point (LMP) agarose plugs (Zhang et al. 1995).

Preparation of BIBAC2 vector

The BIBAC2 vector (Hamilton 1997) was purified by the alkaline-lysis method, followed by two rounds of cesium chloride/ethidium bromide ultra-centrifugation as described by Sambrook et al. (1989). The vector DNA was completely digested and dephosphorylated as described by Zhang et al. (1996).

Library construction

The HMW DNA plugs were pre-run on a 1.0% agarose gel in 0.5 × TBE (Sambrook et al. 1989) by pulsed-field gel electorephoresis (PFGE) to remove the small fragments and some substances that may inhibit restriction enzymes for partial digestion in the plugs. PFGE was conducted with a 5-s pulse time at 4 V/cm, 14 °C for 8 h, under which the DNA fragments smaller than 300 kb that were not suitable for large-insert (>150 kb) cloning were run into the gel, whereas the fragments larger than 300 kb remained in the plugs. The plugs were collected from the gel, dialyzed in 50 ml of TE (10 mM of Tris–HCl, 1 mM of EDTA, pH 8.0) on ice for 4 h, with one change of TE per hour, and stored at 4 °C.

Eight 100-µl plugs containing HMW DNA were sliced into small pieces of equal size at nine pieces per plug. The plug slices were equilibrated in 24 ml of 1 × reaction buffer 3 (Life Technologies, USA) containing 2 mM of spermidine and 1 mM of dithiothreitol (DTT) on ice for 1 h, with one change after 30 min. The plug slices were then transferred into 24 1.5-ml tubes, three slices per tube, each tube containing 170 µl of fresh 1 × reaction buffer containing 0.5 mg/ml of BSA, 2 mM of spermidine and 1 mM of DTT. The 24 tubes were divided into three groups on ice, eight tubes per group, and to each group, 0.4, 0.6 and 0.8 units of *Bam*HI were added per tube, respectively. The tubes were incubated on ice for 60 min to allow the enzyme access to the DNA in the plug slices. The digestion was conducted at 37 °C for 8 min and stopped by adding 1/10 vol of 0.5 M EDTA, pH 8.0. DNA fragments ranging from 100 to 400 kb were selected by PFGE using the CHEF DRIII apparatus (BioRad, USA). The size-selection was first conducted on a 1% agarose gel in 0.5 × TBE buffer (Sambrook et al. 1989) with a 90-s pulse time at 6 V/cm, 14 °C for 20 h. The agarose-gel zones containing DNA fragments from 100 to 200 kb, 200 to 300 kb, and 300 to 400 kb, respectively, were excised. The DNA was eluted from the gel slices into dialysis tubing (molecular-weight exclusion limit = 12,000–14,000 Daltons) (Life Technologies, USA) by PFGE at 6 V/cm, 40-s pulse time and 14 °C for 4 h, followed by reversing the polarity of the currency for 60 s. The DNA was collected carefully with wide-pore tips and then size-selected again on a 1% agarose gel, 0.5 × TBE, using 4 V/cm, and a 5-s pulse time for 12 h at 14 °C. Gel regions containing DNA of each size range were excised, and the DNA was electroeluted as above. The eluted DNA was dialyzed in the same tubing in 0.5 × TE buffer (5 mM of Tris. HCl, pH 8.0, 0.5 mM of EDTA, pH 8.0) at 4 °C for 4 h, with one buffer change per hour. The DNA was collected and ligated to the vector DNA at a molar ratio of 1:4 at 16 °C for 18 h using T4 DNA ligase (Life Technologies, USA). The recombinant DNA was transformed into *Escherichia coli* DH10B competent cells using the Cell Porator and Voltage Booster electroporation system (Life Technologies, USA) using 375 V, 330 uF capacitance, low ohms impedance and a fast charge rate, and 4 kΩ resistance. The electroporated cells were incubated in 1 ml of SOC medium (Sambrook et al. 1989) at 37 °C for 1 h, plated onto LB agar medium containing 50 mg/l of kanamycin and 5% (w/v) sucrose, and then incubated at 37 °C for 24 h. After analyzing the insert sizes of clones from each ligation, the ligation from the gel zone containing DNA fragments from 200 to 300 kb were selected for library assembly. Clones were arrayed in 384-well microtiter plates containing a mixture of LB medium, 1 × freezing buffer [36 mM of K_2HPO_4, 13.2 mM of KH_2PO4, 1.7 mM of Na citrate, 0.4 mM of $MgSO_4$, 6.8 mM of $(NH_4)_2\ SO_4$, 4.4% (v/v) glycerol], and 50 mg/l of kanamycin. The clones were incubated at 37 °C for 20 h and then stored at –80 °C.

