Genetica (2013) 141:217–226
DOI 10.1007/s10709-013-9720-y

Construction and characterization of a BAC library from the *Coffea arabica* genotype Timor Hybrid CIFC 832/2

S. M. B. Cação · N. V. Silva · D. S. Domingues · L. G. E. Vieira · L. E. C. Diniz · F. Vinecky · G. S. C. Alves · A. C. Andrade · V. Carpentieri-Pipolo · L. F. P. Pereira

Received: 4 September 2012 / Accepted: 2 May 2013 / Published online: 16 May 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Most of the world's coffee production originates from *Coffea arabica,* an allotetraploid species with low genetic diversity and for which few genomic resources are available. Genomic libraries with large DNA fragment inserts are useful tools for the study of plant genomes, including the production of physical maps, integration studies of physical and genetic maps, genome structure analysis and gene isolation by positional cloning. Here, we report the construction and characterization of a Bacterial Artificial Chromosome (BAC) library from *C. arabica* Timor Hybrid CIFC 832/2, a parental genotype for several modern coffee cultivars. The BAC library consists of 56,832 clones with an average insert size of 118 kb, which represents a dihaploid genome coverage of five to sixfold. The content of organellar DNA was estimated at 1.04 and 0.5 % for chloroplast and mitochondrial DNA, respectively. The BAC library was screened for the NADPH-dependent mannose-6-phosphate reductase gene (*CaM6PR*) with markers positioned on four linkage groups of a partial *C. arabica* genetic map. A mixed approach using PCR and membrane hybridization of BAC pools allowed for the discovery of nine BAC clones with the *CaM6PR* gene and 53 BAC clones that were anchored to the genetic map with simple sequence repeat markers. This library will be a useful tool for future studies on comparative genomics and the identification of genes and regulatory elements controlling major traits in this economically important crop species.

Keywords BAC clones · Genetic map · Linkage groups · Genetic markers · M6PR · Coffee leaf rust

Electronic supplementary material The online version of this article (doi:10.1007/s10709-013-9720-y) contains supplementary material, which is available to authorized users.

S. M. B. Cação · N. V. Silva · D. S. Domingues · L. G. E. Vieira · L. F. P. Pereira
Laboratory of Plant Biotechnology, Instituto Agronomico do Paraná, CP 481, Londrina, Paraná 86001-970, Brazil

L. E. C. Diniz
Laboratory of Molecular Biology, Embrapa Tabuleiros Costeiros, Aracaju, Sergipe, Brazil

F. Vinecky · G. S. C. Alves · A. C. Andrade
Laboratory of Molecular Genetics, Embrapa Recursos Genéticos e Biotecnologia, Brasília, Distrito Federal, Brazil

V. Carpentieri-Pipolo
Agronomy Department, CCA, Universidade Estadual de Londrina, Londrina, Paraná, Brazil

L. F. P. Pereira (✉)
Embrapa Café, Brasília, Distrito Federal, Brazil
e-mail: filipe.pereira@embrapa.br

Introduction

Coffee is one of the most important beverage crops in the world with over 10 million ha of agricultural land dedicated to its production in tropical and subtropical regions (FAO 2011). In total, 124 species populate the genus *Coffea* L. (Rubiaceae) (Davis et al. 2011). The most commercially important species are *C. arabica* L., also known as Arabica coffee, and *C. canephora* (Pierre ex Froehner), known as Robusta coffee, representing approximately 70 and 30 % of the total coffee market, respectively (International Coffee Organization, http://www.ico.org). *C. arabica* L. ($2n = 4x = 44$), the only polyploid species in the *Coffea* genus, is an allotetraploid containing two diploid sub-genomes originating from the ancestral *C. canephora* and *C. eugenioides* species (Lashermes et al. 2000). For most organisms, including plants, a large

insert genomic library is crucial for physical mapping, map-based gene cloning and the analysis of gene structure and function. The Bacterial Artificial Chromosome (BAC) vector has become a preferred tool for cloning large DNA fragments in genomic research, mainly because of its simple manipulation and the stability of its clones (Gonthier et al. 2010). In higher plants, numerous BAC libraries have been constructed and successfully used for a variety of applications including gene isolation (Liang et al. 2004), comparative genomics (Ilic et al. 2003; Wang et al. 2005; Sehgal et al. 2012), map-based or positional cloning of genes (Patocchi et al. 1999; Jander et al. 2002; Xu et al. 2007; Li et al. 2012) and genome-wide physical map construction (Shultz et al. 2006; Han et al. 2007; Lacombe et al. 2008; Na et al. 2012). For polyploid species such as *C. arabica*, BAC libraries are important for understanding the phylogeny of homeologous loci by aiding in the characterization of the genomic changes that occurred in the early stages of polyploidization (Yu et al. 2011; Cenci et al. 2012).

Noir et al. (2004) constructed and characterized a library of *C. arabica* with 88,320 BAC clones. For the selection of BAC clones, colonies on membrane filters were hybridized with RFLP probes of *C. arabica* linkage groups of the *C. canephora* genome. These probes revealed two loci in *C. arabica*, one in each of the two sub-genome diploid members (*C. canephora* and *C. eugenioides*) (Lashermes et al. 2000). This BAC library was used to study synteny between *C. arabica* and *Arabidopsis thaliana* as well as to characterize a gene cluster for coffee leaf rust disease resistance (Mahé et al. 2007; Ribas et al. 2011). Leroy et al. (2005) characterized a *C. canephora* BAC library with genetically mapped RFLP probes and identified BAC clones containing genes coding for enzymes of sucrose metabolism. However, none of these BAC libraries are currently available for use by coffee research groups.

Timor Hybrid (TH) is a hybrid between *C. arabica* and *C. canephora* with a *C. arabica* phenotype. It has been intensively used in coffee breeding programs for introgression of resistance against pests and diseases including coffee leaf rust (*Hemileia vastatrix*), coffee berry disease (*Colletotrichum kahawae*), bacterial disease (*Pseudomonas syringae* pv. *garcae*) and root-knot nematode (*Meloidogyne exigua*) (Orozco-Castillo et al. 1994; Charrier 1997). The first introduction of TH accessions in Brazil date back to 1976 via vegetative propagation and seeds from the CIFC (Coffee Rust Research Center, Portugal), IIAA (Institute of Agronomy Research, Angola) and ERU (Uige Regional Station, Angola) (Pereira et al. 2002). Exploitation of the TH germplasm has so far relied on conventional breeding. Due to its genetic compatibility with cultivars of *C. arabica*, the TH genotype was used for the production of coffee germplasm with leaf rust resistance such as Sarchimor (*C. arabica* cv. Villa Sarchi × TH CIFC 832/2) and Catimor (*C. arabica* cv. Caturra × TH CIFC 832/1), (Mendes et al. 2008). Based on this strategy, improved cultivars have already been released in several important coffee-producing countries such as Brazil, Kenya and Colombia (Lashermes et al. 1999). For example, Brazilian cultivars IAPAR 59, IPR107, IPR100 and Tupi originated from Sarchimor germplasm (Carvalho 2008).

