Theor Appl Genet (2005) 111: 1032–1041
DOI 10.1007/s00122-005-0018-z

ORIGINAL PAPER

T. Leroy · P. Marraccini · M. Dufour · C. Montagnon
P. Lashermes · X. Sabau · L. P. Ferreira · I. Jourdan
D. Pot · A. C. Andrade · J. C. Glaszmann
L. G. E. Vieira · P. Piffanelli

Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes

Received: 3 March 2005 / Accepted: 14 June 2005 / Published online: 17 August 2005

Abstract The first bacterial artificial chromosome (BAC) library of Robusta coffee (*Coffea canephora*) was constructed, with the aim of developing molecular resources to study the genome structure and evolution of this perennial crop. Clone 126, which is highly productive and confers good technological and organoleptic qualities of beverage, was chosen for development of this library. The BAC library contains 55,296 clones, with an average insert size of 135 Kb per plasmid, therefore representing theoretically nine haploid genome equivalents of *C. canephora*. Its validation was achieved with a set of 13 genetically anchored single-copy and 4 duplicated RFLP probes and yielded on average 9 BAC clones per probe. Screening of this BAC library was also carried out with partial cDNA probes coding for enzymes of sugar metabolism like invertases and sucrose synthase, with the aim of characterizing the organization and promoter structure of this important class of genes. It was shown that genes for both cell wall and vacuolar forms of invertases were probably unique in the Robusta genome whereas sucrose synthase was encoded by at least two genes. One of them (*CcSUS1*) was cloned and sequenced, showing that our BAC library is a valuable tool to rapidly identify genes of agronomic interest or linked to cup quality in *C. canephora*.

Communicated by H. Nybom

T. Leroy (✉) · P. Marraccini · M. Dufour · C. Montagnon
X. Sabau · I. Jourdan · D. Pot · J. C. Glaszmann · P. Piffanelli
Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), TA 80/03, Avenue d'Agropolis, 34398 Montpellier Cedex 5, France
E-mail: thierry.leroy@cirad.fr
Tel.: +33-4-67615690
Fax: +33-4-67615793

L. P. Ferreira · L. G. E. Vieira · P. Marraccini
Instituto Agronômico do Paraná (IAPAR), LBI-AMG, CP 481, 86001-970 Londrina (PR), Brazil

P. Lashermes
Institut de Recherche pour le Développement (IRD), 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France

A. C. Andrade
Parque Estação Biológica, Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) -Cenargen, CP 02372, 70770-900 Brasilia (DF), Brazil

Introduction

Coffee is an extremely important crop with more than 7 million tons of green beans produced every year on about 11 million hectares worldwide. In terms of economic importance on the international markets, it is second only to oil and contributes to more than US $9 billions. Two species are mainly cultivated all over the tropical world: Arabica produced from *Coffea arabica* ($2n = 4x = 44$), which grows in highlands, and Robusta produced from *Coffea canephora* ($2n = 22$), which represents 34% of the global coffee production and grows in lowlands. Recent studies have shown that *C. arabica* is an amphidiploid resulting from a natural cross between the diploid species *C. eugenoides* and *C. canephora* (Lashermes et al. 1999). At the agronomical level, *C. canephora* is particularly interesting to study because it presents some traits of resistance to pests and diseases not encountered in Arabica. At the cup quality level, the latter is more appreciated by the consumers, due to a lower bitterness and a better flavor. However, breeding strategies were recently implemented for Robusta to improve organoleptic and technological characteristics (Guyot et al. 1988; Moschetto et al. 1996).

Molecular markers have been widely applied over recent years to provide a better understanding of the genetic basis of important agronomic traits in plants (Masojc 2002). In coffee, several genetic maps were constructed at the intraspecific or interspecific levels (Paillard et al. 1996; Ky et al. 2000; Lashermes et al. 2001), using RFLP, AFLP (Pearl et al. 2004) and, more recently simple sequence repeats (SSR) markers (Combes et al. 2000; Dufour et al. 2001). Over the last few

years, BAC libraries also became a central tool to identify and characterize genes responsible for important traits in perennial crops (Clément et al. 2004). In that sense, a BAC library was recently constructed from the allotetraploid coffee species *C. arabica* var. IAPAR 59 (Noir et al. 2004), which carries resistance to leaf rust and root-knot nematodes probably coming from the genome of *C. canephora*, through the use of the Timor hybrid as one of the parents (Sera 2001).

Another application of such a BAC library could be to identify genes underlying quantitative trait loci (QTL) linked to cup quality, and particularly those involved in sugar accumulation during coffee fruit development. Indeed, carbohydrates represent about half of the seed dry weight, and among them, sucrose is considered to be the most important precursor of coffee flavor and aroma (Clifford 1985). Sucrose has a central role in the chemical changes occurring during roasting (i.e. Maillard reactions), that lead to a wide range of compounds (i.e. aliphatic acids, hydroxymethyl furfural and other furans, pyrazine and carbonyl compounds), considered together as essential contributors to coffee flavor (Clifford 1985; De Maria et al. 1994). Despite the importance of sucrose as a quality precursor, little is known about its metabolism in coffee plants, particularly concerning key enzymes such as sucrose synthase (SUSY: EC 2.4.1.13), sucrose phosphate synthase (EC 2.4.1.14) and invertase (EC 3.2.1.26) (Zrenner et al. 1995; Roitsch and Ehneß 2000; Winter and Huber 2000). The recent cloning of partial cDNA sequences coding for these enzymes (Marraccini et al. 2003) opens the way to analyze their gene expression during coffee fruit development, to determine their genetic organization in the *Coffea* genus, as well as to modify the sucrose content in coffee Robusta green beans either by conventional breeding (Montagnon et al. 1998) or by genetic engineering (Leroy et al. 2000).

Therefore, the objectives of the present work were: (1) to construct a BAC library of *C. canephora*, and to characterize this library for insert size, chloroplast and mitochondrial contamination, (2) to validate this library with genetically-mapped RFLP probes, (3) to identify and characterize by sequencing BAC clones containing genes coding for enzymes of the sucrose metabolism, and (4) to identify and map SSR markers associated to the BAC clones containing sucrose biosynthetic genes.

Materials and methods

Plant material

Young leaves were collected from *C. canephora* genotype 126 (Moschetto et al. 1996) grown in greenhouse (daytime temperature of 25°C, 70% humidity and 16 h photoperiod of 500 µE s^{-1} m^{-2}) and were used as starting material for high molecular weight (HMW) DNA isolation.