Analysis of BIBAC clones

BIBAC clones were randomly selected from the library and grown in LB medium containing 50 mg/l of kanamycin at 37 °C overnight. DNA was isolated according to Zhang et al. (1996). The insert DNA was released from the vector by digestion with *Not*I (New England Biolabs, USA) and subjected to PFGE on 1% agarose gels in 0.5 × TBE at a 5-s initial pulse time, a 15-s final pulse time, at 14 °C, and 6 V/cm for 18 h. The insert size of each clone was estimated using lambda-ladder PFG markers (New England Biolabs, USA).

Library characterization

The library was screened with chloroplast DNA, mitochondrial DNA, 18S–25S rDNA, 5S rDNA, 180-bp centromeric repeat and

23 RFLP marker probes, respectively. The three chloroplast DNA probe clones, *ndhA*, *rbcL* and *psbA*, were derived from the barley chloroplast genome (Baumgarther et al. 1993). The mitochondrial DNA probe was generated from four mt DNA-derived IGF BAC clones (F1O22, F2A13, F2J23 and F4I8), and the 18S–25S rDNA probe from one 18S–25S rDNA-containing IGF BAC clone (F1A18) of *A. thaliana* (Mozo et al. 1998). The 5S rDNA probe was isolated from wheat DNA (Gerlach and Dyer 1980). The 180-bp centromeric repeats were cloned from *Arabidopsis* genomic DNA BACs in our laboratories (unpublished). The RFLP markers, mi32, mi51, mi54, mi61, mi79a, mi112, mi116, mi123, mi128, mi148, mi167, mi204, mi238, mi277, mi279, mi301, mi306, mi310, mi320, mi330, mi398, mi422 and mi465, were selected from the Arabidopsis genetic map (Liu et al. 1996) and obtained from the Arabidopsis Biological Resource Center (Ohio State University, USA). The chloroplast and mitochondrial DNA, and the rDNA probes were labeled with ^{32}P-dCTP by random hexamer priming (Feinberg and Vogelstein 1983). The 180-bp repeats and RFLP markers were labeled using the Dig Probe Synthesis Kit (Roche Molecular Biochemicals, USA) by PCR (25 cycles of 40 s at 95 °C, 70 s at 43 °C, 3 min at 72 °C, followed by a 7-min extended incubation at 72 °C).

The library was double-grided onto Hybond N+ membrane in a 3 × 3 format using the Biomek 2000 Automated Workstation (Beckman, USA). The filters were processed as described by Zhang et al. (1996). The filters were hybridized with the 180-bp repeat and RFLP probes, and the hybridization signals were detected as described by the manufacturer (Roche Molecular Biochemicals, USA). The chloroplast DNA, mitochondrial DNA, 5S rDNA and 18S–25S rDNA probes were hybridized to filters at 65 °C overnight. After hybridization, filters were washed once in 2 × SSC, 0.1% (w/v) SDS at 65 °C for 20 min, once in 1 × SSC, 0.1% (w/v) SDS at 65 °C for 20 min, and once in 0.5 × SSC, 0.1% (w/v) SDS at 65 °C for 20 min. The filters were exposed to Kodak X-OMAT films to visualize the hybridization results.

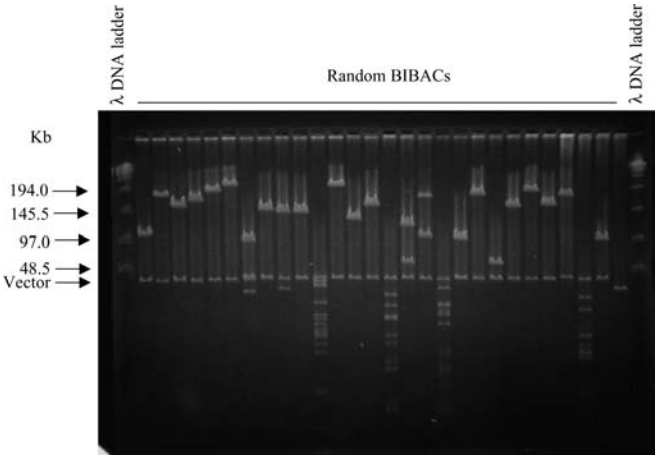

Fig. 1 Analysis of insert size in BIBAC clones randomly selected from the Landsberg library. BIBAC DNA digested with *Not*I and separated on a pulse-field gel was used to estimate the insert size in each clone

