UNIVERSIDADE ESTADUAL DE LONDRINA

SANDRA MARIA BELLODI CAÇÃO

CONSTRUÇÃO E CARACTERIZAÇÃO DE UMA BIBLIOTECA GENÔMICA DE *C.arabica* EM CROMOSSOMO ARTIFICIAL DE BACTÉRIA

Londrina
2012

SANDRA MARIA BELLODI CAÇÃO

CONSTRUÇÃO E CARACTERIZAÇÃO DE UMA BIBLIOTECA GENÔMICA DE *C. arabica EM* CROMOSSOMO ARTIFICIAL DE BACTÉRIA

Tese apresentada ao Programa de Pós graduação em Agronomia da Universidade Estadual de Londrina, como requesito parcial para a obtenção do título de Doutor.

Orientadora: Profa. Valéria Carpentieri-Pípolo.
Co-Orientador: Dr. Luiz Filipe Protassio Pereira.

Londrina
2012

Dados Internacionais de Catalogação-na-Publicação (CIP)

C118c Cação, Sandra Maria Bellodi.
Construção e caracterização de uma biblioteca genômica de *c.arabica* em cromossomo artificial de bactéria / Sandra Maria Bellodi Cação. – Londrina, 2012.
77 f.: il.

Orientadora: Valéria Carpentieri-Pípolo.
Dissertação (Mestrado em Agronomia) – Universidade Estadual de Londrina, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia, 2012.
Inclui bibliografia.

1. Café – Melhoramento genético – Teses. 2. Café – Resistência a doenças e pragas – Teses. 3. Fungos-da-ferrugem – Teses. 4. Mapeamento cromossômico – Teses. 5. Marcadores biológicos – Teses. 6. Biotecnologia vegetal – Teses. I. Carpentieri-Pípolo, Valéria. II. Universidade Estadual de Londrina. Centro de Ciências Agrárias. Programa de Pós-Graduação em Agronomia. III. Título.

CDU 631.52:633.73

SANDRA MARIA BELLODI CAÇÃO

CONSTRUÇÃO E CARACTERIZAÇÃO DE UMA BIBLIOTECA GENÔMICA DE *C. arabica* EM CROMOSSOMO ARTIFICIAL DE BACTÉRIA

Tese apresentada ao Departamento de Pós graduação em Agronomia da Universidade Estadual de Londrina, como requesito parcial para a obtenção do título de Doutor.

BANCA EXAMINADORA

Prof. Dr. Luiz Filipe Protassio Pereira
Embrapa Café – Brasília - DF

Prof. Dr. Douglas Silva Domingues
IAPAR – Londrina - PR

Prof. Edison Miglioranza
UEL – Londrina - PR

Dr. Alexandre Garcia
TMG – Cambé - PR

Dr. Laurival Villas-Boa
UEL – Londrina - PR

Londrina, 24 de maio de 2012.

Ao meu esposo José Marcos, aos meus filhos Gustavo e Gabriela e aos meus pais José Roberto e Maria Ioli dedico.

AGRADECIMENTO

A Deus, pela sua constante presença nos momentos mais difíceis.

Ao co-orientador Dr. Luiz Filipe P. Pereira, pelo constante incentivo e pela confiança depositada, foram 11 anos de trabalho juntos, sem o qual eu não poderia ter alcançado mais este objetivo.

A Prof. Dr. Valéria Carpentieri-Pípolo pela amizade e oportunidade de orientação.

A secretária Weda da Pós-graduação, que sempre me ajudou quando necessário.

Ao INCT Café, pela concessão da bolsa de estudo.

Ao Consórcio de Pesquisa dos Cafés do Brasil pelo apoio financeiro neste trabalho.

Ao programa de Pós-Graduação em Agronomia da Universidade Estadual de Londrina.

A amiga Dra. Alessandra Ferreira Ribas pelo longo tempo de amizade e principalmente pela ajuda nos momentos de correria e dúvidas.

Ao Pesquisador do IAPAR, Dr. Douglas Silva Domingues pelas sugestões.

Ao Tiago pela amizade de longos anos de IAPAR.

As minhas amigas Gislaine e Má Dirça (Marília) que sempre me incentivaram nessa caminhada.

Ao Pesquisador Dr. Luiz Gonzaga Esteves Vieira, pelas contribuições profissionais, no cafezinho.

A todos os amigos do Laboratório de Biotecnologia Vegetal (Léo, Rafinha, Kenia, Julia, Carol, Suzana, Priscila, Rafael, João, Arthur, Fernando, Maiara, Andrea, Pierre, Firmino), a todos que passaram pelo laboratório, Dudu, Renata, Nathalia Volpi, Silvia, Lidia.

Aos colegas do Laboratório de Cultura de Tecidos (Lucinéia, Suely, Seu Arlindo, Cicera) e aos estagiários pelo convívio e amizade,

A todos que de alguma forma, contribuíram para a realização deste trabalho.

Muito Obrigada.

CAÇÃO, Sandra Maria Bellodi. **Construção e caracterização de uma biblioteca genômica de *Coffea arabica* em Cromossomo Artificial de Bactéria.** 77 f. 2012. Doutorado (Agronomia – Fitotecnia) – Universidade Estadual de Londrina, Londrina, 2012.

RESUMO

O Brasil é o principal produtor e exportador mundial de café e possui o segundo maior mercado consumidor. Entre os principais objetivos do melhoramento genético de cafeeiros estão os estudos visando à melhoria de cultivares resistentes à ferrugem utilizando como fonte de resistência, progênies oriundas do Hibrido de Timor, planta derivada do cruzamento interespecífico natural entre *Coffea arabica* e *C. canephora*, resistente a maioria das raças fisiológicas do fungo *Hemileia vastratix*. Bibliotecas baseadas nos vetores do tipo Cromossomo Artificial de Bactéria (*BAC*) são um importante recurso para clonagem posicional, analise comparativa de genomas e construção de mapas físicos. Neste trabalho foi realizada a construção e caracterização de uma biblioteca *BAC* com 56,832 clones com um tamanho médio de insertos de 118 kb, representando 5 a 6 vezes a cobertura do genoma haplóide de *C.arabica*. A contaminação foi estimada como 1,04% para cloroplasto e 0,5% para DNA mitocondrial. Combinadas, as técnicas de seleção por PCR de *pools* representando vários grupos de clones da biblioteca propiciou uma forma rápida e econômica de seleção da biblioteca, pois com um pequeno número de PCR foi possível encontrar clones BAC para os marcadores ancôras e o gene M6PR. As placas foram selecionadas a partir de *superpools* e *pools* de placa. Um total de 54 clones BAC foram selecionados, 10 para o M6PR, 11 para o SSR-16 e CCG-3, 12 para o SSR-18 e 10 para ACCG-1. As análises de Southern Blot para o gene M6PR demonstraram que os *BAC* isolados apresentaram um fragmento relacionado com o fragmento presente no genoma de HT CIFC 832/2, e semelhante no *C. eugenioides*. No estudo de seleção de clones para as marcas de resistência a ferrugem foram identificados 10 clones positivos para a marca SAT-244, seis para o BA-124 e sete para o M-8. Análises de *Fingerprinting* mostraram padrões de sobreposição de *BAC* que deve permitir a formação de contigs *BAC* abrangendo a região de resistência à ferrugem para o gene SH3. Através do *Southern Blot* usando sonda M-8 foram identificados clones BAC em dois grupos que correspondem aos subgenomas do HT (C^{C}-P89O21 e C^{e}-P135L24). Os clones positivos foram classificados em dois grupos que correspondem aos subgenomas do HT. Os clones selecionados para as marcas de resistência a ferrugem, foram submetidos ao seqüenciamento e serão utilizados em estudos de sintenia e comparação de sequencias entre espécies. A biblioteca BAC construída é uma importante ferramenta disponível aos laboratórios de pesquisa em genômica de cafeeiros, visando uma maior integração de informações genômicas e genéticas para aplicação futura em programa de melhoramento.

Palavras-chave: Hibrido de timor. Mapa físico. Ferrugem. Marcadores.

CAÇÃO, Sandra Maria Bellodi: **Construction and characterization of a genomic library of *Coffea arabica* in Bacterial Artificial Chromosome.** 77 f. 2012. Doutorado (Agronomia – Fitotecnia) – Universidade Estadual de Londrina, Londrina, 2012.

ABSTRACT

Brazil is the largest producer and exporter of coffee in the world, besides having the second largest consumer market. Among the major goals of genetic breeding of tree coffee are the studies aimed at improving rust resistant cultivars used as sources of resistance, progenies of “Hibrido Timor”, plant-derived natural interespecific cross between *C. arabica* and *C. canephora,* resistant to most races of fungus *H. vastratix.* Libraries based in Bacterial Artificial Chromossome (BAC) vectors have become an important tool for genomic research. The BAC libraries are an important resource for applications including gene isolation, comparative genomics, map-based or positional cloning of genes and genome-wide physical map construction. Furthermore, they are important genomic resources for sequencing projects, map-based cloning of either genes or QTLs for important agronomic characteristics. In this work we described the construction and characterization a BAC library consisting of 56.832 clones with an average size of inserts 118 kb, representing 5 to 6 times the coverage *C. arabica* haploid genome. Based on the percentage of positive clones identified, the contamination was estimated as 1.04% to chloroplast and 0.5% mitochondrial DNA. Combined, the PCR and pools of library clones representing various groups of clones provide an economical form to screening of library, because with a small number of PCRs is possible to find a single target sequence. The plates were selected from of superpools and plate of pools. A total of 54 BAC were selected, 10 for M6PR, 11 for SSR-16 and CCG-3, 12 for SSR-18 and 10 for ACCG-1. In Southern blot analysis for gene M6PR, the genomic DNA of HT showed four bands, probably related with the corresponding bands in *C.canephora* and *C. eugenioides*. The BAC clones 1-6 and 8 showed the size same fragment present in the genome of CIFC HT 832/2, and with a similar size in *C. eugenioides*. The BAC library was used to select clones related with rust resistance markers. Twenty three BACs were selected with three markers for SH3 resistant locus: 10 for the mark-SAT 244, six to BA-124 and seven for M-8. Fingerprinting analysis showed overlapping patterns of those BACs which should allow the formation of BAC contigs spanning the rust resistance region for SH3. Southern Blot analysis using M-8 probe identified BAC clones in two groups that correspond to the HT subgenomes (C^c- P89O21 and C^e- P135L24). Further characterization of those clones will allow a better comprehension of the mechanisms of genetic resistance for rust in coffee, as well as the identification other molecular markers, are useful for further work in developing markers for assisted selection (MAS), and cloning of genes involved in rust resistance.

Keywords: Hibrido de timor. Physical map. Rust. Markers.

LISTA DE ILUSTRAÇÕES

ARTIGO A

ARTIGO B

LISTA DE TABELAS

ARTIGO A

ARTIGO B

LISTA DE ABREVIATURAS E SIGLAS

AFLP	Amplified Fragment Length Polymorphism
BAC	Bacterial Artificial Chromossome
CIFC	Centro da Investigação das Ferrugem do Café
COSII	Conserved Orthologous Sequence
Cox1	Citocromo Oxidase
DNA	Ácido Desoxirribonucleico
EDTA	Ácido Etilenodiamino Tetra-Acético
ERU	Estação Regional de Uige
EST	Express Sequence Tags
FISH	Fluorescente in Situ Hybridization
FM	Functional Markers
HCl	Ácido Clorídrico
HT	Hibrido de Timor
HMW	High Molecular Weight
IIAA	Instituto Investigação Agronômica de Angola
IPTG	isopropylthio-β-D-galactosidase
LB	Luria Bertani
LG	Linkage Group
M6PR	Manose-6-Phosphate-Oxireductase
NaCl	Cloreto de Sódio
NADPH	Nicotinamide Adenine Dinucleotide Phosphate-Oxidase
NaHCO3	Bicarbonato de Sódio
NaOH	Hidróxido de Sódio
PCR	Polimerase Chain Reaction
PFGE	Pulse Field Gel Electrophoresis
PM	Peso Molecular
QTL	Quantitative Trait Locus
RAPD	Random Amplified Polymorphic
RbcL	Ribulose-1-5-Biphosphatocarboxylae/Oxygenase
RDM	Random DNA Markers
RFLP	Restriction Fragment Length Polymorphism
SCAR	Sequence Characterized Amplified Regions

SSR	Simple Sequence Repeats
TAE	Tris-Acido Acetico EDTA
TBE	Tris Borato EDTA
TE	Tris EDTA
X-Gal	5-bromo-4 - chloro-3-indolyl-β-D-galactosidase
YAC	Yeast Artificial Chromossome

SUMÁRIO

1 INTRODUÇÃO

Com uma produção anual média de 1,3 milhões de toneladas, e uma parcela de aproximadamente 30% do mercado internacional, o Brasil é o principal produtor e exportador de café do mundo, seguido pelo Vietnã, Colômbia e Indonésia. A cadeia produtiva movimenta em torno de US$ 1,6 bilhões por ano (4,5% do PIB agropecuário) e é responsável por cerca de 8 milhões de postos de trabalho diretos e indiretos.

As duas espécies economicamente importantes são *Coffea arábica* L. e *Coffea canephora* (Pierre ex. Froehner), com um total de 4 bilhões de cafeeiros plantados no país. Em Minas Gerais está concentrada a maior área com 1 milhão de hectares em produção, e 50,2% da produção do país. O Espírito Santo é o segundo maior estado produtor e o principal produtor de *C. canephora*, com 479.798 hectares plantados e uma produção de 26,8%. São Paulo detém 7,7% da produção nacional e o Paraná 4,2% (CONAB, 2011).

A produção do café é frequentemente reduzida devido a estresses bióticos e abióticos. A principal doença na cafeicultura é a ferrugem, causada pelo fungo *Hemileia vastatrix* Berk. et Brome. A ferrugem ataca as folhas, provocando lesões que levam à morte dos tecidos (KUSHALAPPA e ESKES, 1989).

A obtenção de cultivares de cafeeiros com resistência durável à ferrugem tem sido dificultada pela grande variabilidade genética das populações de *H. vastatrix*, característica muito comum à maioria dos patógenos biotróficos, especialmente os causadores de ferrugens (Várzea e Marques, 2005). No Brasil, 13 raças já foram confirmadas, sendo a raça II, predominante em cultivos comerciais. Assim, alguns programas de melhoramento genético voltados ao desenvolvimento de cultivares resistentes à ferrugem utilizam como fonte de resistência, progênies oriundas de Hibrido de Timor, planta derivada do cruzamento interespecífico natural entre *C. arabica* e *C. canephora*, resistentes a maioria das raças de fisiológicas de *H. vastratix*.

Para assegurar a produtividade e potencializar o poder competitivo do país na comercialização do café, é necessário constantemente buscar soluções para problemas que podem prejudicar a cafeicultura. A identificação molecular dos fatores de resistência é uma tarefa necessária, para gerar informações úteis para sua aplicação em programas de melhoramento propiciando maior rapidez na seleção de plantas com a característica de interesse e consequente a aceleração do processo de obtenção de cultivares. Estudos em relação à resistência de Híbrido de Timor a *H. vastatrix* utilizando marcadores *SSR* para identificar *QTLs* associados à ferrugem do cafeeiro vem sendo desenvolvidos (LASHERMES et al., 2000). Esses QTLs uma vez identificados são utilizados para a introgressão de genes de resistência à ferrugem em genótipos comerciais, como por exemplo, Catimor, Sachimor, Blumor, Tupi, IAPAR-59, entre outros. Assim a elaboração de estratégias que envolvam o melhoramento convencional associado a técnicas moleculares é de extrema importância para obtenção de marcadores moleculares relacionados ao gene de resistência à ferrugem. (HERRERA et al., 2009).

O Brasil possui uma grande tradição na pesquisa genética cafeeira, principalmente na área de melhoramento. A pesquisa na área biotecnológica de cafeeiros também vem ganhando espaço com recentes trabalhos na área de genoma, fruto do projeto inicial de seqüenciamento de 200.000 *Expressed Sequence Tags* (*ESTs*) (VIEIRA et al., 2006; MONDEGO et al., 2011) sendo a maioria de *C. arabica.* Também foi realizado outro trabalho de transcriptoma pela Universidade de Cornell, onde as maiorias dos *ESTs* são de *C. canephora* (LIN et al., 2005). As informações geradas pelos projetos de *ESTs* vêm sendo utilizadas para identificação de genes, promotores, e regiões de *SSRs* (*Simple Sequence Repeats*) que são empregadas como marcadores em populações de mapeamento genético. Assim o desenvolvimento de tecnologias aplicadas à biologia molecular propicia a produção de informações na área de genomas e mapeamento.

O mapeamento genético de ligação é uma importante ferramenta para auxílio no melhoramento de plantas, permitindo a localização de genes e/ou de locos que controlam características qualitativas ou quantitativas. Em *C. arabica,* há dificuldade de produção de mapas genéticos de ligação devido à baixa diversidade genética na espécie, o que faz que ainda hoje os mapas de ligação sejam parciais, necessitando um aumento de número de marcas e de grupos de ligações a serem cobertos. (Vidal et al., 2010).

Complementando os mapas genéticos e as informações de bancos de trasncriptoma existente, as bibliotecas genômicas baseadas nos vetores do tipo cromossomo artificial de bactéria (*Bacterial Artificial Chromosome-BAC)* tornaram-se uma ferramenta valiosa em estudos genômicos devido a sua capacidade de manter estável fragmentos grandes de DNA e sua facilidade de manipulação (GONTHIER et al., 2010). Essas bibliotecas são recursos importantes para caracterização de regiões genômicas contendo genes de interesse através da construção de mapas fisicos e clonagem posicional. Além disso, as bibliotecas *BAC* são importantes recursos que têm sido utilizados para desenvolvimento marcadores microssatélites (SSR), suporte no seqüenciamento de genomas, estudos evolutivos das famílias multigênicas e caracterização citogenética de genomas através de *BAC-FISH*.

A caracterização dessas bibliotecas para produção de mapas físicos pode ser feita, por exemplo, através de hibridizações de clones *BAC* com *ESTs* gerados pelo Genoma Café e também utilizando marcadores existentes de *SSRs, SCARs, RFLPs* e *AFLPs*. Esses marcadores quando integrados aos mapas físicos, são ferramentas úteis para o entendimento da função e estrutura do genoma, quando o sequenciamento completo do genoma ainda não existe.