BAC library construction

Nuclei were isolated from leaves according to Zhang et al. (1995) with some modifications to eliminate polysaccharides and reduce polyphenol oxidase activity in the coffee cell extracts. Young-adult leaves were ground in liquid nitrogen and nuclei were liberated by incubating the cell extracts at 4°C in homogenization buffer, followed by several rinses in washing buffer plus 1% of Polyvinylpyrrolidone (PVP) 40. This PVP was added to the extraction-washing buffer to reduce the content of polyphenolics in the coffee leaf homogenate. Between the washing steps leaf homogenate was filtered through nylon mesh (250, 100 and 40 µm) to reduce cell debris. The nuclei were resuspended in 1 ml filtered extraction buffer without β-mercaptoethanol and PVP-40, and embedded in 1.2% low-melting-point agarose plugs (In Cert Agarose, BMA, Rockland, USA). Agarose plugs were incubated for 16 h in lysis buffer, and stored in TE 10:10 buffer (10 mM Tris-HCl, 10 mM EDTA, pH 8.0) at 4°C. HMW DNA integrity was tested by pulsed-field gel electrophoresis (PFGE) using a CHEF MAPPER apparatus (Bio-Rad, U.K.) at 6 V / cm, with a 1–50 s pulse, for 16 h at 14°C in 0.5 TBE buffer (0.09 M Tris-borate, 0.09 M boric acid, 0.002 M EDTA). Before digestion with *Hin*dIII restriction enzyme, each agarose plug was chopped into small pieces with a sterile razor blade. Chopped plugs were incubated in *Hin*dIII restriction buffer (Gibco BRL, USA) with 4 mM spermidine for 30 min on ice. Ten units of *Hin*dIII were added to the chopped plugs and allowed to diffuse for 30 min on ice. For partial digestions, reactions were incubated for 2.5 min at 37°C and then stopped by adding one-tenth of the total volume of 0.5 M EDTA, pH 8.0. Partially digested DNA was size-selected in 1% agarose gels in three steps at 14°C in 0.5 TBE buffer: 6 V/cm for 1 h with 90 s of switch time and an angle of 60° followed by 18 h with 2 s switch time and 10 h with 20–50 s of switch time. The region of the gel containing DNA between 125 kb and 350 kb was cut out and the DNA was electro-eluted from the agarose blocks following the method described by Strong et al. (1997) using 100 µl of TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA). After the electro-elution, DNA concentration was estimated in agarose gel and 100 ng of coffee HMW purified DNA was ligated to pCC1BAC *Hin*dIII-Cloning Ready vector (Epicentre, USA).

After ligation, 1 ml of the ligation was used to electroporate 20 µl of *E. coli* ElectroMAX DH10B competent cells (Gibco BRL) using a BRL Cell-Porator System according the manufacturer's recommendations, but reducing the voltage (charge rate) from 400 to 330 V. After transformation, the cells were resuspended in 1,000 µl of SOC medium (2% Bacto tryptone, 0.5% Bacto Yeast extract, 10 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose, pH 7.0) and incubated for 1 h at 37°C with shaking at 225 rpm. The SOC medium containing the cells was plated on LB medium supplemented with 12.5 µg ml^{-1} chloramphe-

Table 1 *Coffea canephora* BAC library filter-hybridization results

Probes	LG	N° hits	BAC addresses
gA67A	1	13	08-C7, 09-I14, 16-N05, 36-C16, 53-O20, 54-J03, 56-P01, 59-P15, 60-P21, 110-A04, 127-I13, 128-N24, 134-E16
gA13	10	5	04-B16, 22-I10, 45-M19, 52-K01, 76-N06
gA14	3	6	17-F14, 20-P10, 71-O01, 117-G15, 118-B19, 140-L11
gR1210	11	9	37-E07, 53-O21, 68-O05, 89-A22, 95-E17, 113-P11, 140-M05, 144-P11, 123-P7
cR167*	4	5	03-O09, 68-K01, 80-O13, 104-D12, 130-C17
gA61	7	11	04-F24, 11-O23, 14-A07, 50-O22, 50-O24, 78-L12, 83-C01, 112-D08, 113-F22, 115-E15, 130-B08
gR19	2	6	15-M07, 39-B08, 50-J12, 62-F01, 85-P15, 133-F17
gA1	9	7	02-I02, 11-E01, 33-O15, 45-B05, 82-L18, 91-F15, 122-J13
gA11	6	9	16-L19, 23-P06, 31-G10, 54-P20, 89-B20, 111-N18, 122-I12, 131-E16, 134-L11
gA71*	3	13	07-L03, 38-G07, 50-E15, 81-D07, 83-J04, 89-L13, 89-N14, 91-B11, 94-B08, 95-C05, 96-O05, 99-A19, 106-E06
gA10	3	8	04-H20, 06-N18, 08-D16, 89-J02, 106-C09, 107-J24, 120-P04, 138-E16
gA25	3	6	33-G07, 40-F05, 104-G18, 115-P18, 123-C09, 126-P07
gA29	5	7	13-K20, 35-L23, 78-N08, 82-H03, 100-L16, 104-O14, 144-C15
gA59*	6	16	06-A13, 08-P03, 29-G05, 30-L06, 40-F10, 42-G11, 43-C09, 46-A06, 62-E23, 91-J23, 105-P06, 106-L18, 120-P05, 126-F05, 137-F07, 140-O10
gR13*	2	6	01-B22, 49-L05, 52-H20, 106-D13, 113-H11, 123-K23
gA38	9	9	06-D01, 25-E11, 44-K10, 50-L11, 54-D24, 75-H14, 85-N11, 90-M24, 99-D04
gR109	9	9	32-F04, 37-A13, 46-C02, 66-N07, 94-E04, 111-N12, 112-F12, 122-I02, 143-J01
CaSUS1		17	15-C17, 16-J23, 40-J16, 52-N18, 54-B22, 71-P05, 81-D24, 82-L06, 83-F22, 83-G13, 85-I02, 102-B24, 117-H15, 121-F7, 123-N03, 123-O14, 137-A24
CaVAC1		6	28-J20, 52-O12, 65-F15, 100-L24, 104-M05, 143-E18
CaCWI1		9	37-L19, 43-A18, 45-C05, 50-M01, 58-K23, 64-L02, 110-K17, 117-N22, 129-I17
Mitochondrial (0.03%)		17	08-N11, 10-B06, 17-A16, 31-D05, 33-B20, 33-N23, 37-F23, 53-J07, 61-M05, 63-E21, 90-G04, 93-F19, 100-A02, 100-E01, 120-D11, 123-K9, 126-H06
Chloroplastic (4.92%)		2710	NI

Thirteen single and four duplicated (indicated by a star) RFLP probes from coffee were used to anchor different linkage groups (LG) of the *C. canephora* genetic map. Addresses localize the clone by the plaque number, column (letter) and line (number)

nicol, 50 μg ml^{-1} X-Gal and 25 μg ml^{-1} IPTG (isopropyl-thiogalactoside), and incubated in the dark at 37°C overnight. White recombinant colonies were transferred to 384-well plates (Genetix) containing 80 μl of LB freezing buffer using a Flexys colony picker robot. The plates were incubated 16 h at 37°C, duplicated and stored at −80°C.