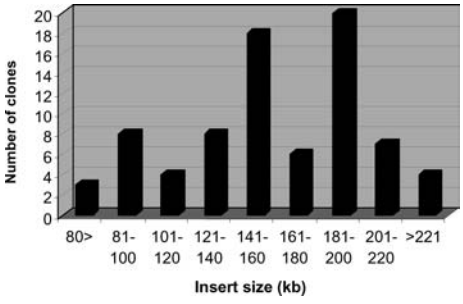

Fig. 2 Insert size distribution of 78 clones from the Landsberg BIBAC library

Results and discussion

Construction of a BIBAC library

We constructed a BIBAC library for the *A. thaliana* Landsberg ecotype in the *Bam*HI site of BIBAC2, an *Agrobacterium*-mediated, plant-transformation-competent binary BAC vector (Hamilton 1997). BIBAC2 was designed to transform high-molecular-weight DNA into plants to facilitate map-based gene cloning, functional analysis of genomic sequence, and genetic engineering (Hamilton et al. 1996, 1999; Hamilton 1997). To construct the BIBAC library, we isolated nuclei from plant tissues and embedded them into 100-µl plugs of LMP agarose (Zhang et al. 1995). We pre-electrophoresed the agarose gel-embedded DNA on a pulsed-field gel to eliminate small DNA fragments, resulting in high-molecular-weight, readily digestible DNA that facilitated subsequent BIBAC cloning. Because this treatment significantly decreased the abundance of small fragments in the plugs, it increased the efficiency of cloning large inserts. Furthermore, after partially digesting the DNA with *Bam*HI, we size-selected the restricted fragments twice on pulsed-field gels, further increasing the efficiency of producing clones with large inserts. The large DNA fragments were ligated to the completely restricted and dephosphorylated BIBAC2 vector. BIBAC2 contains the *sac*B gene, which encodes levansucrase, allowing selection for recombinants. In the presence of sucrose, the levansucrase catalyzes hydrolysis of sucrose into levan, a compound that is lethal to *E. coli* and *Arabidopsis tumefaciens* (Pierce et al. 1992). Therefore, only clones containing DNA inserts that interrupt the *sac*B gene are expected to form colonies on medium containing sucrose. All clones that grew on the selective medium were arrayed in 384-well plates. For the Landsberg *erecta* library, we arrayed a total of 11,520 clones in 30 384-well microplates, duplicated and stored at –80 °C.

Analysis of BIBAC clones

To estimate the insert sizes of the clones in the Landsberg *erecta* BIBAC library, DNA was isolated from 102 randomly selected clones, digested with *Not*I to release the inserts from the vector, and subjected to PFGE. The average insert size of randomly sampled clones was 162 kb (range 50 to 300 kb), with the majority of clones having insert sizes between 140 to 200 kb (see Figs. 1 and 2). Thus, on average, the insert sizes of the clones in this library are much larger than those in the previously described tomato libraries constructed in the BIBAC2

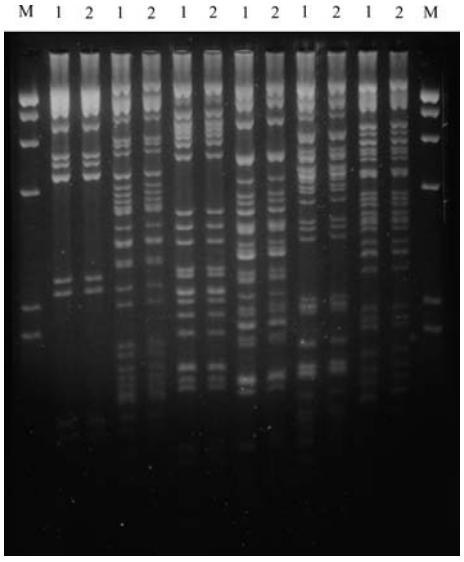

Fig. 3 Stability of Landsberg BIBAC clones. DNA isolated from 1-day (1) and 7-day (2) cultures of each clone was digested with *Hin*dIII and fractionated on an agarose gel. "M" indicates λ/*Hin*dIII DNA markers

vector (120 kb, Hamilton et al. 1999). Of the 102 clones analyzed, 18 (17.6%) had no inserts, suggesting that 83.4% (approximately 9,607) of the 11,520 clones of the library have DNA inserts. The number of clones without an insert is higher than that typically found (<5%) in plant BAC or BIBAC libraries constructed in the vectors that rely on a *lac*Z blue/white selection for insert-containing clones. However, it is consistent with the two tomato BIBAC libraries constructed in the same vector, where approximately 13% of the clones lack an insert (Hamilton et al. 1999). Surprisingly, we found that six of the 102 clones produced many *Not*I DNA fragments, each smaller than 50 kb (Fig. 1). Some of the bands from these clones hybridized to total genomic DNA and thus may have originated from the G/C-rich portions of the Landsberg genome. Similar results were also observed in a new BIBAC library of the Columbia ecotype, for which the tissues were grown and collected under sterile conditions (unpublished). Alternatively, these clones may be derived from abundant, contaminating microorganisms. Of the remaining 78 clones (Fig. 2), 66 (85%) yielded one fragment of insert DNA, 11(14%) yielded two fragments and one yielded three fragments (1%) when cut with *Not*I (Fig. 1) flanking the cloning site of BIBAC2.

