Theor Appl Genet (2001) 102:892–899

R. Ming · P.H. Moore · F. Zee · C.A. Abbey
H. Ma · A.H. Paterson

Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome

Received: 11 April 2000 / Accepted: 28 July 2000

Abstract A bacterial artificial chromosome (BAC) library was constructed from high-molecular-weight DNA isolated from young leaves of papaya (*Carica papaya* L.). This BAC library consists of 39168 clones from two separate ligation reactions. The average insert size of the library is 132 kb; 96.5% of the 18700 clones from the first ligation contained inserts that averaged 86 kb in size, 95.7% of the 20468 clones from the second ligation contained inserts that averaged 174 kb in size. Two sorghum chloroplast probes hybridized separately to the library and revealed a total of 504 chloroplast clones or 1.4% of the library. The entire BAC library was estimated to provide 13.7× papaya-genome equivalents, excluding the false-positive and chloroplast clones. High-density filters were made containing 94% or 36864 clones of the library with 12.7× papaya-genome equivalents. Eleven papaya-cDNA and ten *Arabidopsis*-cDNA probes detected an average of 22.8 BACs per probe in the library. Because of its relatively small genome (372 Mbp/1 C) and its ability to produce ripe fruit 9 to 15 months after planting, papaya shows promise as a model plant for studying genes that affect fruiting characters. A rapid approach to locating fruit-controlling genes will be to assemble a physical map based on BAC contigs to which ESTs have hybridized. A physical map of the papaya genome will significantly enhance our capacity to clone and manipulate genes of economic importance.

Keywords *Carica* · Genome · Physical mapping · Positional cloning

Communicated by F. Salamini

R. Ming (✉) · P.H. Moore · H. Ma
Hawaii Agriculture Research Center, Aiea, HI 96701, USA
e-mail: rming@harc-hspa.com

P.H. Moore
USDA-ARS, Tropical Fruit,
Vegetable and Sugarcane Research Unit, Aiea, HI 96701, USA

F. Zee
USDA-ARS, National Clonal Germplasm Repository, Hilo,
HI 96720, USA

C.A. Abbey · A.H. Paterson
Dept Soil and Crop Science, Texas A&M University,
College Station, TX 77843, USA

H. Ma
Dept Molecular Biosciences and Biosystems Engineering,
University of Hawaii, Honolulu, HI 96822, USA

Present address:
A.H. Paterson, Applied Genetic Technology Center,
University of Georgia, Athens, GA 30602, USA

Introduction

Papaya (*Carica papaya* L.) is a principal crop of tropical and subtropical regions worldwide. Papaya trees are grown for fruit that is consumed, and for stems, leaves and roots that have a wide range of applications in folk medicine (Purina and Sandhya 1988; Ockerman et al. 1993; Osato et al. 1993). In addition, milky juice from unripe papaya fruit is the primary source of papain, a widely used proteolytic enzyme (Dunne and Horgan 1992).

Papaya is a perennial crop that fruits throughout the year. It is one of the few tree crops that produce ripe fruit as quickly as 9 months from planting. It is polygamous with three sex forms: male, female, and hermaphrodite. Controlled crossing of papaya is easy, using pollen from male or hermaphrodite flowers to fertilize stigmata of female or hermaphrodite flowers. Year-round flowering and fruiting of papaya plants produce a continuous supply of fruit for the evaluation of quality and productivity. The high number of seeds per fruit (about 800–1200), coupled with the perennial fruiting of papaya, can provide virtually unlimited numbers of offspring from a controlled cross. An alternative to sexual crossing for manipulating genes in papaya is based on an efficient system for tissue culture and genetic transformation (Fitch et al. 1992). The short generation time (9–15 months) of papaya facilitates timely evaluation of transgenic lines and promotes rapid production and evaluation of breeding and mapping populations. In addition, a

clonal propagation system has been established to allow testing of individual plants in multiple environments. The availability of genetic transformation, the efficiency of making crosses, the production of large numbers of offspring, and having continuous flowering throughout the year make papaya an attractive model plant for fruit tree genetic and genomic research.

Molecular techniques are beginning to contribute to our understanding of papaya genetics. Anonymous DNA markers were used to distinguish and estimate genetic relationships among papaya cultivars and *Carica* species (Sharon et al. 1992; Stiles et al. 1993). A first-generation linkage-map of papaya was constructed based on randomly amplified polymorphic DNA (RAPD) markers (Sondur et al. 1996). Quantitative trait loci (QTLs) were mapped using RAPD markers for plant height, stem diameter, and the number of the node at which the first flower appears. A total of 2–4 QTLs were detected for each trait (Sondur et al. 1995).