Portanto, este trabalho visou à construção e validação de uma biblioteca *BAC* de Hibrido de Timor acesso 832/2. Foram iniciados trabalhos de integração mapas através da identificação de clones *BAC* com marcadores moleculares de mapas de grupos de ligação. Também foram identificados e caracterizados clones *BAC* relacionados a região cromossômica de resistência à ferrugem.

2 REVISÃO DE LITERATURA

2.1 CAFÉ

2.1.1 Aspectos Econômicos

O Brasil é o maior produtor mundial de café, sendo responsável por 30% do mercado internacional, volume equivalente à soma da produção dos demais seis maiores países produtores. É também o segundo mercado consumidor, atrás somente dos Estados Unidos. As áreas cafeeiras estão concentradas no centro-sul do país, onde se destacam quatro estados produtores: Minas Gerais, São Paulo, Espírito Santo e Paraná. A região Nordeste também tem plantações na Bahia e na região norte pode-se destacar o estado de Rondônia. A produção de café arábica se concentra em São Paulo, Minas Gerais, Paraná, Bahia e parte do Espírito Santo, enquanto o café robusta e/ou conilon é cultivado, principalmente, no Espírito Santo e Rondônia (CIC, 2011). Dados da CONAB (Companhia Nacional de Abastecimento) estimaram para a safra de 2011/2012 uma produção entre 43,48 milhões de sacas beneficiadas.

O café arábica representa 74,5% (39,2 milhões de sacas) da produção nacional, e sendo Minas Gerais o maior produtor, com um volume de produção de 26,82 milhões de sacas. No Paraná, apesar das transformações ocorridas na cafeicultura estadual a partir dos anos 70, o café ainda constitui uma importante atividade agrícola para geração de renda e emprego, sendo cultivado nas regiões Norte, Noroeste e Oeste. A produção de café por ser também uma atividade de agricultura familiar tem além da grande importância econômica uma importância social pela mão de obra empregada no campo, constituindo-se, portanto em fonte de renda e emprego para o Brasil.

2.1.2 Características Botânicas do Cafeeiro

O cafeeiro é uma planta de clima tropical úmido, de temperaturas amenas, que encontra no Brasil extensas áreas apropriadas para seu cultivo (TOMAZIELLO, 1996). É uma planta perene pertencente à família Rubiaceae, gênero *Coffea* (VEIRA et al., 2006). Embora mais de 100 espécies desse gênero já tenham sido identificadas (BRIDSON; VERDCOURT, 1988), do ponto de vista comercial, somente as espécies: *C. arabica* L. e *C. canephora* Pierre ex Froehner são extensivamente cultivadas (KUMAR; NAID; RAVISHANKAR, 2006). *C. arabica* L. é a de maior significância econômica (MIRANDA; PEREIRA; BERGO, 1999). Cerca de 70% do café negociado mundialmente é *C.arabica* e 30% *C. canephora* (YU et al., 2011). Espécies como *C. liberica*, *C. dewevrei* e *C. racemosa* são cultivadas para satisfazer a demanda de consumo local (KUMAR; NAID; RAVISHANKAR, 2006) e para a realização de pesquisas relacionadas à introdução de características genéticas de resistência a patógenos.

C. arabica L. é preferido às demais espécies de café devido à melhor qualidade da bebida (DE CASTRO; MARRACCINI, 2006). É um arbusto polimorfo, introduzido no norte do Brasil em 1727 e evidências indicam que é oriunda do sudeste da Etiópia, de regiões montanhosas, entre 1.000 a 2.500 m de altitude (MIRANDA; PEREIRA; BERGO, 1999). *C. arabica* é um alotetraplóide (2n=4x=44), que se reproduz em 90% dos casos por autofecundação (CHARRIER; BERTHAUD, 1985), originado a partir da hibridação natural entre *C. eugenioides* (2n=2x=22) e *C. canephora* (2n =2x =22) (LASHERMES et al., 1997). O *C. arabica* é caracterizado por apresentar baixa variabilidade genética (LASHERMES et al., 1996), que é atribuída a sua recente origem e a sua biologia reprodutiva, isto é, uma planta autógama (PRAKASH et al., 2002).

2.1.3 Ferrugem do Cafeeiro

A ferrugem do cafeeiro tem como agente causal o fungo *Hemileia vastatrix* Berk. et Br., pertencente à ordem Uredinales e família Pucciniaceae, um parasita biotrófico exclusivo do gênero *Coffea*. Esse patógeno do grupo das ferrugens produz os estágios de urédia, télia e basídio, mas somente uredosporos dicarióticos são responsáveis pela doença. A ferrugem do cafeeiro foi descrita em 1869 por Berkeley e Broome, próximo ao lago Vitória, no Leste Africano (Silva et al., 2006). O fungo se desenvolve a partir da superfície abaxial foliar do cafeeiro pela germinação de uredosporos na presença de água e temperatura favorável (18-24ºC), em 6-8 horas, podendo emitir de um a três tubos germinativos. Na extremidade do tubo germinativo ocorre a formação de um apressório sobre um estômato, dando origem, em seguida, a hifa de penetração, a qual, após atravessar o ostíolo do estômato, diferencia-se em vesícula sub-estomática (Guzzo, 2004). O desenvolvimento subseqüente da hifa de infecção na câmara sub-estomática leva à colonização das células subsidiárias e do mesófilo foliar do cafeeiro, com a formação de micélio intercelular e, em seguida, de haustórios (Zambolim et al., 2002). Em cafeeiros susceptíveis, a colonização do mesófilo acaba dando origem aos soros uredospóricos, que esporulam na forma de buquê através dos estômatos (Silva et al., 2006; Rijo & Rodrigues Jr., 1978).

A ferrugem alaranjada é considerada a doença mais importante em *C. arabica* podendo resultar em perdas de até 35% da produção (GUZZO et al., 2009). A ferrugem ataca as folhas, provocando lesões que levam à morte dos tecidos (KUSHALAPPA; ESKES, 1989). Inicialmente, confinada na África e Ásia, esta doença chegou ao Brasil em 1970, mas atualmente encontra-se presente em todos os países produtores de café, exceto no Havaí e Austrália (BRITO et al., 2007). Os sintomas da doença são observados na face inferior das folhas, onde surgem manchas de coloração amarelo-pálida de aspecto pulverulento (uredosporos). Na face superior das folhas observam-se manchas cloróticas amarelas, correspondendo aos limites da pústula na face inferior, que, posteriormente, necrosam e causam a queda da folha infectada (Zambolim *et al*., 1999).

O emprego de controle químico para esta doença mostra-se eficaz pela utilização de fungicidas protetores (ZAMBOLIM et al., 2002). Embora eficientes, os efeitos causados ao meio ambiente e aos organismos não alvos podem conduzir a aumentos populacionais de pragas e/ou de outras doenças do cafeeiro. Adicionalmente, a pressão de seleção exercida sobre o patógeno com esta estratégia de controle, predispõe ao surgimento de novas raças resistentes aos produtos aplicados (ZAMBOLIM et al., 2002). A alternativa mais indicada ao controle químico tem sido a obtenção de cultivares portadores de genes de resistência ao patógeno, a qual é possível ser obtida pelo melhoramento convencional, auxiliado por técnicas moleculares (FAZUOLI et al., 2002; PEREIRA et al., 2002; FAZUOLI et al., 2005). A importância da obtenção de resistência durável a este patógeno reside no fato de que a maioria dos cafeeiros cultivados atualmente são suscetíveis ao fungo causador desta doença (VAN DER VOSSEN, 2001). Assim, o uso de cultivares resistentes é a melhor maneira de controlar esta doença tanto do ponto de vista econômico como ambiental.

C. canephora é uma das principais fontes de genes de resistência para pragas e doenças nos programas de melhoramento do cafeeiro. Por volta de 1917, em uma população de *C. arabica* no Timor Português, atual Timor Leste, foram identificadas plantas proveniente do cruzamento natural entre *C. arabica* e *C. canephora*, (VÁRZEA et al., 2002), denominado de Híbrido de Timor. As plantas provenientes deste híbrido passaram a ser cultivadas comercialmente, em sua localidade de origem, na década de 1940, em substituição às variedades locais. Elas não apresentavam sintomas de ataque do patógeno *H. vastratix*, em contraste com as outras plantas afetadas pela ferrugem do cafeeiro (PEREIRA et al., 2002). Devido a essa característica, o Híbrido de Timor constituiu-se em uma importante fonte de resistência às doenças, para os programas de melhoramento genético do cafeeiro (PEREIRA et al., 2002).

Estudos sobre a herança da resistência à ferrugem realizados no Centro de Investigações das Ferrugens do Cafeeiro (CIFC), em Portugal, demonstraram que a hipótese de Flor (1971) da teoria gene-a-gene é aplicável à interação cafeeiro-ferrugem. Para cada gene que condiciona uma reação de resistência no hospedeiro existe um gene complementar no patógeno que condiciona a avirulência onde a resistência completa é condicionada por pelo menos nove genes dominantes de efeito maior (SH1-SH9) (BETTENCOURT; NORONHA-WAGNER, 1980; BETTENCOURT et al., 1981). Os genes SH1, SH2, SH4 e SH5 foram encontrados em *C. arabica* puros de origem etíope. O gene SH3 é considerado como sendo derivado de *C. liberica*, enquanto SH6, SH7, SH8 e SH9 são oriundos de *C. canephora* (BETTENCOURT; RODRIGUES, 1988). Entretanto, quando alguns genes SH estão suplantados, os cafeeiros podem apresentar resistência incompleta ou parcial a infecções (ESKES et al., 1990).

As plantas derivadas do Híbrido de Timor são tetraplóide, assim como *C. arabica*, o que viabiliza o cruzamento entre estas espécies, permitindo que arábica receba genes de acessos de Híbrido de Timor. Segundo Várzea et al. (2002), muitas das variedades melhoradas e cultivadas comercialmente, tiveram como fonte de resistência o Híbrido de Timor, selecionado no Centro de Investigações das Ferrugem do Cafeeiro (CIFC) com as designações CIFC 832/1, CIFC 832/2 e CIFC 1343. O cruzamento do Híbrido de Timor com variedades comerciais, como Caturra e Villa-Sarchi deu origem as linhagens conhecidas como Catimor (Oeiras, Sábia e Paraíso) e Sarchimor (Tupi, Obatã, IAPAR-59).

Poucos marcadores moleculares ligados a genes de resistência a ferrugem foram obtidos até o momento. Através de estudos de genes envolvidos na resistência a ferrugem, foi observado que marcas *AFLP* co-segregaram com o gene de resistência e foram convertidos em *SCAR* (*Sequence Characterized Amplified Regions*) e usados para seleção de clones BAC (LASHERMES et al., 2001, PRAKASH et al., 2004, MAHÉ et al., 2007).

Visando também estudar os genes da ferrugem do cafeeiro, Diola et al. (2011) construíram um mapa genético de alta densidade com marcadores *AFLP* e *SCARs* flanqueando o gene de resistência a ferrugem para a raça II.

A busca de marcadores moleculares para resistência ferrugem também foi realizada através da mineração em banco de dados de *ESTs*. Caixeta et al. (2007) analisaram 8.968 *ESTs* potencialmente envolvidas com a resistência do cafeeiro, selecionando 831 sequências com diferentes prioridades. Deste trabalho foi obtido somente um marcador potencialmente associado à resistência do cafeeiro à ferrugem.

Para aumentar o número de marcas a serem utilizadas na saturação de mapas físicos de *C. arabica* Hibrido de Timor, a construção de bibliotecas *BAC* se faz interessante, pois a identificação de clones *BAC* com os marcadores moleculares possibilitam a geração de novas marcas para as regiões cromossômicas de resistência à ferrugem do cafeeiro. Os *BAC* identificados além de propiciar o adensamento dos mapas de ligação, podem ser utilizados na clonagem posicional do gene de resistência.

2.1.4 Bibliotecas *BAC*

O trabalho de construção de mapas físicos de genomas completos era difícil principalmente pela falta de sistemas de clonagem de DNA que acomodassem fragmentos grandes onde fosse possível uma análise detalhada através de *fingerprinting*. Essa primeira barreira foi inicialmente superada com a utilização dos vetores do tipo cromossomo artificial de levedura (*Yeast Artificial Chromossome-YAC*). Entretanto, os *YAC* tem algumas desvantagens como o nível de recombinação e formação de quimeras que dificultam o mapeamento. Além disso, os métodos de isolamento de DNA para análise de um grande número de clones são demorados. O desenvolvimento de vetores de clonagem baseados em

BAC veio solucionar em parte os problemas encontrados com os *YAC*. Estes vetores praticamente não apresentam rearranjos cromossomais, e se comportam como um vetor plasmidial em *E. coli*, facilitando a extração e manipulação do DNA, apesar do baixo rendimento quantitativo.

Os vetores do tipo *BAC* permitem a inserção e clonagem de fragmentos de DNA genômico de 60 a 300 Kb (SHIZUYA et al., 1992). Ao contrário dos *YAC*, ele apresenta uma série de vantagens, como estabilidade dos clones, baixo quimerismo e melhor eficiência de transformação (FRITGERS et al., 1997). A estabilidade e o baixo quimerismo ocorre devido ao fator F que permite que o vetor seja mantido em pequeno número de cópias na bactéria (LUO et al., 2003; CHALHOUB et al., 2004).

Bibliotecas *BAC* tem se tornado uma ferramenta central na pesquisa genética devido a sua facilidade de manipulação. Estas bibliotecas já foram construídas para diferentes organismos onde destacam-se: *Arabidopsis* (MOZZO et al., 1999), sorgo (KLEIN et al., 2000), arroz (JENA; Mackill, 2008), soja (WU et al., 2004), maçã (HAN et al., 2007), algodão (KELLEHER et al., 2007) e tem sido empregadas em uma variedade de aplicações como o sequenciamento de clones que possibilita a estratégia de *BAC-end sequencing* (VENTER et al., 1998). Também tem facilitado o mapeamento de regiões de genes de cópia simples possibilitando uma forma de buscar essas regiões no genoma (PETERSON et al., 2000). O uso de clones *BAC* combinado ao desenvolvimento de métodos de *fingerprinting*, sequenciamento de *BAC-ends* e busca de genes de cópia única, tem possibilitado o preenchimento de lacunas no mapeamento de espécies de genoma grande. Clones *BAC* também podem ser utilizados em experimentos de *FISH* (*Fluorescence in situ hybridization*). A localização de DNA clonado como sonda diretamente nos cromossomos através de *FISH* possibilita a superposição direta de mapas genéticos e físicos (HERRERA et al., 2007).

Uma das formas de construção de mapas físicos a partir de bibliotecas *BAC* é através de técnicas de seleção por PCR de *pool* de *BAC*. Para tanto é necessário que a biblioteca seja condensada em *pools* envolvendo combinações somente de placas ou combinações de colunas e linhas. Essa combinação é realizada de forma que um mínimo de reações de PCR seja necessário para identificação de um clone específico dentro da biblioteca (GREEN; OLSEN 1990; BRUNO et al., 1995). Combinadas, as técnicas de PCR e de *pools* da biblioteca representando vários grupos de clones propiciam uma forma econômica de seleção de clones, pois com um pequeno número de reações é possível encontrar uma única sequência alvo (GARDINER et al., 2004). A partir de uma biblioteca *BAC* de algodão organizadas em pools e superpools, Hu et al. (2010) realizaram a triagem com vários marcadores baseados em PCR, para ser utilizada em mapeamento físico e análise comparativa de genes. Yim et al. (2007), desenvolveram uma estratégia de *BAC pooling* para seleção de clones BAC de milho utilizando PCR baseado em marcadores *STS.*

Bibliotecas *BAC* foram construídas tanto para *C. arabica* e *C. canephora* servindo como base para o estudo da estrutura do genoma do café, genômica comparativa, biologia do desenvolvimento e evolução (NOIR et al., 2004; KOCHKO et al., 2010), no entanto nenhuma utilizou a técnica de pools e superpools de BAC para seleção de clones, além disso nenhuma das bibliotecas de café estão disponíveis.

Yu et al. (2011), identificaram dois *BAC* homeólogos de uma biblioteca *BAC* de arábica Mokka Tall, e compararam as seqüências com o ortólogo em *C. canephora* para investigar os padrões e grau de divergência entre as seqüências dos genomas de *C. arabica* e *C. canephora*. Através dos clones *BAC* dessa biblioteca foi possível realizar um estudo comparativo entre *C. arabica* (tetraplóide) e *C. canephora* (diplóide) fornecendo uma oportunidade para estudar a história evolutiva desses dois genomas.

Nas análises de genoma poliplóide Mahé et al., (2008) utilizou uma biblioteca de *C. arábica* para reconstruir a filogenia de lócus homólogos em genoma alotetraplóide para melhor investigar os mecanismos de especiação de *C. arabica*.

Utilizando uma biblioteca de *C. canephora* Guyot et al. 2009 realizou o sequenciamento completo do clone *BAC* 46C02, visando a identificação do gene CcEIN4, que codifica um receptor de etileno. A análise da sequência deste clone *BAC* de 160 Kb indicou uma alta densidade de genes ativos e com poucos elementos transponíveis. A análise comparativa da sequência indicou uma extensiva conservação entre *C. canephora* e da maioria dos genomas referências de dicotiledôneas estudados como, *Solanum Lycopersicum, Vitis vinifera, Medicago Truncatula, Populus trichocarpa* e *Arabidopsis thaliana.* Alto grau de microlinearidade foi encontrado entre *C. canephora* e *V. Vinifera*, que pertencem respectivamente aos

Asterids e Rosids, dois clados que divergem em mais de 114 milhões de anos atrás. (GUYOT et al., 2009).

Lashermes et al. (2010) mapearam fisicamente o gene SH3 para resistência a ferrugem do cafeeiro, usando uma biblioteca *BAC*. Este mapa físico consiste em dois contigs homólogos de 1.170 e 1.208 Kb, correspondente aos subgenomas de *C. canephora* e *C. eugenoides*, respectivamente. O segmento do cromossomo *C. liberica* que carrega o gene SH3 parece introgredidos no sub-genoma de *C. canephora*.

Vários trabalhos vêm utilizando *ESTs* para seleção de bibliotecas *BAC* e mapeamento físico. Os marcadores selecionados a partir de *ESTs* vêm sendo denominados de marcadores funcionais (FM) e a vantagem sobre marcadores randômicos de DNA (RDM) como *RFLP*, *SSRs* e *RAPD*, é a sua ligação com os loci dos alelos de características desejáveis (ANDERSEN; LÜBBERSTEDT, 2003). No mapeamento de milho, mais de 9000 *ESTs*, em forma de *overgo,* foram utilizados como “âncoras” para produção de contigs e mapeamento físico, através da técnica de hibridização multiplex de filtros de membranas de colônia de *BAC* (GARDINER et al., 2004). Em arroz, cerca de 1800 marcadores *ESTs* foram utilizados para mapeamento *in silico* juntamente com outros marcadores do tipo RDM (YUAN et al., 2000).