Stability of BIBAC clones

The stability of large-insert clones in the *E. coli* host is critical for long-term maintenance and use of a library for genome research. We examined the stability of the Landsberg BIBAC2 inserts, by growing 24 randomly selected clones for 7 consecutive days. DNA was isolated from each clone on the 1st and 7th days of growth, digested to completion with *Hin*dIII, and fractionated on agarose gels. The restriction patterns of each clone did not change between the 1st and 7th days of culture (Fig. 3), indicating that the large-insert BIBAC2 clones are highly stable in *E. coli* host cells.

Genome origin of the large-insert BIBAC2 clones

In order to assess the extent to which a large-insert library represents the entire genome, it is important to ascertain the fraction of clones that are derived from organellar DNA. Libraries with few or no clones originating from organelle genomes are typically preferred. Screening the Landsberg BIBAC library with three barley chloroplast DNA clones (*ndh*A, *rbc*L and *psb*A) showed that 287 clones originated from the chloroplast genome, roughly 2.99% of the insert-containing clones in the library (Table 1). The percentage of clones with a chloroplast origin is similar to that found in BAC libraries derived from nuclear DNA (e.g., Choi et al. 1995; Moullet et al. 1999), but up to 10-fold less than that found in the BAC or YAC libraries derived from protoplast DNA (e.g., Grill and Somerville 1991; Martin et al. 1992; Creusot et al. 1995). This supports the previous conclusions that nuclear DNA preparations efficiently reduce the contribution of chloroplast DNA to a large-insert library (Zhang et al. 1995). Screening the library with mitochondrial DNA-derived BAC probes identified 25 BIBAC clones (0.26% of the library, Table 1). These clones most likely originated from the mitochondrial genome, or from large insertions of mitochondrial DNA into the nuclear genome, such as those characterized in the Columbia ecotype (Lin et al. 1999; Stuper et al. 2001). This percentage is about the same as that found in the Arabidopsis IGF BAC library (0.3%, Mozo et al. 1998). Taken together, these results predict that approximately 3.25% of the estimated 9,607 insert-containing clones in this library correspond to organelle DNA. Thus, the Landsberg BIBAC library contains approximately 9,295 clones derived from the nuclear genome, representing 11.5 × haploid genomes of *A. thaliana*. With a library of this size, the probability of obtaining a particular nuclear DNA clone is greater than 99%.

Identification of the clones containing nuclear repetitive elements

In the *Arabidopsis* genome, several repetitive sequences have been characterized, including 180-bp repeats (Martinez-Zapater et al. 1986), 5S rDNA (Campell et al. 1992) and 18S–25S rDNA (Pruitt and Meyerowitz 1986). We screened the Landsberg BIBAC library with probes corresponding to these repeat elements. Ninety two clones hybridized to the 180-bp repeats, 78 to 5S rDNA and 95 to 18S–25S rDNA, representing 0.99%, 0.84% and 1.02% of the nuclear clones, respectively

Table 1 Clones of the Landsberg BIBAC library originated from chloroplast and mitochondrial DNA

