Theor Appl Genet (2004) 108:1627–1634
DOI 10.1007/s00122-004-1593-0

ORIGINAL PAPER

D. Clément · C. Lanaud · X. Sabau · O. Fouet · L. Le Cunff · E. Ruiz · A. M. Risterucci · J. C. Glaszmann · P. Piffanelli

Creation of BAC genomic resources for cocoa (*Theobroma cacao* L.) for physical mapping of RGA containing BAC clones

Received: 24 April 2003 / Accepted: 5 January 2004 / Published online: 16 March 2004

Abstract We have constructed and validated the first cocoa (*Theobroma cacao* L.) BAC library, with the aim of developing molecular resources to study the structure and evolution of the genome of this perennial crop. This library contains 36,864 clones with an average insert size of 120 kb, representing approximately ten haploid genome equivalents. It was constructed from the genotype Scavina-6 (Sca-6), a Forastero clone highly resistant to cocoa pathogens and a parent of existing mapping populations. Validation of the BAC library was carried out with a set of 13 genetically-anchored single copy and one duplicated markers. An average of nine BAC clones per probe was identified, giving an initial experimental estimation of the genome coverage represented in the library. Screening of the library with a set of resistance gene analogues (RGAs), previously mapped in cocoa and co-localizing with QTL for resistance to *Phytophthora* traits, confirmed at the physical level the tight clustering of RGAs in the cocoa genome and provided the first insights into the relationships between genetic and physical distances in the cocoa genome. This library represents an available BAC resource for structural genomic studies or map-based cloning of genes corresponding to important QTLs for agronomic traits such as resistance genes to major cocoa pathogens like *Phytophthora* spp (*palmivora* and *megakarya*), *Crinipellis perniciosa* and *Moniliophthora roreri*.

Introduction

Theobroma cacao L. (*T. cacao*) is native of Central and South America and was domesticated in pre-Columbian times by the Mayans and Aztecs (Paradis 1979). *T. cacao* L. (2*n*=20), used in the production of chocolate, is an important crop for several tropical countries, particularly in Africa, where the main cocoa producing countries are located. To improve the adaptation of *T. cacao* to various environments, particularly to areas of strong disease pressure, the accumulation of favourable alleles into new varieties is needed.

Molecular markers have provided a clearer understanding of the genetic basis of important agronomic traits and their use has been developed considerably over recent years. The first cocoa linkage map, constructed by Lanaud et al. (1995) was recently enriched with microsatellite markers (Pugh et al. 2004). Studies of the genetic bases of agronomic traits were carried out using plant materials derived from different genetic groups of *T. cacao*: Criollo, Forastero and Trinitario (Lanaud et al. 1999). Several regions of the genome involved in yield components and *Phytophthora* spp. resistance were identified (Crouzillat et al. 2000; Flament et al. 2001; Lanaud et al. 2002; Clément et al. 2003a, 2003b; Risterucci et al. 2003) and confirmed the polygenic nature of this quantitative resistance in cocoa. A region of chromosome 6 appears particularly interesting and gave QTL for yield components and resistance to *Phytophthora.*

Construction of large-insert size BAC libraries and map-based cloning are becoming an efficient strategy for structural studies and gene isolation in plant species (Lahaye et al. 1998; Shirasu et al. 1999; Frary et al. 2000). While many important traits (e.g. disease resistance) have been genetically mapped in cocoa, until now no large-insert size BAC libraries were available to start structural analyses or map-based cloning. Over the last few years BAC libraries have become a central tool to characterize and clone resistance genes (Shirasu et al. 1999;Ballvora et al. 2001; Marek et al. 2001). Numerous genes conferring resistance to different pathogens have been identified in different plant species (Dangl and Jones 2001; Riely and Martin 2001; Jones et al. 2002). These genes contain highly conserved protein domains and have been classified into five general classes (Hammond-Kosack and Jones 1996). Using degenerate primers across highly

Communicated by J.W. Snape

D. Clément (✉) · C. Lanaud · X. Sabau · O. Fouet · L. Le Cunff · E. Ruiz · A. M. Risterucci · J. C. Glaszmann · P. Piffanelli
TA 43/02, Centre de coopération internationale en recherche pour le développement (CIRAD),
Avenue d'Agropolis, 34398 Montpellier, Cedex 5, France
e-mail: didier.clement@cirad.fr

conserved domains [e.g. nucleotide binding site (NBS), leucine rich region (LRR), kinase domains] candidate disease resistance genes from a large set of plants were isolated and referred as resistance gene analogues (RGAs) (Leister et al. 1998; Shen et al. 1998; Pflieger et al. 1999; Peneula et al. 2002). Mapping of RGA sequences in various plant species indicated that they cluster at multiple locations in the genome (Botella et al. 1997, 1998; Wei et al. 1999; Graham et al. 2000, 2002) and are linked to active resistance genes (Zhou et al. 2001; Chauhan et al. 2002). Recently, both RGAs and defence gene analogues (DGA) were isolated from cocoa genomic DNA using degenerate primers designed from conserved domains of several plant resistance and defence genes including the NBS motif (Kuhn et al. 2003, Lanaud et al. 2004), present in a number of resistance genes such as the tobacco *N*; the STK (serine/threonine kinases) subdomains of plants such as the *Pto* tomato gene; and conserved domains of two defence gene families, the pathogenesis-related proteins (PR) of classes 2 and 5 (Lanaud et al. 2004). Most of the RGA and PR2-DGA clones, isolated by Lanaud et al. (2004), were mapped in a segregating population and co-localizations were observed between some RGA and DGA, and QTL for resistance to *Phytophthora* previously identified in several progenies (Crouzillat et al. 2000, Lanaud et al. 2002, Clément et al. 2003b; Risterucci et al. 2003).