In this work, we have constructed and characterized a BAC library from the Timor Hybrid CIFC 832/2, which is the parental genotype of the Sarchimor germplasm. To demonstrate the utility of this genomic library, we used a cDNA probe corresponding to NADPH-dependent mannose-6-phosphate reductase (*CaM6PR*), a single copy gene coding for a key enzyme in mannitol synthesis, and simple sequence repeat (SSR) markers to screen the library.

Materials and methods

Plant material

C. arabica Timor Hybrid CIFC 832/2 adult trees were grown in a greenhouse at the Agronomic Institute of Paraná (Londrina-Brazil). The plants were covered with dark plastic bags for five days before young leaves were collected for high molecular weight (HMW) DNA isolation.

BAC vector preparation

The plasmid DNA vector pCC1BAC^{TM} was extracted by alkaline lysis and purified by a cesium chloride gradient (Sambrook et al. 1989). The vector (10 μg) was digested with 7 U *Hind*III in a final volume of 75 μL at 37 °C for 1 h. The inactivation of *Hind*III was performed at 65 °C for 20 min. The plasmid was dephosphorylated using 5 U of alkaline phosphatase at 30 °C for 1 h 30 min. Alkaline phosphatase was inactivated at 65 °C for 20 min, and the dephosphorylated vector was electrophoresed through a 0.8 % agarose gel in 1X TAE buffer. Linear plasmid bands were extracted and purified from the gel and used for ligation with HMW DNA.

BAC library construction

Nuclei were isolated from leaves according to Peterson et al. (2000) with some modifications to eliminate polysaccharides and to reduce polyphenol oxidase activity in the coffee cell extracts. The leaves were ground in liquid nitrogen followed by incubation of the cell extracts at 4 °C in homogenization buffer. The leaf homogenate was filtered through nylon meshes (250 and 100 μm) to reduce cell debris followed by three centrifugations at 4 °C and 650*g* for 15 min in washing buffer with 0.2 % β-mercaptoethanol

and 2 % polyvinylpyrrolidone (PVP 40). The nuclei were resuspended in 1 mL filtered extraction buffer without β-mercaptoethanol and embedded in 1.0 % low-melting-point agarose plugs. The agarose plugs were then incubated in the lysis buffer (1 % w/v sodium lauryl sarcosinate, 0.3 mg/mL proteinase K and 0.13 % diethyl sodium dissolved in 0.5 M EDTA, pH 9.1) for 72 h, changing the lysis solution every 24 h without proteinase K. The plugs were stored in 70 % ethanol at −20 °C.

To prepare DNA for digestion, the agarose plugs containing HMW DNA were washed as described previously (Peterson et al. 2000) with two modifications. Ascorbic acid was not added, and PVP-40 was lowered from 2 to 0.25 % at the end of the washing step. The plugs were dialyzed for 2 h in 0.5× TBE and subjected to pre-electrophoresis to remove potential inhibitors (Osoegawa et al. 1998) on a 1 % Agarose Pulse Field (USB-Affymetrix, Cleveland, OH, US) gel in a GN Controller-PFGE—Pharmacia Biotech (GE Life Sciences, Little Chalfont, UK) with the following conditions: pulse ramping from 1 to 4 s, 120° angle, 6 V/cm current and 8 h run time at 10 °C. The plugs were removed from the gel, washed three times in TE (10 mM Tris, 10 mM EDTA) for 1 h each, chopped into small pieces and partially digested with *Hind*III (Peterson et al. 2000). DNA was separated by PFGE in a 1 % agarose gel at pulse times ramping from 1 to 45 s with 120° angles and 6 V/cm at 10 °C in 0.5× TBE for 10 h, followed by a second run performed for 8 h at 5 V/cm with a pulse time ramp of 4–5 s and an angle of 120°. The flanking marker lanes were removed from the gel and stained with ethidium bromide to indicate the size ranges. Fragments from 100 to 200 Kb were selected, excised from the gel and subjected to a second size selection on a low melting point 1 % agarose gel (Invitrogen®) at 10 °C in 0.5× TBE at 5 V/cm with ramping times from 3 to 10 s for 24 h with 120° angles. The fragments were excised from the gel and electroeluted in 1X TAE buffer (40 mM Tris–acetate, 1 mM EDTA, pH 8.0) using dialysis tubes pretreated for 10 min at 90 °C in a solution of 1 mM EDTA and 2 % $NaHCO_3$ followed by several washes with distilled water in individual dialysis tubes (Strong et al. 1997).

After the electroelution of DNA fragments, ligations were performed using 25 ng of the vector, 100 ng of DNA fragments, 6 U of T4 DNA ligase (high concentration; Invitrogen, Carlsbad, CA) and 10 μL of 5X reaction buffer and incubated at 16 °C overnight. The ligation reactions were desalinated for 1 h at 4 °C in 1.5 mL tubes containing 1.8 % glucose and 1.0 % ultra-pure agarose (Invitrogen, Carlsbad, CA). Twelve microliters of the ligations were added to 50 μL of DH10B competent cells, placed in cold cuvettes (0.1 cm electrode gap) and electroporated at 1.6 kV using an Electroporator 2510™ apparatus (Eppendorf, Hamburg, Germany) and placed in microtubes with 1 mL of SOC medium (Sambrook et al. 1989). The transformed cells were incubated with agitation at 37 °C for 90 min and plated on LB medium with 12.5 μg/mL chloramphenicol, 0.40 mM IPTG (isopropylthio-β-D-galactosidase) and 80 μg/mL of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase). Plates were incubated overnight in the dark at 37 °C, and white recombinant colonies were hand-picked in 96-well plates and subsequently arrayed in 384-well plates containing freezing media (Peterson et al. 2000) using a Genetix Q-Bot (Genetix Ltd., Christchurch, Dorset, UK). The plates were incubated for 16 h at 37 °C, duplicated and stored at −80 °C. The BAC library was replicated twice using the Genetix Q-Bot apparatus and stored at −80 °C.

Estimation of insert sizes of BAC clones

One hundred twenty-two BAC clones were randomly selected and grown for 24 h at 37 °C in 1.5 mL LB medium containing chloramphenicol (12.5 mg/mL). The BAC DNA was extracted using an alkaline lysis procedure (Sambrook et al. 1989) and digested with *Not*I. The products of digestion were separated by PFGE in a 1 % agarose gel in 0.5× TBE. The size of the inserts was estimated using PFG Marker I (New England Bio Labs).