Portanto é de grande importância a aplicação de projetos integrados de mapeamento para aumentar a saturação destes mapas. A utilização de clones *BAC* e de regiões contíguas de DNA por mapas físicos, pode auxiliar na saturação dos mapas de ligação através da descoberta de novas marcas mais próximas a genes e/ou QTLs de interesse.

2.1.5 Mapeamento Físico e Integração de Mapas

Os mapas genéticos de ligação são um importante recurso para auxílio no melhoramento de plantas, permitindo a localização de genes e/ou de locos que controlam características qualitativas ou quantitativas (QTL) (BLINDER et al., 2011). Em *C. arabica* a dificuldade de produção de mapas genéticos de ligação é grande, devido ao fato de ser uma planta perene como baixa diversidade genética da espécie, com uma baixa frequência de polimorfismos. Isto faz com que os mapas existentes sejam parciais, necessitando um aumento de número de marcas e de grupos de ligações a serem cobertos para saturação e formação dos grupos de ligação. Desta forma, a partir do mapa genético e da identificação de clones BAC que contenham a mesma sequência do marcador, é possível obter o mapa físico. (DIAZ et al., 2011).

A produção de mapas físicos ancorados com marcadores moleculares tem o potencial de ampliar os conhecimentos da estrutura genômica, bem como da localização, funcionamento e regulação dos genes. A construção de mapas integrados com marcadores randômicos de DNA e marcadores funcionais é uma das mais úteis ferramentas para o entendimento da função e estrutura do genoma, quando o sequenciamento completo não for possível (HAN, et al., 2011). Sondas baseadas em marcadores *RFLP* e *SSRs* também vêm sendo utilizadas para seleção dos *BAC* e formação dos contigs, essas sondas servem de âncoras para comparação dos mapas.

No caso de marcas associadas à QTLs, a integração permite gradualmente saturar uma região através do mapeamento físico com genes candidatos a partir de *ESTs*, de genes *COS*II (*Conserved Orthologous Sequence*) e de seqüenciamento, além da utilização das marcas provenientes dos mapas de ligação obtidos por *AFLPs, SSRs* e *SCARs* (WU et al., 2006). Esta estratégia integrada é atualmente um dos principais alvos de pesquisa de diversos grupos de melhoramento das principais espécies cultivadas, sejam anuais ou perenes.

O sucesso do desenvolvimento de um mapa físico baseado em *BAC* associado a uma grande cobertura do genoma é essencial para completar o sequenciamento de todo o genoma e depende de vários fatores, incluindo o tamanho dos insertos, a representação do genoma na biblioteca, o tamanho e a complexidade do genoma alvo. O tamanho dos insertos dos clones *BAC* reduzem não só o número de clones necessários para desenvolvimento do mapa, mas também assegura o sucesso no desenvolvimento de um mapa físico de alta qualidade baseado em *BAC*. A característica linear do mapeamento físico também abre portas para pesquisas em genoma estrutural, organização genômica e evolução, regulação e interação gênica. Permite também que sejam identificados genes e clusters de genes de interesse, através de técnicas de clonagem baseadas em mapeamento (YIM et al., 2007). Portanto, o mapa físico pode centralizar diversos estudos genéticos e moleculares que por sua vez deverão propiciar uma maior informação genética e sua aplicação para o melhoramento da espécie. Assim, a partir da identificação e sequenciamento de clones *BAC* que contenham a mesma sequencia do marcador, é possível realizar a integração de mapas fisicos e genéticos. Uma vez analisadas, essas marcas podem ser empregadas como marcadores moleculares em populações de mapeamento e por meio destes clones, apontar os primeiros possíveis loci dentro da biblioteca *BAC*.

3 REFERÊNCIAS

ANDERSEN, R.; LUBBERSTEDT, T. Functional markers in plants. **TRENDS in Plant Science**, v. 8 p. 554-560, 2003.

BETTENCOURT, A. J.; RODRIGUES, J. R. C. J. Principles and practice of coffee breeding for resistance to rust and other disease. **Elsevier Applied Science** In: Clarke, R. J,; Macrae, R. (Org.). **Coffee**. Agronomy, v. 4, p. 199–234, 1988.

BETTENCOURT, A. J. **Melhoramento genético do cafeeiro: transferência de factores de resistência à Hemileia vastatrix Berk. et Br. para as principais cultivares de *Coffea arabica* L.** Tese de Doutorado, Centro de Investigação das Ferrugens do Cafeeiro, Lisboa, Portugal. 1981.

BETTENCOURT, A. J.; NORONHA-WAGNER, M.; LOPES, J. Factor genético que condiciona a resistência do clone 1343/269 (Híbrido de Timor) à *Hemileia vastatrix* Berk. et Br. **Broteria Genética**, v. 77, p. 53-58. 1980.

BINDLER, G.; PLIESKE, J.; BAKAHER, N.; GUNDUZ, I.; IVANOV, N.; VAN DER HOEVEN, R.; GANAL, M.; DONINI, P. A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. **Theorical Applied Genetics,** v.123, p. 219–230. 2011.

BRIDSON, D. M.; VERDCOURT, B. **Flora of tropical East Africa:** Rubiaceae. (Part 2). Cape Town: Iziko Museums of Cape Town,. p. 415-747, 1988.

BRITO, G. G. **Mapeamento genético de marcadores AFLP ligados ao gene de resistência do cafeeiro a *Hemileia vastatrix*.** Tese de doutorado, Fisiologia Vegetal. Universidade Federal de Viçosa, Viçosa-MG. 98 p. 2007.

BRUNO W. J.; KNILL E.; BALDING D. J.; BRUCE D. C.; DOGGET N. A.; SAWHILL W. W.; STALLINGS R. L.; WHITTAKER C. C.; TORNEY D. C. Efficient pooling designs for library screening. **Genomics,** v. 26, p. 21-30, 1995.

CABRAL, T. A. T.; SAKIYAMA, N. S.; ZAMBOLIM, L.; PEREIRA, A. A.; SCHUSTER, I. Single-locus inheritance and partial linkage map of Coffea arabica L. **Crop breeding and applied biotechnology**, Viçosa, v. 4, p. 416-421, 2004.

CAIXETA, E. T.; ALVARENGA, S. M.; THIEBAUT, F.; HUFNAGEL, B.; ZAMBOLIM, E. M.; XAVIER, K. V.; SAKIYAMA, N. S. (2007). **Anais do V Simpósio de Pesquisa dos Cafés do Brasil**. Brasília: Embrapa CD-ROM. Localização de marcadores moleculares derivados de EST potencialmente associado ao domínio LRR em mapa genético de *Coffea arabica*. In: V Simpósio de Pesquisa dos Cafés do Brasil, 2007, Águas de Lindóia.

CAPUCHO, A. S. **Herança e mapeamento de qtls da resistência do hibrido de Timor à ferrugem do cafeeiro**. Dissertação de Mestrado. Universidade Federal de Viçosa, 74 p. 2008.

CHARRIER, A.; BERTHAUD, J. Botanical classification of coffee. In: Clifford, M.N. & Wilson, K.C. (ed.). **Coffee: botany, biochemistry and production of beans and beverage**. London: Croom Helm, p.13-47, 1985.

CIC. Centro de Intelegência do Café. 2011. Disponível em: http://www.cicbr.org.br/. Acesso em: 21 de dezembro de 2011.

CONAB. Companhia Nacional de Abastecimento. 2010. Disponível em: <http://www.conab.gov.br>. Acesso em: 12 ago 2011.

COULIBALY, I.; REVOL, B.; NOIROT, M.; PONCET, V.; LORIEUX, M.; CARASCO-LACOMBE, C.; MINIER, J. DUFOUR, M.; HAMON, P. AFLP and SSR polymorphism in a *Coffea* interespecif backcross progeny [(*C. heterocalyx* x *C. canephora*) x *C. canephora*. **Theorical Applied Genetics**, v. 107, p.1148-1155, 2003.

DE CASTRO, R. D.; MARRACCINI, P. Cytology, biochemistry and molecular changes during coffee fruit development. **Brazilian Journal of Plant Physiology**, v.18, n.1, p.175-199, 2006.

DIAZ, A.; FERGANY, M.; FORMISANO, G.; ZIARSOLO P. A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.). **BMC Plant Biology**, v.11 p. 1-14, 2011.

DIOLA, V.; BRITO, G. G.; CAIXETA, E. T.; MACIEL-ZAMBOLIM, E.; SAKIYAMA, N. S.; LOUREIRO, M. E. High-density genetic mapping for coffee leaf rust resistance. **Tree Genetics and Genomes,** v.*7*, p. 1199-1208, 2011.

ESKES, A. B.; HOOGSTRATEN, J. G. J.; TOMA-BRAGINI, M.; CARVALHO, A. Race-specificity and inheritance of incomplete resistance to coffee leaf rust in some Icatu coffee progenies and derivatives of Hibrido de Timor. **Euphytica,** v. 47, p.11-19. 1990.

ESKES, A. B.; HOOGSTRATEN, J. G. J.; TOMA-BRAGINI, M; CARVALHO, A; FAZUOLI, L. C.; MEDINA FILHO, H. P.; GONZALVES, W.; GUERREIRO FILHO, O.; SILVAROLLA, M. B. Melhoramento do Cafeeiro: variedades tipo arábica obtidas no Instituto Agronômico em Campinas. In: Zambolin, L. (ed). **O estado da arte de tecnologias na produção de café**. UFV, Viçosa, p. 63-215, 2002 .

FAZUOLI, L.C.; MEDINA, FILHO, H. P.; GONZALVES, W.; GUERREIRO, FILHO, O.; SILVAROLLA, M.B. Melhoramento do Cafeeiro: variedades tipo arábica obtidas no Instituto Agronômico em Campinas. In: Zambolin, L. (ed). **O estado da arte de tecnologias na produção de café**. UFV, Viçosa, p.63-215, 2002.

FAZUOLI, L.C.; OLIVEIRA, A.C.B., TOMA-BRAGUINI, M., SILVAROLLA, M.B.; Identification and use of sources of durable resistence to coffee leaf rust at the IAC. In: Zambolin, L. (ed). **Durable resistance to coffee leaf rust.** UFV, Viçosa, p.137-185, 2005.

FRITGERS, A.C.J.; ZHANG, Z.; VAN DAMME, M.; WANG, G.L.; RONALD, P.C.; MICHELMORE, R. W. Construction of a bacterial chromosome library containing large *Eco*RI and *Hin*dIII genomic fragments of lettuce. **Theorical Applied Genetics,** v.94, p.390-399, 1997.

GARDINER, J.; SCHROEDER S.G; 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. Anchoring 9.371 maize expressed sequenced tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. **Plant Physiology**, v. 134, p.1317-1326, 2004.

GONTHIER, L.; BELLEC, A.; BLASSIAU, C.; PRAT, E.; HELMSTETTER, N.; RAMBAUD, C.; HUSS, B.; HENDRIKS, T.; BERGÈS, H.; QUILLET, M-C. Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). **BMC Research,** v.3, p. 225, 2010.

GREEN, E. D.; OLSON, M. V. Systematic screening of yeast artificial-chromossome libraries by use of the polymerase chain reaction. **Proceedings of the National Academy of Sciences USA, v.** 87 p. 1213-1217, 1990.

GUYOT, R.; DE LA MARE, M.; VIADER, V.; HAMON, P.; CORITON, O.; BUSTAMANTE-PORRAS, J.; PONCET, V.; CAMPA, C.; HAMON, S.; DE KOCHKO, A. 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 Biology**, p. 9-22, 2009.

GUZZO, S. D.; **Aspectos bioquímicos e moleculares da resistência sistêmica adquirida em cafeeiro contra *Hemileia vastatrix***. Tese de doutorado, Piracicaba, Centro de Energia Nuclear na Agricultura, 236 p. 2004.

GUZZO, S. D.; HARAKAVA, R.; TSAI, S. M. Identification of coffee gene expressed during systemic acquired resistance and incompatible interaction with *Hemileia vastatrix*. **Journal Phytopathology**, v. 1, p.1-14, 2009.

HAN, Y.; GASIC, K.; MARRON, B.; BEEVER, J. E.; KORBAN, S. S. A BAC-based physical map of the apple genome. **Genomics,** v. 89. p. 630-637, 2007.

HAN, Y.; ZHENG, D.; VIMOLMANGKANG, S.; KHAN, M. A.; BEEVER, J. E.; KORBAN S. S. Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome. **Journal of Experimental Botany,** v. 62, p. 5117–5130, 2011.

HERRERA, P. J. C.; ALVARADO, A. G; CORTINA, G. H A.; COMBES, M-C; ROMERO, G. G; LASHERMES, P. Genetic analysis of partial resistance to coffee leaf rust (*Hemileia vastatrix* Berk & Br.) introgressed into the cultivated *Coffea arabica* L. from the diploid *C. canephora* species. **Euphytica**, v. 167, p. 57-67, 2009.

JENA, K. K.; MACKILL, D. J. Molecular markers and their use in marker-assisted Selection in Rice. **Crop Science***,* v. 48, p. 1266-1276, 2008.

KELLEHER, C. T.; CHIU, R.; SHIN, H.; BOSDET, I. E.; KRZYWINSKI, M. I.; FJELL, C. D.; WILKIN, J.; et al. A physical map of the highly heterozygous *Populus* genome: integration with the genome sequence and genetic map and analysis of haplotype variation. **Plant Journal***,* v. 50, p. 1063-1078, 2007.

KLEIN, P. E.; KLEIN, R. R.; CARTINHOUR, S. W.; ULANCH, P. E.; DONG, J.; OBERT, J. A.; MORISHIGE, D. T.; SCHLUETER, S. D.; CHILDS, K. L.; ALE, M.; MULLET, J. E. A highthroughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. **Genome Reserch**, v. 10, p.789-807, 2000.

KUMAR, V.; NAIDU, M. M.; RAVISHANKAR, G. A. Developments in coffee biotechnology - in vitro plant propagation and crop improvement. **Plant Cell, Tissue and Organ Culture**, v. 7, p.49–65, 2006.

KUSHALAPPA, A. C.; ESKES, A.B. **Coffee rust: epidemiology, resistance and management**. Boca Raton: CRC Press. 345 p. 1989.

KUSHALAPPA, A. C.; ESKES A. B. Advances in coffee rust research. **Annu Review Phytopathology,** v. 27, p. 503-531, 1989.

KY, C. L.; BARRE, P.; LORIEUX, M.; TROUSLOT, P.; AKAFFOU, S.; LOUARN, J.; CHARRIER, A.; HAMON, S.; NOIROT, M. Interespecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theorical Applied Genetics**, v. 101, p. 669-676, 2000.

LASHERMES, P.; PACZEK, V.; TROUSLOT, P.; COMBES, M-C.; COUTURON, E. E; CHARRIER, A. Single-*locus* inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C arabica x C. canephora*. **Journal Heredity,** v. 91, p. 81-85, 2000.

LASHERMES, P.; COMBES, M. C.; TROUSLOT, P.; CHARRIER, A. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. **Theoretical Applied Genetics**, v. 94, p. 947–955, 1997.

LASHERMES, P.; COMBES, M.C.; PRAKASH, N.S.; TROUSLOT, P.; LORIEUX, M.; CHARRIER, A. Genetic linkage map of *Coffea canephora*: effect of segregation rate in male and female meioses. **Genome**, v. 44, p. 589-596, 2001.

LASHERMES, P.; COUTURON, E.; MOREAU, N.; PAILLARD, M.; LOUARN, J. Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. **Theorical Applied Genetics,** v. 93, p. 458-462, 1996.

LASHERMES, P.; COMBES, M. C.; RIBAS, A.; CENCI, A.; MAHÉ, L.; ETIENNE H. Genetic and physical mapping of the SH3 region that confers resistance to leaf rust in coffee tree (*Coffea arabica* L). **Tree Genetics and Genomes**, v. 6, p. 973-980, 2010.

LEROY, T.; MARRACCINI, P.; DUFOUR, M.; MONTAGNON, C.; LASHERMES, P.; SABAU, X.; FERREIRA, L. P.; JOURDAN, I.; POT, D.;, ANDRADE, A. C.; GLASZMANN, J. C.; VIEIRA, L. G. E.; PIFFANELLI, P. Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes, **Theorical Applied Genetics**, v. 111, p. 1032-1041, 2005.

LIN, C.; MUELLER, L. A.; MCCARTHY, J.; CROUZILLAT, D.; PÉTIARD, V.; TANKSLEY, S. D. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. **Theorical Applied Genetics,** v. 112, p. 114–13, 2005.

LUO, M. C.; THOMAS, C.; YOU, F. M.; HSIAO, J.; OUYANG, S.; BUELL, C.R.; MALANDRO, M.; MCGUIRE, P. E.; ANDERSON, O. D.; DVORAK, J. High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis, **Genomics**, v. 82, p. 378-389, 2003.

MAHÉ, L.; COMBES, M. C.; VÁRZEA, V. M. P.; GUILHAUMON, C.; LASHERMES, P. Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica*). **Molecular Breeding**, v. 21, p.105–113, 2008.

MAHÉ, L.; COMBES, M. C.; LASHERMES, P. 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 Molecular Biology,** v. 64, p. 699–711, 2007.

MIRANDA, E. M.; PEREIRA, R. C. A.; BERGO, C. L. Comportamento de seis linhagens de café (*Coffea arabica* L.) em condições de sombreamento e a pleno sol no estado do Acre. **Ciência e Agrotecnologia**, v. 23, p. 62-69, 1999.

MONDEGO, J. M. C.; VIDAL R. O.; CARAZZOLLE, M. F.; TOKUDA, E. K.; PARIZZI, L. P.; COSTA, G. G. L.; PEREIRA, L. F. P.; ANDRADE, A. C.; COLOMBO, C. A.; VIEIRA, L. G. E.; PEREIRA, G. A. G. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*, **BMC Plant Biology,** v. 11, p. 1-30, 2011.

MOZO, T.; DEWAR, K.; DUNN, P.; ECKER, J. R.; FISCHER, S.; KLOSKA, S.; LEHRACH, H.; MARRA, M.; MARTIENSSEN, R.; MEIER-EWERT, S.; ALTMANN, T. A complete BAC-based physical map of the *Arabidopsis thaliana* genome. **Nature Genetics,** v. 22, p. 271-2750, 1999.