BAC library screening

High-density Hybond N+ filters were made using a Flexys robot (Genomic Solutions, UK). Each high-density filter contains 18,432 double-spotted clones. Hybridization and washes were performed as described in the Clemson BAC protocols web page (http://www.genome.clemson.edu/protocols). Filters were exposed for 5–8 days using Fuji Medical X-Film (Super RX-100 NIF).

BAC DNA isolation

Individual BAC clones were cultivated overnight in 3 ml LB cultures with 12.5 μg ml^{-1} chloramphenicol. Supercoiled BAC DNA was isolated using QIAGEN BIO ROBOT 9600 (Qiagen GmbH, Germany). BAC DNAs were digested with *Not*I to release the BAC insert. The digested clones were separated by PFGE at 6 V/cm, switch time from 5 to 15 s, angle 120° and run of 5 h.

Genomic probes

Probes used for screening are presented in Table 1. Thirteen single-copy and four duplicated RFLP probes, previously mapped in *C. canephora* (Lashermes et al. 2001) were used to validate the genome coverage of the BAC library. Estimation of chloroplast and mitochondrial DNA contaminations was carried out using specific probes previously used for validation of the Arabica BAC library (Noir et al. 2004). Three partial cDNAs from *C. arabica* var. IAPAR 59 coding for sugar metabolizing enzymes (Marraccini et al. 2003) were used to screen the Robusta BAC library. The first one corresponded to a 750 bp fragment coding for one isoform of Sucrose Synthase (*CaSUS1*: AJ575256) and the two others (490 bp) coded for two different types of invertase: an acid vacuolar form (*CaVAC1*: AJ575258) and a cell wall form (*CaCWI1*: AJ575257). In all the cases, these probes used were internal to each cDNA, corresponding to part of the coding sequence.

BAC ends' sequencing

Ends of 59 BAC clones hybridizing with the Arabica genomic probes gA10, gA14, and gA71, or with *CaSUS1*,*CaVAC1* and *CaCWI1* cDNA sequences, were sequenced (Genome Express) using universal forward

Table 2 List of microsatellites (SSR) identified by BAC ends' sequencing

AC number	Probe	SSR motif	Allele size (bp)	Forward primer 5′ → 3′	Reverse primer 5′ → 3′
AJ 871880	gA71	T^{11}	120	ATCAAACTCCCATTTCCT	TCCTCCACAATACAACCT
AJ 871881	gA14	A^{20}	113	CATCTAATAATCCAGAGAA	CCTTGAAACCAACCAC
AJ 871882*	gA71	$(CAAT)^5$	162	TAACAGAAGCACCAAAACC	TCTAAACCCACCTCACAAC
AJ 871883	*CaCWI1*	$(AT)^6$	123	GTTTTTGTTTTTGGGTCT	ACTTATGGGGCATTTCT
AJ 871884	gA14	$(CTAG)^3$	222	TAGGGTCGGTTCTACCA	TAGCAGCATACAAATCCA
AJ 871885	gA71	$(TAT)^4$	223	TGTAAATGGCTAAAACT	TGTGTAATGTAAGCAAGA
AJ 871886	*CaSUS1*	A^{24}	113	CCACTTGCTACTACATCC	TACAGTTTATGGCTGACAC
AJ 871887	gA14	A^{20}	171	TATCTTAGCGTCTTCCCTTT	GCATAATCCTTCGCACTC
AJ 871888	*CaSUS1*	T^{14}	148	GGAATAGCCTCAGAAAAG	GTTGTAAGTCGGTGGTTC
AJ 871889*	*CaSUS1*	A^{14}	149	TAGTCCCTTTTCAGTGGT	TTTCTTTGTTACGGAGTG
AJ 871890*	*CaCWI1*	$(GCT)^4 + (CAT)^8$	264	ATACATAAGCAAGCACTGA	CAGAACAAATGAAATGGA
AJ 871891	*CaSUS1*	A^{12}	210	TCCTCCATACAGCAATAA	GATTTCAGCCTTTGTCTC
AJ 871892*	*CaCWI1*	$(CA)^6 + (CT)^8$	267	AGAGGGATGTCAGCATAA	ATTTGTGTTTGGTAGATGTG
AJ 871893	*CaSUS1*	A^{14}	174	GGCAGGGATGTGAAA	TTGGTATGTTTGGTTCTG
AJ 871894	*CaSUS1*	$(TA)^5$	178	GTCCTTTGTTCTCCTTCTC	AAGTTCCTCTTCCCATTT
AJ 871895	*CaVAC1*	T^{12}	158	ATTTGGTGTTAGGTGTTG	TATTGCGACTTCTTTGAG
AJ 871896	gA71	C^{12}	120	GTTTGTGCGGTTGATT	AATGATGGGGTAGAAGTG
AJ 871897	*CaSUS1*	$(AC)^{11}$	100	AGGACAAAATGGGAATAA	ATTAGGAAATGAGGATGG
AJ 871898	*CaCWI1*	$A^{20} + A^{18}$	139	ATACACGGCTCCTATTC	CCTTTCACCATTTTTGTT
AJ 871899*	gA71	T^{24}	238	TGCTCAAACTTCTTGCT	CGCCAACTCTAATGTGT
AJ 871900	gA10	T^{12}	208	TTTTCTACCCCAAGAGAG	GAAGCAACCATTTCATTT
AJ 871901	*CaSUS1*	A^{16}	118	GCTTTTAGTTGCTTTTCC	ATGGATGGTGTTTTGGT
AJ 871902	*CaVAC1*	T^{16}	132	ATTTCCGTATTTGTGTGG	TCTATCCTTTACATCCTTCAC
AJ 871903	gA14	$(AAT)^4$	133	CGAGCGTAGAGGAACA	ACATGATACAAGAATGGAGA
AJ 871904*	*CaSUS1*	C^{18}	197	TTGTTGAGAGTGGAGGA	CCAAAGACAGTGCAGTAA
AJ 871905*	*CaSUS1*	A^{18}	124	CGAGACGAGCATAAGAA	GCTGGAATGAAGAATGTAG
AJ 871906	gA71	$T^{14} + T^{14}$	220	TTTCAATCGGCAAA	GCAGGCAAAGAAGG
AJ 871907*	*CaSUS1*	C^{18}	180	CCCAAACACGCATACT	CAGCAAAGGCAGGTT
AJ 871908	gA14	$(AACA)^3$	182	GCTATGGGATTCAGGAC	GCAGCAGTTCAAGCA
AJ 871909*	*CaVAC1*	T^{14}	157	CGTTCAATGCTTCCA	AGGCTTCCCACCAC
AJ 871910	*CaCWI1*	$(CTC)^7$	220	TTGGCACCTGAATG	AGGGAAAGAAGCAGA
AJ 871911*	gA10	$(CATG)^3$	237	TGTTGGTGAAGAAATCC	ATGGAGACAGGAAATAAAC
AJ 871912	*CaSUS1*	T^{24}	117	TTGACCAGTTTCTGATG	TCCCGACCTGTTTAG
AJ 871913	gA71	$(TA)^6$	181	CGACCACCAGGAATG	CGAGCCAGTTAGGG
AJ 871914	*CaVAC1*	T^{20}	212	ACTTCCACGGTCTATC	CGTCAGTTCTCAAGG