The objectives of the present study were: (1) to establish the protocol to successfully isolate cocoa high-molecular-weight (HMW) DNA and construct a bacterial artificial chromosome (BAC) library of *T. cacao* L, (2) characterize the BAC library for insert size, and chloroplast and mitochondrial DNA contamination, (3) validate the BAC library with genetically mapped RFLP probes and (4) start the structural analyses of RGA colocalized with QTL for resistance to *Phytophthora* in cocoa by the physical mapping of BAC clones containing these RGAs.

Materials and methods

Plant material

Leaves at the young adult stage were selected as the starting material for HMW DNA isolation. Leaf material was collected from a Scavina-6 clone implanted in a germplasm collection of the Centre National de Recherche Agronomique (CNRA) on the Ivory Coast. Scavina-6 (Sca6) is an upper Amazon Forastero clone collected in Peru (Pound 1945).

HMW DNA isolation

Nuclei were isolated from leaves according to Zhang et al. (1995) with some modifications to eliminate pectin contamination (Ademako 1972) and reduce polyphenol oxidase activities present in the cocoa cell extracts. Young adult leaves were ground in liquid nitrogen and nuclei were liberated by incubating the cell extracts at 4°C in HB 1×buffer, plus 1% of PVP40. Polyvinylpyrrolidone was added to the extraction washing buffer to reduce the oxidation of polyphenolic substances in the cocoa leaf homogenate. Then the leaf homogenate was filtered through nylon filters (250, 100 and 40 μm) to reduce cell debris contamination. Centrifugation at low speeds (57 *g* for 2 min) followed by a second centrifugation through a Percoll gradient (37.5%) enabled the successful separation of nuclei from the pectin matrix. The nuclei were then resuspended in approximately 1 ml filtered HB 1× extraction buffer without β-mercaptoethanol and PVP-40, and embedded in 1.2% low-melting-point agarose plugs (In Cert Agarose, BMA, Rockland, Md.). 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, UK) 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).

BAC library construction

Chopped plugs were incubated in 1× *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 min 30 s at 37°C and then stopped by adding one-tenth of the total digestion volume of 0.5 M EDTA, pH 8.0. Partially digested HMW 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 switch time and an angle of 60° followed by 18 h with switch time 2 s, then 10 h with switch times from 20 to 50 s. 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 1× TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA). After electro-elution, DNA concentration was estimated in agarose gels and approximately 100 ng of purified genomic DNA was ligated to pCC1BAC *Hin*dIII Cloning Ready vector (Epicentre, USA).

After ligation, 1 μl of the ligation reaction was used to electroporate 20 μl o*f Escherichia coli* ElectroMAX DH10B cells (BRL) using a BRL Cell-Porator System according to the manufacturers' recommendations, but reducing the voltage of transformation (charge rate) from 400 to 330 V. After transformation, the cells were resuspended in 1 ml of SOC medium (2% Bacto tryptone, 0.5% Bacto Yeast extract, 10 mM KCl, 10 mM $MgCl_2$, 10 mM MgSO4.2 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 2YT plates containing 12.5 μg/ml chloramphenicol, 50 μg/ml X-Gal and 25 μg/ml IPTG (isopropylthiogalactoside), and incubated at 37°C overnight. White recombinant colonies were transferred to 384-well plates (Genetix) containing 70 μl of 2YT freezing buffer using a Flexys colony picker robot (Genomic Solutions, UK) The plates were incubated 16 h at 37°C, duplicated and stored at −80°C.

BAC library screening

High-density filters were made using a robot Flexys (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 (http://www.genome.clemson.edu/protocols). Filters were exposed for 5 days using Fuji Medical X-Film (Super RX-100 NIF). Probes used for BAC library screening are listed in Tables 1 and 2

BAC DNA isolation

Individual BAC clones were inoculated in 3 ml overnight 2YT cultures with 12.5 μg/ml chloramphenicol. Supercoiled BAC DNA was isolated using an Qiagen Bio Robot 9600 (Qiagen GmbH, Germany). BAC DNAs were restricted with *Not*I to release the insert. The digested clones were separated by PFGE at 9 V/cm, with a switch time from 5 to 15 s, an angle of 120° and a run time of 5 h.

Table 1 Results of filter hybridizations of the cocoa BAC library using 13 cocoa genomic single-copy probes and one duplicated probe (gTcCIR145) anchored to the 10 linkage groups (*LG*) of the reference genetic map of cocoa