NOIR, S.; PATHEYRON, S.; COMBES, M. C.; LASHERMES, P.; CHALHOUB, B. Construction and characterization of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.) **Theorical Applied Genetics**, v. 109, p. 225-230, 2004.

PAILARD, M.; LASHERMES, P.; PETIARD, V. Construction of a molecular linkage map in coffee. **Theorical Applied Genetics,** v. 93, p. 41-47, 1996.

PEARL, H. M.; NAGAI, C.; MOORE, P. H.; STEIGER, D. L.; OSGOOD, R. V.; MING, R. Construction of a genetic map for coffee. **Theorical Applied Genetics,** v. 108, p. 829-835, 2004.

PEREIRA, A. A.; MOURA, W, M.; ZAMBOLIM, L.; SAKIYAMA, N. S.; CHAVES, G. M. 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, p. 253-295, 2002.

PETERSON, D. G.; TOMKINS, J. P.; FRISCH, D. A.; WING, R. A.; PATERSON, A. H. Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrade guide. **Journal of Agriculture Genomics**. Disponível em: <http://www.ncgr.org/research/jag>. Acesso em: 2000.

PRAKASH, N. S.; COMBES, M. C.; SOMANNA, N.; LASHERMES, P. AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. **Euphytica**, v. 124, p. 265–271, 2002.

PRAKASH, N. S.; MARQUES, D. V.; VARZEA, V. M. P.; SILVA, M. C.; COMBES, M. C.; LASHERMES, P. Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *Coffea arabica* L. **Theoretical Applied Genetics**, v. 109, p. 1311-1317, 2004.

RIJO, L.; RODRIGUES, C. J. Processo de infecção da *Hemileia vastatrix* Berk. et Br. em cultivares susceptíveis e resistentes de *Coffea rábica* L. **Garcia de Orta** v. 5, p. 23-24, 1978.

SHIZUYA, H.; BIRREN, B.; KIM, U. J.; MANCINO, V.; SLEPAK, T.; TACHIRI, Y; SIMON, M. Cloning and stable maintenance of 300-kilobase–pair fragments of human DNA in *E. coli* using an F-factor-based vector. **Proceeds National Academic Science USA**, v. 89, p. 8794-8797, 1992.

SILVA, M. D. C.; VÁRZEA, V.; GUIMARÃES, L. G.; AZINHEIRA, H. G.; FERNANDEZ, D.; PETITOT, A.S.; BERTRAND, B.; LASHERMES, P.; NICOLE, M. Coffee resistance to the main diseases: leaf rust and coffee berry disease. **Brazilian Journal of Plant Physiology.** v. 18, p.119-147, 2006.

TANKSLEY, S. D.; YOUNG, N. D.; PATERSON, A. H.; BONIERBALE, M. W. RFLP mapping in plant breeding: New tools for an old science. **Biotechnology**, v. 7, p. 257–264, 1989.

TOMAZIELLO, R. A.; OLIVEIRA, E. G.; TOLEDO FILHO, J. A.; COSTA, T. E. **Cultura do Café**, 2. ed. Campinas: CATI, 1996, p. 69.

VAN DER VOSSEN, H. Coffee breeding practices. In: Clarke, R. J.; Vitzthum, O. G (eds) Coffee Recent Developments. **Agronomy**, v. 1, p. 184-201, 2001.

VÁRZEA, V. M. P.; RODRIGUES JUNIOR, C. J.; SILVA, M DO C. M. L.; GOUVEIA, M.; MARQUES, D. V.; GUERRA-GUIMARÃES, L.; RIBEIRO, A. Resistência do cafeeiro a *Hemileia vastatrix.* **O estado da arte de tecnologias na produção de café.** Viçosa: UFV, 2002. cap.8, p.297-320.

VÁRZEA, V. M. P.; MARQUES, D.V. Population variability of *Hemileia vastatrix* vs. coffee durable resistance. p.53-74 In: ZAMBOLIM, L., ZAMBOLIM, E.M.; VÁRZEA, V.M.P. (Eds.) Durable Resistance to Coffee Leaf Rust. Viçosa: Universidade Federal de Viçosa. 2005.

VENTER, J. C.; SMITH, H. O.; HOOD, L. A new strategy for genome sequencing. **Nature**, v. 381, p. 364-366, 1996.

VENTER, J. C.; ADAMS, M. D.; SUTTON, G. G.; KERLAVAGE, A. R.; SMITH, H. O.; HUNKAPILLER, M. Shotgun sequencing of the human genome. **Science,** v. 280, p. 1540-1542. 1998.

VIDAL, R. O.; MONDEGO, J. M. C.; POT, D.; AMBRÓSIO, A. B.; ANDRADE, A. C.; PEREIRA, L. F. P.; COLOMBO, C. A.; VIEIRA, L. G. E.; CARAZZOLLE, M. F. ; PEREIRA G. A. G. 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 Physiology,** v. 154, p. 1053-1066, 2010.

VIEIRA, L. G. E.; ANDRADE, A. C.; COLOMBO, C. A.; MORAES, A. H. A.; METHA, A., et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazillian Journal Plant Physiology***,* v. 18, p.95-108, 2006.

WU, C.; SUN, S.; NIMMAKAYALA, P.; SANTOS, F. A.; MEKSEM, K.; SPRINGMAN, R.; DING, K.; LIGHTFOOT, D.A.; ZHANG, H. B. A BAC- and BIBAC-based physical map of the soybean genome. **Genome Reserch**, v. 14, p. 319-326, 2008.

WU, F.; MUELLER L. A.; CROUZILLAT, D.; PÉTIARD, V.; TANKSLEY, S. D. Combining Bioinformatics and Phylogenetics to Identify Large Sets of Single-Copy Orthologous Genes (COSII) for Comparative, Evolutionary and Systematic Studies: A Test Case in the Euasterid Plant Clade. **Genetics**, v. 174, p. 1407–1420, 2006.

YIM, Y-S.; MOAK, P.; SANCHEZ-VILLEDA, H.; MUSKET, T.A.; CLOSE, P.; KLEIN, P.E.; MULLET, J.E. 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,** v. 8, p. 47, 2007.

YU, Q.; GUYOT, R.; KOCHKO, A.; DE BYERS, A.; NAVAJAS-PÉREZ, R.; LANGSTON B, J.; DUBREUIL-TRANCHANT, C. D.; PATERSON, A. H.; PONCET, V.; NAGAI, C.; MING, R. Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*). **Plant Journal,** v. 67, p. 305–317, 2011.

YU, Y.; TOMKINS, J. P.; WAUGH, R.; FRISCH, D. A.; KUDRNA, D.; KLEINHOFS, A.; BRUEGGEMAN, R. S.; MUEHLBAUER, G. J.; WISE, R. P.; WING, R. A bacterial artificial chromosome library for barley (Hordeum vulgare L.) and the identification of clones containing putative resistance genes. **Theoretical Applied Genetics**, v. 101, p. 1093-1099, 2000.

YUAN, Q.; LIANG F.; HSIAO J.; ZISMANN V.; BENITO M. I.; QUACKENBUSH J.; WING, R.; BUELL, R. Anchoring of rice BAC clones to the rice genetic map *in silico*. **Nucleic Acids Research**, v. 28, p. 3636-3641, 2000.

ZAMBOLIM, L.; VALE, F. X. R. D.; PEREIRA, A. A.; CHAVES, G. M.; Manejo integrado das doenças do cafeeiro. p.134-215 In: ZAMBOLIM, L. (Ed.) Produção de café com qualidade. Viçosa-MG: Universidade Federal de Viçosa. 1999.

ZAMBOLIM, L.; VALE, F. X.R.; COSTA, H.; CHAVES, G.M. Epidemiologia e controle integrado da ferrugem do cafeeiro In: new technique for DNA fingerprint. **Nucleic Acids Research,** v. 23, p. 4407-4414, 2002.

4 ARTIGO A

CONSTRUCTION AND CHARACTERIZATION OF A *Coffea arabica* BAC LIBRARY

4.1 ABSTRACT

Most of the coffees produced in the world are from *Coffea arabica,* an allotetraploid specie, with low genetic diversity and few genomic resources available. Genomic libraries with large insert-DNA are useful tools for studies of the plant genome, including the production of physical maps, integration studies of physical and genetic maps, genome structure analysis and gene isolation by positional cloning. As an initial step towards the integration of strategies for physical and genetic mapping, as well as to perform map based cloning for identification of genes of interest, a BAC library of Hibrido Timor access CIFC 832/2 was constructed. We performed the characterization of this library for insert size, chloroplast and mitochondrial contamination. The BAC library was screening to select BAC within *M6PR* gene and with markers of four linkage groups from a *C. arabica* genetic map. We report the construction and characterization of a BAC library consisting of 56,832 clones with an average insert size of 118 kb, representing 5.6-fold haploid genome coverage. Based upon the percentage of positive clones identified, the contamination was estimated in 1.04% for chloroplast and 0.5% for mitochondrial DNA respectively. A mixed approach using PCR selection of BAC pools and membrane hybridization, allowed the selection of 53 BAC clones which mapped four anchor genetics markers and nine BAC clones with M6PR gene related to response to abiotic and biotic stress. The high-quality and large fragments of HT CIFC 832/2 BAC library is a valuable resource for further studies in *Coffea arabica* genome characterization.

4.2 INTRODUCTION

Coffee is one of the most important beverage crops in the world. It is cultivated in over 60 countries in the tropical and subtropical regions. Approximately 100 species of the genus *Coffea* L. (Rubiaceae) are known. The most commercially important species are *C. arabica* L., also known as arabica coffee, and *C. canephora* (Pierre ex Froehner), known as robusta or conillon coffee, which represent, respectively, about 70 % and 30 % of the total coffee market (MONDEGO et al., 2011). *C. arabica* L. ($2n=4x=44$), the only polyploid species in the *Coffea* genus, is an allotetraploid containing two diploid subgenomes, originated from two ancestral species of *C. canephora* and *C. eugenioides* (LASHERMES et al., 2000).

"Híbrido de Timor" (HT) is a hybrid *Coffea arabica* genotype derived from a spontaneous interspecific cross between *C. arabica* and *C. canephora* and has been used intensively in coffee breeding programs for introgression of resistance to pests and diseases including coffee leaf rust (*Hemileia vastatrix*), Coffee Berry Disease (CBD - *Colletotrichum kahawae*), bacterial disease (*Pseudomonas syringae* pv. *Garcae*) and root-knot nematode (*Meloidogyne exigua*) (OROZCO-CASTILLO et al., 1994; CHARRIER; ESKES, 1997). The first introduction of HT accessions in Brazil date back from 1976 via vegetative propagation and seeds from CIFC (Centro da Investigação das Ferrugem do Café, Portugal), IIAA (Instituto Investigação Agronomia de Angola) and ERU (Estação Regional de Uige) (PEREIRA et al., 2002). Exploitation of HT populations has so far relied on conventional breeding, in which a hybrid is produced with an outstanding arabica genotype, and the progeny is selfed and selected over three to four generations. Due the genetic compatibility with the cultivars of *C. arabica* and, mainly, due the resistance to rust, the genotypes of HT have been widely used for production of coffee populations with rust resistant, such as Catimor, Sarchimor, Cavimor, Cachimor and Blumor (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, Sarchimor derived cultivars such as IAPAR 59, IPR107, IPR100, and Tupi were originated from a cross between *C. arabica* Villa Sarchi and HT CIFC 832/2.

The Bacterial Artificial Chromosome (BAC) vectors has become a preferable tool for cloning large DNA fragments in genome research, mainly because of its simple manipulation and the stability of the clones, without DNA rearrangements (GONTHIER et al., 2010). The availability of a large-insert genomic DNA library is important for physical mapping, map-based gene cloning, and analysis of gene structure and function in most plant species. In higher plants, numerous bacterial artificial chromosome 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), map-based or positional cloning of genes (PATOCCHI et al., 1999; JANDER et al. 2002; XU et al. 2007), and genome-wide physical map construction (SHULTZ et al., 2006; HAN et al., 2007; LACOMBE et al., 2008). Noir et al. (2004) constructed and characterized a library of *C. arabica* with 88,320 BAC clones. For selection of BAC clones, membrane filters colonies were hybridized with RFLP probes of *C. arabica* linkage groups of genome of *C. canephora*. These probes revealed two locus in *C.arabica*, one in each of two sub-genome diploid (*C. canephora* and *C. eugenioides*) member *C. arabica*. Leroy et al. (2005) also characterized a BAC library of *C. canephora*, with genetically-mapped RFLP probes and identified by sequencing BAC clones containing genes coding for enzymes of the sucrose metabolism.

In *C. arabica* the difficulty for production of linkage genetics maps is enormous, due to the fact to being a perennial plant, with low genetic diversity and low frequency of polymorphisms. For those reasons, most available genetic maps are partials. Thus, from genetic maps and identification of BAC clones containing the same sequence of marker, it is possible obtain the physical map. Further studies in sequencing selected BAC clones with molecular anchors markers are also under way in order to integrate genomic resources with the genetic linkage existing maps.

Since several modern coffee breeding programs uses HT CIFC 832/2, Sarchimor and Catimor introgression, further genomic information about this library can be quickly applied to breeding programs. So, the BAC library has been replicated and distributed for others coffee research groups in Brazil. Successful reports in gene identification (ALVARENGA, 2011) and physical map of rust resistance using the BAC clones has been reported (Diola, 2009).

In this work we have constructed and characterized a BAC library from HT CIFC 832/2, the HT genotype which originated Sarchimor cultivars. In order to characterize the BAC library we have identified BAC clones containing a single copy gene such as mannose 6 phosphate reductase NADPH dependent (*M6PR*), a key enzyme in the production of mannitol, and anchor markers from linkage genetic maps. The library has been multiplied and was used by other t research groups to characterize genes related to genomic regions related to rust resistance, positional cloning of resistance genes, as well as to identify gene and promoters related to abiotic stress.

4.3 MATERIALS AND METHODS

4.3.1 Plant Material

Coffee leaves from Hybrid Timor access CIFC 832/2 grown in greenhouse were covered with dark plastic bag during five days, after were collected and used as starting material for high molecular weight (HMW) DNA isolation.

4.3.2 BAC Vector Preparation

The plasmid DNA vector pCC1BAC^{TM} was extracted by alkaline lysis and purified by gradient cesium chloride to obtain high quality plasmidial DNA free of contaminating bacterial DNA (SAMBROOK, 1989). The vector (10 μg) was digested with *Hind* III 7U in 75 mL at 37 °C for 1 hour. The inactivation of the restriction enzyme was performed at 65 °C for 20 minutes. The plasmid was dephosphorylated using 5U of alkaline phosphatase, 30 °C for 1 hour 30 minutes. Alkaline phosphatase was inactivated at 65 °C for 20 minutes. The vector digested and dephosphorilated was submitted at electrophoresis in 1X TAE buffer in 0.8% agarose gel. Linear plasmid bands were extracted and purified from agarose gel (PETERSON et al., 2000).

4.3.3 BAC Library Construction

Nuclei were isolated from leaves according to Peterson et al. (2000) with some modifications (described below) to eliminate polysaccharides and reduce polyphenol oxidase activity in the coffee cell extracts. The leaves were ground in liquid nitrogen follow by incubation of the cell extracts at 4 °C in homogenization buffer, followed by several centrifugations in washing buffer plus 0,2% 0f β-mercaptoetanol and 2% of Polyvinylpyrrolidone (PVP 40). Between the washing steps, the leaf homogenate was filtered through nylon mesh (250 and 100 μm) to reduce cell debris. The nuclei were resuspended in 1 ml filtered extraction buffer without β-mercaptoetanol and embedded in 1.0% low-melting-point agarose plugs. Agarose plugs were incubated in lysis buffer (sodium lauryl sarcosinate 1% w/v; 0,3 mg/ml proteinase K and 0.13% of diethyl sodium dissolved in 0.5 M EDTA, pH 9.1) for 72 hours, changing the lysis solution every 24 hours without proteinase K. The plugs were stored in ethanol 70% at -20 °C.

In order to prepare DNA for digestion, 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 to 2% for 0,25% at the end of the washing step. The plugs were dialysed for 2 hours in 0.5 x TBE and subjected to pre-electrophoresis to remove inhibitors of the cloning process (OSOEGAWA et al.,1998) on a 1% Agarose-Pulse Field USB® gel in GN Controller-PFGE (Pharmacia Biotech®) with the following conditions: pulse ramping from 1 a 4s, angle 120°, current 6V/cm and 8 hours run time at 10 ºC. The plugs were removed from the gel, washed three times in TE (Tris10mM-EDTA 10mM) for 1 hour each, chopped into small pieces and partially digested with *Hind* III (Peterson et al.; 2000). DNA was separated by PFGE in 1% agarose gel at pulse ramping from 1-45 s, 10 hours, 120° and 6 V/cm at 10 °C in 0,5x TBE, followed by a second run at: pulse ramping from 4-5 s, 8 hours, angle 120°, 5 V/cm. The flanking marker lanes were removed from the gel and colored with ethidium bromide to indicate to location of the size ranges. Fragments from 100 to 200 Kb were selected, excised from the gel and subjected to second size selection on 1% agarose gel of low melting point (Invitrogen®) at 10 °C in 0.5x TBE: 3 a 10 s, 24 hours, angle 120 °C, 5 V/cm . The fragments were excised from the gel and electro-eluted in 1X TAE buffer (40 mm Tris-acetate, 1 mM EDTA pH 8.0) using dialysis tubes pre-treated in a solution of 1 mM EDTA, 2% NaHCO3 to 90 °C for 10 minutes, followed by several washes of individual dialysis tubes in distilled water (STRONG et al., 1997).