Hybridization of membranes with chloroplast and mitochondrial probes

High density filters were made using the Genetix Q-Bot with 18,432 double spotted clones per membrane to avoid false positives. The filters were incubated at 37 °C on Q-trays containing LB agar with 12.5 mg/mL chloramphenicol. Colony filters were processed as follows: denaturation on Whatman paper soaked in a solution of 0.5 M NaOH and 1.5 M NaCl for 4 min at 100 °C; neutralization on Whatman paper soaked in 1 M Tris–HCl (pH 7.4) and 1.5 M NaCl for 4 min at 100 °C; and incubation in a solution of 0.25 mg/mL proteinase K for 50 min at 37 °C and baking for 3 h at 80 °C. To screen for organelle DNA in the library, fragments of the mitochondrial *cox1* (cytochrome oxidase subunit 1) and chloroplast *rbcL* (large subunit of the ribulose-1, 5-bisphosphate carboxylase-oxygenase) genes were used as probes. The probe fragments were generated by PCR amplification of total DNA using the primers *cox*-1 forward, 5′-AGC AGC TTT TCC AGA CGA CC-3′ and reverse, 5′-CGG TTG GTA TTG GAA CTG-3′; and *rbcl* forward, 5′ ATG TCA CCA CAA ACA GAA AC 3′ and reverse 5′TCC TTT TAG TAA AAG ATT GGG CCG AG 3′. The amplification conditions were 5 min at 94 °C for one cycle, followed by 30 cycles of 94 °C for 45 s, 60 °C for 20 s and 72 °C for 2 min and a final extension at 72 °C for 10 min. The fragments generated from the two reactions were gel-purified with the QIAquick gel extraction kit (Qiagen, Hilden, Germany)

and labeled with [α-P^{32}]-dCTP. The hybridization was performed overnight at 42 °C using UltraHyb^{R} hybridization buffer (Ambion, Austin, TX) according to the procedure by Sambrook et al. (1989). Hybridization membranes were sequentially washed twice in 2× SSC and 0.1 % SDS for 10 min at 42 °C and twice in a solution of 0.1× SSC and 0.1 % SDS for 10 min at 42 °C. The membranes were exposed to the image plate BAS-IP MS 2340, and the images were captured using the fluorescent image analyzer FLA 3000 series (Fuji Photo Film Co, Ltd., Tokyo, Japan). Positive clones were identified by a duplicate signal pattern.

Library pooling

The BAC library was clustered into plate pools with 384 clones each. BACs were grown individually in plates with 200 µL enriched LB medium [1 % Bacto-tryptone, 0.5 % yeast extract, 10 mM NaCl, 13 mM KH_2PO_4, 36 mM K_2HPO_4, 1.7 mM sodium citrate, 6.8 mM $(NH_4)_2SO_4$ and 0.4 mM $MgSO_4$] in selective medium with 12.5 mg/mL chloramphenicol. DNA isolation from BAC pools was performed by alkaline lysis (Sambrook et al. 1989). For the formation of superpools, DNA from the plate pools was grouped with equal concentrations of DNA (5 ng/µL) from 15 plates to form 10 superpools with approximately 5,760 clones each.

Selection of BAC clones by PCR and hybridization membranes

The BAC pools and superpools were screened by PCR. The amplifications were performed using 50 ng DNA from the pools and superpools. To amplify the *CaM6PR* gene, specific primers were designed (*CaM6PR* forward 5′-AGA AGC ACG GCA TCT GTGTCACTG-3′ and reverse 5′-TCC ATG TCC TCT TCG CTG AGA TCG-3′) using Primer Designer 2.0. PCR was performed under the following conditions: initial denaturation at 94 °C for 2 min, followed by 40 cycles of 94 °C for 1 min, 64 °C for 45 s and 72 °C for 45 s and a final extension step of 72 °C for 3 min. BAC clones were also screened with the SSR primers **SSR-18** (LG1), **SSR-16** (LG2), **ACGG-1** (LG3) and **CCG-3** (LG6) obtained from enriched genomic libraries of the *C. arabica* Bourbon amarelo UFV 570 strain (Missio et al. 2009). The amplification conditions were as follows: denaturation at 94 °C for 2 min; 30 cycles of 94 °C for 30 s, 57 °C/66 °C for 30 s, 57 °C for 30 s and 72 °C for 30 s; and a final extension at 72 °C for 8 min. The PCR products were separated by electrophoresis in a 10 % polyacrylamide gel.

The plate pools selected for the *CaM6PR* and SSR markers were used for filter colony hybridization. The clones were inoculated onto positively charged nylon membranes on top of plates containing LB and incubated at 37 °C for 16 h. After the growth of clones on the membranes, they were subject to denaturation and neutralization followed by a treatment with proteinase K and fixed on the membrane at 80 °C for 3 h. Probes from *CaM6PR* and SSR markers were labeled with [α-P^{32}]-dCTP by random priming. The membrane hybridization, washing and image capture procedures are described above. The BAC clones selected by hybridization were confirmed by PCR and Southern blot (*CaM6PR* BAC clones).

Restriction enzyme digestion of BAC clones (BAC fingerprinting)

BAC clones in the four linkage groups deemed positive by PCR and hybridization were subjected to fingerprinting to observe different restriction patterns between BAC clones. Agarose gel-based BAC DNA fingerprinting was performed as described by Marra et al. (1997) with some modifications. Five µg of DNA and 8 U/µL of *Hind*III were used for restriction digestion. DNA fragments were separated in a 1 % agarose gel by electrophoresis at 80 V in 1X TAE buffer at 4 °C for 16 h. The gel was stained in 1 L of a 1:20,000 dilution of Vistra Green (Molecular Probe, USA) in 1× TAE for 30 min. Gel images were captured using the Kodak Digital Science Electrophoresis Documentation and Analysis System 120.

Southern blot analysis

Five micrograms of BAC clone plasmid DNA and 15 µg of genomic DNA from *C. canephora* and *C. eugenioides* were digested with *Hind*III. DNA was separated on a 0.7 % agarose gel and transferred to a nylon membrane by capillarity and subsequently hybridized with the 450-bp *CaM6PR* genomic probe radiolabelled with [α-P^{32}]-dCTP by random priming. The Southern membrane hybridization, washing and image capture procedures were performed as described above.

Results

BAC library construction and characterization

A BAC library of Timor Hybrid CIFC 832/2 was constructed by partial digestion of total genomic DNA with *Hind*III and linked in a pCC1BAC vector. The library contains 56,832 clones in 148 384-well microtiter plates.