Storey (1938) and Hofmeyr (1938) established the genetic principle of sex determination in papaya as a multi-allelic locus (M_1, M_2, and m) controlling all three sex types. The constitutions of the three sex types are male (M_1 m), hermaphrodite (M_2m), and female (mm), while the combination of dominant alleles (M_1M_1, M_1M_2, and M_2M_2) is lethal. Although a more complex model for sex determination in papaya was later published (Storey 1953), the original hypothesis remains valid and is supported by accumulated field data. The sex-determination process in papaya is unique in flowering plants (Dellaporta and Calderon-Urrea 1993), and positional cloning of this gene might be the only valid approach to uncover the mechanism controlled by the above three allelic forms. The construction of physical maps is a valuable step for the identification and isolation of genes of interest by map-based or positional cloning. Yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are the two most commonly employed vectors for carrying large DNA fragments. YAC vectors have been extensively used in human and *Arabidopsis* projects (Burke et al. 1987; Grill and Somerville 1991) and for positional cloning of a number of plant disease resistance genes (Martin et al. 1993; Bent et al. 1994; Yoshimura et al. 1996; Tai and Staskawicz 2000). However, chimerism and difficulty of manipulation are limitations in the YAC vector systems. BAC vectors, an *Escherichia coli* F-factor-based system, overcome the chimerism problem and speed up the construction of large insert DNA libraries. However, the average size of BAC inserts falls below 200 kb in general (Shizuya et al. 1992). BAC clones showed high efficiency for generating contig maps by BAC fingerprinting with RFLP analysis on agarose gels (Marra. et al. 1997). An entire BAC containing a candidate gene(s) can be transformed into a plant nuclear genome for complementation tests (Hamilton. et al. 1996). BAC clones are suitable for fluorescence in situ hybridization (FISH) analyses to enable direct gene localization on chromosomes (Jiang et al. 1995). BAC libraries have been generated for important plant species, including *Arabidopsis* (Choi et al., 1995), rice (Wang et al., 1995; Zhang et al. 1996), sorghum (Woo et al. 1994; Lin et al. 1999), sugarcane (Tompkins et al. 1999), wheat (Moullet et al. 1999), soybean (Marek and Shoemaker 1997), and cotton (Abbey et al. 2000). The small genome (372 mbp) of papaya should be relatively easy to characterize, particularly for physical mapping (Arumuganathan and Earle 1991).

The long-term objective of our work is to study the genome structure, organization, and content of papaya using a large insert DNA library as one of the major tools. The specific objectives of this study are: (1) to construct a papaya BAC library with at least four haploid-genome equivalents, (2) to characterize the library for insert size, as well as chloroplast and ribosomal DNA content, (3) to screen the library with homologous and heterologous cDNA probes, and (4) to identify BAC clones containing a DNA marker linked to the sex-determination gene.

Materials and methods

Chemicals and enzymes

All chemicals were obtained from Sigma (St. Louis, Mo., USA) or Fisher Scientific (Pittsburgh, USA) unless otherwise noted. Restriction and other required enzymes were obtained from New England Biolabs (Beverly, Mass., USA). Low-melting-point agarose was from FMC Bioproducts (Rockland, Md., USA).

Plant material

Papaya cultivar SunUp was used as the source of genomic DNA for BAC library construction. SunUp is one of the parents of the papaya hybrid cultivar UH Rainbow currently in production in Hawaii. SunUp carries genes conferring resistance to anthracnose (chocolate spot) and Broad mite, and possesses a number of favorable agronomic traits, such as low bearing height, short juvenile phase, high sucrose content, floral fragrance of fruit, and proper fruit size for shipping. In addition, it also carries the transgenic coat-protein gene conferring resistance to papaya ring-spot virus disease,

Young leaves were collected from hermaphroditic SunUp trees growing at Waiakea Station, Hilo, Hawaii, and shipped on dry ice to Texas A&M University. Samples were stored at –80°C until extracted.

Papaya nuclear isolation

Papaya high-molecular-weight DNA was isolated from young leaves according to the procedure described H. Zhang et al. (1995). Approximately 5 g of leaves were homogenized by grinding in a mortar containing liquid nitrogen. The powdered leaf tissue was transferred and gently stirred for 10 min into 200 ml of HB (homogenization buffer, 10 mM Tris, 80 mM KCl, 10 mM EDTA, 1 mM spermine, 1 mM spermidine, and 0.5 M sucrose, pH9.4–9.5) plus 0.15% β-mercaptoethanol and 0.5% Triton X-100 min. The homogenate was filtered into a beaker through one layer of miracloth and two layers of cheesecloth. The filtrate was divided into four 45-ml aliquots placed in 50-ml tubes and centrifuged at 1000 g for 20 min at 4°C. The supernatant was decanted and the pellet was re-suspended in 45 ml of HB plus Triton and β-mercaptoethanol. The re-suspended nuclei were filtered through a #50

mesh fine filter, rinsed with 200 ml of HB, divided into 45-ml fractions in each of four 50-ml tubes, and centrifuged at 1000 g for 20 min at 4°C. The washing process was repeated three times. After the third wash, the pelleted nuclei in each tube were re-suspended in 10 ml of HB without Triton and β-mercaptoethanol, and combined into one tube. The nuclei were then mixed with an equal volume of 1% low-melting-point agarose pre-warmed to 45°C and poured into plug molds on ice, 100 µl per plug. When the agarose was completely solidified, the plugs were transferred into 5–10 vol of lysis buffer (0.5 M EDTA pH 9.0–9.3, 1% sodium laurylsarcosine, and 0.1 mg/ml proteinase) and stored at 4°C.