After the electroelution of fragments, ligations were performed using 25 ng of the vector, 100 ng of DNA fragments electro-eluted, 6 units of T4 DNA ligase (high concentration-Invitrogen®), 10 µL of 5x reaction buffer and incubated at 16 °C overnight. The ligations reactions were desalinated for 1 hour at 4 °C in 1.5 mL tubes containing 1.8% glucose and 1.0% agarose ultra-pure (Invitrogen®). Twelve microliters of ligation were added to 50µL of DH10B competent cells, placed in cold cuvettes of 0.1 cm and electroporated at 1.6 kV using Eletroporator 2510 (Eppendorf®) and placed in microtubes with 1 mL of SOC medium (SAMBROOK, 1989). Transformed cells were incubated in agitation at 37 °C for 90 minutes and plated on LB medium (SAMBROOK, 1989) with 12.5 µg/mL chloramphenicol, 0.40 mM of IPTG (isopropylthio-β-D-galactosidase) and 80 µg/mL of X-Gal (5-bromo-4 -

chloro-3-indolyl-β-D-galactosidase). Plates were incubated in the dark at 37 °C overnight. White recombinant colonies were handly picked in 96 well plate and compressed in 384 well plates using a Genetix Q-Bot. At the end of transfer for the 384 plates containing 70 µL freezing media (PETERSON et al., 2000). The plates were incubated 16 hours at 37 °C, duplicated and stored at -80 °C. BAC library was multiplied twice using Genetix Q-Bot and store at -80 °C.

4.3.4 Estimation of Insert Sizes of BAC Clones

BAC clones were randomly selected and grown for 24 hours at 37 °C in 1.5 mL LB medium containing chloramphenicol (12.5 mg/mL). The BAC DNA was extracted using alkaline lysis procedure (SAMBROOK et al., 1989) and digested with *Not* I. The products of digestion were separated by PFGE in agarose gel 1% in TBE 0.5x.

4.3.5 Hybridization of Membranes with Probes of Choroplast and Mitochondria

High density filters were made using a Q-bot (Genetix, Dorset, UK), each membrane containing 18,432 double spotted clones to avoid false positives. The filters were incubated on Q-trays containing LB/agar with 12.5 mg/mL of chloramphenicol at 37 °C. Colony filters were processed as follow: denaturation on Whatman paper soaked with a solution of 0.5 M NaOH and 1.5 M NaCl for 4 minutes at 100 °C, following the neutralization on Whatman paper soaked with 1 M Tris-HCl pH 7.4, and 1.5 M NaCl for 4 minutes at 100 °C, and incubation in a solution of 0.25 mg/mL proteinase K for 50 minutes at 37 °C, baking for 3 hours at 80 °C.

Fragments of the mitochondrial (*cox1* - citochromo oxidase) and the chloroplast (*rbcL* – ribulose-1-5-biphosphate carboxylase/oxygense) were used as probes to screen for organelle DNA in the library. The probe fragments were generated by PCR amplification using total DNA as template. The amplification conditions were 5 min 94 °C for one cycle followed by 30 cycles of 94 °C 45 sec, 60 °C 20 sec, 72 °C 2 min and a final extension at 72 °C for 10 min. Fragments generated from the two reactions were gel-purified with the QIAquick gel extraction kit (Qiagen, Germany) and labeled with [α-P^{32}]-dCTP. The hybridization was performed at 42 ºC overnight using UltraHyb^{R} buffer (Ambion). After hybridization membranes were initially washed 2X for 10 min (SSC 1X; SDS 0.1%) with a final wash of 2X for 10 min (SSC 1X; SDS 0.1%). Positive clones are identified by a duplicated signal pattern.

4.3.6 Pooling of Library

The BAC library was clustered into plate pools with 384 clones. BACs were grown individually in plates with 200 uL LB enriched based on recommendations of Qatibi (1991) [1%Bacto-triptone, 0.5% yeast extract, 10 mM NaCl, 13 mM KH_2PO_4, 36mM de K_2HPO_4, 1.7 mM de sodium citrate, 6.8mM $(NH_4)_2SO_4$, 0.4 mM $MgSO_4$ in selective medium with chloramphenicol (*Sigma-Aldrich*) (12.5 mg/mL). The isolation DNA from BACs pools was performed by alkaline lysis (SAMBROOK, 1989). For the formation of superpools DNA from the plate pools were grouped on the equal concentrations of DNA (5ng/ul) of each 15 plates, forming 10 superpools, each representing approximately 5760 clones.

4.3.7 Selection of BAC Clones by PCR / Hybridization Membranes

The superpools and pools of BAC clones were screening by PCR technique. The amplifications were performed using 50 ng DNA of the pools/superpools. To amplify the M6PR gene, specific primers were designed using Primer Designer 2.0: **M6PR** sense 5’-CTC ATC GTT GGA AGT GGT GGT GG-3’; and antisense 5’-CTG TAA TCT CCG CTG CCA TCC T-3’. PCR was performed under the following conditions: 94 °C for 2 minutes, 94 °C for 1 minute, 64 °C for 45 seconds, 72 °C for 45 seconds, 35 cycles from 94 °C for one minute and 72 °C for 3 minutes. The screening of superpools and pools of BAC clones also was performed with primers for SSR: **SSR-18** (LG1), **SSR-16** (LG2), **ACGG-1** (LG3), **CCG-3** (LG6) of a partial genetic map of *C. arabica* (Missio et al. 2009). To amplify the markers SSR specific primers were used: **SSR-16** sense 5’-CCA CAC AAC TCT CCT CAT TC-3’; anti-sense 5’-ACC CGA AAG AAA GAA CCA AG-3’; **SSR-18** sense 5’-GGC TCG AGA TAT CTG TTT AG-3’; antisense 5’-TTT AAT GGG CAT AGG GTC C-3’; **ACGG-1** sense 5’-GTG AAC CTC CCT TTC CCT TG -3’; antisense 5’-ACT GGT CTC TCG TCT GTG AA-3’; **CCG-3** sense 5’-CGC CCG AAG ATC AAA CAA-3’; antisense 5’-TTA TAT CCC GCG GCA AGT CC-3’. The amplification conditions were: 94 °C for 2 minutes, 94 °C for 30 seconds, 66 °C-57 °C for 30 seconds, 72 °C for 30 seconds, 10 cycles from 94 °C for 30 seconds, 57 °C for 30 seconds, 72 °C for 30 seconds, 30 cycles from 72 °C for 8 minutes. The PCR products were separated by electrophresis in 10% poliacrylamide gel.

The plate pools selected for the M6PR gene and SSR markers were used for filter colony hybridization of the plate. They were inoculated onto positively charged nylon membranes, on top of plates containing LB and incubated the 37 °C for 16 hours. The clones grown in membranes were submitted to denaturation and neutralization process, followed by treatment with proteinase K and fixed in membrane at 80 °C for three hours. Membranes were hybridized with DNA fragment from cloned M6PR gene and SSR markers. Probes were labeled with [α-P^{32}]-dCTP by random priming. The hybridization was performed at 42 ºC overnight using UltraHyb^{R} buffer (Ambion). After hybridization membranes were initially washed 2X for 10 min (SSC 1X; SDS 0.1%) with a final wash of 2X for 10 min (SSC 1X; SDS 0.1%). After membranes were exposed in image plate BAS-IP MS 2340. The images were captured using fluorescent image analyzer FLA 3000 – series (Fuji Photo Film CO, Lts. Tokyo – Japan). BAC clones selected in hybridization were re-confirmed by PCR, and Southern blot (M6PR BAC clones).

4.3.8 Restriction Enzyme Digestion of BAC Clones (BAC Fingerprinting)

BAC clones that were selected as positive by PCR and hybridization membranes for the four linkage groups, were subjected to fingerprinting technique to observe different restriction patterns between BAC clones. The DNA was extracted using alkaline lysis procedure (SAMBROOK et al., 1989). Agarose gel-based BAC DNA fingerprinting was performed as described by Marra et al. (1997) with some modifications. Five µg of DNA and 8U/ul of *Hind* III were used for restriction. DNA fragments were separated into a 1% agarose gel by electrophoresis at 80 V in 1× TAE buffer at 4°C for 16 h. The gel was stained in 1 liter of a 1:20,000 dilution of Vistra Green (Molecular Probe, USA) in 1× TAE for 30 min. Gel images were captured using a scanner (Typhoon 9400, Amersham Biosciences).

4.3.9 Southern Blot Analysis

For Southern Blot analysis 5.0 µg of plasmid DNA of BAC clones and 15 µg genomic DNA from *C. canephora, C. eugenióides* e HT access CIFC 832/2, were digested with *Hind* III. Genomic DNA was separated on 0.7% agarose gel and transferred by capillarity to positively charged nylon membrane according to Sambrook et al*,* (1989). The gel of DNA BAC clones for M6PR gene digested with *Hind* III, also were transfer to nylon membranes after agarose gel electrophoresis. Membranes were hybridized with the 450-bp genomic DNA probe from the *CaM6PR* gene. After transfer the membranes were pre-hybridized at 42 °C for 2 hours using the buffer UltraHyb^{R} (Ambion). The M6PR probe was radiolabelled with [α-P^{32}]-dCTP by random priming and membranes were hybridized at 42 °C overnight. After hybridization membranes were initially washed 2X for 10 min (SSC 1X; SDS 0.1%) with a final wash of 2X for 10 min (SSC 1X; SDS 0.1%). Membranes were exposure to film BAS-IP MS 2340 and the autoradiogram images were captured using the fluorescent image analyzer FLA 3000 – series (Fuji Photo Film CO, Lts. Tokyo – Japan).

4.4 RESULTS

4.4.1 BAC Library Construction

A BAC library of HT accession CIFC 832/2 was constructed by partial digestion of total genomic DNA using *Hind* III and linked in pCC1BAC vector. The BAC clones were handly picked in 96 well plate and compressed in 384 well plates using a Genetix Q-Bot. At the end of transfer for the 384 plates there was 56,832 clones. The original library was multiplied twice using Genetix Q-Bot and store at – 80 °C.

4.4.2 Insert Size Distribution

To examine the quality of the BAC library, 200 BAC clones were randomly selected. The DNA was extracted and digested with *Not I*, presented in the flanking regions of the cloning site of pCC1BAC vector. Size was estimated using

PFGE eletrophoresis. The average insert size was 118 Kb, ranging from 45 to 250 Kb, 74% of the inserts were more than 100 Kb and 41% showed inserts more than 120Kb (Figures 1 and 2). The *C. arabica* genome size is estimated in 1.3 pg or 1254 Mb. Based on the number of clones and average insert sizes of library, we estimated that the genome coverage of library ranged from 5-6x genome equivalents (56,832 x 0,118 Mb = 6,706). Therefore, based in insert size our BAC library should be well-suited for many kinds of molecular and genomic research, including map based cloning, whole genome physical mapping and sequencing.

Figure 1 – DNA analysis of 16 BAC clones from the HT CIFC 832/2 BAC library by pulse-field gel electrophoresis. DNA samples digested with Not I were separated on 1% agarose gelin 0.5x TBE buffer for 18 h under the following conditions: ramp pulse time 5-15s at 6 V/cm, temperature at 10 °C. Markers used are Lambda Ladder PFGE (New England Biolabs). The shared 8 Kb band represents the pCC1 BAC Vector.

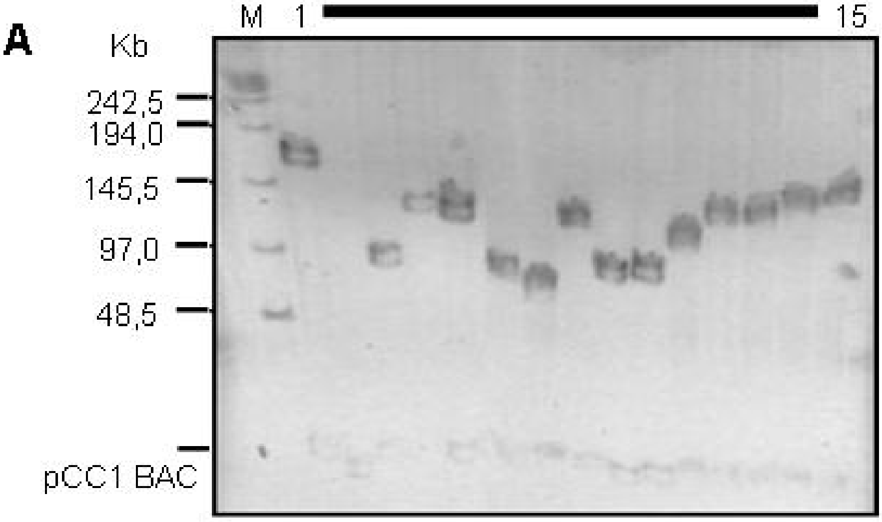
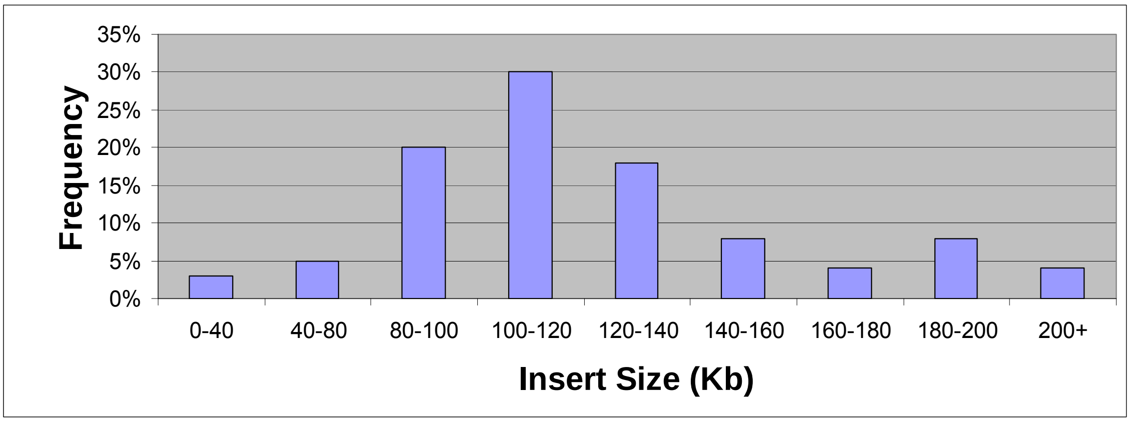


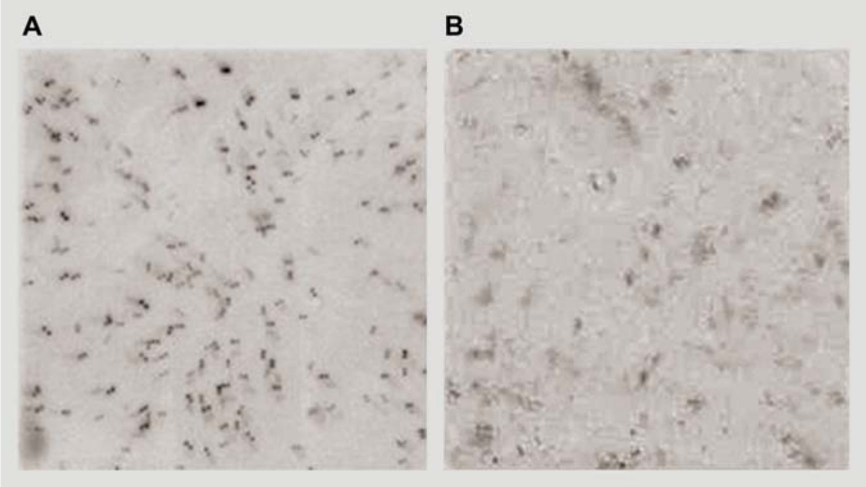
Figure 2 – Insert size distribution of 200 BAC clones of library HT CIFC 832/2. DNA samples of the 200 clones randomly picked the BAC library were analyzed and grouped according their size. Results indicate that the average insert size is 118 Kb.



4.4.3 Determination of Organelle DNA Contamination in Library

Nuclei preparation of from leaf tissue extracts by differential centrifugation and washes does not fully eliminate chloroplast and mitochondria present in the initial cell extract. Thus, it is expected that some organellar DNA will be present in some clones of the library. To evaluate the degree of contamination of chloroplast and mitochondrial DNA contamination in the BAC library, we hybridized high-density membranes, carrying 56,832 colonies in total, with probe that derived from ribulose-1-5-biphosphatocarboxylae/oxygenase (*rbcL*-chloroplast) and citocromo oxidase (*cox1*-mitochondrial)*.* Based upon the percentage of positive clones identified, the contamination was estimated as 1.04% for chloroplast and 0.5% for mitochondrial-specific DNA, respectively (Figure 3, Table 1).

Figure 3 – Determination of organelle DNA contamination present in BAC library HT CIFC 832/2. Hybridization signals obtained from high-density filters carrying 18432 double-spotted clones after hybridization to (a) ribulose-1-5-biphosphatocarboxylae/oxygenase (*rbcL*) and (b) citocromo oxidase (*cox1*) probes. Positive clones are identified by a duplicated signal pattern.



4.4.4 BAC Library Screening

A mixed approach using PCR for selection of BACs in superpools, pools and membrane hybridization was used. This methodology allowed the identification of BAC clones with genetically mapped SSR markers and a M6PR gene related to abiotic and biotic stress tolerance. Ten superpools, each one representing 5,760 BACs were screened by PCR and five positive for M6PR (Figure 4). The superpools were dismembered in order to identify the plate pools containing BAC clones of interest (Figures 4B, 4C, 4D, 4E and 4F). In order to optimize the work, each plate pool identified by PCR was subjected to colony hybridization technique. Through colony hybridization of six plates, it was possible to identify directly the eleven BAC clones positive for M6PR gene (Figure 5). The clones found by colony hybridization were subjected to PCR to verify the occurrence of false positives. From 11 BAC clones indentified by hybridization, 10 were positives for M6PR gene (Table 1, Figure 6). During the selection of BAC clones with SSR markers, eight superpools amplified for SSR-16, six for SSR-18, five for the ACGG-1 and CCG-3 (data not shown). Nine colony membranes were hybridized with probes from the SSR markers. A total of 44 BACs were selected, 11 for SSR-16 and CCG-3, 12 for SSR-18 and 10 for ACCG-1 (Table 1). Those BACs were selected for fingerprinting analysis to observe throughout their restriction pattern, duplications of the clones and/or similar pattern of the bands. There were nine different clones for SSR-16; ten for each SSR-18 and CGC-3 and seven for marker CCG-3. A similar band pattern was observed between the BACs clones selected, within the same probe, indicating a formation of BAC contigs.

Table 1 – Results of hybridization of plates containing 384 clones each, plotted in membranes. The plates were selected from of superpools and pools. Eighteen membranes were probing using seven probes.

Probe	**Linkage groups (LG)**	**Hits of number**	**BAC clones addresse***
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	------	284 (0.5%)	NI
Chloroplastic	------	591 (1.04%)	NI

*Addresse localizes the clone by Plate number, column (letter) and line (number) NI:Not Identified.