To examine the quality of the BAC library, 200 BAC clones were randomly selected, digested with *Not*I to release the inserts from the pCC1BAC vector and analyzed

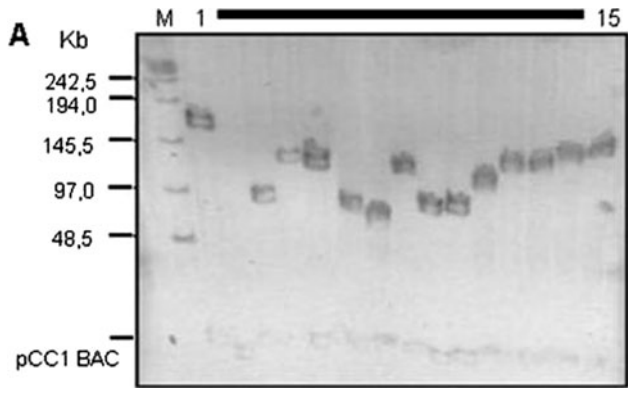

Fig. 1 DNA analysis of 15 BAC clones from the TH CIFC 832/2 BAC library by pulse-field gel electrophoresis. DNA samples digested with *Not*I were separated on a 1 % agarose gel in 0.5x TBE buffer for 18 h under the following conditions: 5–15 s ramp pulse time at 6 V/cm at 10 °C. The shared 8 Kb band represents the pCC1 BAC Vector

using PFGE electrophoresis (Fig. 1). The estimated average insert size was 118 Kb, ranging from 45 to 250 Kb, with 74 % of the inserts larger than 100 Kb and 41 % above 120 Kb (Fig. 2). Considering the *C. arabica* genome size (1.3 pg or 1,254 Mb) (Arumuganathan and Earle 1991), the number of clones and the average insert size, we estimate that the genome coverage of the library ranged from 5 to 6 genome equivalents (56,832 × 0.118 Mb = 6,706).

Nuclei preparation by differential centrifugation and washes does not fully eliminate the chloroplasts and mitochondria present in the initial leaf tissue cell extract. Thus, it is expected that organellar DNA will be present in some library clones. To evaluate the degree of chloroplast and mitochondrial DNA contamination in the BAC library, we hybridized high-density membranes carrying 56,832 colonies with probes derived from the large subunit of ribulose-1, 5-bisphosphate carboxylase-oxygenase (chloroplast *rbcL*) and cytochrome oxidase (mitochondrial *cox1*). Based on the percentage of positive clones identified, the contamination was estimated to be 1.04 % chloroplast DNA and 0.5 % mitochondrial DNA (Table 1).

BAC library screening

To select BAC clones in superpools and pools, a mixed approach using PCR and membrane hybridization was applied. This methodology allowed the identification of BAC clones with genetically mapped SSR markers and the *CaM6PR* gene, which is related to abiotic and biotic stress tolerance (Zhifang and Loescher 2003).

Ten superpools encompassing 5,760 BAC clones each were screened for *CaM6PR* by PCR and yielded five positive pools (Fig. 3a). The superpools were subdivided into primary pools in order to identify the plate pools containing the clones of interest (Fig. 3b–f). To optimize the work, each plate pool identified by PCR was subjected to colony hybridization. In this manner, it was possible to identify eleven BAC clones that were positive for the *CaM6PR* gene (Fig. 4). The clones found by colony hybridization were subjected to PCR to test for the occurrence of false positives. From the BAC clones identified by hybridization, ten were confirmed positive for the *CaM6PR* gene (Table 1; Fig. 5). The screen for SSR markers identified eight positive superpools for SSR-16, six for SSR-18 and five each for ACGG-1 and CCG-3 (data not shown). Nine colony membranes were hybridized with the SSR marker probes, and a total of 44 BAC clones were selected as follows: eleven each for SSR-16 and CCG-3, twelve for SSR-18 and ten for ACCG-1 (Table 1). The BACs were selected for *Hind*III restriction analysis to examine their restriction profile, duplications of the clones or similar band patterns. Nine different clones for SSR-16, ten each for SSR-18 and CGC-3 and seven for SSR marker CCG-3 were found.

Southern blot of BAC clones and coffee plants

Further analysis of the selected *CaM6PR* BAC clones was performed by Southern blot hybridization. Of the ten positive clones selected by PCR and membrane hybridization from the BAC pools, seven clones showed an

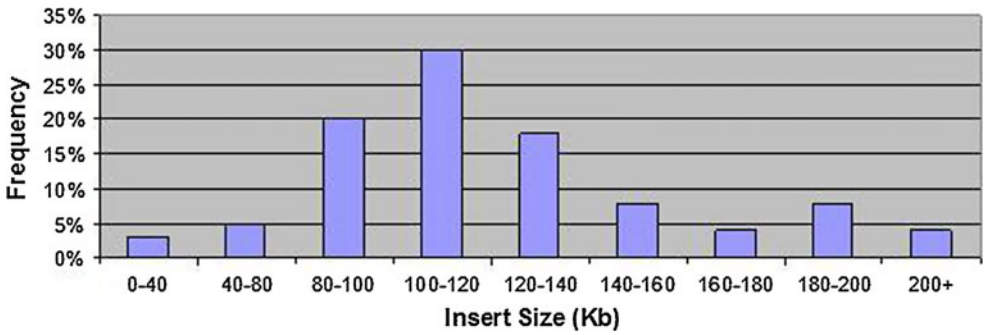

Fig. 2 Insert size distribution of 200 BAC clones in the TH CIFC 832/2 library. DNA samples of 200 randomly selected clones were analyzed and grouped according to their size. The results indicate an average insert size of 118 Kb

Table 1 Results of membrane hybridization of plates containing 384 clones using seven different probes

Probe	Linkage groups (LG)	Number of hits	BAC clone address[a]
CaM6PR	–	10	P60M16, P89H2, P116M2, P123G3, P123K9, P123L9, P133O9, P133P9, P138N1
SSR-16	1	11	P5A18, P5E1, P5O17, P56 L12, P75F10, P75G18, P85J12, P101N11, P101O2, P105I22, P105L21
SSR-18	2	12	P4D17, P4H18, P56F19, P56J23, P67F7, P67K3, P67K5, P75J19, P75L16, P101L19, P101D15, P126K18, P126N5
ACGG-1	3	10	P4K14, P4M11, P4M20, P56A2, P56B15, P56I23, P75A11, P75E11, P101E21, P101H15
CCG-3	6	11	P56E20, P56F16, P67G8, P67J19, P75P16, P75A15, P85C13, P85E4, P85F11, P126H3, P126K1
Mitochondrial[b]	–	284 (0.5 %)	NI
Chloroplast	–	591 (1.04 %)	NI