SSR markers from BAC clones hybridizing with some RFLP probes (i.e. gA10, gA14 and gA71) and with cDNA probes for sucrose metabolism (*CaVAC1*, vacuolar invertase; *CaCWI1*, cell wall invertase; *CaSUS1*, sucrose synthase) are given with their corresponding GenBank accession number (AC number). *Stars* indicate microsatellites that have been mapped on an intraspecific genetic map of *Coffea canephora*

and reverse M13 primers of the pCC1BAC vector. For those containing SSR markers, sequences were submitted to GenBank and accession numbers are given in Table 2. Primers were defined in the SSR flanking regions using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), and their polymorphism was studied.

CcSUS1 gene cloning

The *CcSUS1* gene from *C. canephora* clone 126 was cloned using the primers UPM2 (5′-CCGTGAGTTGAAACCTGTCTGAGATTT-3′) and REVM3 (5′-TACAAAATGACATTTGAAAA-TGCCCAATT-3′), flanking the *SUS1* cDNA from *C. arabica* (P. Marraccini, unpublished results) and one of the BAC clones hybridizing with the *CaSUS1* probe as a matrix. PCR reaction was performed in a PTC-100 Thermocycler (MJ Research, USA), using Advantage2 *Taq* DNA polymerase (BD Biosciences, Clontech), with an initial denaturation step (94°C, 1 min) followed by 35 cycles (94°C, 1 min, 68°C, 6 min) and ending by an extension step (68°C, 6 min). The fragment amplified was cloned in the pTOPO2.1 (Invitrogen, USA) before being double-strand sequenced (Fig. 3).

Southern-blot analysis

Genomic DNA (15 µg) of *C. canephora* clone 126 was digested with *Hin*dIII and *Dra*I enzymes at a concentration of 10 U μg^{-1} (Fig. 4). After concentration, and separation on a 0.8% agarose gel, DNA was transferred to Hybond N+ membranes (Appligene), and hybridized with an internal fragment of 750 bp of the *CaSUS1* cDNA sequence (AJ575256) labeled with ^{32}P (Sambrook et al. 1989).

Results

Construction and characterization of the BAC library

The genotype 126 of *C. canephora* has been chosen to construct the BAC library due to its good agronomic and organoleptic traits (Moschetto et al. 1996). The BAC library constructed consists of a total of 55,296 recombinant *E. coli* clones stored in 144 trays of 384-wells. One hundred BACs were randomly selected and analyzed by a *Not*I digestion in order to estimate the insert sizes (Fig. 1). Based on this analysis, over 85% of the BAC clones were shown to carry a DNA insert greater than 100 kb, with a reasonable fraction (11%) carrying inserts larger than 175 kb. No clone was found without any insert. Based on the mean insert size of the library (135 kb), on the estimated size of 800 Mb per haploid genome of *C. canephora* (Cros et al. 1995) and its diploid nature, the coverage of our library was estimated to equal nine haploid genome equivalents.

To evaluate the degree of contamination of chloroplast and mitochondrial DNA, a mixture of two mitochondrial (*Cox*III and *NAD*3/*rps*12) and three chloroplast (*trn*D/*trn*T, *trn*K/*mat*K and *trn*Q/*Rps*16) probes was used as specific probes to screen the library. Based upon the percentage of positive clones identified, the contamination was estimated as 4.9% for chloroplast and 0.03% for mitochondrial-specific DNA, respectively.

To experimentally confirm the estimation of nuclear genome coverage of the BAC library, high-density filters were screened with 4 duplicated (cR167, gA71, gA59 and gR13) and 13 single-copy RFLP probes, previously mapped and representing 10 of the 11 linkage groups (LGs) of *C. canephora* (Table 1, Fig. 2).

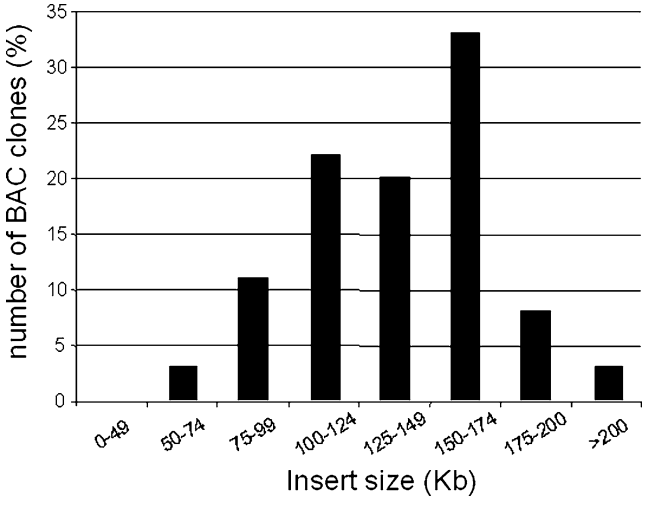

Fig. 1 Insert size distribution of Robusta BAC clones. The insert size was estimated after *Not*I digestion of 100 randomly taken BAC plasmids

For all of the single-copy probes tested, hybridization results (ranging from 5 to 13 hits) are in accordance with the theoretical estimation of BAC library genome coverage with an experimentally derived mean of nine BAC clones per single-copy probe. For gA59 and gA71 duplicated probes, 13 and 16 BAC clones, respectively, were identified again confirming the duplicated nature of these markers. However, only five and six BAC clones were identified, respectively, with the cR167 and gR13 nuclear probes. Except for LG8, for which no RFLP probe was tested, these results showed that all other Canephora LGs are represented in the present BAC library.