Probes	Size (bp)	LG	No. of hits	Addresses
gTcCIR104	3,200	I	15	1P-11, 6C-14, 8M-15, 12O-10, 14O-21, 28L-12, 30P-11, 31L-09, 56H-09, 57L01, 62J-12, 64I-14, 66O-11, 76E-03, 92I-19
gTcCIR110	2,000	II	9	3F-13, 13D-07, 24J-10, 28O-09, 33O-03, 46F-22, 49H-06, 78M-24, 86I-04
gTcCIR122	2,500	III	10	6I-21, 19E-22, 20B-17, 36G-14, 51L-18, 63K-15, 67N-03, 81F-08, 81J-15, 87H-16
gTcCIR112	2,300	IV	10	12E-04, 15C-09, 16D-17, 16L-02, 37B-11, 43F-10, 44E-19, 60C-01, 80F-09, 96F-15
gTcCIR129	1,000	IV	4	36B-24, 38I-03, 66K-05, 94I-06
gTcCIR106	1,600	V	5	35C-20, 55B-17, 55H-19, 74B-09, 74N03
gTcCIR145	2,000	V	24	2D-20, 2K-06, 5L-24, 7K-19, 12C-12, 12G-23, 13D-12, 17E-02, 28D-18, 37A-03, 46J-12, 49F-09, 51C-16, 66C-04, 68O-21, 73O-16, 75O-19, 76J-05, 78D-12, 81A-19, 81P-10, 87O-14, 92C-07, 95K-19
gtcCIR160	1,600	VI	8	4F-05, 6J-02, 7F-04, 9O-15, 35D-07, 36E-09,46L-20,64O-23
gTcCIR139	600	VI	5	9E-20, 25N-15, 58D-24, 62B-02, 79J-17
gTcCIR138	2,000	VII	10	10L-03, 17F-17, 25O-18, 35K-18, 47B-22,58I-03, 66C-19, 71O-14, 75C-10, 94N-24
gTcCIR114	2,200	VIII	20	7I-08, 15J-22, 22M-13, 23D-15, 24M-08, 27I-12, 28I-12, 47D-19, 51N-19, 52A-03, 52O-20, 62G-14, 63C-21, 64M-18, 65G-20, 66E-12, 67J-02, 81N-07, 90C-23, 92F-15
gTcCIR159	2,000	IX	4	4C-15, 65F-13, 74I-07, 86D-08
gTcCIR102	1,900	IX	11	9M-02, 18O-12, 19P-02, 26C-15, 38E-02, 59C-16, 70B-16, 76G-07, 76I-14, 90M-03
gTcCIR103	1,800	X	9	28E-17, 43E-18, 44J-17, 57M-02, 57N-20, 63A-17, 69F-22, 69I-07, 79L-09

Table 2 Results of filter hybridizations to the cocoa BAC library using RGAs and one DGA probe that have been previously genetically mapped

Probes	No. of hits	LG	Addresses
PtoSca6-A1	5	IV	4C-19, 36D-11, 67N-05, 67O-22, 76P-21
Pto172/A1	11	IV	22G-24, 23J-12, 25J-05, 32F-20, 31D-18, 36D-11, 54J-17, 64H-11, 78E-19, 94C-02, 94H-17
PtoSca6-A3	16	IV	8I-16, 17E-04, 18F-06, 22L-13, 23L-19, 34H-13, 54N-09, 58F-07, 64I-21, 67C-17, 78E-19, 78F-08, 89E-07, 89E-16, 94C-02, 94H-17
PtoSca6-B2	6	IV	14F-24, 24C-08, 33M-14, 50H-24, 82C-01, 90L-24
Pto172/B9	7	IV	1F-02, 18D-06, 64A-24, 65K-18, 65N-07, 84H-10, 87N-03
Pr2/18	5	V	7E-05, 56H-18, 59B-18, 66J-07, 74O-13
N142/C12	10	VII	10C-14, 23F-24, 23M-21, 34G-23, 42P-23, 43B-23, 50K-20, 66N-06, 74I-05, 81B-18
NSca6-A12	7	VII	2D-15, 10C-14, 23F-24, 23M-21, 43B-23, 66N-06, 81B-18
N172/F12	21	X	1A-07, 4G-04, 9A-13, 9L-22, 12J-08, 12K-21, 18D-09, 19A-10, 20E-17. 22C-23, 23H-21, 36K-21, 55C-11, 57G-09, 64D-21, 71L-24, 71P-09, 75C-03, 78D-11, 86N-08, 96E-19
NSca6-A1	13	X	03J-17, 17P-15, 25C-04, 26M-04, 32H-20, 65A-11, 66K-09, 71O-06, 73F-18, 73K-20, 74J-04, 76J-06, 90J-20

Genomic probes

The 14 genomic probes used to identify BAC clones were previously mapped in cocoa (Lanaud et al. 1995). Estimation of chloroplast DNA contamination was carried out using cocoa probes obtained by PCR amplification as described in Demesure et al. (1995), and using a maize plasmid clone encoding the large subunit of the Rubisco protein. To identify BACs containing mitochondrial DNA, two clones encoding the subunits I and II of cytochrome oxidase (*coxI*, *coxII*), respectively, were used as probes to screen the cocoa BAC library. The cocoa RGAs used in this study were isolated, characterized and genetically mapped by Lanaud et al. (2004).

Fingerprints and contig analyses

Five hundred nanograms of each BAC DNA sample was restricted with 10 U of *Hin*dIII. Restricted DNAs were separated in 0.7% ultra-pure agarose gels during 22 h at 50 V. The fingerprint gels were transferred onto nylon membranes and hybridized with the RGA probes according to Risterucci et al. (2000). The fingerprint analyses were made using FPC software (Soderlund et al. 1997). Some bands appeared more intense in the fingerprints and were considered as repeated sequences in the genome. The number of repetitions could be evaluated with the im3 module of the FPC software and were included in the analyses.

Results

BAC library construction and validation

To successfully purify HMW DNA and construct the cocoa BAC library, a modified nuclei isolation protocol (see Materials and methods) was used to efficiently eliminate contamination from the pectin matrix and polyphenols present at high levels in the Sca6 cocoa leaves used as starting material (Blakemore et al. 1966; Adomako 1972). The cocoa BAC library was constructed

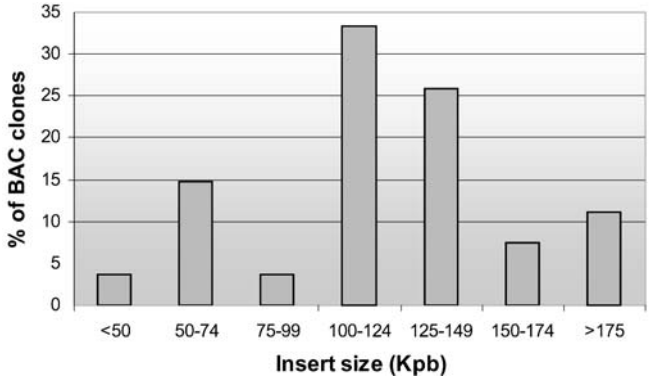

Fig. 1 Insert size distribution of BAC clones in the *Theobroma cacao* Scavina-6 BAC library. To estimate the insert size range, BAC DNA from 192 randomly selected clones were analysed by *Not*I digest and CHEF gel analysis. The BAC insert sizes were grouped and plotted against the frequency of each group. The estimated average BAC insert size is 120 kb

using *Hin*dIII partially-digested genomic DNA and consists of 36,864 clones stored in 96 384-well plates.