Figure 4 – PCR screening of the superpools and pools of the HT CIFC 832/2 BAC library. A- First round PCR screening in 10 superpools representing the entire library or 148 384-well microtiter plates. Lanes 1–10: superpools, PM: DNA ladder 1 Kb plus, G: genomic DNA as positive control and B: Blank. Each superpool contains DNA of 15 plates or 5,760 individual BAC clones. In six superpools: 3 (P60), 5 (P89), 6 (P116), 7 (P123), 9 (P133) and 10 (P138), the gene M6PR was detected by PCR. B to F: Second round PCR screening of plates pools. Ex: In B superpool 3, only one plate showed a signal amplified by the gene M6PR.

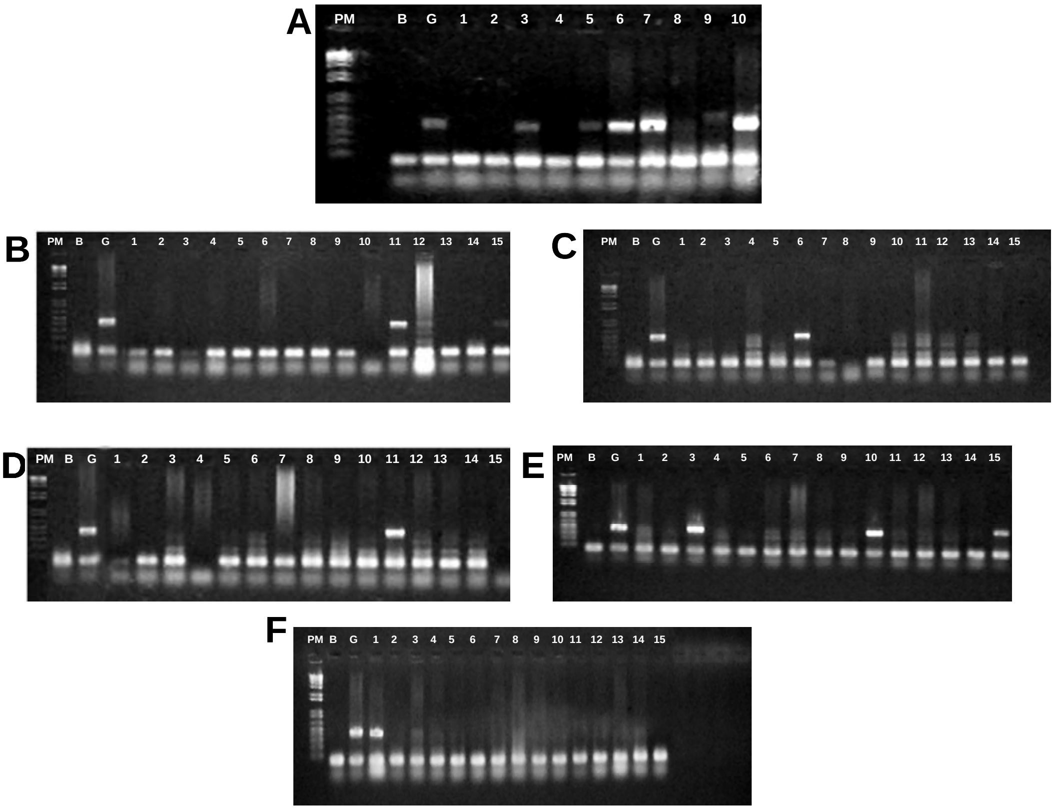


Figure 5 – Colony hybridization of BAC membranes. Clones were spotted in nylon membrane and hybridized with the single-copy probe M6PR (A) and with markers SSR: SSR-16 (B), SSR-18 (C), ACGG-1 (D) and CCG-3 (E); arrows indicate the positive clones.

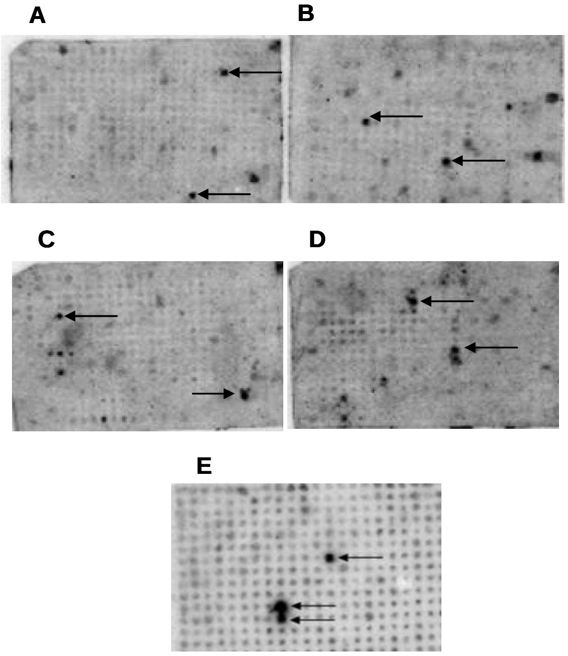


Figura 6 – Amplification of M6PR gene in BAC clones identified by PCR pool and colony hybridization. Ten BAC clones were identified for the gene interest. PM: DNA ladder 1 Kb plus, G: genomic DNA as positive control and B: Blank.

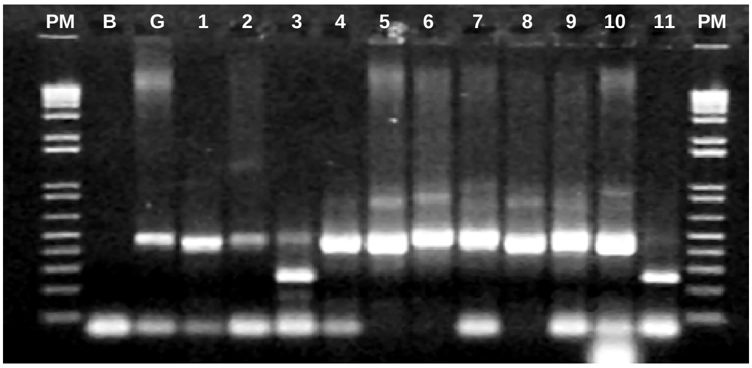
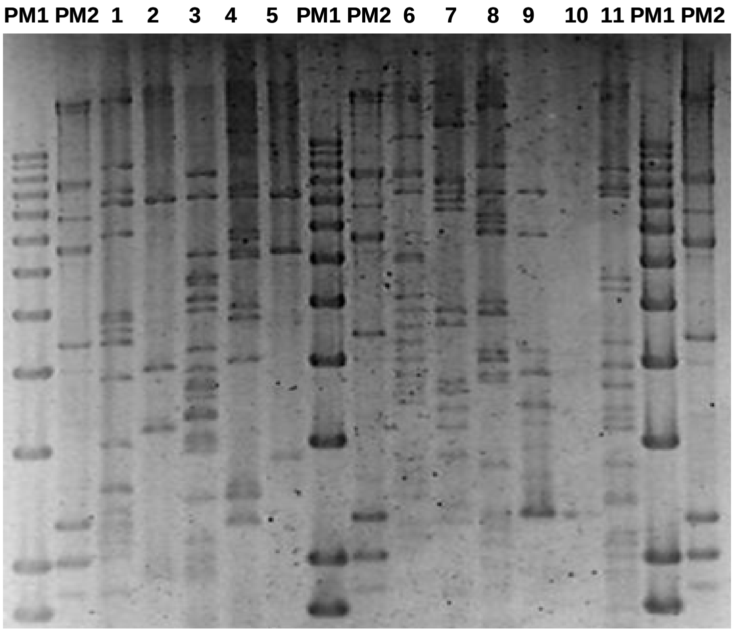


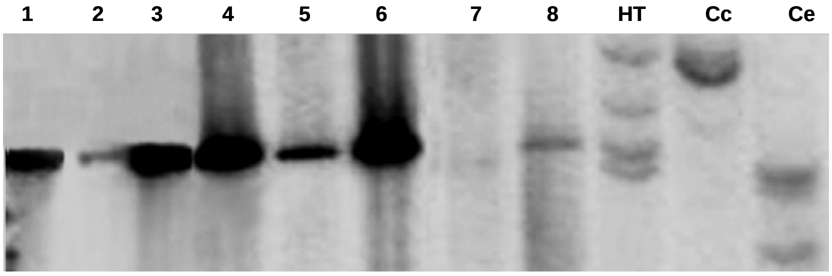
Figure 7 – Agarose gel showing restriction perfil of BAC clones positives for linkage groups digested with Hind III. 1 to 11 BAC clones. The molecular weight (PM) used (PM1) 1 Kb plus DNA ladder, (PM2) Lambda Hind III DNA.



4.4.5 Southern Blot Coffee Plants and Clones BAC

Further analysis of the selection of M6PR BAC clones was performed by Southern Blot hybridization. Of the ten positive clones, selected by PCR from BAC pools and membrane hybridization, seven clones showed an intense signal. Hybridization was also observed in genomic DNA from *C. arabica* HT CIFC 832/2, *C. canephora* and *C. eugenioides* digested with *Hind* III. In HT CIFC 832/2 genomic DNA there was four hybridization bands, which are probably related to the corresponding bands observed in *C. canephora* and *C. eugenioides*. An extra band was observed in *C. eugenioides,* possible by the hybridization with another M6PR isoform. The BAC clones 1 to 6 and 8 showed a fragment that is related for the same one present in the genome of HT CIFC 832/2, and with a similar band size presented in *C. eugenioides.*

Figure 8 – Southern hybridization of BAC clones using the M6PR probe. BAC and Genomic DNA were digested with Hind III. Lanes 1-8 clones BAC, genomic DNA from Hibrido Timor accession CIFC 832/2 (HT), *C. canephora* (*Cc*) and *C. eugenióides* (*Ce*).



4.5 DISCUSSION

BAC libraries are important tools for genomic and mapping studies. We constructed a BAC library for *C. arabica*, containing 56.832 clones with an average insert size of 118 Kb ranging from clones 45 to 250Kb, indicating that this library provided a plus 5x coverage of the HT CIFC 832/2. We have reported that 74% of the inserts in our BAC library were more than 100 Kb and 41% showed inserts more than 120Kb. A library with a larger insert size offers advantage by reducing the number of clones required for adequate coverage of the genome, and consequently simplifies the screening and generation of BAC contigs as demonstrated in the BAC screening of this study. Larger inserts can span larger gaps and can cover more space with less overlap, reducing cost of sequencing. This phenomenon has been observed in several BAC libraries, such as in tomato (BUDIMAN et al., 2000), common bean (Liu, 2010) and papaya (GSCHWEND et al., 2011).

Analysis of BAC clones through digestion with *Not* I followed by electrophoresis showed that most clones contained one fragment. Maughan et al. (2008) identified in *Amaranthus* 14% of the clones with one or two inserts, result that is consistent with the low percentage of sites for *Not* I recognition site (GCGGCCGC) found in BAC libraries of other dicotyledonous species. Thus, the coffee genome apparently contain fewer sites for *Not* I, a common aspect observed in the genomes the others dicotiledoneous species.

Contamination of library with the mitochondrial DNA was low compared with the results obtained in rice and soybean with 0,8% and 1,85 % of contamination respectively (WANG et al. 2005; TOMKINS et al. 1999).

Contamination with mitochondrial DNA was lower than with the chloroplast DNA suggesting that the coffee mitochondrial DNA is easier to spin down along with nuclei than chloroplast (RATNAYAKA et al., 2005).

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 chloroplast contamination. This confirms that the extraction procedure effectively purified nuclei and that the library comprises predominantly nuclear sequences.

The strategy of screening the library from pools representing groups of clones, normally require a reduce number of PCR to find a single clone. The M6PR gene is present in single copy at plant genomes (DELAVAULT et al., 2002), thus the PCR screening of pooled BACs was used to find the BAC clones containing this gene, as well as to validate our strategy combining PCR and colony hybridization of 384 plates. With only two rounds of PCR reactions, it was possible to identify 384 plates containing BAC clones of interest. The use of PCR and colony hybridization techniques were complementary and very effective in BAC clone screening for both M6PR gene and to SSR markers related to linkage groups. These techniques decreases the work involved to identify BAC clones as well the possibility of selecting false positives clones. Yim et al, (2007), described an approach combining a BAC pooling strategy with PCR-based primer screening that can aid in physical map construction and provide anchor points to the genetic map. 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 pool DNAs with PCR based primers include a low rate of false positives, low cost, and increased throughput compared to conventional hybridization techniques.

The screening the BAC library with markers SSR, indicates that this library BAC system can be useful tool to generate a high-resolution physical map. The identification of BAC clones related to linkage group markers, combined with the development of fingerprinting methods, sequencing BAC-ends, contigs formation or full BAC sequencing will contribute to fill gaps in physical and genetic mapping as well.

As expected, in Southern blot experiments, the bands found in *C. canephora* and *C. eugenioides* are presents also in HT CIFC 832/2 genome, suggesting the presence of this ancestral haplotype within the *C. arabica* genome.

4.6 CONCLUSION

A BAC library of HT CIFC 832/2 with more than 5x coverage of the haploid genome has been constructed and characterized. Screening the library with PCR in BAC pools associated to plate colony hybridization allowed a more reliable and faster searching of BAC clones. This BAC library is an important resource towards physical mapping, integration of physical and linkage maps, and positional cloning for *C. arabica* genes.

4.7 REFERENCES

ALVARENGA, S. M. (2011). **Caracterização funcional e identificação de SNPs e de genes de resistência a doenças do cafeeiro.** Tese de Doutorado, Universidade Federal de Viçosa, Viçosa, Brasil. 2011.

BUDIMAN, M. A.; MAO, L.; WOOD, T. C.; AND WING, R. A. A deepcoverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. **Genome Research.** v. 10, p. 129-136, 2000.

CHARRIER, A.; ESKES, A. Coffee. In: Charrier A, Jacquot S,Hamon D and Nicolas C (eds.). Cirad,Repères. **Tropical plant breeding.** p. 128-152, 1997.

DELAVAULT, P.; SIMIER, P.; THOIRON, S.; VERONESI, C.; FER, A., THALOUARN, P. Isolation of mannose 6-phosphate reductase cDNA, changes in enzyme activity and mannitol content in broomrape (*Orobanche ramosa*) parasitic on tomato roots. **Physiol Plantarum.** v. 115, p. 48– 55, 2002.

DIOLA, V.; BRITO, G. G.; CAIXETA, E. T.; MACIEL-ZAMBOLIM, E.; SAKIYAMA, N.S.; LOUREIRO, M.E. High-density genetic mapping for coffee leaf rust resistance. **Tree Genetics and Genomes.** v. *7*, p. 1199-1208, 2011.

GONTHIER, L.; BELLEC, A.; BLASSIAU, C.; PRAT, E.; HELMSTETTER, N.; RAMBAUD, C.; HUSS, B.; HENDRIKS, T.; BERGÈS, H.; QUILLET, M-C. Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). **BMC Research.**v. 3, p. 225, 2010.

GSCHWEND, A. R.; YU, Q.; MOORE, P.; SASKI, C.; CHEN, C.; WANG, J.; NA, J-K.; MING, R. Construction of Papaya Male and Female BAC Libraries and Application in Physical Mapping of the Sex Chromosomes. **Journal of Biomedical and Biotechnology.** v. 1, p. 7, 2011.

HAN, Y.; GASIC, K.; MARRON, B.; BEEVER, J. E.; KORBAN, S. S. A BAC-based physical map of the apple genome. **Genomics.** v. 89. p. 630-637, 2007.

ILIC, K.; SANMIGUEL, P. J.; BENNETZEN, J. L. A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. **Nucleic Acids Research.** v. 100, p. 12265–12270, 2003.

JANDER, G.; NORRIS, S. R.; ROUNSLEY, S. D.; BUSH, D. F.; LEVIN, I. M.; ROBERT, L. L., Arabidopsis map-based cloning in the post-genome era. **Plant Physiology.** v. 129, p. 440–450, 2002.

LACOMBE, E.; COSSEGAL, M.; MIROUZE, M.; ADAM, T.; VAROUQUAUX, F.; LAOUBET, S. Construction and characterization of a BAC library from *Arabidopsis halleri*: evaluation of physical mapping based on conserved synteny with *Arabidopsis thaliana*. **Plant Science.** v. 174, p. 634-640, 2008.

LASHERMES, P.; COMBES, M. C.; ROBERT, J.; TROUSLOT, P.; D'HONT, A.; ANTHONY, F., CHARRIER, A. Molecular characterisation and origin of the Coffea arabica L. genome. **Molecular and General Genetics.** v. 2, p. 259-266, 1999.

LASHERMES, P.; PACZEK, V.; TROUSLOT, P.; COMBES, M. C.; COUTURON, E. E.; CHARRIER, A. Single-*locus* inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C arabica x C. canephora*., **Journal Heredity.** v. 91, p. 81-85, 2000.

LEROY, T.; MARRACCINI, P.; DUFOUR, M.; MONTAGNON, C.; LASHERMES, P.; SABAU, X.; FERREIRA, L.P.; JOURDAN, I.; POT, D.; ANDRADE, A. C. Construction and characterization of a *Coffea canephor* BAC library to study the organization of sucrose biosynthesis genes. **Theorical Applied Genetics.** v. 6, p. 1032–1041, 2005.

LIANG, C.; XI, Y.; SHU, J.; LI, J.; YANG, J.; CHE, K. Construction of a BAC library of *Physcomitrella patens* and isolation of LEA gene. **Plant Science.** v. 167, p. 491-498, 2004.

LIU, H.; JIANG, Y.; WANG, S.; NINWICHIAN, P.; SOMRIDHIVEJ, B.; XU, P.; ABERNATHY, J.; KUCUKTAS, H.; LIU, Z. Comparative analysis of catfish BAC end sequences with the zebrafish genome. **BMC Genomics.** v. 10, p. 706-712, 2009.

LIU, S. Y.; YU, K.; HUFFNER, M. S.; PARK, J.; BANIK, M.; PAULS, K. P.; CROSBY W. Construction of a BAC library and a physical map of a major QTL for CBB resistance of common bean (*Phaseolus vulgaris* L.) **Genomics.** v. 138, p. 709-716. 2010.

MARRA, M. A.; KUCABA, T. A.; DIETRICH, N. L.; GREEN, E. D.; BROWNSTEIN, B.; WILSON, R. K.; MCDONALD, K. M. High throughput fingerprint analysis of large-insert clones. **Genome Research**. v. 7, p. 1072-1084, 1997.