After selection by PCR the plates were selected from superpools and pools

[a] Addresses localize the clone by plate number, column (letter) and line (number) NI: Not Identified

[b] For Mitochondrial and chloroplast hybridization, membranes with 18,432 double spotted clones were used

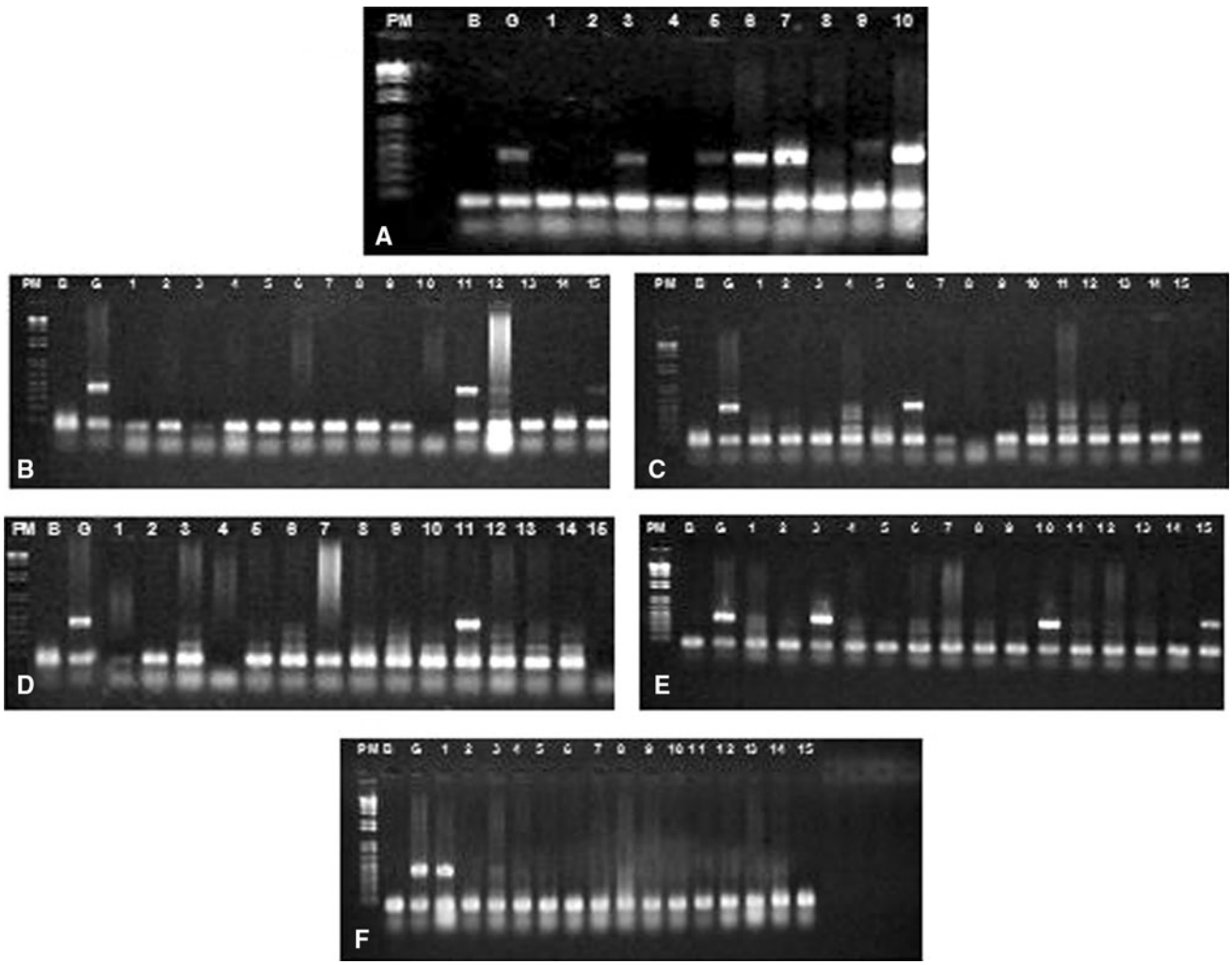

Fig. 3 PCR screening of the superpools and pools of the TH CIFC 832/2 BAC library. **a** First round of PCR screening in 10 superpools representing the entire library or 148 plates (384-well). *Lanes 1–10*: superpools; PM: 1 Kb plus DNA ladder; G: genomic DNA (positive control); and B: blank (negative control). Each superpool contains DNA from 15 plates or 5,760 individual BAC clones. The *CaM6PR* gene was detected by PCR in the following six superpools: 3 (P60), 5 (P89), 6 (P116), 7 (P123), 9 (P133) and 10 (P138). **b–f** Second round of PCR screening of plate pools. Ex: In B superpool 3, only one plate showed a signal amplified by the gene *CaM6PR*

intense signal with the *CaM6PR* probe. Hybridization with the *CaM6PR* gene was also observed in *C. arabica* TH CIFC 832/2, *C. canephora* and *C. eugenioides* genomic DNA digested with *Hind*III (Fig. 6). Four hybridization bands in the TH CIFC 832/2 genomic DNA are likely related to the corresponding bands observed in *C. canephora* and *C. eugenioides*. An extra band was observed in *C. eugenioides,* possibly due to differences in its enzyme restriction profile or hybridization with another *M6PR* isoform. The BAC clones 1 through 6 and 8 exhibited a fragment that is related to the same one present in the TH CIFC 832/2 genome and with a similar band size to that present in *C. eugenioides* (Fig. 6).

Discussion

BAC libraries are important tools for genomic and mapping studies. We constructed a BAC library of the *C. arabica* TH CIFC 832/2 strain containing 56,832 clones with an average insert size of 118 Kb, indicating that this library represents from 5 to 6 dihaploid genome equivalents for this species. The protocol resulted in a library with 74 % of the clones containing inserts larger than 100 Kb and 41 % greater than 120 Kb. A large-insert library is important because the number of clones required for adequate coverage of the genome is reduced, and it consequently simplifies the screening and generation of BAC contigs as demonstrated by the BAC screening in this study. Larger inserts can span larger gaps, covering more space with less overlap and reducing the costs of sequencing as reported for several BAC libraries such as in tomato (Budiman et al. 2000), common bean (Liu et al. 2010) and papaya (Gschwend et al. 2011).