A random sample of 384 BACs were analysed by *Not*I digestion, and their insert sizes were grouped and plotted against the frequency of each group (Fig. 1). Based on this analysis, over 80% of the BAC clones in the library have an insert size greater than 100 kb with a sizeable fraction (13%) being over 150 kb. Based on the library average insert size (120 kb) and that of the cocoa haploid genome, equivalent to 390 Mb (Figueira et al. 1992; Lanaud et al. 1992), the estimated coverage of the BAC library is of ten haploid genome equivalents. To experimentally confirm the theoretical estimate of nuclear genome coverage in the library, high-density filters were screened with 14 RFLP probes that have been previously mapped (Lanaud et al. 1995). Southern blot analysis using several restriction enzymes showed that 13 of the RFLP probes used in this study hybridized to a single locus in the Sca6 nuclear genome and one (gTcCIR145) to two loci. These 14 RFLPs identified on average nine BAC clones per probe (see Table 1 and Fig. 2) and provided a satisfactory initial experimental confirmation of the genome coverage in the cocoa library. To investigate the percentage of BAC clones containing sequences hybridizing with DNA derived from chloroplast and mitochondrial genomes, we screened the BAC library with chloroplast- and mitochondrial-specific probes, respectively. Based upon the number of positive BAC clones, we estimated that the chloroplast DNA contamination is of the order of 1.510^{-3} % and that of mitochondrial DNA is 2.1×10^{-4} %. These values reflect the high level of purification of cocoa nuclei obtained with the protocol we used, which com-

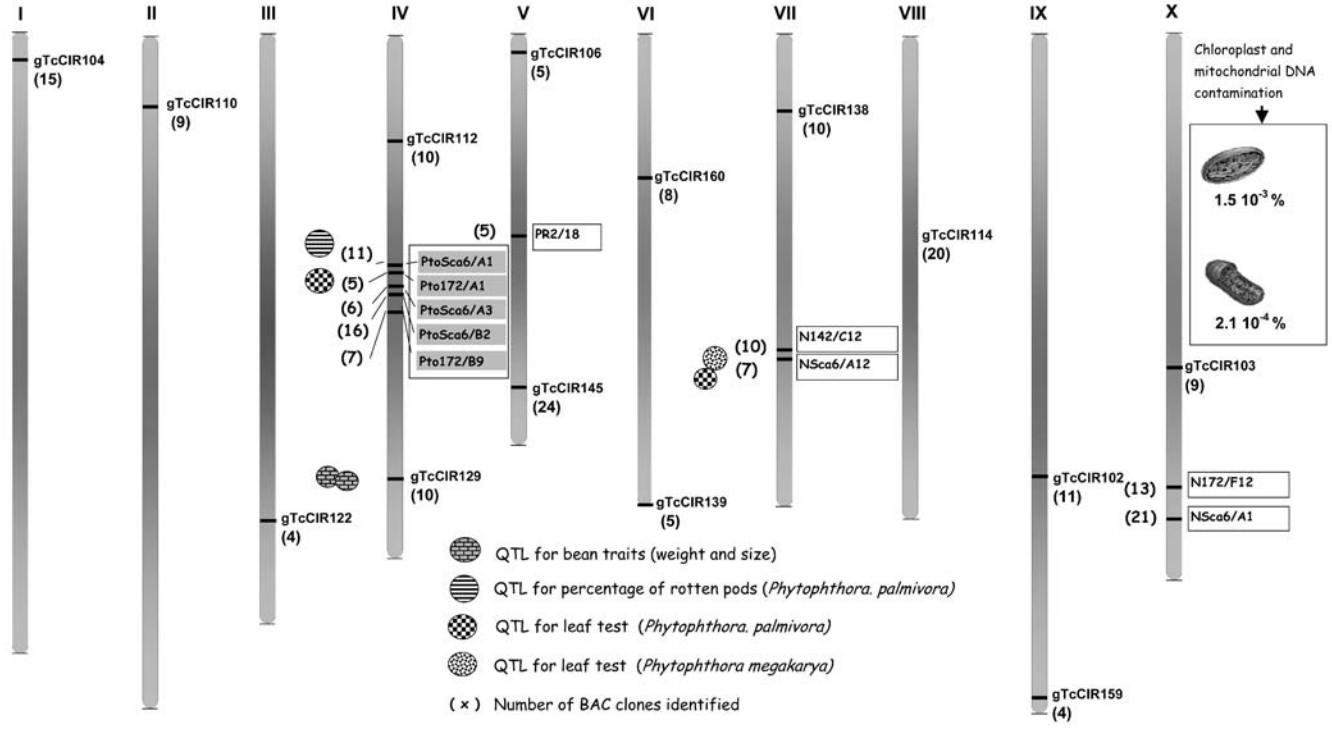

Fig. 2 Cocoa BAC library characterization. Results obtained using 14 RFLP probes, (gTcCIR probes), previously anchored to the cocoa consensus genetic map, are summarized and the number of BAC clones isolated using each RFLP probes is indicated in *parentheses*. The map position of the nine RGA and one DGA genes used to screen the Scavina-6 BAC library, are indicated by *boxes* and the corresponding number of identified BAC clones in *parentheses*. The map position of identified QTLs for resistance to *Phytophthora* spp. and the most important QTL for bean weight and size are indicated as *circles*. Results from the screening of the BAC library with chloroplast and mitochondrial-specific probes are indicated as percentages of BACs over the total number of clones in the BAC library