Partial digestion and size selection

Partial digestion conditions used in cotton BAC library construction were tested in papaya (Abbey et al. unpublished). Different enzyme concentrations of *Hin*dIII, ranging from 0.1 to 0.5 units/mg per plug (32 ng DNA/mg per plug) were used for 2–10 min incubations to determine the optimal conditions for partially digesting papaya high-molecular-weight DNA and producing DNA fragments larger than 300 kb in size. Four sets of about 60 mg of papaya nuclei in agarose plugs (approximately 2 µg of DNA/plug) were sliced into small pieces. The sliced plugs were equilibrated on ice for 1.5 h with one change of buffer after 1 h in 1 ml of *Hin*dIII reaction buffer (NEB, Beverly, Mass., USA) supplemented with 1 mM of bovine serum albumin, 1 mM of DTT, and 4 mM of spermidine. The supernatant was replaced with 500 µl of fresh *Hin*dIII reaction buffer containing 0.2 units/mg per plug of enzyme for the four sets of plugs. The reactions were incubated on ice at 4°C for 5 h to allow time for the enzyme to diffuse into the plugs. The tubes containing the reactions were then transferred into a 37°C water bath and incubated for 3 min in the reaction solutions. The reactions were stopped by adding 1/10 vol (50 µl) of 0.5 M EDTA at pH 8.0 to each tube. The partially digested DNA was separated by pulsed-field gel electrophoresis (PFGE) in a 1% agarose gel with TAE buffer at 14°C and a 8-s switch time for 16 h. The compression zone of the gel, containing DNA fragments from 50 to >300 kb, was excised into nine thin pieces (5 mm each). Each gel piece was digested with Gelase at 45°C for 1 h. The DNA solution was tested for concentration by running the DNA samples along with a series of λ standards in a gel containing ethidium bromide, and then used in ligation.

BAC vector preparation

BAC vector *pBeloBAC11* (Shizuya et al. 1992**)** was isolated by alkaline lysis (Sambrook et al. 1989) followed by two cesium chloride density-gradient centrifugations. The vector was completely digested with *Hin*dIII and self-ligated to test for false-positive white colonies. The digested BAC vectors were de-phosphorylated with shrimp alkaline phosphatase.

Ligation and transformation

Size-selected *Hin*dIII DNA fragments were ligated into the dephosphorylated *pBeloBAC11* vector in a ratio of 1 ng of vector: 4 ng of insert in a total volume of 300 µl. Each ligation was performed in ligation buffer with 80 units of T4 DNA ligase and incubated at 16°C overnight. The ligation mix was placed on a Millipore 0.025-µM filter (BRL, Rockville, Md., USA) against 0.5× TE for 30 min to 1 h to remove salts in the ligation buffer. Three samples of ligation mix were used to transform 25 µl of *E. coli* ElectroMAX (BRL, Rockville, Md., USA) DH10B cells by electroporation at the following optimized conditions (Abbey et al. unpublished): capacitance 25 µF, resistance 100 Ohms, and a voltage gradient of 1.8 kV/cm. After electroporation, the cells were resuspended in SOC medium (2% bacto-trypton, 0.5% bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose, pH 7.0) and incubated at 37°C for 45 min while shaking at 200 rpm. The cells were then plated on LB medium (1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl pH 7.5, 1.5% agar) containing chloramphenicol (12.5 µg/ml), X-GAL (3%) and IPTG (20%), and grown at 37°C for 35–48 h. At this time, the recombinant clones (BACs) could be clearly identified as either blue (non-recombinant) or white (recombinant) phenotypes. After checking the insert size of positive clones, the remaining ligation mix with suitable-sized DNA fragments was transformed. White colonies were picked by hand (ligation 1) or by a Q-BOT (Genetix, UK) (ligation 2) into 384-well plates containing 150 µl of LB freezing media [36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM Na citrate, 0.4 mM $MgSO_4$, 6.8 mM $(NH_4)_2SO_4$, 4.4% (v/v) glycerol, 12.5 µg/ml chloramphenicol, LB], incubated overnight at 37°C, and then stored at –80°C.

Individual BAC analysis

Ten BAC clones were randomly chosen from each transformation for determining the insert size. Individual BAC clones were grown in 5 ml of LB with 12.5 µg/ml of chloramphenicol at 37°C while shaking at 200 rpm for 18–20 h. Supercoiled plasmid DNA was isolated by the miniprep alkaline-lysis method (Sambrook et al. 1989). *Not*I-digested BAC clones were separated by pulsed-field gel electrophoresis (PFGE) in 1% agarose gel with TBE buffer at 14°C, 6.0 V/cm with a 20-s pulse time for 16 h. The gel was stained with ethidium bromide and photographed. The insert size of each BAC clone was determined by the sum of the sizes of all bands compared with a λ mid-range size standard (NEB, USA), excluding the vector band.

High-density filter preparation and screening

High-density filters were made by replicating 48 of the 384-well microtiter plates onto a single 22.5×22.5 cm nitrocellulose (Hybond N+) filter using Q-BOT. Clones were double-spotted using a 4×4 array with six fields per filter. The grid pattern allowed 18432 clones of each library to be double-spotted on one filter. Filters were incubated overnight on Q-trays containing LB/Agar with 25 µg/ml of chloramphenicol. The fixation of plasmid DNA to the membrane was performed according to the manufacturer's recommendation. Colony filters were processed and hybridized using standard techniques (Sambrook et al. 1989).

Papaya and *Arabidopsis* cDNA probes were selected to screen the BAC library. Four cloned papaya genes were also hybridized to the BAC library. Probes were amplified using M13 universal (SP10) and reverse (SP30) primers (Operon, USA), and digested with *Eco*RI and *Xho*I for papaya cDNA clones and with *Sal*I and *Not*I for *Arabidopsis* cDNA clones to eliminate the vector sequences. The insert DNA was excised and digested with gelase before being used for hybridization. Two sorghum chloroplast probes, *ropB* and *trunk*, were prepared by amplifying sorghum genomic DNA with specific primers designed for these two probes. One cotton rDNA probe, *pXP108*, was also used to screen the BAC library.