Analysis of BAC clones through digestion with *Not*I followed by electrophoresis demonstrated that most clones contained one fragment. Maughan et al. (2008) found that 14 % of clones contain one or two inserts in *Amaranthus*, which is consistent with the low percentage of *Not*I recognition sites (GCGGCCGC) found in BAC libraries of other dicotyledonous species (Tomkins et al. 1999; Hong et al. 2004).

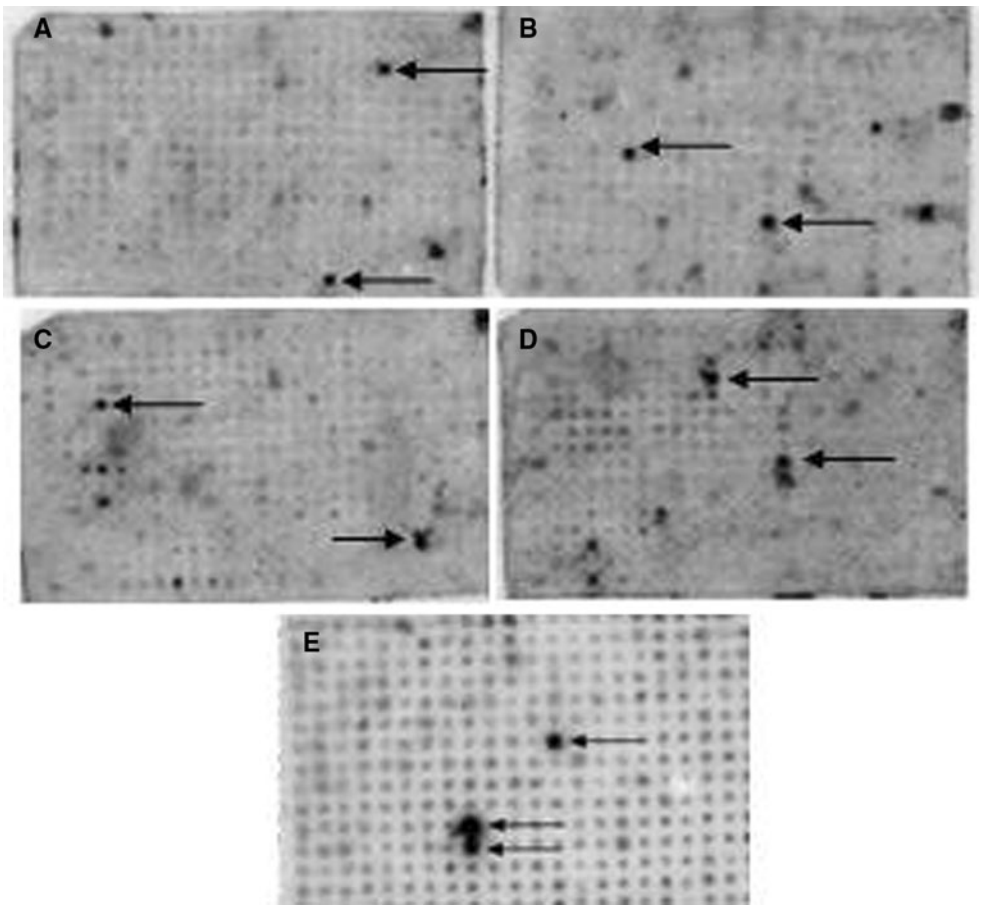

Fig. 4 Colony hybridization of BAC membranes. Clones were spotted on a nylon membrane and hybridized with the single-copy probe *CaM6PR* (**a**) and with the SSR markers SSR-16 (**b**), SSR-18 (**c**), ACGG-1 (**d**) and CCG-3 (**e**). *Arrows* indicate positive clones

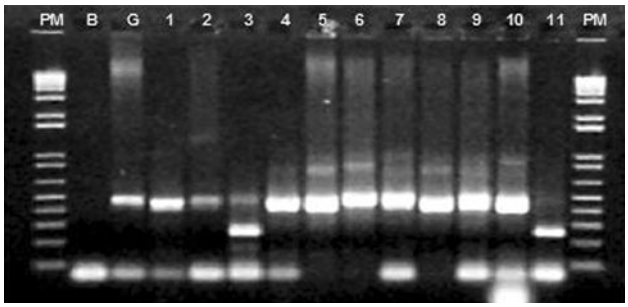

Fig. 5 Amplification of the *CaM6PR* gene in BAC clones identified by PCR and colony hybridization. Ten BAC clones were identified for the gene of interest. PM: 1 Kb plus DNA ladder; G: genomic DNA (positive control); and B: blank (negative control)

Mitochondrial DNA contamination was low in this library when compared to the results obtained in rice and soybean with 0.8 and 1.85 % contamination, respectively (Wang et al. 2005; Tomkins et al. 1999). In comparison to the libraries of *C. arabica* cv. IAPAR-59 (Noir et al. 2004) and *C. canephora* (Leroy et al. 2005), our library showed a lower presence of chloroplast and mitochondrial DNA sequences. This finding confirms that the extraction procedure effectively purified nuclei and that the library contains predominantly nuclear sequences.

The strategy of screening a library from pools representing groups of clones normally requires a reduced number of PCR amplifications to find a single clone (Gardiner et al. 2004). The *M6PR* gene is present as a single copy in many plant genomes. BLAST searches in the Phytozome database (www.phytozome.net) have shown M6PR to be a single-copy gene in tomato, potato, grapevine, *Populus trichocarpa* and *Medicago truncatula*, which are species used extensively for comparative sequence analysis with coffee (data not shown) (Guyot et al. 2009; Guyot et al. 2012). Experimental evidence also confirmed this fact in *Orobanche ramosa* (Delavault et al. 2002). Thus, we used *CaM6PR* to validate our strategy combining PCR and colony hybridization. The two rounds of PCR screening from the superpool (10 PCRs) and the BAC pools (150 PCRs) and the colony hybridization techniques were complementary and very effective in screening BAC clones for both the *CaM6PR* gene and SSR markers. The combined use of these techniques decreased the work to identify BAC clones and reduced the possibility of finding false positives. The Southern blot result for *CaM6PR* indicates that this gene is present at a single or low copy number. This result, together with the number of BACs identified for *CaM6PR,* not only demonstrates the quality of our library but also establishes the feasibility of using a PCR and hybridization selection approach to identify single or low copy number genes. Using this library and a similar BAC pooling and PCR screening approach, two BAC clones containing a disease resistance gene have been identified (Alvarenga et al. 2011). Yim et al. (2007) described an approach combining a BAC pooling strategy with PCR-based primer screening that may assist in physical map construction and may provide anchor points in genetic maps. This methodology allows identification of overlapping BAC clones while simultaneously establishing links between the BAC contigs and the genetic map. The advantages of screening BAC pools with PCR-based primers include a low rate of false positives, low cost and increased throughput compared to conventional hybridization techniques.