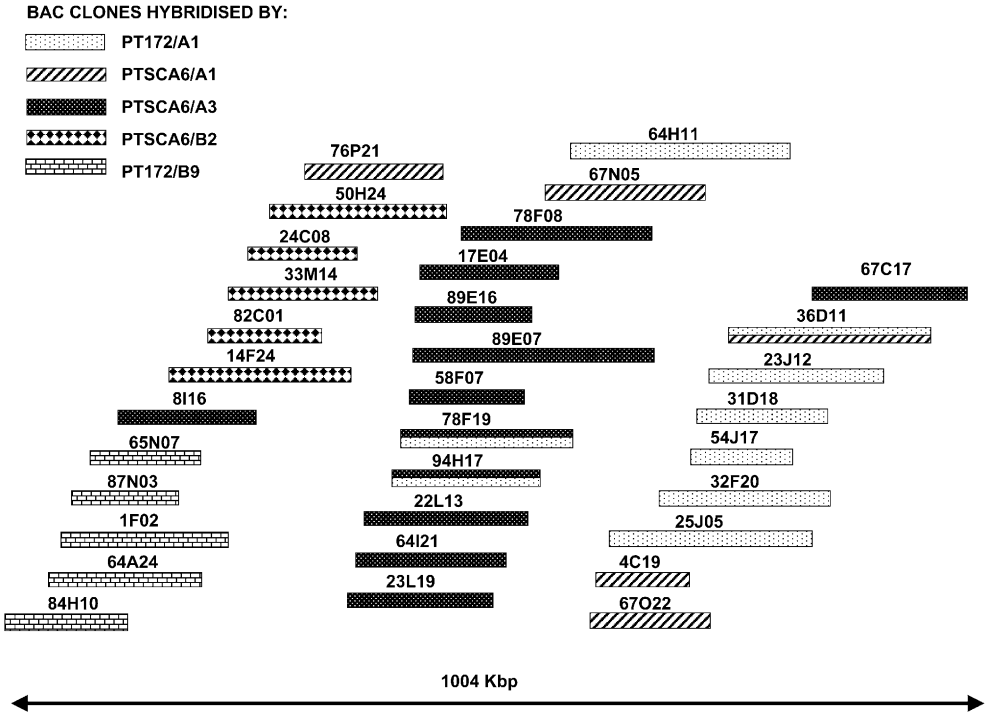

Fig. 3 Contig of BAC clones hybridized by five RGA homologous to *Pto* and *LRK10* resistance genes and clustered in a region of 3.4 cM of chromosome 4. The total contig length is 1,004 kb

bined an extended incubation of the cell extracts with Triton X-100, the use of 40 μm nylon filters and low speed centrifugations to eliminate intact organelles and other cellular contaminants (e.g. cell wall fragments). The low levels of chloroplast DNA contamination are also in agreement with microscopic observations of DAPI-stained isolated cocoa nuclei (data not shown) and the report of Baker et al. (1975) of the presence of only three chloroplasts per cell in *T. cacao* leaf mesophyll.

Physical mapping of RGAs and DGAs

BAC libraries represent a valuable tool to characterize genomic regions containing resistance genes. Over the last 5 years, resistance genes and RGA sequences have been characterized and mapped in various plant species (Collins et al 1998; Chauhan et al. 2002). These analyses indicate that the resistance genes are frequently located in clusters in the genome (Kunkel 1996; Holub et al. 1998; Michelmore and Meyers 1998; Meyers et al. 1999, 2002).

To characterize, at the physical level, genomic regions associated with disease resistance and refine the relationships between physical and genetic distances, the BAC library was screened with an array of nine cocoa RGA and one DGA probes (Table 2), previously isolated, characterized and genetically mapped in cocoa (Lanaud et al. 2004). Contigs were built from BAC clones identified using RGAs clustered in two genomic regions and colocalizing with QTL for resistance to *Phytophthora* (Lanaud et al. 2004). These were the five *Pto*-like RGA clustered in a region of 3.4 cM of chromosome 4, and the two RGAs containing an NBS motif clustered in a region

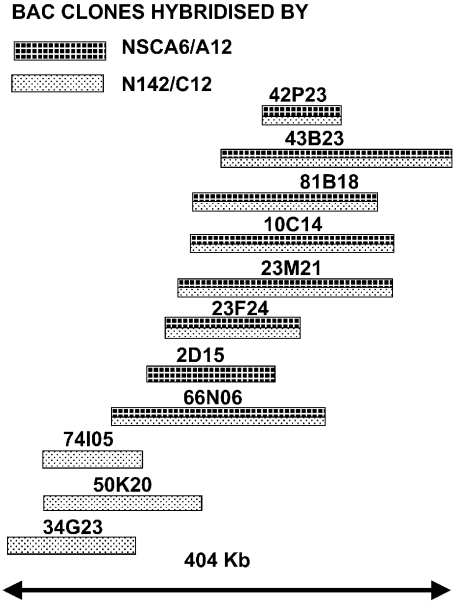

Fig. 4 Contig of BAC clones hybridized by two RGAs homologous to *Pto* and *LRK10* resistance genes and clustered in a region of 1.1 cM of chromosome 7. The total contig length is 404 kb

of 1.1 cM of chromosome 7. The BAC clones identified after hybridization with these probes were restricted by *Hin*dIII. The fingerprints were analysed with FPC software and the DNA blotted on nylon membranes and hybridized with the corresponding RGA probes. Using FPC software, two contigs were built: (1) a contig of 1,004 kb derived from 33 BAC clones identified using the five *Pto*-like RGA probes (obtained at tolerance 1 and