Southern-blot analysis

The CHEF gels containing the BAC clones were photographed then placed in 0.25 N HCl for two 5-min incubations, with shaking, to fragment the large insert DNA by de-purination. CHEF gels were then placed in 0.4 N NaOH for two 20-min incubations with shaking to denature the DNA. The Southern transfer onto Hybond N+ membrane was performed in 0.4 N NaOH buffer for 20 h. The membranes were treated with 2× SSC after blotting. The Southern blots were hybridized with papaya total DNA to ensure that each BAC contains papaya DNA fragments. Genomic DNA of papaya varieties SunUp and Kapoho was extracted, digested with *Eco*RI, *Hin*dIII, and *Xba*I, and transferred to Southern blots for labeling the probes used in BAC screening.

Results

Construction of the BAC library

Based on the DNA size standard, most partially digested DNA fragments were concentrated on the size-selection gel in the region above 300 kb, while a faint smear of DNA was spread to the 25-kb region. Nine segments were cut every 5 mm from the size-selection gel-yielding fragments ranging from 50 kb to more than 300 kb in size. These plugs were digested with gelase and ligated to *pBeloBAC11* vectors to test the ligation efficiency. The DNA fragments ranging from 50 to 125 kb showed the highest ligation and transformation efficiencies. No positive clone was generated from DNA fragments larger than 200 kb. Forty BAC clones were randomly picked and tested for insert size (Fig. 1a). Ligation mixes with DNA fragments ranging from 75 to 125 kb were transformed into host cell DH10b. A total of 18700 BAC clones were generated from this first ligation.

The optimal partial-digestion condition was repeated for a second set of plugs. Nine segments were also cut out from the size-selection gel containing DNA fragments ranging from 50 kb to larger than 300 kb. Initial transformation showed the highest ligation and transformation efficiency for DNA fragments larger than 175 kb and between 75 and 125 kb. No positive clone was generated from the ligation reaction for DNA fragments between 125 and 175 kb. Forty BAC clones were randomly picked and tested for insert size (Fig. 1b). DNA fragments larger than 175 kb were transformed into host cell DH10b. A total of 20468 BAC clones were generated from this second ligation.

Fourteen BAC clones, each from ligations 1 and 2 that had been tested for insert sizes, were transferred onto Hybond N+ membranes. Total papaya genomic DNA was used as a probe to hybridize to the BAC blots. The results confirmed that the BAC clones contain papaya insert DNA (Fig. 1c, d). BAC clone samples 2 and 3 in Fig. 1a showed a similar DNA concentration with the other clones. However, very faint signals were generated from the Southern hybridization of total papaya genomic DNA (Fig. 1c), indicating a rare sequence of insert DNA in these two clones.

BAC library screening

Sixty sets of high-density filters, each set consisting of one filter each from ligations 1 and 2, were made using