In *C. arabica*, the production of linkage maps is difficult because this allotetraploid self-pollinated species exhibits limited genetic diversity and a low frequency of polymorphisms (Vidal et al. 2010). Screening the BAC library with SSR markers indicated that this BAC library system can be useful to generate genetic and high-resolution physical maps. The identification of BAC clones associated with linkage group markers combined with the development of fingerprinting methods, sequencing BAC-ends, contig formation or full BAC sequencing will contribute to filling gaps in physical and genetic maps. Using AFLP and SCAR markers, Diola et al. (2011) reported a high density map for a coffee leaf rust resistance gene. The SCAR markers were used to select 24 BAC clones from this library and were mapped to a 4.5 cM region surrounding the resistance gene (Diola et al. 2009). Sequencing and characterization of the BAC clones are in progress.

In conclusion, a BAC library of the TH CIFC 832/2 with more than 5× coverage of the dihaploid genome has been constructed and characterized. Screening the library in

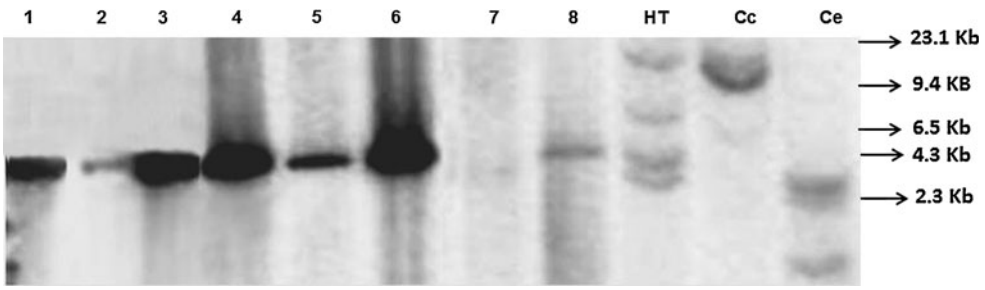

Fig. 6 Southern hybridization of BAC clones using an *CaM6PR* probe. BAC and genomic DNA were digested with *Hind*III. Legend: BAC clones (1–8), genomic DNA from Timor Hybrid accession CIFC 832/2 (TH), *C. canephora* (*Cc*) and *C. eugenioides* (*Ce*)

BAC pools with PCR and plate colony hybridization allowed a reliable and faster BAC clone screen. The BAC library has been replicated, distributed and successfully used by other coffee research groups in Brazil. Due to the significance of the TH CIFC 832/2 for breeding purposes, the BAC library described here will be an important resource in further research toward physical mapping, integration of physical and linkage maps and positional cloning for *C. arabica* genes.

Acknowledgments This project was supported by the Brazilian Consortium for Coffee Research, EMBRAPA MP2 and National Council for Scientific and Technological Development (CNPq). S.M.B. Cação received a fellowship from CNPq-INCT Café. L.F.P. Pereira and L.G.E. Vieira received CNPq research fellowships. The authors also acknowledge Dr. Eveline Caixeta (Embrapa Café), who provided SSR markers used in this study.

References

Alvarenga SM, Pinto GB, Caixeta ET, Diola V, Zambolim EM, Zambolim L, Sakiyama NS (2011) Identificação de cromossomo artificial de bactéria contendo gene de resistência a doenças em cafeeiro. Anais VII Simpósio de Pesquisa dos Cafés do Brasil

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

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

Carvalho CHS (2008) Cultivares de Café: origens, características e recomendações. Embrapa Café, Brasília, p 334

Cenci A, Combes MC, Lashermes P (2012) Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. Plant Mol Biol 78:135–145

Charrier A (1997) A coffee In: Charrier A, Jacquot S, Hamon D, Nicolas C (eds) Cirad, Repères Tropical Plant Breeding 128–152

Davis AP, Tosh J, Ruch N, Fay MF (2011) Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. Bot J Linn Soc 167:357–377

Delavault P, Simier P, Thoiron S, Veronesi C, Fer A, Thalouarn P (2002) Isolation of mannose 6-phosphate reductase cDNA, changes in enzyme activity and mannitol content in broomrape (*Orobanche ramosa*) parasitic on tomato roots. Physiol Plant 115:48–55

Diola V, Brito GG, Caixeta ET, Maciel-Zambolim E, Sakiyama NS, Loureiro ME (2011) High-density genetic mapping for coffee leaf rust resistance. Tree Genet Genomes 7:1199–1208

Diola V, Caixeta ET, Zambolin, EM, Sakiyama NS, Pereira LFP, Loureiro ME (2009) Construção do mapa genético e físico para um gene de resistência a *Hemileia vastatrix* em café. Anais XII Congresso Brasileiro de Fisiologia Vegetal

FAO (2011) FAOSTAT. Land-use statistics. Food and Agricultural Organization, Rome

Gardiner J, Schroeder SG, Polacco M, Sanchez-Villeda H, Fang Z, Morgante M, Landewe T, Fengler K, Useche F, Hanafey M, Tingey S, Chou H, Soderlund C, Wing R, Coe E (2004) Anchoring 9.371 maize expressed sequenced tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol 134:1317–1326

Gonthier L, Bellec A, Blassiau C, Prat E, Helmstetter N, Rambaud C, Huss B, Hendriks T, Bergès H, Quillet MC (2010) Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). BMC Res 3:225

Gschwend AR, Yu Q, Moore P, Saski C, Chen C, Wang J, Na J-K, Ming R (2011) Construction of papaya male and female BAC libraries and application in physical mapping of the sex chromosomes. J Biomed Biotechnol 2011:1–7

Guyot R, De La Mare M, Viader V, Hamon P, Coriton O, Bustamante-Porras J, Poncet V, Campa C, Hamon S, De Kochko A (2009) Microcollinearity in an ethylene receptor coding gene region of the *Coffea canephora* genome is extensively conserved with *Vitis vinifera* and other distant dicotyledonous sequenced genomes. BMC Plant Biol 9:1–15

Guyot R, Lefebvre-Pautigny F, Rigoreau M, Hamon P, Leroy T, Hamon S, Poncet V, Crouzillat D, de Kochko A (2012) Ancestral synteny shared between distantly-related plant species from theasterid (*Coffea canephora* and *Solanum* sp.) and Rosid (*Vitis vinifera*) Clades. BMC Genomics 13:1–14

Han Y, Gasic K, Marron B, Beever JE, Korban SS (2007) A BAC-based physical map of the apple genome. Genomics 89:630–637