Organization of sucrose metabolism genes

With the aim of characterizing the genomic organization of genes encoding sugar-metabolizing enzymes in Robusta, we used partial sequences of *CaSUS1, CaVAC1* and *CaCWI1* cDNAs from *C. arabica* cv. IAPAR 59 (Marraccini et al. 2003) as specific probes to screen our Canephora BAC library. For *CaSUS1*, 17 BAC clones gave strong hybridization signals, whereas 6 and 9 positive BAC clones were revealed respectively with the *CaVAC1* and *CaCWI1* probes (Table 1).

Isolation of the *SUS1* gene from *C canephora*

The present BAC library was used to rapidly clone the complete *SUS1* gene from *C. canephora*. Using BAC 52-N18 previously identified as strongly hybridizing to the *CaSUS1* probe, as a template for PCR with primers located at the extremities of the *SUS1* cDNA sequence from *C. arabica* (P. Marraccini, unpublished results), a fragment of around 4 kb was amplified. Sequence analysis revealed that it contained an open reading frame of 2,418 bp (806 amino acid residues) coding for a SUSY isoform with a deduced MW of 92.6 kDa and a theoretical pI of 6.73 (Fig. 3). The highest percentage of amino acid identity/similarity (89/95%) of this protein was found with the SUS2 (P49039) and SUS4 (AAA97571) isoforms of SUSY from *Solanum tuberosum* (Fu and Park 1995). The Canephora gene revealed the presence of 12 small introns ranging from 79 to 168 bp, with all of them bordered by the 5′-GT/3′-AG consensus. This gene was called *CcSUS1* because it presented a predicted exon/intron structure strictly identical to that observed for the large group of dicot SUS1 genes (Komatsu et al. 2002).

A Southern-blot of Robusta genomic DNA digested with *Dra*I and *Hin*dIII, and hybridized with the 750-bp from the *CaSUS1* cDNA, was also performed. Two bands were observed with *Hin*dIII digestion, one at 9.0 kb which hybridized strongly to the probe and a faint signal at 5.8 kb (Fig. 4). When digested by *Dra*I, several discrete signals were detected at 15, 11 and 8 kb.

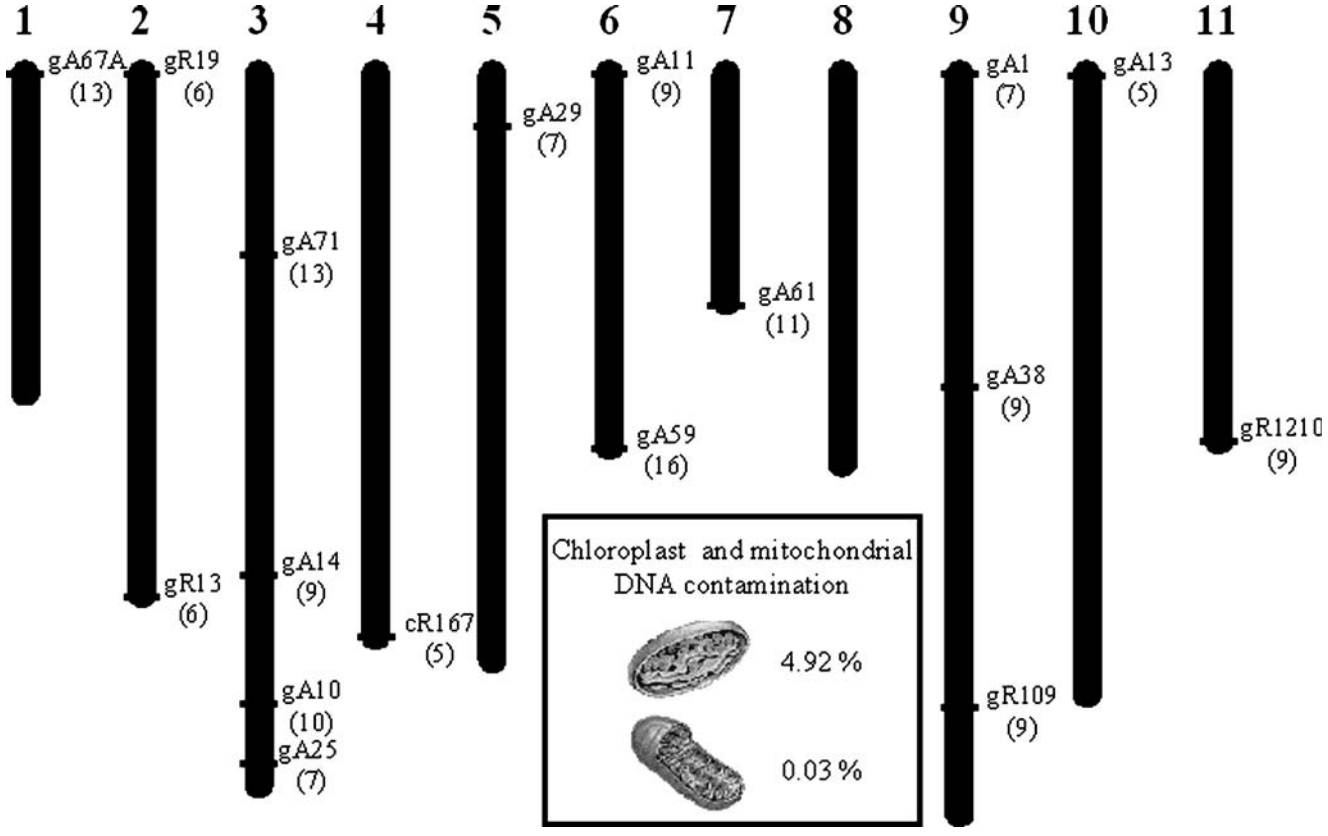

Fig. 2 Genetic linkage map of Robusta coffee. Numbers indicate linkage group (LG). The quantity of BAC clones identified with nuclear probes is indicated in *parentheses* below each probe

Isolation of microsatellites from BAC end sequences

Thirty-five microsatellites were identified in the sequences from the 59 BAC-ends of selected BAC clones (Table 2). Twenty of them were mononucleotide microsatellites, which might make it difficult to use them for discrimination of allele polymorphisms. When tested with *C. canephora* genotypes from guinean (G), congolese (SG1 and SG2 sub-groups) and inter-group (G×SG1) origins (Montagnon 2000), 25 SSR markers out of the 35 firstly identified, showed polymorphism. Using our ongoing genetic intra-specific Robusta map, ten of them were mapped (K. Avia, personal communication).