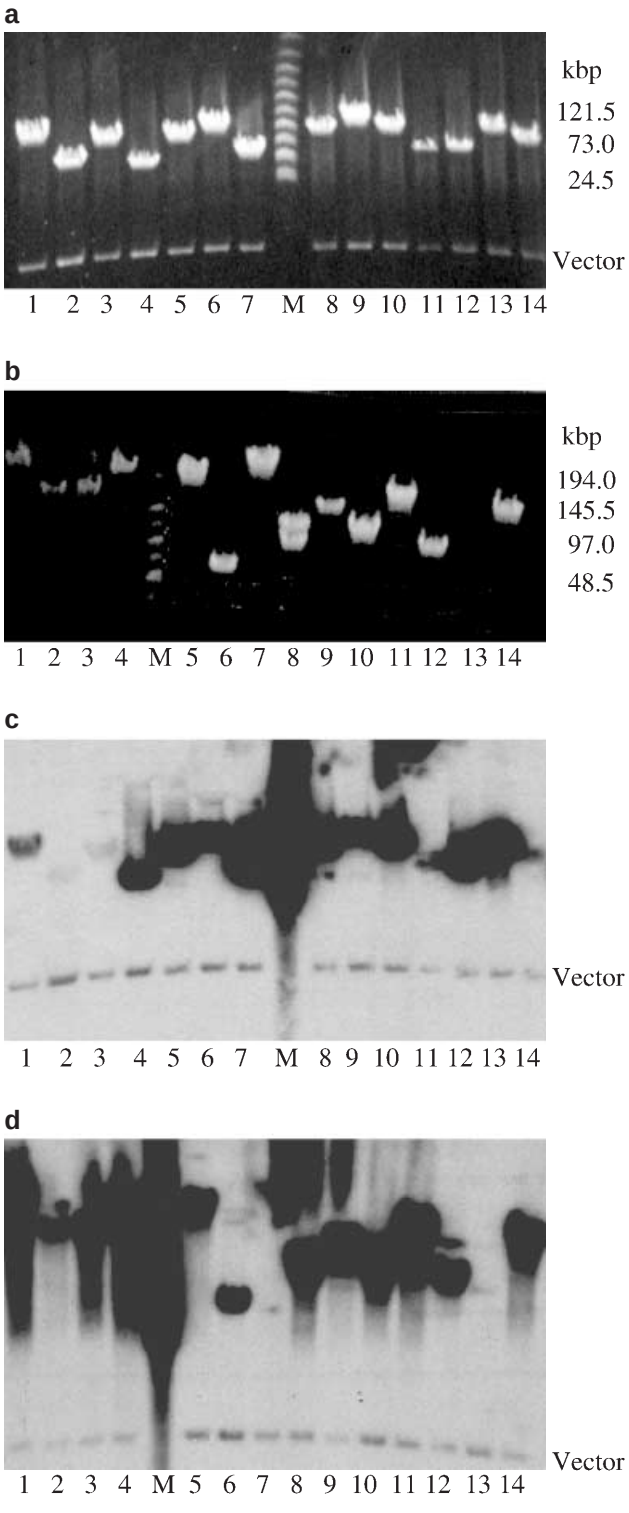

Fig. 1 a CHEF gel to test the insert sizes of BAC clones generated from the first batch of ligation. *Lane M* molecular-weight standard. *Lanes 1–14* individual BAC clones showing the insert and vector (bottom) DNA bands. *Samples 1 to 10* were randomly selected clones from the ligation reaction with the 6th largest DNA fragments cut out of the size-selection gel (see Materials and methods). Samples from the ligation reaction with the 5th largest DNA fragments not shown. Ligation mixes of the 5th and 6th largest DNA fragments were used in full-scale transformation to produce the first batches of BAC clones in the library. *Samples 11 to 14* were from the ligation reaction with the 7th largest DNA fragments, and were later discarded. **b** CHEF gel to test the insert sizes of BAC clones generated from the second batch of ligation. *Samples 1 to 7* were randomly selected clones from the ligation reaction with the largest-sized fragments. *Samples 8 to 14* were from the ligation reaction with the second largest-sized fragments. *Sample #13* was a false positive clone without insert DNA. Ligation mixes of both sized fragments were used in full-scale transformation to produce the second batches of BAC clones. **c**, **d** Individual BAC clones were hybridized with total papaya genomic DNA. The Southern blots were made from the gels in **a** and **b**, respectively

Table 1 Results of screening the papaya BAC library with homologous and heterologous cDNA, rDNA, and chloroplast DNA probes

Probes	No. of bands[a]	No. of hits		
		Ligation 1	Ligation 2	Total
AEST9	2	5	8	13
AEST18	2	5	7	12
AEST36	2	6	14	20
AEST37	6	0	0	0
AEST47	2	10	17	27
AEST48	3	6	17	23
AEST63	4	8	17	25
AEST64	2	9	9	18
AEST69	4	16	47	63
AEST127	2	19	29	48
CPF9A1	5	8	7	15
CPF9A2	3	0	0	0
CPF9A3	2	6	5	11
CPF9A4	1	9	11	20
CPF9A5	2	4	6	10
CPF9A6	2	3	25	28
CPF9A7		10	11	21
CPF26A3		3	3	6
CPF26A7		18	19	37
CPF26A4&5		19	27	46
Total	**44**	**154**	**279**	**433**
18sPXP108		18	43	61
ropB and *trunk*		211	293	504

[a] Number of bands detected by Southern hybridization of genomic DNA digested with *Eco*RI, *Hin*dIII, and *Xba*I. The number of bands recorded here was the maximum number detected by one of the three enzymes

Q-BOT. Each filter contained 18432 BAC clones (48 of the 384-well microtiter plates).

Ten *Arabidopsis* and 11 papaya cDNA clones were isolated and digested with restriction enzymes as described. The insert DNA of these clones was used to screen the BAC library. Nine *Arabidopsis* and ten papaya probes detected a total of 433 BAC clones, 154 clones from ligation 1 and 279 clones from ligation 2 (Table 1). Two probes, *AEST-37* and *CPF 9A2*, failed to hybridize to the BAC filters. The number of clones identified by each probe varied from 3 to 19 for ligation 1 and from 3 to 47 for ligation 2, with averages of 8.1 and 14.8, respectively (Table 1). The number of clones identified by each probe was significantly correlated between ligations 1 and 2 (r=0.72, P=0.0005), indicating that the ligation events were random and genome coverage was nearly complete (Fig. 2).

Most BAC clones hybridizing to each of the two sorghum chloroplast probes *ropB* and *trunk* overlapped. The two sorghum chloroplast probes together identified 211 BAC clones from ligation 1 and 293 from ligation 2.

To estimate the number of clusters of rDNA genes in papaya, a single rDNA probe from cotton was hybridized to the library and identified 18 clones from ligation 1 and 43 clones from ligation 2. This result suggested a minimum of five clusters of the rDNA gene family in the papaya genome based on the genome equivalents of the library (Table 1).

A total of 311 clones were digested with *Not*I, and the vector and insert were separated by CHEF electrophoresis

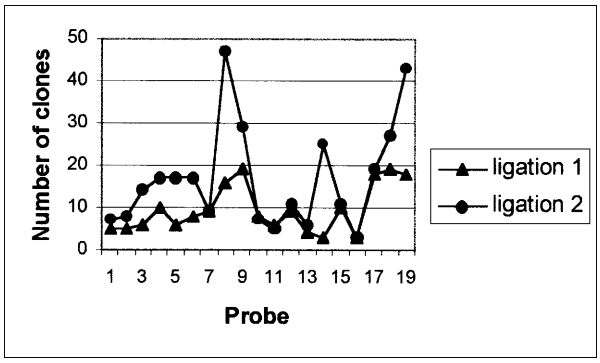

Fig. 2 Number of clones identified by individual probes in the library. These probes are (*from 1 to 19*) nine *Arabidopsis* EST clones: AEST 9, AEST 18, AEST 36, AEST 37, AEST 48, AEST 63, AEST 64, AEST 69 and AEST 127; and ten papaya cDNA clones: CPF9A1, CPF9A2, CPF9A3, CPF9A4, CPF9A5, CPF9A6, CPF9A7, CPF9A12, CPF9A16, and CPF9A13 + CPF9A14. The last two papaya cDNA probes were combined and hybridized to a set of BAC filters