A. Chloroplast DNA-derived clones:

1B13, 1H4, 1I12, 1I15, 1J17, 1K8, 1L13, 1M19, 2D3, 2J8, 2K19, 2L19, 2N17, 3A12, 3A23, 3B21, 3C24, 3I22, 3J18, 3K17, 3M1, 3M18, 3O1, 3O17, 3P9, 4A8, 4B23, 4G17, 4J9, 4L2, 4O22, 4O24, 4P8, 4P21, 4P22, 5C6, 5F10, 5H10, 5I21, 5O13, 5P5, 6B22, 6D13, 6D24, 6E2, 6E11, 6J4, 6L3, 6N3, 6N11, 7A4, 7B8, 7D15, 7E20, 7F21, 7F24, 7I15, 7K13, 7L6, 7M10, 7N5, 8B21, 8C7, 8C14, 8D11, 8E13, 8E18, 8G5, 8I11, 8K1, 8K19, 8N23, 8N24, 8P1, 8P23, 9A13, 9B10, 9B23, 9H2, 9I5, 9I19, 9L4, 9M1, 9O13, 10B17, 10D5, 10D14, 10E6, 11A7, 11A16, 11B5, 11D20, 11D23, 11E5, 11E11, 11F3, 11H1, 11I11, 11J3, 11K1, 11K15, 11K22, 11L11, 11M22, 11P12, 11P17, 12A14, 12B24, 12C23, 12J10, 12J14, 12K2, 12K3, 12M12, 12O7, 13B3, 13D20, 13K10, 13M9, 13O4, 14D5, 14F4, 14G2, 14N22, 14O3, 14P21, 15B5, 15B9, 15C22, 15H7, 15H11, 15J8, 15K12, 15K19, 15L12, 15M21, 16D12, 16D17, 16H13, 16I6, 16I22, 16J23, 16K6, 16L7, 16N6, 16P11, 17C19, 17D19, 17E18, 17E24, 17H20, 17K4, 17L13, 17N22, 17O6, 17O24, 17P10, 18A21, 18B12, 18C24, 18D8, 18D13, 18E8, 18F7, 18H10, 18I21, 18L5, 18L12, 18M10, 18M17, 18P1, 18P9, 19A4, 19A6, 19A16, 19B13, 19G9, 19M6, 19M22, 19O20, 19P5, 20A19, 20E1, 20F21, 20K19, 20O10, 20P10, 20P17, 21A3, 21C1, 21C14, 21E18, 21F22, 21M7, 21N10, 22B3, 22B11, 22D5, 22F8, 22G2, 22L5, 22O5, 22O23, 22P1, 23A17, 23D15, 23E14, 23E22, 23I16, 23I24, 23P18, 24A3, 24A22, 24B6, 24E2, 24E13, 24F1, 24G3, 24H1, 24I17, 24J16, 24K18, 24L17, 24M17, 25C10, 25D1, 25D4, 25D9, 25E14, 25J11, 25L9, 25N24, 25O5, 25P4, 26C10, 26F7, 26F8, 26G11, 26I23, 26J19, 26M15, 26P4, 27A1, 27A20, 27D3, 27D10, 27F11, 27G4, 27G16, 27H20, 27H24, 27J2, 27J16, 27K1, 27L11, 27O1, 27P15, 28A17, 28B24, 28E11, 28I8, 28J3, 28J9, 28L13, 28L21, 28N15, 28O9, 29D8, 29D19, 29E16, 29E20, 29F16, 29O10, 29O20, 29P6, 29P12, 30B6, 30B19, 30C13, 30D8, 30E9, 30I9, 30I23, 30J19, 30L6, 30L14, 30N2

B. Mitochondrial DNA-derived clones:

2L5, 6N2, 7I3, 9B7, 10A10, 11F17, 12M1, 13N13, 14A13, 14K12, 15P21, 16E9, 16P20, 18A9, 18A11, 18L23, 18P4, 20J1, 23M24, 23P11, 25K3, 27J22, 29C11, 29I18, 30C10

(Table 2). The combined percentage (2.85%) of clones containing these abundant repeats is dramatically less than that observed in the large insert libraries from the Columbia ecotype (15.6%, IGF BAC library, Mozo et al. 1998; 18.3%, CIC YAC library, Creusot et al. 1995). This disparity is most-likely a consequence of the restriction enzymes used for construction of the libraries. *Bam*HI, used to construct the library described here, is G/C-rich in its recognition sequence, whereas the CIC YAC and IGF BAC libraries were constructed with *Eco*RI that is A/T-rich in its recognition sequence. In addition, there may be sequence differences between the Landsberg and Columbia libraries that contribute to repeat representation. For example, Southern blotting of 5S rDNA repeats in the Columbia and Landsberg ecotypes revealed heterogeneity (Campell et al. 1992).

Representation of single-copy, nuclear sequences in the Landsberg BIBAC library

To estimate the representation of single-copy nuclear sequences in the Landsberg BIBAC library, we hybridized the filters of the library with probes corresponding to 23 RFLP markers from the Arabidopsis genetic map (Liu et al. 1996; Table 2). Between 3 and 11 positive clones (seven on average) were identified with each DNA marker. Although this number is lower than the predicted 11.5 × haploid genome coverage, it is in agreement with studies from other similar libraries (e.g., Choi et al. 1995; Zhang et al. 1996; Mozo et al. 1998; Budiman et al. 2000; Meksem et al. 2000; Tao et al. 2001). Genome coverage estimates that are based on results from library screening with single-copy sequences often differ from the estimates based on the number of clones in the library, the average insert size of the clones and the species genome size. Such differences may result from incomplete gridding of the library onto high-density colony membranes, plate wells that lack clones, and/or from an inaccurate estimate of the species genome size.