MAUGHAN, P. J.; SISNEROS, N.; LUO, M.; KUDRNA, D.; AMMIRAJU, J. S. S.; WING, R. A. Construction of an Amaranthus hypochondriacus Bacterial Artificial Chromosome Library and Genomic Sequencing of Herbicide Target Genes. **The Plant Genome**. v. 1, p. 85-94, 2008.

MENDES, A. N. G.; CARVALHO, G. R.; BOTELHO, C. E.; FAZUOLI L. C.; SIVAROLLA, M. B. Mutante de *C. arábica* L. E cultivares de importância histórica. IN:

Ed Carvalho C. H. S. Cultivares de café: origens características e recomendações. Brasília: Embrapa café, p. 334, 2008.

MISSIO, R. F.; CAIXETA, E. T.; ZAMBOLIM, E. M.; ZAMBOLIM, L.; SAKIYAMA, N. S. Development and validation of SSR markers for Coffea arabica L. **Crop Breeding and Applied Biotechnology.** v. 9, p. 361-371, 2009.

MONDEGO, J. M. C.; VIDAL R. O.; CARAZZOLLE, M. F.; TOKUDA, E. K.; PARIZZI, L. P.; COSTA, G. G. L.; PEREIRA, L. F. P.; ANDRADE, A. C.; COLOMBO, C. A.; VIEIRA, L. G. E.; PEREIRA, G. A. G. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*, **BMC Plant Biology.** v. 11, p. 30, 2011.

NOIR, S.; PATHEYRON, S.; COMBES, M. C.; LASHERMES, P.; CHALHOUB, B.. Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). **Theoretical Applied Genetics** v. 1, p. 225-230, 2004.

OROZCO-CASTILLO, C.; CHALMERS, K. J.; WAUGH, R.; POWELL, W. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. **Theorical Applied Genetics.** v. 87, p. 934–940, 1994.

OSOEGAWA, K.; TATENO, M.; WOON, P. Y.; FRENGEN, E.; MAMMOSER, A. G.; CATANESE, J. J.; HAYASHIZAKI, Y.; JONG. P. J. Bacterial Artificial Chromosome Libraries for Mouse Sequencing and Functional Analysis. . **Genome Research.** v. 10. p. 116-128, 1998.

PATOCCHI, A.; VINATZER, B. A.; GIANFRANCESCHI, L.; TARTARINI, S.; ZHANG, H. B.; SANSAVINI, S.; GESSLER, C. Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene. **Molecular Genome Genetic.** v. 262, p. 884-891, 1999.

PEREIRA, A. A.; MOURA, W. M.; ZAMBOLIM, L.; SAKIYAMA, N. S.; CHAVES, G. M. 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, p. 253-295, 2002.

PETERSON, D. G.; TOMKINS, J. P.; FRISCH, D. A.; WING, R. A.; PATERSON, A. H. Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrade guide. **Journal of Agriculture Genomics**. http://www.ncgr.org/research/jag. 2000.

RATNAYAKA, I.; BAGA, M.; FOWLER, D.; RAVINDRA, B.; CHIBBAR, N. Construction and Characterization of a BAC Library of a Cold-Tolerant Hexaploid Wheat Cultivar. **Genomics, Molecular Genetics Biotechnology**. v. 45, p.1571–1577, 2005.

SAMBROOK, J.; FRITSCH, E. F. E.; MANIATIS, T. Molecular Cloning - A Laboratory Manual, **Cold Spring Habour Laboratory Press.** New-York. 2nd Edition, 1989.

SHULTZ, J. L.; KURUNAM, D.; SHOPINSKI, K.; IQBAL, M.J.; KAZI, S.; ZOBRIST, K. 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 Research**. v. 34, p. 758-765, 2006.

SHULTZ, J. L.; KAZI, S.; BASHIR, R.; AFZAL, J. A.; LIGHTFOOT, D. A. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. **Theoretical Applied Genetics**.v. 6, p. 1081-1090, 2007.

STRONG, S. J.; OHTA, O.; LITMAN, G. W.; AMEMIYA, C. T. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. **Nucleic Acids. Research.** v. 25, p. 3959–3961, 1997.

TOMKINS, J. P.; MAHALINGHAM, R.; MILLER-SMITH, H.; GOICOECHEA, J. L.; KNAPP, H. T.; WING, R. A. A soybean bacterial artificial chromosome library for PI 437654 and the identification of clones associated with cyst nematode resistance. **Plant Molecular Biology.** v. 41, p. 25–32, 1999.

WANG, W.; TANURDZIC, M.; LUO, M.; SISNEROS, N.; KIM, H. R.; WENG, J. Construction of a bacterial artificial chromosome library from the spike moss *Selaginella moellendorffii*: a new resource for plant comparative genomics. **BMC Plant Biology.** v. 5, p.10, 2005.

XU, P.; WANG, S.; LIU, L.; THORSEN, J.; KUCUKTAS, H.; LIU, Z. A BAC based physical map of the channel catfish genome. **Genomics**. v. 90, p. 380-388, 2007.

YIM, Y-S.; MOAK, P.; SANCHEZ-VILLEDA, H.; MUSKET, T. A.; CLOSE, P.; KLEIN, P. E.; MULLET, J. E. 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.** v. 8, p. 47, 2007.

5 ARTIGO B

IDENTIFICATION OF BAC CLONES WITH MARKERS LINKED WITH RESISTANCE TO COFFEE RUST

5.1 Abstract

The coffee leaf rust, caused by the biothrophic fungus *Hemileia vastatrix* is one of the most important coffee diseases. “Híbrido de Timor” (HT) is a *C. arabica* hybrid genotype derived from a spontaneous interspecific cross between *C. arabica* and *C. canephora* In recent decades, HT have been used intensively in coffee breeding programs as the main source of resistance to pests and diseases including coffee leaf rust (*Hemileia vastatrix*). Over the last few years, BAC libraries also became an important tool to identify and characterize genes responsible for important traits in perennial crops.Therefore, exploitation of a BAC library from HT can be a valuable tool in coffee research in order to identify genomic regions related to rust resistance. Therefore the objective this work was to identify BAC clones for rust resistance markers linked to SH genes, using a BAC library of HT access CIF 832/2. The BAC library was screening by PCR of superpools and pools of BAC clones using previously identified rust resistante markers (SAT-244, BA-124 and M-8). The BA-124 marker showed amplification for six superpools, marker SAT-244 was positive for five, while M-8 marker showed amplification for two superpools. These superpools were dismembered in pooled plates. Through colony hybridization followed by PCR of individual clones, 23 BACs were selected. Fingerprinting analysis showed overlapping patterns of those BACs which should allow the formation of BAC contigs spanning the rust resistance region for SH3. Southern Blot analysis using M-8 probe identified BAC clones in two groups that correspond to the HT subgenomes (C^c- P89O21 and C^e- P135L24). Further characterization of those clones will allow a better comprehension of the mechanisms of genetic resistance for rust in coffee, as weel as the identification other molecular markers, are useful for further work in developing markers for assisted selection (MAS), and cloning of genes involved in rust resistance.

5.2 Introduction

Coffee is an important international traded commodity, being ranked as the second most valuable primary commodity exported by developing countries (Pendergrast, 2009). Although there are more than 100 species in the genus *Coffea* L., only two species are widely used in commercial production. *C. arabica* L., also known as Arabica coffee, accounts for 75–80% of the world’s production *C. canephora* Pierre ex Froehner (Robusta coffee) represents approximately 20% of the world production (International Coffee Organization, http:// www.ico.org). *C. arabica* is a tetraploid plant, characterized by low genetic diversity due its recent origin as well as their botanic characteristics. Therefore, the transfer of desired characteristics such as disease resistance from related diploid species into cultivars of *C. arabica*, is an ongoing priority in coffee breeding (VAN DER VOSSEN, 2001).

The Hibrido Timor (HT) is genotype which was identified in a *C. arabica* field (planted in 1927) on the island of Timor (BETTENCOURT, 1973). Based on information relating to coffee germplasm introduced into Timor at the beginning of the century, the limited fertility of the original plant, characteristics of disease resistances and preliminary molecular investigations, it is believed that the HT originated from a spontaneous interspecific cross between *C. arabica* and *C. canephora* (LASHERMES et al., 1996; OROZCO-CASTILLO et al., 1994). Progenies of the HT have been distributed worldwide and, when observed, show 2n=44 chromosomes (RIJO, 1974). In recent decades, they have been used intensively (CHARRIER; ESKES, 1998) in coffee breeding programs as the main source of resistance to pests and diseases including coffee leaf rust (*Hemileia vastatrix*).). In Brazil, the main source of genes for resistance to all races of *H. vastatrix* is Hibrido Timor (VARZEA; MARQUES, 2005). Hibrido Timor CIFC 832/2 is a genotype that was used as a parental line for Sarchimor coffee cultivars, such as Tupi and IAPAR 59.

Coffee Leaf Rust (CLR), caused by the fungus *Hemileia vastatrix*, is one of the important coffee diseases. CLR is found in almost all coffee growing countries around the world and causes large annual economic losses estimated at more than US$ 1 billion (VAN DER VOSSEN, 2001). Host resistance to CLR is conditioned primarily by a number of dominant resistance factors, and coffee genotypes are classified in groups based on their interaction with physiologic races of the rust pathogen. Resistance to coffee leaf rust appears to be determined by at least nine putative resistance genes, referred to as SH1 to SH9, either singly or in combination. The genes SH1, SH2, SH4, and SH5 were identified in *C. arabica* (BETTENCOURT; NORONHA-WAGNER, 1971), whereas genes SH6, SH7, SH8, and SH9 were identified in *C. canephora* (BETTENCOURT; RODRIGUES, 1988) and SH3 in *C. liberica* (PRAKASH et al., 2004).

The number of known markers and molecular information on coffee leaf rust-resistance genes are limited (HERRERA et al., 2009). Prakash et al. (2004) screened 80 AFLP primer combinations and identified 21 AFLP markers linked to the

SH3 gene in *C. arabica* S.288 genotype (derivied from a *C. liberica* introgression). Using plants derived from a HT progeny, Brito et al. (2010) and Diola et al. (2010) showed that a dominant gene confers resistance to *H. vastatrix* race II in this genotype.

Mahé et al. (2008), used segregating population for SH3 and identified AFLP markers that were converted into SCARS. With those new markers BAC from unrelated *C. arabica* genotype were select (NOIR et al., 2004, MAHÉ et al., 2008). Ribas et al. (2011) characterized those BAC clones spanning the SH3 locus and identified a disease-resistance gene cluster, indicating that this genomic region is shared among coffee trees.

In a previous work we describe the construction and characterization of a HT CIFC 832/2 BAC library. Here, we identify and validate BAC clones spanning the SH3 locus with molecular markers previously identified (MAHÉ et al., 2008) in order to further characterize his important genomic region for rust resistance.

5.3 MATERIAL AND METHODS

For identification of the BAC clones we used HT access CIF 832/2 BAC library described previously in Article 1. The library was screened using PCR primers and probes from rust resistance markers (SAT-244, BA-124, M-8) (Table 1) near the to SH3 gene (MAHÉ et al., 2008), using the same rational of Cenci et al. (2011) and Ribas et al. (2011).

Table 1 – Sequence-characterized DNA markers SCARs to the rust resistance gene

Markers	Primers sequences Foward: (5'>3')	Primers sequences Reverse: (5'>3')
SAT-244	GCATGTGCTTTTTGATGTCGT	GCATACTAAGGAAATTATCTGACTGCT
BA-124	TGATTTCGCTTGTTGTCGAG	TGCAGATTGATGGCACGTTA
M-8	GAATTCAGCGACGATTG	GATTTGGTGGAAGGGAGC

5.3.1 Selection of BAC Clones by PCR

For identification superpools and pools of BAC library the same strategy described in article 1 was used. The amplifications were performed using 50 ng DNA of the pools/superpools. PCR was performed under the following conditions: an initial denaturation step at 95 °C for 5 min, followed by 35 denaturation cycles at 94 °C for 45 s, primer annealing at 50 °C for 45 s, elongation at 72 °C for 45 s, and a final extension step at 72 °C for 10 min. The PCR products were separated by electrophoresis on 10% polyacrylamide gel.

5.3.2 Hybridization Membranes BAC Clones

The selection rust resistance probes (SAT-244, BA-124 and M-8) (MAHÉ et al., 2008) was performed by PCR using genomic DNA of HT access CIF 832/2. DNA extraction was according Doyle and Doyle, 1989. Amplification was performed in a 20 µl mixture containing 2 µl of the DNA solution, 1x buffer, 2.0 mM of MgCl2, 10 mM of dNTPs, 0.4 µM of each of the primers and 0.5 U of Taq DNA polymerase (Invitrogen). The PCR program consisted of an initial denaturation step at 95 °C for 5 min, followed by 35 denaturation cycles at 94 °C for 45 s, primer annealing at 50 °C for 45 s, elongation at 72 °C for 45 s, and a final extension step at 72 °C for 10 min.

In order to eliminate components of the PCR reaction, the amplification product was applied in agarose gel (1%) containing ethidium bromide 0,02% and subjected to electrophoresis in TAE buffer (1x). The fragment for each rust resistance marker (SAT-244: 300 pb; BA-124: 350 pb; M-8: 240 pb) were cut and eluted from gel using the Kit illustra GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare). Fragment recovered was quantified cloned into vector pCR 2.1-TOPO using a TOPO TA Cloning®. The ligation was transformed by eletroporation in *E. coli* TOP10 eletrocompetent cells. Plasmid DNA of positives clones for rust markers were extracted according to Sambrook, 1989 and used as probe in experiments of hybridization.

The positive plate pools were inoculated onto positively charged nylon membranes, from plates containing LB and incubated the 37 °C for 16 hours. The clones grown in membranes were submitted to denaturation and neutralization process, followed by treatment with proteinase K. DNA was fixed by baking the membranes at 80 °C for three hours. Membranes were hybridized with [α-P^{32}]-dCTP labeled probes for BA-124, SAT-244 and M-8 markers. After hybridization, the membranes were exposed to imaging plate BAS-IP MS 2340. Images were captured using a fluorescent image analyzer FLA 3000. The selected BACs in hybridizations were confirmed by extracting the BAC DNA of each clone followed by PCR to check the fragment of interest.

5.3.3 BAC Clone Fingerprinting and Analysis

BAC clones were extracted using an alkaline lysis procedure (SAMBROOK, 1989). Agarose gel-based BAC DNA fingerprinting was performed as described by Marra et al. (1997) with some modifications. DNA from each BAC clone was digested using *Hind* III in a 50 µl solution containing 15 µl BAC DNA (5.0 µg/µl), 5 µl 10x React 2 buffer, 0.4 µl of BSA, 8U of *Hind* III (Invitrogen®). The reaction was incubated 37 °C for 5 h DNA and the fragments were separated into a 1% agarose gel by electrophoresis at 40 V in 1× TAE buffer at 4 °C for 16 h. The gel was stained in Vistra Green (Molecular Probe, USA) in 1× TAE for 30 min. Gel images were captured using Kodak Digital Science. Following agarose gel electrophoresis, was performed washing of gel with solutions of depurination, denaturation and neutralization. The DNA was transferred by capillarity to positively charged nylon membrane according to Sambrook (1989).

5.3.4 Southern Blot Analysis

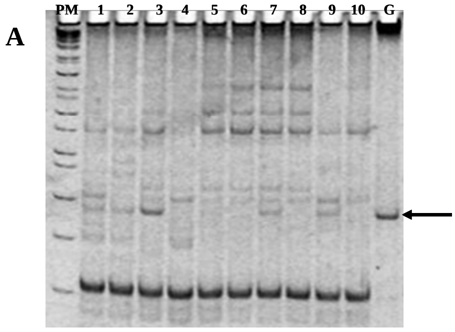
For Southern Blot analysis 15 µg genomic DNA from *C. canephora, C. eugenioides* e *C. arabica* HT access CIFC 832/2, were digested with *Hind* III in a 50 µl solution containing 20 µl DNA (15.0 µg), 5 µl 10x React 2 buffer, 0.4 µl of BSA, 8U of *Hind* III (Invitrogen®). Genomic DNA and BAC DNA digested as described previously were separated on 0.7% agarose gel by electrophoresis at 40 V for 16 hours. The DNA was transferred to nylon membranes by capillarity according SambrooK et al. 1989. BAC clones and genomic DNA, were fixed in membrane at 80 °C for three hours and hybridized with the probe SAT-244 (300-bp), BA-124 (350-bp) and M-8 (250-bp) radiolabelled with [α-P^{32}]-dCTP by random priming and membranes were hybridized at 42 °C overnight. The membranes were pre-hybridized at 42 °C for 2 hours using the buffer UltraHyb^R (Ambion). The hybridization was performed at also 42 ºC overnight using same buffer. After hybridization membranes were initially washed 2X for 10 min (SSC 1X; SDS 0.1%) with a final wash of 2X for 10 min (SSC 1X; SDS 0.1%). After membranes were exposed in image plate BAS-IP

MS 2340. The images were captured using fluorescent image analyzer FLA 3000 – series (Fuji Photo Film CO, Lts. Tokyo – Japan).

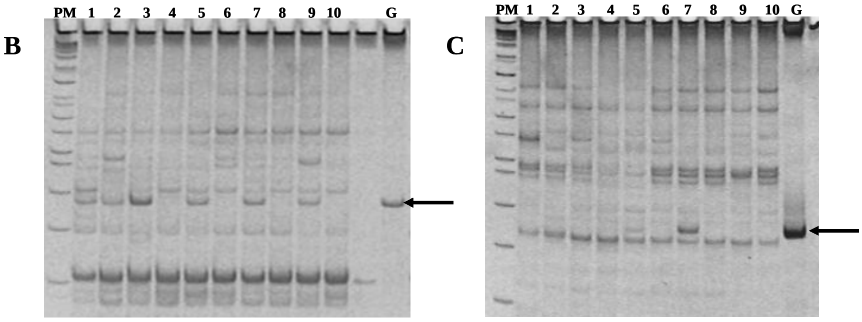
5.4 RESULTS

A genomic DNA HT 832/2 BAC library was screened by PCR of BAC clones using three rust resistance markers (SAT-244, BA-124 and M-8) in superpools and pools. The initial amplification allowed the identification of six superpools for BA-124; five for SAT-244 and two for M-8 (Figure 1).

Figure 1 – PCR screening of superpools HT CIF 832/2 BAC clones with primers for rust resistance markers: A- SAT-244; B- BA-124; C- M-8. Each superpool contains DNA of 15 plates or 5,760 individual BAC clones.Lanes 1–10: superpools, G: genomic DNA as poas positive control. The molecular weight (PM) used was DNA ladder 1Kb plus.