Discussion

The experimental validation of the BAC library was performed with a set of 13 single-copy and 4 duplicated RFLP probes, previously mapped in *C. canephora* (Lashermes et al. 2001). For most of the probes tested, the hybridization results obtained agreed with a 9-fold covering of the Robusta genome by the BAC library. However, for the cR167 and gR13 nuclear probes, normally considered as duplicated sequences in the genome, only five and six BAC clones were identified, respectively. This could indicate that these alleles were present in a single locus, as also suggested after testing the Arabica BAC library with the cR167 probe (Noir et al. 2004). The fact that these probes could be carried by genomic regions that may be under-represented in the BAC library (P. Piffanelli, unpublished results) or that they could be duplicated and organized in tandem (P. Lashermes, unpublished results) could not be ruled out.

In this work, a particular effort was devoted to use this library to gain insights into the structure and evolution of genes coding for enzymes of sucrose metabolism in coffee, particularly invertases and SUSY, which are known to play key roles in controlling source/sink traits during plant and seed development (Roitsch and Ehneß 2000; Winter and Huber 2000). In *C. arabica*, the activities of both enzymes were studied during the maturation of fruits and beans (Marraccini et al. 2003; Geromel et al. 2004), revealing a precise temporal regulation and clear-cut correlation to the sugar accumulation. We used partial cDNAs from Arabica coding for cell wall invertase (*CaCWI1*), vacuolar invertase (*CaVAC1*) and SUSY (*CaSUS1*) enzymes to screen the present BAC library in order to evaluate the genetic organization and copy number of these genes in the Robusta genome.

For invertases, the two cDNA probes only shared 57% of identity and, as expected, different BAC clones were identified. From the number of hits detected, we suggested that genes encoding both cell wall and

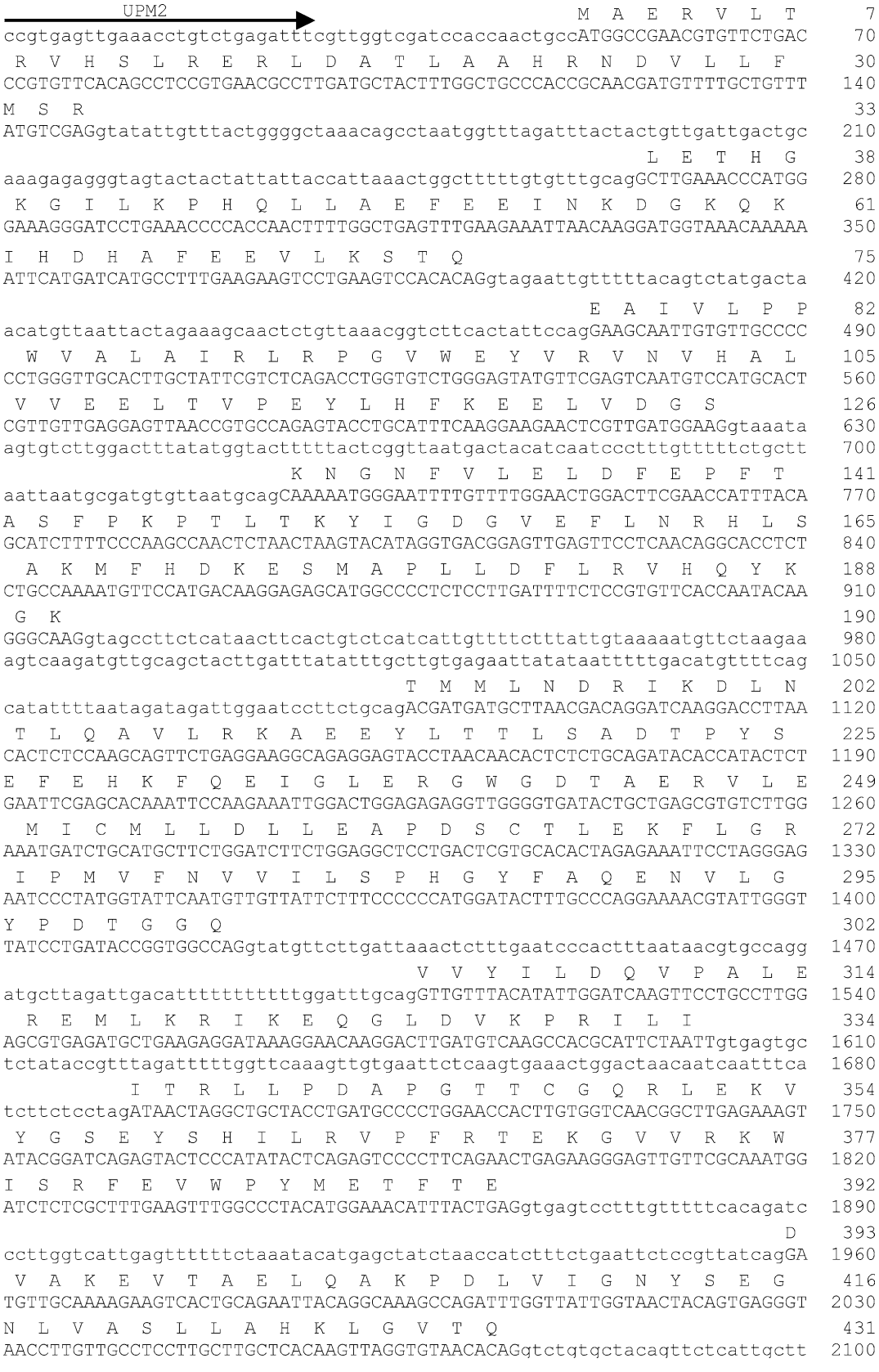

Fig. 3 Complete nucleotide sequence of the *CcSUS1* gene. Amino acid and nucleotide numbers are indicated at the *right of each lane*. Introns are in lower case and exons in upper case. *Horizontal arrows* indicate primers used to amplify the gene. A *star* marks the stop codon. The accession number of the *CcSUS1* deposited in the EMBL/GenBank database is AJ880768