The BIBAC library described here represents the first large-insert DNA library for the *Arabidopsis* Landsberg ecotype. It contains very large inserts (162 kb on average), is stable in *E. coli*, has sufficient genome coverage (11.5 ×) and contains a low fraction of clones with mitochondrial or chloroplast DNA. These features indicate that the BIBAC library is suitable for many applications in genome research, including physical mapping, map-based cloning, and functional and comparative genomics. The 162-kb insert size of the library is larger than the BAC and P1 libraries used to complete the genome sequence in the Columbia ecotype (Choi et al. 1995; Liu et al. 1995; Mozo et al. 1998). It is also larger than other BIBAC libraries reported to-date (Hamilton et al. 1999; Moullet et al. 1999; Meksem et al. 2000; Wu et al. 2000). Importantly, the clones of the library were constructed in a binary vector (BIBAC2) (Hamilton 1997), which is competent for plant transformation (Hamilton et al. 1996, 1999). Therefore, use of the library in *Arabidopsis* genomic research will streamline positional cloning and functional analysis of the genome sequence. Additionally, this library was constructed in the *Bam*HI site of the BIBAC2 vector. The G/C-rich recognition site of *Bam*HI provides a complement to the A/T-rich recognition sites of *Hin*dIII and *Eco*RI used for the TAMU (Choi et al. 1995) and IGF (Mozo et al. 1998) BAC libraries. The Landsberg BIBAC library may, therefore, provide an alternative tool for closing the remaining gaps in the Arabidopsis genome sequence map (The Arabidopsis Genome Initiative 2000).

The Landsberg BIBAC library is maintained at the GENE*finder* Genomic Resources, Texas A&M University, and at the Arabidopsis Biological Resource Center, Ohio State University, USA, and is available to the pub-

Table 2 Clones of the Landsberg BIBAC library identified by screening with 23 RFLP markers, 180-bp repeat, 5S rDNA and 18S–25S rDNA

Probe	Positive clones
A. RFLP marker-containing clones:	
mi32:	3C6, 4D1, 7O11, 23N15
mi51:	1H22, 3O14, 11F17, 18A11, 22L21, 27J22, 28F10
mi54:	11A17, 13F16, 13K22, 15A1, 19E1
mi61:	4C5, 4G22, 4M2, 6K22, 12J5, 14I24, 21I5, 25B5, 27I6, 29M12, 29M13
mi79a:	4B11, 7F20, 8N9, 12E16, 17C10, 28J13, 29F22
mi112:	4B12, 4C1, 5K11, 13B24, 16M12, 24M13, 29C1, 29F15, 29I9
mi116:	3O9, 15K20, 16E2, 26O2, 27D8
mi123:	17E13, 22C3, 23E4, 24O8, 25O4, 26K12, 27G24
mi128:	3I5, 5M22, 9M10, 18N6, 18N15, 18P11, 20J3, 21N20, 23M23, 25L6
mi148:	6L20, 7F10, 7O5, 11F18, 12P14, 16I11, 26H22, 27G1, 27N17
mi167:	1A15, 3K5, 7P10, 8E4, 13I2, 21F9, 24G9, 25D18, 25G1, 27M5, 28L24
mi204:	3O14, 5K10, 21J6, 24I16, 24O23
mi238:	10O2, 18B17, 22O2, 23J6, 26F18, 29L7
mi277:	6M24, 24O8, 30H13
mi279:	5N3, 10H3, 12C9, 18G14, 21F3, 24O8, 26H17, 26M20, 29N9, 30H8
mi301:	2K3, 2L6, 2M13, 6N2, 18J22, 23F4, 25O13, 26K3
mi306:	2G5, 9L17, 25D11, 29H6
mi310:	2D20, 8M2, 10E3, 12M22, 15I10, 24O8, 25N22, 26P14, 27E23, 27H3, 30A7
mi320:	9B2, 9B3, 16F10, 25K19, 27P19, 28H5
mi330:	10P6, 15E24, 16L19, 29L10
mi398:	4E11, 5J22, 9E12, 11A3, 16D13, 18P5, 29K11
mi422:	1H19, 7E23, 19A12, 28M19
mi465:	2F1, 8G6, 9B2, 13B6, 18O7, 23I14, 23M9, 24A21, 28G10
B. 180-bp repeat clones:	
2A14, 2D7, 2G8, 2J1, 2J16, 2K24, 2N12, 2P17, 3B9, 4A2, 4E10, 5G13, 6G2, 6I20, 6L16, 6M9, 7E23, 7M20, 8J5, 8K9, 9A3, 9C20, 10H4, 10O4, 10O20, 11C18, 11D4, 11I9, 11J10, 11L15, 13A18, 13C20, 13I1, 13I21, 14F8, 14H13, 14J16, 14O7, 15D5, 16A9, 16L10, 16L16, 16N16, 16O1, 17B12, 17H22, 17L11, 19A10, 20B11, 20B17, 20C1, 21F11, 21H15, 21H16, 21L13, 21L19, 21M10, 22A24, 22I7, 22I21, 22O16, 22O19, 22P16, 23O6, 24I22, 24J8, 24N1, 24O8, 24O18, 25E16, 25E23, 25H16, 25J13, 25O24, 26A7, 26B13, 26B19, 26F21, 26H22, 26L4, 27D21, 27K10, 28A19, 28L20, 29G16, 29G17, 29M15, 30A2, 30A9, 30B21, 30J9, 30K7	
C. 5S rDNA clones:	
1A11, 1G17, 2P17, 3C14, 3H13, 4C1, 4N16, 5G14, 5I23, 6H22, 6M9, 7C17, 7P16, 8H11, 8L17, 8P5, 9G12, 9M19, 9N14, 9O12, 10E5, 10H4, 11F7, 11M5, 11M12, 12A6, 12K6, 13D13, 13I21, 13N9, 14A13, 14B21, 14C11, 14E18, 14O7, 15D16, 15I19, 15K7, 16B3, 16I8, 16L8, 16L15, 17H22, 17I5, 17L11, 18L23, 18O17, 19G1, 19L1, 20B17, 20F6, 21H13, 21H15, 22I21, 22J17, 22N19, 23F11, 23G7, 23J22, 24F19, 24I22, 24M12, 25B1, 25J13, 25K3, 25K16, 26L18, 26P5, 27C22, 28E3, 28E18, 28I5, 28J5, 29E12, 29H21, 30I1, 30J2, 30J15	
D. 18S–25S rDNA clones:	
1J13, 1L3, 1P6, 2A18, 2O8, 2P3, 3C3, 3K11, 3L21, 4G6, 4I3, 4O2, 4O17, 4P2, 6B16, 6N8, 7A10, 7N19, 8C3, 8H2, 8P6, 9D17, 9E24, 9G18, 9J19, 9M5, 9N17, 9P16, 10E9, 10F11, 10I12, 10M21, 10N20, 12A18, 12F8, 12M1, 12O9, 12O20, 13E23, 13F20, 13G7, 14F3, 14K4, 15A21, 15C19, 15H16, 15J5, 16H12, 17D23, 17F23, 17G13, 17L15, 17O2, 19H6, 19N8, 19O12, 20A5, 20L24, 20P11, 21D20, 21G19, 22L21, 22M8, 22M11, 23A5, 23G8, 23M5, 23O16, 23O20, 24D14, 24E17, 24P22, 25B6, 25E12, 25F8, 25F20, 25G5, 25H13, 26D10, 26E14, 26F2, 26M14, 26O9, 26P3, 27O5, 27P4, 28F2, 28F10, 29D13, 29I1, 29L15, 29P23, 30B18, 30B24, 30O8	