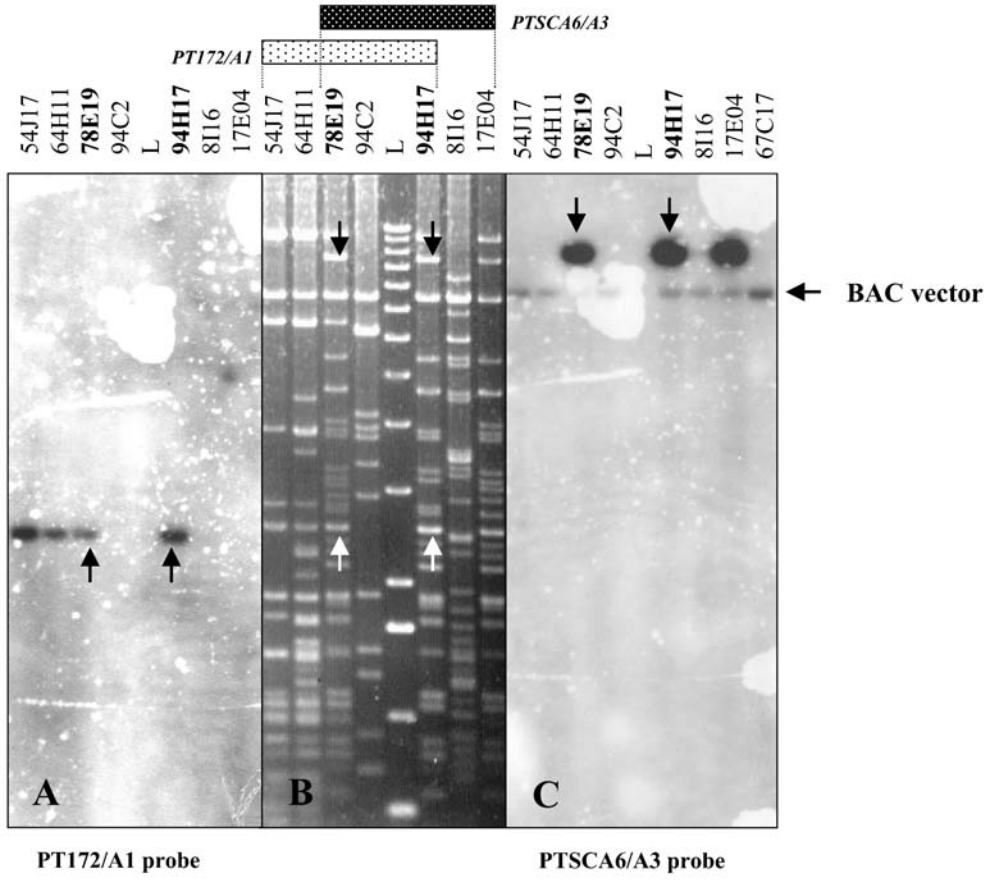

Fig. 5A–C Hybridization of BAC clones from the cluster located on chromosome 4, with two RGA probes: PT172/A1 and PTSCA6/A3. **A, C** Hybridization of the same groups of BAC clones with the RGA probes (PT172/A1 and PTSCA6/A3) after restriction of the BAC DNA by *Hin*dIII and transfer onto nylon membranes. Two BAC clones, 78E19 (135 kb) and 94H17 (112 kb), are positive with both RGA probes. Three BACs (94C2, 8I16 and 67C17) identified with PTSCA6/A3 as the probe on high density filters did not hybridize under more stringent conditions following *Hin*dIII restriction. After fingerprint analysis, the BAC clone 94C2 appeared not to contig with the other BAC clones identified, and probably corresponds to another genomic region. The two other BACs, 8I16 and 67C17, clustered with the whole contig, but independently from the other BACs identified by the same probe, revealing their sequence divergence. **B** Fingerprint (*Hin*dIII) of BAC clones identified after hybridization of high density filters with PT172/A1 and PTSCA6/A3. Groups of BAC clones hybridizing with each probe are indicated at the *top* of panel **B**

cutoff e^{-12}) and located in chromosome 4. (Fig. 3), and a contig of 404 kb derived from 11 BAC clones identified using two RGA probes containing an NBS motif (obtained at tolerance 1 and cutoff e^{-10}) and located in chromosome 7 (Fig. 4).

Two BAC clones (94C2 and 18D6), identified using the *Pto* like RGA probes, did not fit into the built contig even if a lower stringency was used with the FPC software. These BACs could correspond to other regions of the genome containing *Pto*-like RGAs.

Within the whole contig of chromosome 4, most of the BAC clones are continuously contiged inside each BAC group identified using a unique RGA probe. Exceptions were observed for one BAC identified by the probe PTSCA6/A1 (76P21) and two BACs identified by PTSCA6/A3 (8I16 and 67C17). These three BACs did not fit with the others identified by the same RGA probe. After hybridization of the fingerprints with the RGA probe, using more stringent conditions than during the screening of the library, these BACs did not hybridize with the probe initially used to identify them (Fig. 5). These three BACs could contain sequences homologous to the probes, but be located in different regions and less similar to the probe sequence than the other BAC clones identified and clustered. They could correspond to several members of a same gene family.