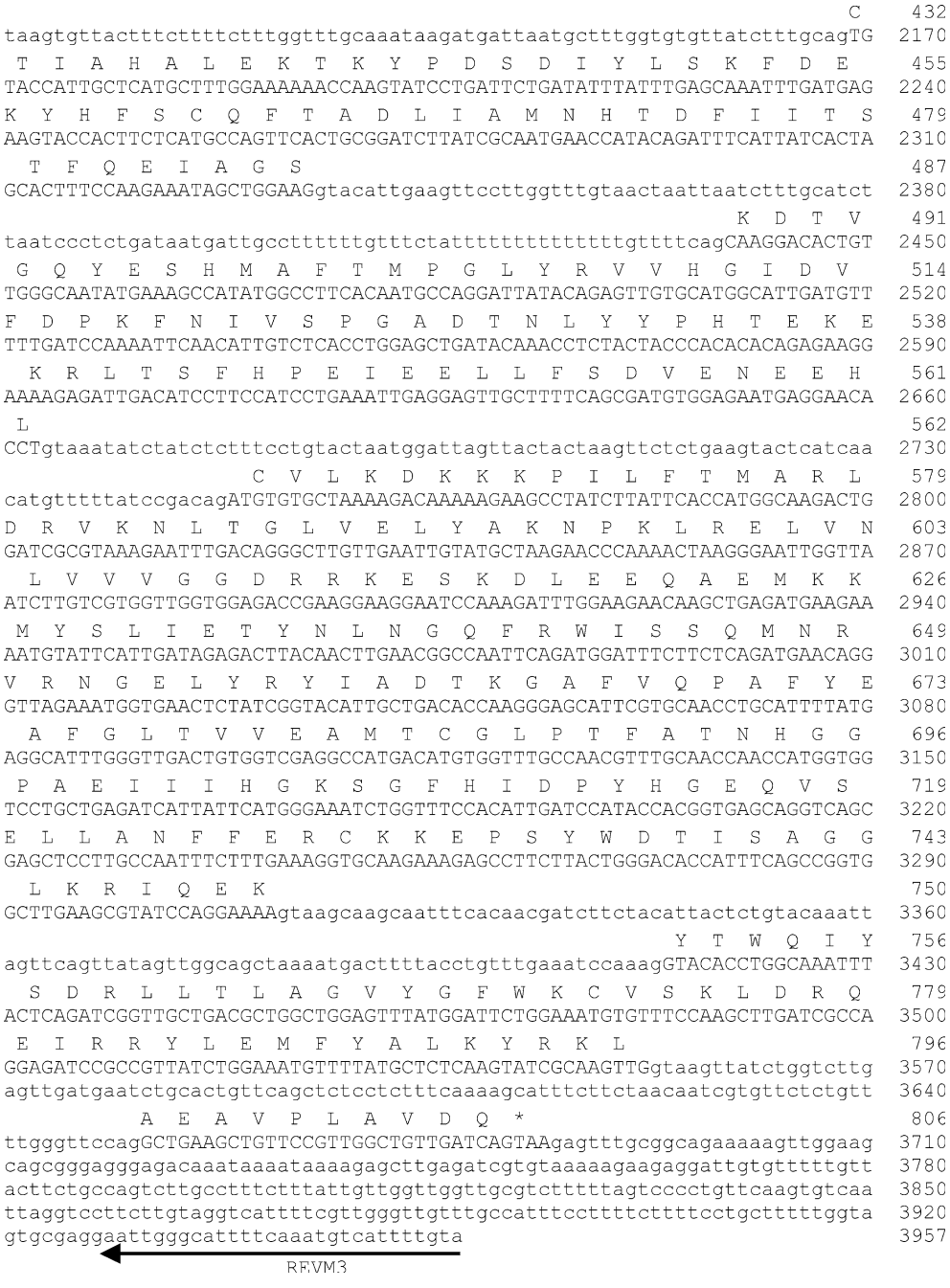

Fig. 3 (Contd.)

vacuolar invertases exist as a single-copy in the Robusta genome, as also observed by Southern-blot experiment of the Arabica genome (P. Marraccini, unpublished results). Several microsatellite (SSR) markers were identified by end sequencing of BAC clones selected with invertase probes, which showed the value of the present BAC library as a source of these markers. Some of them were validated and enabled us to map these genes using an intra-specific genetic map of *C. canephora* (K. Avia, unpublished results). For example, the SSRs (AJ871890 and AJ871892) linked to the BACs hybridizing with the *CaCWI1* cDNA probe, were mapped very closely (two cM) on the LG1, confirming that this type of invertase was encoded by a single copy gene in Robusta.

Concerning SUSY, three SSR markers (AJ871904, AJ871905 and AJ871907) were mapped quite closely on the LG1, whereas a fourth one (AJ871889) was mapped on the LG5. Together with the results of filter hybridizations and southern-blot experiments using the *CaSUS1* cDNA as a probe, we suggested the existence of a multigenic family for the SUS1 isoform with at least two genes in Robusta, as also observed by Southern-blot

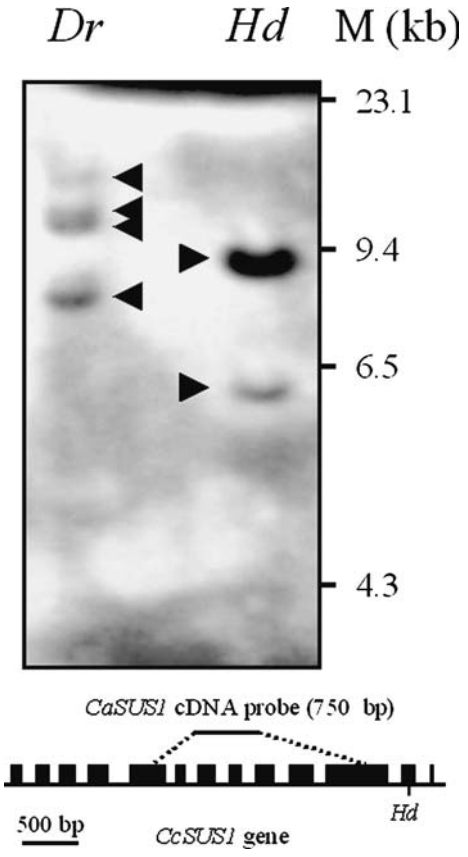

Fig. 4 Southern-blot analysis of *C. canephora* genomic DNA digested with *Dra*I (lane *Dr*) and *Hin*dIII (lane *Hd*) restriction enzymes. Molecular size markers are indicated. Schematic organization of *CcSUS1* gene (*black boxes* = exons) is also presented along with the *CaSUS1* cDNA probe used

experiment of the Arabica genome (P. Marraccini, unpublished results) . One of them was cloned and sequenced from a BAC clone hybridizing with the *CaSUS1* probe. It was called *CcSUS1* because its structure appeared strictly identical to the structure of sucrose synthase plant genes of the dicot SUS1 group (Baud et al. 2004), including the *CitSUS1* gene of *Citrus unshiu* (Komatsu et al. 2002) and the *Sus3*-65 and *Sus4*-16 genes of *Solanum tuberosum* (Fu and Park 1995). It is interesting to note that in these plants, these genes are directly related to the control of the sink function, as deduced by their expression patterns (Fu and Park 1995; Komatsu et al. 2002). This also seems to be the case for the *SUS1* gene in Arabica, where its expression was observed at the time of the rapid expansion of both perisperm and endosperm tissues during bean development (C. Geromel, manuscript in preparation). Together, this suggests a key role of the *CaSUS1* in defining the volume of the locular space and therefore the final size of coffee beans (Wormer 1964).