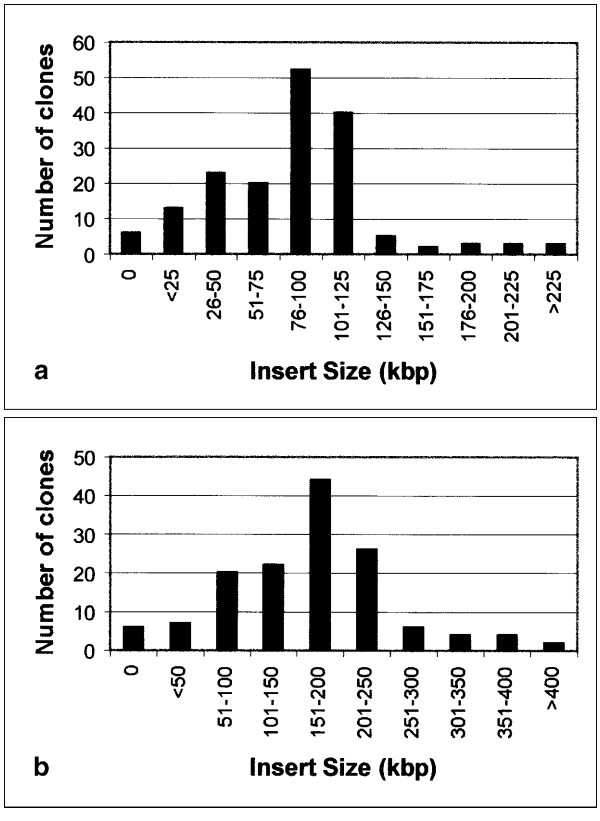

Fig. 3a, b Insert size distribution of BAC clones in the BAC library. BAC clones were randomly selected and digested with *Not*I before running the CHEF gels. **a** 170 clones from the first batch of BAC clones generated from ligation 1; **b** 141 clones from ligation 2

to estimate the average insert size of the library. Among the 170 clones from ligation 1 examined, six (3.5%) were false positives (no insert). The insert sizes ranged from 10 to 340 kb and averaged 86 kb; 63% of the clones were between 75 to 125 kb in size (Fig. 3a). The

18700 clones of ligation 1 were estimated to contain 4.5× haploid-genome equivalents of papaya excluding the false-positive and chloroplast clones.

Among the 141 clones from ligation 2 examined, six (4.3%) were false positives. The insert sizes ranged from 15 to 565 kb and averaged 174 kb; 53% of the clones were between 150 to 250 kb in size (Fig. 3b). The 20468 clones of ligation 2 covered 9.2× haploid-genome equivalents of papaya excluding the false-positive and chloroplast clones.

The 39168 clones of the entire library covered 13.7× haploid-genome equivalents. However, only 18432 clones each from ligations 1 and 2 were spotted onto high-density filters. The first filter containing BAC clones from ligation 1 covered 4.4× of the papaya genome while the second filter containing BAC clones from ligation 2 covered 8.3× of the papaya genome. These two high-density filters contained 36864 BAC clones or 94% of the BAC library with 12.7× papaya-genome equivalents.

Seventeen of the 21 probes used to screen the BAC library hybridized to a total of 44 bands on the SunUp genomic DNA blots. These 17 probes detected 114 BAC clones from ligation 1 and 219 from ligation 2. The average number of BAC clones per band ranged from 3 to 24 with an average of 9.5 (Table 1). The number of bands detected by a given probe was not correlated with the number of BAC clones hit by the same probe (r=0.27, P=0.35). If each band represents one homologous region of the probe, the genome equivalents of this BAC library are calculated to be 3.5× for ligation 1 and 6.3× for ligation 2. This estimate represents a 20% and 24% reduction compared with the 4.4× and 8.3× for ligations 1 and 2 estimated from the average insert size and number of clones, respectively.

Preliminary library screening for BAC clones containing sex determination gene

The two filters containing the BAC library were hybridized against an 800-bp papaya DNA fragment that is a sequence-characterized amplified region (SCAR). This SCAR marker, W11, co-segregates with the sex-determination gene when mapped in a papaya F_2 population with 139 individuals (J. I. Stiles, P. H. Moore and R. Ming, unpublished). Four BAC clones were identified by this SCAR marker. BAC end-cloning (Chen and Gmitter 1999; Yang and Mirkov 2000) was carried out on one of the four BAC clones. BAC end-clones were used as probes to iteratively screen the positive BAC clones and determine their order in a contig.

Discussion

Partial digestion of high-molecular-weight DNA is a pivotal step in BAC library construction. Conditions for optimal digestion must be determined experimentally for each individual species. It is known that the purity of nuclear DNA has a significant impact on the degree of digestion. We initially used the digestion conditions that had been used for producing a cotton BAC library (Abbey et al. unpublished). This protocol produced fragments that were too small, apparently because the purity of papaya DNA was higher than that of cotton. After testing a range of lower enzyme concentrations and shorter digestion times, we established the optimal conditions for partial digestion of papaya DNA to generate DNA fragments larger than 300 kb (see Materials and methods).

Second-size selection has been reported to effectively reduce the number of small insert clones in the library (Woo et al. 1994; Zhang et al. 1996). However, for our papaya library, the ligation and transformation efficiencies were too low to be useful when using the DNA fragments generated from second-size selection. Therefore, the insert DNA fragment size for the present BAC library was selected only once on the CHEF gel. This procedure resulted in 21% and 5% of the clones with insert DNA smaller than 50 kb for BAC clones generated from ligations 1 and 2, respectively (Fig. 3a, b).