Two BAC clones, 78E19 (135 kb) and 94H17 (112 kb), hybridized to both RGA probes in both the initial screening on high density filters and subsequent fingerprinting (Fig. 5). Two different size bands from the fingerprints are hybridized by each of the probes, revealing two different localizations of both RGA on the same BAC clones.

When a separate analysis of each group of BAC clones that hybridize to a unique probe is made, their relative positions in the contig may vary as compared to the results obtained when all BAC groups are analysed together. Several hypotheses can be made to explain these

discrepancies. Firstly, the presence of repeated sequences in different regions of the whole contig, or sequences present only in one allelic form, could increase the complexity necessary to build the whole contig. Another cause could be heterozygosity of SCA6, the cocoa genotype used to establish the BAC library. Indeed, Sca6 is known to be partly heterozygous (Lanaud et al. 2001) and different restriction sites in the BAC clones corresponding to homologous chromosome regions could prevent a perfect match between homologous BAC clones being obtained, and could induce more discordances during contig building. The hybridization of BAC ends with neighbouring BAC clones will be necessary to refine the precise localization of each BAC clone in the contig.

Discussion

The value of using RGA sequences to study the molecular basis of resistance traits has recently been demonstrated in rice. Chauhan et al. (2002) used three RGA probes tightly linked to the rice genome locus *Pi-CO39*, which confers resistance to the rice pathogen *Magnaporthe grisea*, to assemble a BAC contig to physically delimit the genomic region containing the resistance gene. In potato, Ballvora et al. (2002) have cloned the *R1* gene, involved in the quantitative resistance to late blight, by combining positional cloning with a candidate gene approach.

Our results indicate that a similar clustered organization of resistance genes may be encountered in cocoa as in other plant species. The results obtained in this study represent a first step towards dissecting out the structure of the resistance genes clusters in cocoa, and identifying genes potentially contributing to quantitative resistance to *Phytophtora* spp. The physical mapping of BAC clones containing RGAs showed that such clusters could be spread in regions as large as 1,000 kbp. Many other serious diseases affect cocoa and the candidate genes physically mapped in this study could present a broader interest for cocoa disease management.

Major QTL were identified in cocoa, particularly for resistance to *Crinipellis perniciosa,* a fungus causing the witches broom disease (Queiroz et al. 2002) or for yield traits such as pod weight (Clement et al. 2003b). Given the small size of the cocoa genome, estimated at 390 Mb (Figueira et al. 1992; Lanaud et al 1992), this BAC library represents a valuable resource for structural genomic studies or map-based cloning of genes corresponding to these important QTLs for agronomic traits in cocoa.

Acknowledgements We thank USDA for their financial contribution to the contig analyses of RGA-containing BAC clones. We are grateful to CNRA (Centre National de Recherche Agronomique) of the Côte d'Ivoire for providing the Scavina-6 leaves and to the Montpellier Languedoc Roussillon Genopole for the robotic platform used to manage the BAC library.

References

Adomako D (1972) Cocoa pod husk pectin. Phytochemistry 11:1145–1148

Baker NR, Hardwick K, Jones P (1975) Biochemical and physiological aspects of leaf development in cocoa (*Theobroma cacao)* II. Development of chloroplast ultrastructure and carotenoids. New Phytol 75:513–518

Ballvora A, Schornack S, Baker BJ, Ganal M, Bonas U, Lahaye T (2001) Chromosome landing at the tomato Bs4 locus. Mol Genet Genomics 266:639–645

Ballvora A, Ercolano MR, Weis J, Meksem K, Bornmann CA, Oberhagemann P, Salamini F, Gebhart C (2002) The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/ NBS/ LRR class of plant resistance genes. Plant J 30:361–371

Blakemore WR, Dewar ET, Hodge RA (1966) Polysaccharides of the cocoa pod husk. J Sci Food Agric 17:558–560

Botella MA, Coleman MJ, Hughes DE, Nishimura MT, Jones JD, Somerville SC (1997) Map positions of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. Plant J 12:1197–1211

Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB, Jones JD (1998) Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell 10:1847–1860

Chauhan RS, Farman ML, Zhang HB, Leong SA (2002) Genetic and physical mapping of a rice blast resistance locus, Pi-CO39(t), that corresponds to the avirulence gene AVR1-CO39 of *Magnaporthe grisea*. Mol Genet Genomics 267:603–612

Clément D, Risterucci AM, Motamayor JC, N'Goran JD, Lanaud C (2003a) Mapping quantitative trait loci for bean trait and ovule number in *Theobroma cacao* L. Genome 46:103–111

Clément D, Risterucci AM, Motamayor JC, N'Goran JD, Lanaud C (2003b) Mapping QTL for yield components, vigor and resistance to *phytophthora palmivora* in *Theobroma cacao* L. Genome 46:204–212

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogues in maize. Mol Plant Microbe Interact 11:968–978

Crouzillat D, Phillips W, Fritz PJ, Pétiard V (2000) Quantitative trait loci analysis in *Theobroma cacao* L. using molecular markers. Inheritance of polygenic resistance to *Phytophthora palmivora* in two related cacao populations. Euphytica 114:25–36

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–131

Figueira AJ, Janick J, Goldsbrough P (1992) Genome size and DNA polymorphism in *Theobroma cacao* L. J Am Soc Hortic Sci 117:673–677

Flament M-H, Kébé I, Clément D, Pierretti I, Risterucci AM, N'Goran JAK, Cilas C, Despréaux D, Lanaud C (2001) Genetic mapping of resistance factors to *Phytophthora palmivora* in cocoa. Genome 44:79–85