Hong CP, Lee SJ, Park JY, Plaha P, Park YS, Lee YK, Choi JE, Kim KY, Lee JH, Jin JH, Choi SR, Lim YP (2004) Construction of a BAC library of Korean ginseng and initial analysis of BAC-end sequences. Mol Genet Genomics 271:709–716

Ilic K, Sanmiguel PJ, Bennetzen JL (2003) A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. Nucleic Acids Res 100:12265–12270

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Robert LL (2002) Arabidopsis map-based cloning in the post-genome era. Plant Physiol 129:440–450

Lacombe E, Cossegal M, Mirouze M, Adam T, Varouquaux F, Laoubet S (2008) Construction and characterization of a BAC library from *Arabidopsis halleri*: evaluation of physical mapping based on conserved synteny with *Arabidopsis thaliana*. Plant Sci 174:634–640

Lashermes P, Combes MC, Robert J, Trouslot P, D'hont A, Anthony F, Charrier A (1999) Molecular characterization and origin of the *Coffea arabica* L. genome. Mol Gen Genet 2:259–266

Lashermes P, Paczek V, Trouslot P, Combes MC, Couturon EE, Charrier A (2000) Single-*locus* inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C arabica x C. canephora*. J Hered 91:81–85

Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, Glaszmann JC, Vieira LGE, Piffanelli P (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. Theor Appl Genet 6:1032–1041

Li J, Hong D, He J, Ma L, Wan L, Liu P, Yang G (2012) Map-based cloning of a recessive genic male sterility locus *in Brassica napus* L. and development of its functional marker. Theor Appl Genet 125:223–234

Liang C, Xi Y, Shu J, Li J, Yang J, Che K (2004) Construction of a BAC library of *Physcomitrella patens* and isolation of LEA gene. Plant Sci 167:491–498

Liu SY, Yu K, Huffner MS, Park J, Banik M, Pauls KP, Crosby W (2010) Construction of a BAC library and a physical map of a major QTL for CBB resistance of common bean (*Phaseolus vulgaris* L.). Genomics 138:709–716

Mahé L, Combes MC, Lashermes P (2007) Comparison between a coffee single copy chromosomal region and Arabidopsis duplicated counterparts evidenced high level synteny between the

coffee genome and the ancestral Arabidopsis genome. Plant Mol Biol 64:699–711

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

Maughan PJ, Sisneros N, Luo M, Kudrna D, Ammiraju JSS, Wing RA (2008) Construction of an *Amaranthus hypochondriacus* bacterial artificial chromosome library and genomic sequencing of herbicide target genes. Plant Genome 1:85–94

Mendes ANG, Carvalho GR, Botelho CE, Fazuoli LC, Sivarolla MB (2008) Mutante de *C. arábica* L. E cultivares de importância histórica. In: Ed Carvalho CHS Cultivares de café: origens características e recomendações. Embrapa, Brasília

Missio RF, Caixeta ET, Zambolim EM, Zambolim L, Sakiyama NS (2009) Development and validation of SSR markers for *Coffea arabica* L. Crop Breed Appl Biotechnol 9:361–371

Na J-K, Wang J, Murray JE, Gschwend AR, Zhang W, Yu Q, Pérez RN, Feltus FA, Chen C, Kubat Z, Moore PH, Jiang J, Paterson AH, Ming R (2012) Construction of physical maps for the sex-specific regions of papaya sex chromosomes. BMC Genomics 13:176

Noir S, Patheyron S, Combes MC, Lashermes P, Chalhoub B (2004) Construction and characterization of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). Theor Appl Genet 1:225–230

Orozco-Castillo C, Chalmers KJ, Waugh R, Powell W (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theor Appl Genet 87:934–940

Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y, Jong PJ (1998) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res 10:116–128

Patocchi A, Vinatzer BA, Gianfranceschi L, Tartarini S, Zhang HB, Sansavini S, Gessler C (1999) Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene. Mol Genome Genet 262:884–891

Pereira AA, Moura WM, Zambolim L, Sakiyama NS, Chaves GM (2002) Melhoramento genético do cafeeiro no estado de Minas Gerais: cultivares lançados em fase de obtenção. In: Zambolim L (ed) O estado da arte de tecnologias na produção de café. UFV, Viçosa, pp 253–295

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

Ribas A, Cenci A, Combes M-C, Etienne H, Lashermes P (2011) Organization and molecular evolution of a disease-resistance gene cluster in coffee trees. BMC Genomics 12:240

Sambrook J, Fritsch EFE, Maniatis T (1989) Molecular cloning - a laboratory manual. Cold Spring Harbour Laboratory Press, New-York

Sehgal SK, Li W, Rabinowicz PD, Chan A, Simková H, Dolezel J, Gill BS (2012) Chromosome arm-specific BAC end sequences permit comparative analysis of homoeologous chromosomes and genomes of polyploid wheat. BMC Plant Biol 12:64

Shultz JL, Kurunam D, Shopinski K, Iqbal MJ, Kazi S, Zobrist K (2006) The Soybean Genome Database (Soy GD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max*. Nucleic Acids Res 34:758–765

Strong SJ, Ohta O, Litman GW, Amemiya CT (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

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

Vidal RO, Mondego JMC, Pot D, Ambrósio AB, Andrade AC, Pereira LFP, Colombo CA, Vieira LGE, Carazzolle MF, Pereira GAG (2010) A high-throughput data mining of single nucleotide polymorphisms in Coffea species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica* L. Plant Physiol 154:1053–1066

Wang W, Tanurdzic M, Luo M, Sisneros N, Kim HR, Weng J (2005) Construction of a bacterial artificial chromosome library from the spike moss *Selaginella moellendorffii*: a new resource for plant comparative genomics. BMC Plant Biol 5:10

Xu P, Wang S, Liu L, Thorsen J, Kucuktas H, Liu Z (2007) A BAC based physical map of the channel catfish genome. Genomics 90:380–388

Yim Y-S, Moak P, Sanchez-Villeda H, Musket TA, Close P, Klein PE, Mullet JE (2007) A BAC pooling strategy combined with PCR-based screenings in a large, highly repetitive genome enables integration of the maize genetic and physical maps. BMC Genomics 8:47

Yu Q, Guyot R, Kochko A, De Byers A, Navajas-Pérez R, Langston BJ, Dubreuil-Tranchant CD, Paterson AH, Poncet V, Nagai C, Ming R (2011) Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (C*offea*). Plant J 67:305–317

Zhifang G, Loescher WH (2003) Expression of a celery mannose 6-phosphate reductase in *Arabidopsis thaliana* enhances salt tolerance and induces biosynthesis of both mannitol and a glucosyl-mannitol dimer. Plant Cell Environ 26:183–337