This BAC library was constructed from two separate ligation reactions. The first ligation produced a smaller insert DNA, while the second ligation produced DNA inserts twice as large. We have no explanation for differences in the efficiencies between the two ligation reactions since the DNA concentration, ligation, and transformation conditions were kept the same. Nevertheless, these two batches of BAC clones with different insert sizes may prove to have advantages for specific purposes. The high-density filter containing the large insert (174-kb) clones of ligation 2 is calculated to have 8.8× genome equivalents, and would entail a lower cost, by one-half for physical mapping, compared with the BAC clones from ligation 1. On the other hand, the BAC clones of ligation 1 could be more efficient for sub-cloning because of their smaller inserts.

Two out of the twenty-three probes tested, one papaya and one *Arabidopsis* cDNA clone, failed to hybridize to the BAC library. Based on the limited data that 11 of the 12 papaya probes tested hybridized to the BAC library, we estimate that this BAC library covers about 91.7% of the papaya genome. The high hybridization rate (91%) of *Arabidopsis* cDNA probes to the papaya BAC library does indicate high homology between these two genomes and supports the phylogenetic reports that the Caricaceae and Brassicaceae are closely related (Bremer et al. 1998).

The average number of clones identified by a papaya or *Arabidopsis* cDNA probe was 19.4 or 26.6, respectively (Table 1). These numbers are higher than would occur given the estimated 12.7× genome equivalents for the two high-density filters used for screening. This suggests a 50% duplication within the papaya genome and a 100% duplication of the papaya genome compared with that of *Arabidopsis*. The duplication might be even higher if 60% of the published sequence of the *Arabidopsis* genome is duplicated, as has been suggested (Michel

Delseny, personal communication). Southern-hybridization results showed that most probes detected two or more bands for most *Arabidopsis* and papaya cDNA clones used in a linkage map project (Table 1, Ming et al. unpublished). Moreover, Only 2 of the 28 BAC (7%) clones that hybridized to total papaya genomic DNA showed a light signal (Fig. 1c, d), suggesting that a large portion of the papaya genome consists of duplicated-sequence and/or repetitive elements.

If one considers the number of bands on Southern blots as the number of copies of a gene or genomic regions, then the genome coverage would be only 9.5× for this library. This is lower than the 12.7× estimated from the average insert size and the number of BAC clones. A possible explanation for this difference is the presence of *Hin*dIII sites within the genomic sequence homologous to the cDNA probes that led to the over-estimation of the gene copy number on Southern blots. Tomkins et al. (1999) suggested the presence of additional *Hin*dIII sites in regions containing single-copy sequences based on characterization of a sugarcane BAC library. This hypothesis is supported by a papaya gene-cloning project that showed such a *Hin*dIII site in the papaya sucrose synthase gene (Robert Paull, personal communication).

The average number of BAC clones hit by papaya cDNA probes was 8.1 for ligation 1 and 14.7 for ligation 2. These differences indicate that the 89% larger inserts from ligation 2 contain larger genome equivalents. The significant correlation between the number of BAC clones hit by a cDNA probe from ligations 1 and 2 suggested that the overall ligation events of the genomic fragment were random (Fig. 2). However, more clones were hit by an *Arabidopsis* probe (29.4) than by a papaya probe (19.4) from ligation 2, while the number of clones identified by homologous or heterologous probes was almost equal from ligation 1 (8.2 for an *Arabidopsis* probe and 8.0 for a papaya probe). Further investigation is needed to determine whether this result is due to sampling error or some unknown factor(s).

Based on 61 BAC clones identified by the cotton 18 S rDNA probe, it appears that about five clusters of rDNA genes are represented in this 12.7× papaya BAC library. Since the number of BAC cones identified by a single gene (per band) ranged from 3 to 24, the number of rDNA gene-clusters in papaya might range from 3 to 20. It is well-documented that rDNA genes occur in tandemly repeated segments in higher plants, and one or several clusters of rDNA genes can be found per haploid (Long and Dawid 1980; Gottlob-McHugh et al. 1990). If there are *Hin*dIII sites within the tandemly repeated segments, more BAC clones would be hit by this 18 S rDNA probe. The actual number of rDNA gene clusters is likely to be close to the lower end of the estimates.

Although we used nuclear isolation and purification to reduce the amount of chloroplast DNA, a small proportion of the BAC library (1.4%) was found to contain chloroplast DNA. All angiosperms have cpDNAs that range in size from 120 to 160 kb (Ozeki et al. 1989); this is a little larger than the average insert size of BAC clones from ligation 1, but smaller than that of BAC clones from ligation 2. It is possible that some of the cpBAC clones that hybridized to both *ropB* and *trunk* might contain the entire genome and could be useful for studying the chloroplast genome.

BAC clones identified in the present BAC library by the sex-linked DNA marker provide us with a starting point for positional cloning of the sex-determination gene. The sex-determination gene was mapped on linkage group 1 (Sondur et al. 1996) and co-segregates with the SCAR marker W11. With the population size of 139 F_2 plants, this marker is linked to the target gene within 0.35 cM. A fine mapping project is underway to further determine the map position of the marker W11 in an F_2 population with over 1000 individual plants. The contig around the sex-determination gene will be constructed from both ends of the BAC clone before the fine mapping is completed. BAC end-DNA fragments will be used to screen the BAC library for overlapping BAC clones.