lic. To request the library, individual clones, or high-density colony filters, please contact either the GENE*finder* Genomic Resources at http://hbz.tamu.edu -BAC Library-Library List, or the Arabidopsis Biological Resource Center at http://aims.cps.msu.edu/aims.

Acknowledgements The authors thank Dr. Carol M. Hamilton, Paradigm Genetics, Inc., for kindly providing the BIBAC2 vector, Chantel Scheuring, Texas A&M University, for critically reading the manuscript, and Jason Wilcox and Kathryn Nickel for technical assistance. This work was supported by a grant from the National Science Foundation-Plant Genome Program (award #: DBI-9872641).

References

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map based cloning of a gene containing omega-3 fatty acid desaturation in Arabidopsis. Science 258: 1353–1355

Azpiroz-Leehan R, Feldmann KA (1997) T-DNA insertion mutagenesis in Arabidopsis: going back and forth. Trends Genet 13:152–156

Baumgarther BJ, Rapp JC, Mullet JE (1993) Plastid genes encoding the transcription/translation apparatus are differentially transcribed early in barley (*Hordeum vulgare*) chloroplast development. Plant Physiol 101:781–791

Bechtold N, Ellis J, Pelletier G (1993) In planta *Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. Life Sci 316:1194–1199

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. Genomics 19:137–144

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

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

Canilleri C, Lafleuriel J, Macadre C, Varoquaux F, Parmentier F, Parmentier Y, Picard G, Caboche M, Bouchez D (1998) A YAC contig map of *Arabidopsis thaliana* chromosome 3. Plant J 15:633–642

Campell BR, Song Y, Posch TE, Cullis CA, Town CD (1992) Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. Gene 112:225–228

Chang C, Bowman J, DeJohn A, Lander E, Meyerowitz E (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85:6856–6860

Chang SS, Park SK, Kim BC, Kang BJ, Kim DU, Nam HG (1994) Stable genetic transformation of *Arabidopsis thaliana* by *Agrobacterium* inoculation in planta. Plant J 5:551–558