The analysis presented here on the organization of genes coding for enzymes of sucrose metabolism in Robusta, is a first example of the powerful use of this BAC library as a tool for further genomic studies in coffee. In fact, the availability of this high-genome coverage (9×) library, in parallel with the access to the Arabica BAC library also recently constructed (Noir et al. 2004), will allow us to perform further comparative studies between these two important coffee species like synteny, allelic variations and gene organization (presence of duplications/deletions). Intraspecific studies will also be facilitated with the present BAC library, particularly to investigate linkage disequilibrium within wild and cultivated populations of *C. canephora* that usually come from a limited number of mother-trees (Montagnon 2000). The Robusta BAC library will not only be very useful to physically delimit quantitative traits loci (QTL) for important agronomic traits but also to clone and study the structure and organization of genes involved in fructification time and resistance to nematodes (Akaffou et al. 2003; Noir et al. 2003).

Acknowledgments We thank the robotic team of CIRAD for the management of the BAC library. We are also very grateful to K. Avia for his help concerning gene mapping. The work was supported by the French Embassy in Brazil (DCSUR-BRE-4C5-008) to P. Marraccini. T. Leroy should be contacted for further information about the access to the BAC library.

References

Akaffou DS, Ky CL, Barre P, Hamon S, Louarn J, Noirot M (2003) Identification and mapping of a major gene (Ft1) involved in fructification time in the interspecific cross *Coffea pseudozanguebariae* × *C liberica* var Dewevrei: impact on caffeine content and seed weight. Theor Appl Genet 106:1486–1490

Baud S, Vaultier MN, Rochat C (2004) Structure and expression profile of the sucrose synthase multigene family in *Arabidopsis*. J Exp Bot 55:397–409

Clément D, Lanaud C, Sabau X, Fouet O, Le Cunff L, Ruiz E, Risterucci AM, Glaszmann JC, Piffanelli P (2004) Creation of BAC genomic resources for cocoa (*Theobroma cacao* L.) for physical mapping of RGA containing BAC clones. Theor Appl Genet 108:1627–1634

Clifford MN (1985) Chemical and physical aspects of green coffee and coffee products. In: Clifford MN, Wilson KC (eds) Coffee, Botany, Biochemistry and Production of Beans and Beverage. Publishing Company Inc., Westport, pp 305–374

Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Grasiozi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. Mol Ecol 9:1171–1193

Cros J, Combes MC, Chabrillange N, Duperray C, Monnot des Angles A, Hamon S (1995) Nuclear DNA content in the subgenus *Coffea* (Rubiaceae): inter- and intra-specific variation in African species. Can J Bot 73:14–20

De Maria CAB, Trugo LC, Moreira RFA, Werneck CC (1994) Composition of green coffee fractions and their contribution to the volatile profile formed during roasting. Food Chem 50:141–145

Dufour M, Hamon P, Noirot M, Risterucci AM, Brottier P, Vico V, Leroy T (2001) Potential use of SSR markers for *Coffea* spp. genetic mapping. In: Proceedings of the 19th international colloquium on coffee science ASIC. CD-ROM, ASIC, Paris (France)

Fu H, Park WD (1995) Sink- and vascular-associated sucrose synthase functions are encoded by different gene classes in potato. Plant Cell 7:1369–1385

Geromel C, Ferreira LP, Cavalari AA, Pereira LFP, Vieira LGE, Leroy T, Mazzafera P, Marraccini P (2004) Sugar metabolism during coffee fruit development. In: Proceedings of the 20th international colloquium on coffee science ASIC. CD-ROM, ASIC, Paris (France)

Guyot B, Petnga E, Vincent JC (1988) Analyse qualitative d'un café *Coffea canephora* var Robusta en fonction de la maturité Partie I Evolution des caractéristiques physiques, chimiques et organoleptiques. Café Cacao Thé 32:127–130

Komatsu A, Moriguchi T, Koyama K, Omura M, Akihama T (2002) Analysis of sucrose synthase genes in citrus suggests different roles and phylogenic relationships. J Exp Bot 53:61–71

Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). Theor Appl Genet 101:669–676

Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. Mol Gen Genet 261:259–266

Lashermes P, Combes MC, Prakash NS, Trouslot P, Lorieux M, Charrier A (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. Genome 44:589–596

Leroy T, Henry AM, Royer M, Altosaar I, Frutos R, Duris D, Philippe R (2000) Genetically modified coffee plants expressing the *Bacillus thuringiensiscry*1Ac gene for resistance to leaf miner. Plant Cell Rep 19:382–389

Marraccini P, Pereira LFP, Ferreira LP, Vieira LGE, Cavalari AA, Geromel C, Mazzafera P (2003) Biochemical and molecular characterization of enzyme controlling sugar metabolism during coffee bean development. In: Proceedings of the 7th international symposium of plant molecular biology ISPMB. Poster S19–14, Barcelona (Spain)

Masojc P (2002) The application of molecular markers in the process of selection. Cell Mol Biol Lett 7:499–509

Montagnon C (2000) Optimisation des gains génétiques dans le schéma de sélection récurrente réciproque de *Coffea canephora* Pierre. Ph. D Thesis, ENSA Montpellier (France)

Montagnon C, Guyot B, Cilas C, Leroy T (1998) Genetic parameters of several biochemical compounds from green coffee, *Coffea canephora*. Plant Breed 117:576–578

Moschetto D, Montagnon C, Guyot B, Perriot JJ, Leroy T, Eskes AB (1996) Studies on the effect of genotype on cup quality of *Coffea canephora*. Trop Sci 36:18–31

Noir S, Anthony F, Bertrand B, Combes MC, Lashermes P (2003) Identification of a major gene (*Mex*-1) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea arabica*. Plant Pathol 52:97–103

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 109:225–230

Paillard M, Lashermes P, Pétiard V (1996) Construction of a molecular linkage map in coffee. Theor Appl Genet 93:41–47

Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV, Ming R (2004) Construction of a genetic map for arabica coffee. Theor Appl Genet 108:829–835

Roitsch T, Ehneß R (2000) Regulation of source/sink relations by cytokinins. Plant Growth Reg 32:359–367

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

Sera T (2001) Coffee genetic breeding at IAPAR. Crop Breed Appl Biotechnol 1:179–199

Strong SJ, Ohta Y, 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

Winter H, Huber SC (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. Crit Rev Plant Sci 19:31–67

Wormer TM (1964) The growth of the coffee berry. Ann Bot 28:47–55

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

Zrenner R, Salanoubat M, Willmitzer L, Sonnewald U (1995) Evidence of a crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum* L). Plant J 7:97–107