Unraveling the sex-determination process could have profound application in papaya production. It is the general practice in Hawaii to plant at least five seeds per hill in order to ensure that the number of female plants in the field is less than 3%. The five plants in a hill are grown for the next 4–6 months before flowering so that the sexes can be determined. This practice is inefficient in respect of time and labor, and results in delayed production. Better understanding of the sex-determination process in papaya may lead to the development of methods to eliminate the undesired seeds, or plants of a particular sex, automatically before seeds are sown or seedlings are transplanted. This papaya BAC library with 13.7× genome equivalents is expected to provide a sufficiently large fraction of the genome to eventually clone the papaya sex-determination gene.

Acknowledgements We thank Denise Steiger for technical assistance; John Stiles for unpublished data; John Stiles and Robert Paull for providing papaya cDNA libraries; the *Arabidopsis* Biological Resource Center for providing *Arabidopsis* cDNA probes; Henrik Albert, K. K. Wu, Robert Osgood, and Stephanie Whalen for reviewing the manuscript. This work was supported by a U.S. Department of Agriculture, Agricultural Research Service Cooperative Agreement (no. CA 58–5320–3-460) with the Hawaii Agriculture Research Center.

References

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

Bent AF, Kunkel BN, Dahlbeck D, Brown K, 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

Bremer KM, Chase W, Stevens PF (1998) An ordinal classification for the families of flowering plants. Ann Missouri Bot Gard 85:531–553

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

Chen C, Gmitter FG (1999) Direct cloning and sequencing of bacterial artificial chromosome (BAC) insert ends based on double digestion. Plant Mol Biol Rep 17:231–238

Choi S, Creelman RA, Mullet JE, Wing RA (1995) Construction and haracterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*. Weeds World 2:17–20

Dellaporta, SL, Calderon-Urrea A (1993) Sex determination in flowering plants. The Plant Cell 5:1241–1251

Dunne J, Horgan L (1992) Meat tenderizers. In: Hui YH (ed) Encyclopedia of Food Science and Technology, vol 3. Wiley, New York, pp 1745–1751

Fitch MMM, Manshardt RM, Gonsalves D, Slightom JL (1992) Virus resistant papaya derived from tissue bombarded with the coat protein gene of the papaya ringspot virus. Bio/Technology 10:1466–1472

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

Gottlob-McHugh SG, Levesque M, MacKenzie, K, Olson M, Yarosh O, Johnson DA (1990) Organization of the 5 S rRNA genes in the soybean *Glycine max* (L.) Merrill and conservation of the 5 S rDNA repeat structure in higher plants. Genome 33:486–494

Hamilton GM, 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

Hofmeyr JDJ (1938) Genetical studies of *Carica papaya* L. I. The inheritance and relation of sex and certain plant characteristics. II. Sex reversal and sex forms. S Afr Dept Agri and Sci Bull No 187

Jiang J, Gill BS, Wang GL, Ronald PC, Ward DC (1995) Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci USA 92:4487–4491

Lin YR, 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

Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49:727–764

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

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald K, Hillier LW, McPherson JD, Waterson RH (1997) High-throughput fingerprint analysis of large-insert clones. Genome Res 7:1072–1084

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

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

Ockerman HW, Harnsawas S, Yetim H (1993) Inhibition of papain in meat by potato protein or ascorbic acid. J Fod Sci 58: 1265–1268

Osato JA, Santiago L, Remo G, Cuadra M, Mori A (1993) Antimicrobial and antioxidant activities of unripe papaya. Life Sci 53:1383–1389

Ozeki H, Umesono K, Inokuchi H, Kohchi T, Ohyama K (1989) The chloroplast genome of plants: a unique origin. Genome 31:169–174

Purina A, Sandhya B (1988) Genotypic differences of in vitro lateral bud establishment and shoot proliferation in papaya. Curr Sci 7:440–442

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

Sharon D, Hillel A, Vainstein A (1992) Applications of DNA fingerprints for identification and genetic analysis of *Carica papaya* and other *Carica* species. Euphytica 62:119–125

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

Sondur SN, Manshardt RM, Stiles JI (1995) Genetics of growth rate and flowering time in papaya (*Caraca papaya* L.). J Quant Trait Loci (http://probe.nalusda.gov:8000/otherdocs/jqtl/1995–04/jqtl8r2.html)

Sondur SN, Manshardt RM, Stiles JI (1996) A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers. Theor Appl Genet 93:547–553

Stiles JI, Lemme C, Sondur S, Morshidi MB, Manshardt RM (1993) Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. Theor Appl Genet 85:697–701

Storey WB (1938) Segregation of sex types in Solo papaya and their application to the selection of seed. Proc Am Soc Hort Sci 35:83–85

Storey WB (1953) Genetics of the papaya. J Hered 44:70–78

Tai T, Staskawicz BJ (2000) Construction of a yeast artificial chromosome library of pepper (*Capsicum annuum* L.) and identification of clones from the Bs2 resistance locus. Theor Appl Genet 100:112–117

Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo SS, Wing RA (1999) A bacterial artificial chromosome library for sugarcane. Theor Appl Genet 99:419–424

Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant J **7**:525–533

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

Yang ZN, Mirkov TE (2000) Isolation of large terminal sequences of BAC inserts based on double restriction enzyme digestion followed by anchored PCR. Genome 43:412–415

Yoshimura S, Umehara Y, Kurata N, Nagamura Y, Sasaki T, Minobe Y, Iwata N (1996) Identification of a YAC clone carrying the *Xa-1* allele, a bacterial blight resistance gene in rice. Theor Appl Genet 93:117–122

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

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