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

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature Genet 2:203–207

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

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

Desprez T, Amselem J, Caboche M, Hofte H (1998) Differential gene expression in Arabidopsis seedlings monitored using cDNA arrays. Plant J 14:643–652

Ecker J (1992) PFGE and YAC analysis of the Arabidopsis genome. Methods 1:186–194

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:235–240

Feldmann KA (1991) T-DNA insertion mutagenesis in Arabidopsis: mutational spectrum. Plant J 1:71–82

Feldmann KA, Marks MD (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. Mol Gen Genet 208:1–9

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

Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM (1992) Isolation of the Arabidopsis *ABI3* gene by positional cloning. Plant Cell 4:1251–1261

Grill E, Somerville C (1991) Construction and characterization of a yeast artificial chromosome library of Arabidopsis which is suitable for chromosome walking. Mol Gen Genet 226:484–490

Hamilton CM (1997) A binary-BAC system for plant transformation with high-molecular-weight DNA. Gene 200:107–116

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

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

Hauge BM, Giraudat J, Hanley S, Hwang I, Kohchi T, Goodman HM (1991) Physical mapping of the Arabidopsis genome and its applications. In: Hermann RG, Larkins B (eds) Plant Molecular Biology 2. Plenum Press, New York, pp 239–248

Helliwell CA, Sheldon C, Olive MR, Walker AR, Zeevaart JAD, Peacock WJ, Dennis E (1998) Cloning of the Arabidopsis ent-kaurene oxidase gene *GA3*. Proc Natl Acad Sci USA 95:9019–9024

Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Z, Korber H, Redei GP, Schell J (1989) High-frequency T-DNA-mediated gene tagging in plants. Proc Natl Acad Sci USA 86:8467–8471

Konieczny A, Ausubel FM (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Koornneef M, Eden JV, Hanhart CJ, Stam P, Braaksma FJ, Feenstra WJ (1983) Linkage map of *Arabidopsis thaliana*. J Hered 74:265–272

Kotani H, Hosouchi T, Tsuruoka H (1999) Structural analysis and complete physical map of *Arabidopsis thaliana* chromosome 5 including centromeric and telomeric regions. DNA Res 6:381–386

Lin X, Kaul S, Rounsley S, Shea TP, Benito M-I, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldman TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Shen M, Pai G, Aken SV, Umayam L, Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Neilman WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter JC (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761–768

Liu Y-G, 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

Liu Y-G, Mitsukawa N, Lister C, Dean C, Whitter R (1996) Isolation and mapping of a new set of 129 RFLP markers in *Arabidopsis thaliana* using recombinant inbred lines. Plant J 10:733–736

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

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

Martienssen RA (1998) Functional genomics: probing plant gene function and expression with transposons. Proc Natl Acad Sci USA 95:2021–2026

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 loci. Mol Gen Genet 233:25–32

Martinez-Zapater JM, Estelle M, Somerville CR (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol Gen Genet 233:25–32

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

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

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

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

Pierce JC, Sauer B, Sternberg N (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. Proc Natl Acad Sci USA 89:2056–2060

Pruitt RE, Meyerowitz EM (1986) Characterization of the genome of *Arabidopsis thaliana*. J Mol Biol 187:169–183

Redei GP (1962) Single locus heterosis. Z Vererbungslehre 93:164–170

Redei GP (1970) *Arabidopsis thaliana* (L.) Heynh. A review of the genetics and biology. Bibliogr Genet 20:1–151

Ruan Y, Gilmore J, Conner T (1998) Towards Arabidopsis genome analysis: monitoring expression profiles of 1,400 genes using cDNA microarrays. Plant J 15:821–833

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

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schmidt R, West J, Love K, Lenehan Z, Lister C, Thompson H, Bouchez D, Dean C (1995) Physical map and organization of *Arabidopsis thaliana* chromosome 4. Science 270:480–483

Somerville C, Dangl J (2000) Plant Biology in 2010. Science 290:2077–2078

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

Sussman MR, Amasino RM, Young JC, Krysan PJ, Austin-Phillips S (2000) The Arabidopsis knockout facility at the University of Wisconsin-Madison. Plant Physiol 124:1465–1467

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

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

Ward ER, Jen GC (1990) Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. Plant Mol Biol 14:561–568

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

Zachgo EA, Wang ML, Dewdney J, Bouchez D, Camilleri C, Belmonte S, Huang L, Dolan M, Goodman HM (1996) A physical map of chromosome 2 of *Arabidoopsis thaliana*. Genome Res 6:19–25

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

Zhang H-B, Choi S, Woo S-S, Li Z, Wing RA (1996) Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 2:11–24