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) Fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Graham MA, Marek LF, Lohnes D, Cregan P, Shoemaker RC (2000) Expression and genome organization of resistance gene analogues in soybean. Genome 43:86–93

Graham MA, Marek LF, Shoemaker RC (2002) Organization, expression and evolution of a disease resistance gene cluster in soybean. Genetics 162:1961–1977

Hammond-Kosack KE, Jones JD (1996) Resistance gene-dependent plant defence responses. Plant Cell 8:1773–1791

Holub EB, Jones JD (1998) Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell 10:1847–1860

Jones PG, Allaway D, Gilmour DM, Harris C, Rankin D, Retzel ER, Jones CA (2002) Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. Planta 216:255–264

Kuhn DN, Heath M, Wisser RJ, Meerow A, Brown JS, Lopes U, Schnell RJ (2003) Resistance gene homologues in *Theobroma cacao* as useful genetic markers. Theor Appl Genet 107:191–202

Kunkel BN (1996) A useful weed put to work: genetic analysis of disease resistance in *Arabidopsis thaliana*. Trends Genet 12:63–69

Lahaye T, Shirasu K, Schulze-Lefert P (1998) Chromosome landing at the barley *Rar1* locus. Mol Gen Genet 260:92–101

Lanaud C, Hamon P, Duperray C (1992) Estimation of nuclear DNA content of *Theobroma cacao* L. by flow cytometry. Café Cacao Thé 36:3–8

Lanaud C, Risterucci AM, N'Goran J, Clément D, Flament M, Laurent V, Falque M (1995) A genetic linkage map of *Theobroma cacao* L. Theor Appl Genet 91:987–993

Lanaud C, Motamayor JC, Sounigo O (1999) Le cacaoyer. In: Diversité génétique des plantes tropicales. Éditions CIRAD, Montpellier, pp 141–169

Lanaud C, Motamayor JC, Risterucci AM (2001) Implications of new insight into the genetic structure of *Theobroma cacao* L. for breeding strategies. In: Proceedings of the international workshop on new technologies and cocoa breeding. INGENIC, London, pp 89–107

Lanaud C, Flament M-H, Nyassé S, Risterucci AM, Fargeas D, Kébé I, Motilal L, Thévenin J-M, Paulin D, Ducamp M, Clement D, N'Goran JAK, Cilas C (2002) Synthesis of studies on genetic basis of cocoa tree resistance to *Phytophthora* using molecular markers. Proceedings of the 13th international cocoa research conference, October 9–14, Kota Kinabalu, Malaysia, pp 127–136

Lanaud C, Risterucci AM, Pieretti I, N'Goran JAK, Fargeas D (2004) Characterization and genetic mapping of resistance and defence gene analogues in cocoa (Theobroma cacao L.). Mol Breed (in press)

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 95:370–375

Marek LF, Mudge J, Darnielle L, Grant D, Hanson N, Paz M, Huihuang Y, Denny R, Larson K, Foster-Hartnett D, Cooper A, Danesh D, Larsen D, Schmidt T, Staggs R, Crow JA, Retzel E, Young ND, Shoemaker RC (2001) Soybean genomic survey: BAC-end sequences near RFLP and SSR markers. Genome 44:572–581

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317–332

Meyers BC, Morgante M, Michelmore RW (2002) TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. Plant J 32:77–92

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Paradis L (1979) Le cacao précolombien: Monnaies d'échange et breuvage des dieux. J Agric Trad Bot Appl 26:3–4

Penuela S, Danesh D, Young ND (2002) Targeted isolation, sequence analysis, and physical mapping of non-TIR NBS-LRR genes in soybean. Theor Appl Genet 104:261–272

Pflieger S, Lefebvre V, Caranta C, Blattes A, Goffinet B, Palloix A (1999) Disease resistance gene analogues as candidates for QTLs involved in pepper-pathogen interactions.Genome Res 42:1100–1110

Pound FJ (1945) A note of the cocoa population of South America. Arch Cocoa Res 1:93–97

Pugh T, Fouet O, Risterucci AM, Brottier P, Deletrez C, Courtois B, Clément D, Abouladze M, Larmande P, N'Goran JNK, Lanaud C (2004) A new codominant markers-based cocoa linkage map: development and integration of 201 new microsatellites markers. Theor Appl Genet (in press)

Queiroz VT, Guimarães CT, Anhert D, Schuster I, Daher RT, Pereira MG, Miranda VRM, Loguercio LL, Barros EG, Moreira MA (2003) Identification of a major QTL in cocoa (*Theobroma cacao* L.) associated with resistance to witches broom disease. Plant Breed 122:268–272

Riely BK, Martin GB (2001) Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. Proc Natl Acad Sci USA 98:2059–2064

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament M-H, Lanaud C (2000) A high-density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C (2003) Identification of QTL related to cocoa resistances to three species of *Phytophthora*. Theor Appl Genet 108:168–174

Shen KA, Meyers BC, Islam-Faridi MN, Chin DB, Stelly DM, Michelmore RW (1998) Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol Plant Microbe Interact 11:815–823

Shirasu K, Lahaye T, Tan MW, Zhou F, Azevedo C, Schulze-Lefert P (1999) A novel class of eukaryotic zinc-binding proteins is required for disease resistance signalling in barley and development in *C. elegans*. Cell 99:355–366

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. CABIOS 13:523–535

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

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240 kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153:1929–1948

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

Zhou F, Kurth J, Wei F, Elliott C, Vale G, Yahiaoui N, Keller B, Somerville S, Wise R, Schulze-Lefert P (2001) Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a Rar1-independent signalling pathway. Plant Cell 13:337–350