SP 1 to 10 SAT-244



SP 1 to 10 BA-124

SP 1 to 10 M-8

These superpools were separated in plates and we identified 12 plates pools for markers SAT-244 and BA-124, but only two plate pools for M-8. After plate colony hybridization, we identified a total 78 BAC clones positive for the markers (Figure 2). To confirm accuracy of data obtained from the membrane hybridization, the clones identified as positive were confirmed individually by PCR.

From them, 10 positives clones were selected for marker SAT-244, six for BA-124 and seven for M-8 (Table 2).

Figure 2 – PCR screening of pools of plates. A- Pool plate of superpool five (marker SAT-244) B- Pool plate of superpool one (marker BA-124) C- Pool plate of superpool 5 (marker M-8). G: genomic DNA as positive control. The molecular weight (PM) used was1Kb plus DNA ladder.

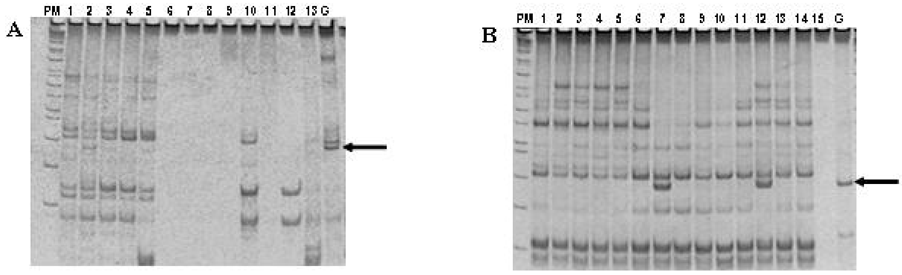
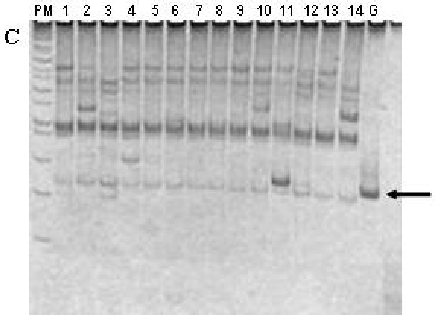


Plate Pool SP5 SAT-244

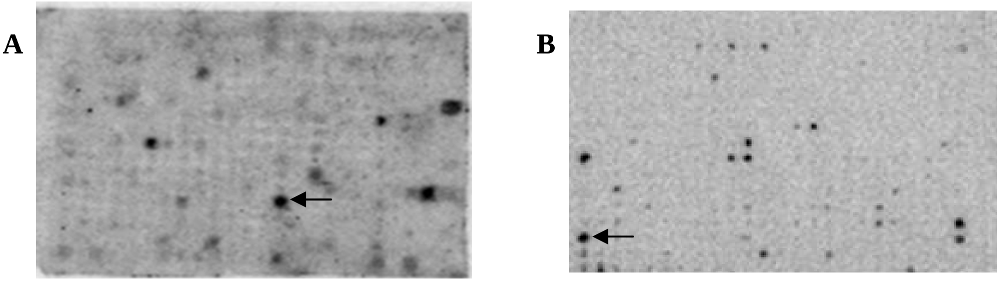
Plate Pool SP1 BA-124



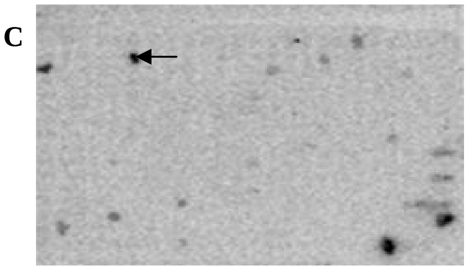
Pool of Plate SP7 M-8

Clones related with rust resistance locus were submitted to fingerprinting. Analysis of restriction patterns for BACs with marker BA-124 indicated that the clones P134D14, P126E2, P126F1 and P126H8 showed different restriction patterns, but with overlapping regions. The clones P126H10 and P126H9 showed the same restriction patterns, indicating that they are the same clone (Figure 5).

Figure 3 – Colony hybridization of BAC filters. SAT-244, BA-124, M-8 probes hybridized with high-density nylon filter; arrows indicate the positive clones.



SAT-244 BA-124



M-8

The clones for marker M-8 showed overlapping regions between clones P89K18; P89L20; P89O21; P135L24 and P135K24 (Table 2 and Figure 4). We also observed that clones P89L17 and P89L18 showed nearly identical restriction patterns, indicating that they are probably redundant (Figure 5).

For marker SAT-244, the restriction pattern were different among the clones but also showed overlapping regions.

The probes from SAT-244 and BA-124 markers hybridized a high number of BAC pools (Table 2) and contained most likely repetitive sequences; but for M8 only two BAC pools were positive, in agreement to Mahé et al. (2008). Subsequently, in order to develop and extend our analysis, five non-redundant M-8 clones based on restriction data were chosen for further analysis using Southern Blot.

Table 2 – Positive BAC clones for rust resistance markers

Marker	**Superpools 1 to 10 (5,760 BAC clones per superpool)**	**Plate Pools(384 BAC clones per Pool)**	**Positive clones hybridization 78 total**	**Positive BAC clones**	**BAC addresses***
BA-124	1, 2, 3, 5, 7, 9	2, 3, 10, 12, 17, 33, 35, 51, 56, 101, 126, 134	42	6	P126E2(1A)* P126F1(2A)* P126H8(3A)* P126H9(4A)* P126H10(5A)* P134D14 (6A)*
M-8	5, 7	89, 135	10	7	P89K1(7A)* P89L17(8A)* P89L18(9A)* P89L20(10A)* P89O21(11A)* P135L24 (12A)* P135K24 (13A)*
SAT-244	1, 2, 3, 7, 9	2, 3, 10, 12, 17, 33, 35, 51, 56, 126, 134, 135	26	10	P10K18(1B)* P10L9(1B)* P51L14(1B)* P56P4(1B)* P126C20(1B)* P26M23(1B)* P134H4(1B)* P134M22 (1B)* P135K24 (1B)* P135N17 (1B)*

Addresses localize the clone by the plaque number, column (letter) and line (number). *****
*Number in parenthesis indicates BAC position in Figure 4.

Figure 4 – Agarose gel showing restriction perfil of BAC clones positives for rust resistance markers digested with Hind III. (A) 1 to 6 BAC clones for marker BA-124 and 7 to 13 BAC clones for marker M-8. (B) 1 to 5 and 8 to 12 BAC clones for marker SAT-244. The molecular weight (PM) used (PM1) Lambda *Hind* III DNA, (PM2) 1 Kb plus DNA ladder.

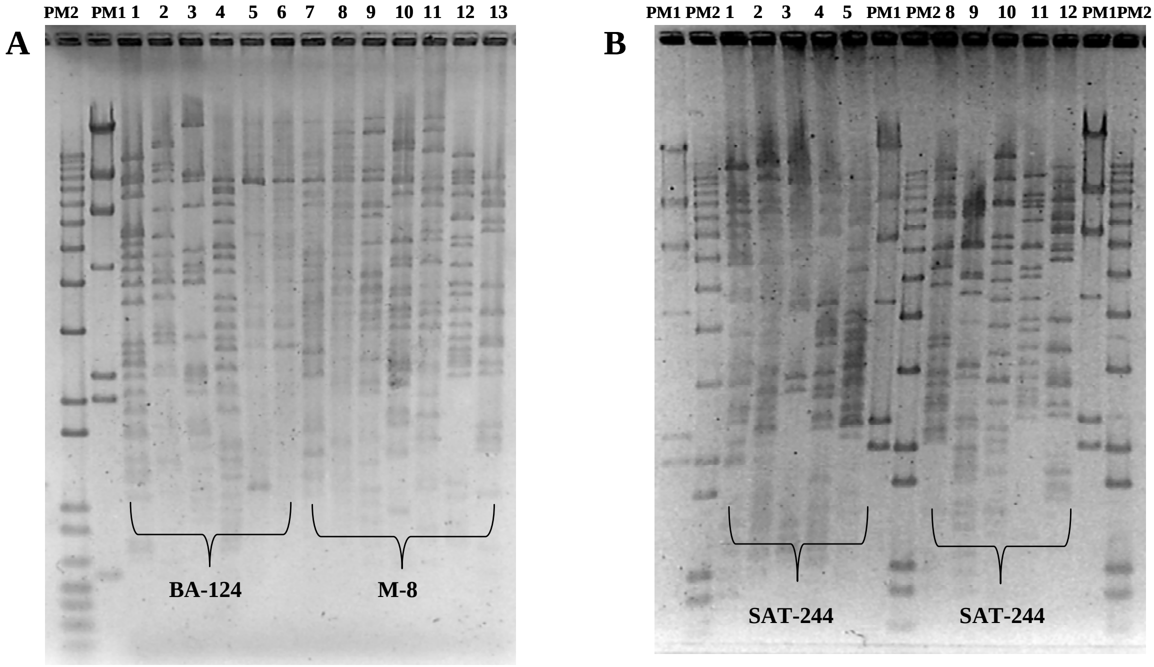
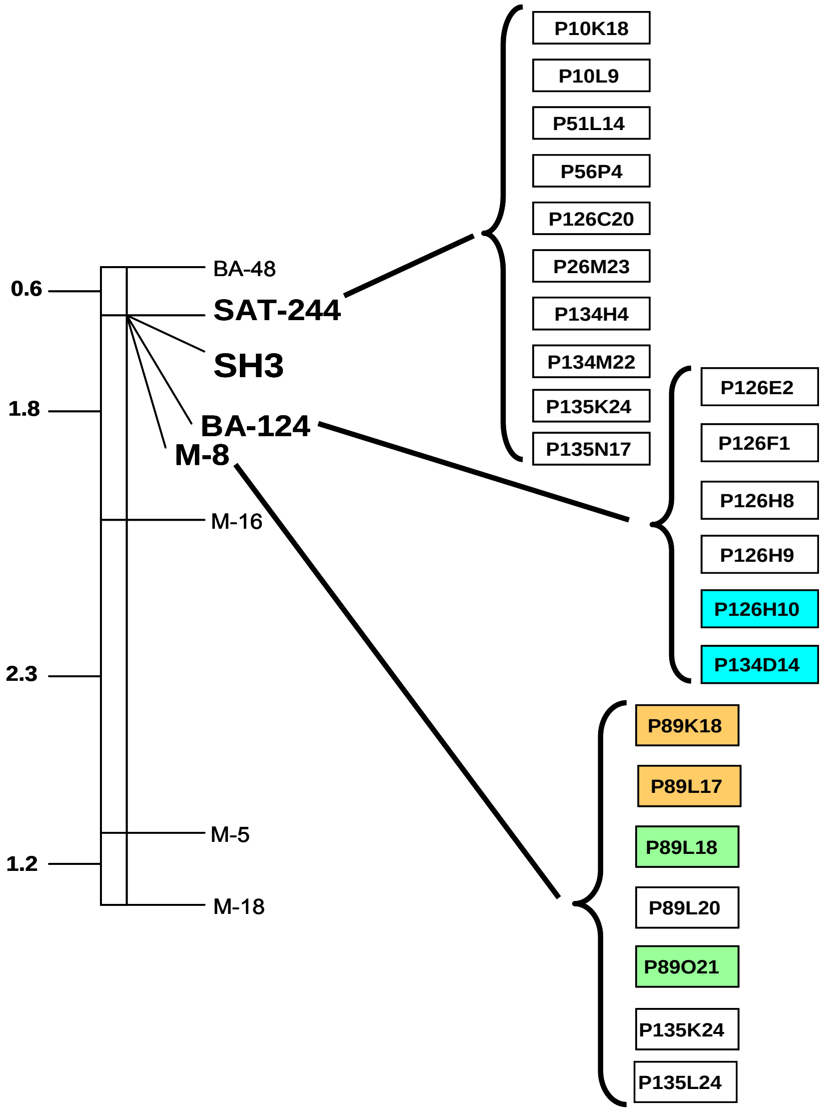


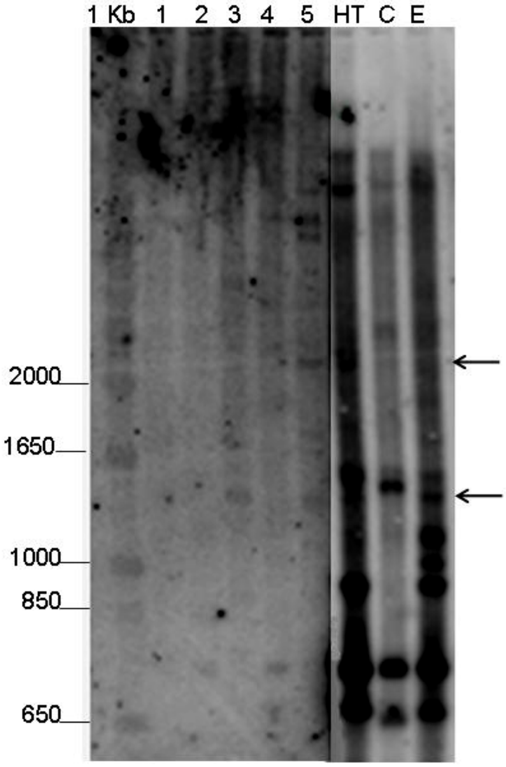
Figure 5 – Linkage map based Mahé et al., 2008. Molecular marker names are on the right, while the estimated map distances in cM are shown on the left. Molecular markers used in ths work are in bold. Clones identified in HT BAC library around SH3 gene locus. Clones with the same colors for the markers BA-14 and M-8 showed the same restriction profile.



5.4.1 Southern Blot Coffee Plants and Clones BAC

Further analysis of BAC clones selected with marker M-8 was performed by Southern Blot Hybridization. At least two BAC clones showed a polymorphism which correspond to the C. arabica subgenomes (C^{c}- P89O21 and C^{e}-P135L24). Hybridization was also observed in genomic DNA from *C. arabica* HT CIFC 832/2, *C. canephora* and *C. eugenioides* digested with *Hind* III. It was observed more hybridization that expected. In HT CIFC 832/2 genomic DNA was possible to observe four bands, three corresponding to the subgenome *C. canephora* and the same bands were showed in *C. eugenioides* genome.

Figure 6 – Southern Blot hybridization of BAC clones using the M-8 probe. BAC and Genomic DNA were digested with Hind III. DNA ladder 1 Kb plus, Lanes 1-5 clones BAC, (1-P89L18; 2-P89L20; 3-P89O21; 4-P135K24; 5-P135L24) genomic DNA from Hibrido Timor accession CIFC 832/2 (HT), *C. canephora (C)* and *C. eugenióides (E)*.



5.5 DISCUSSION

HT is derived of a natural interespecific cross between *C. arabica* and *C. canephora.* It has important agronomic characteristics such as resistance to several races of biotrophic fungus *Hemileia vastatrix* and has been used intensively in coffee breeding programs. Understanding the molecular basis of rust resistance in HT is an important way to speed up rust resistance introgression in coffee breeding. Therefore, we started to select BAC clones spanning rust resistance regions using previous selected genetic markers.

The screening of BAC clones used both PCR and membrane hybridization strategies, in order to guarantee an efficient identification of BAC clones spanning the SH3 disease-resistance gene cluster in HT and eliminate more precisely false positive clones. In this way, from an initial screening that obtained 78 positive BAC clones, we further selected 23 clones.

Restriction analysis of 23 BAC clones indicates that we isolated a contiguous sequence for all three markers evaluated. This way, it we will be possible to ordinate those BAC clones and consolidated a BAC contig region related to the SH3 locus in HT. This assembly would be further confirmed in future studies by both BAC-end sequence analysis and SNP-based fingerprinting, such as Mahé et al. (2007) and Lashermes et al. (2010). Those BAC clones are also being full sequenced to help identify genes as well as other markers related to this genomic region.

Southern blot hybridizations of BAC clones related to marker M-8, allowed the classification of BAC clones according to each *C. arabica* subgenome (C^c P89O21 and C^e P135L24), indicating that this probe is highly specific to SH3 region. Ribas et al. (2011) also confirmed the highly specific hybridization pattern of SH3 region in *C. arabica* IAPAR-59. These observations suggest that in different Arabica genotypes, SH3 region is probably high synthenic.

It is important to highlight that, in *C. arabica* HT, Brito et al. (2010) have already targeted a rust resistance locus that is clearly distinct from SH3 region, based on BAC isolation analysis (DIOLA et al., 2009) and map-based approaches (DIOLA et al., 2011) . In this way, rust resistance in HT may involve at least two distinct chromosomal regions. Since SH3 region is a disease-resistance gene cluster (RIBAS et al., 2011) highly prone to recombination (CENCI et al., 2010, 2011), a fact that is well-known for resistance gene regions (WICKER et al., 2007), it is possible that the locus mapped in HT is derived from a recombination event exclusively found in this genotype and the SH3 region in HT plays importance in resistance to rust caused by *H. vastatrix* races other than race II.

Once we have identified and initially characterized BAC clones spanning SH3 region, future studies involving BAC sequencing and exhaustive gene annotation will allow a comprehensive comparative analysis with previous published data and Diola et al. (2009, 2011). These analyses will be crucial to better mitigate the processes that drive the molecular basis of rust resistance in coffee trees, as well as provide interesting information for genome evolution studies in plants.

5.6 CONCLUSION

BAC clones related SH3 locus were identified and characterized. BAC fingerprinting analyses of those clones indicates a contig assembly for Sh3 region in HT. This region can be employed to discover new molecular makers related to the SH3 region, in order to produce saturate maps which can be used to marker-assisted selection (MAS). The comparison of those regions with previous identified BAC regions in Coffea spp. will allow sinteny studies.

5.7 REFERENCES

BETTENCOURT, A. J.; NORONHA-WAGNER, M. Genetic factors conditioning resistence of *Coffea arabica* L. to *Hemileia vastatrix* Berk. et Br. **Agronomia Lusitana.** v. 31, p. 285-292, 1971.

BETTENCOURT, A. J. Considerações gerais sobre o 'Híbrido de Timor'. Instituto Agronômico de Campinas, Circular v. 23, p. 20, 1973.

BETTENCOURT, A. J.; RODRIGUES, C. J. Principles and practice of coffee breeding for resistance to rust and other diseases. **Elsevier Applied Science.** p. 199-234. 1988.

BRITO, G. G.; CAIXETA, E. T.; GALLINA, A. P.; ZAMBOLIM, E. M.; ZAMBOLIM, L.; DIOLA, V.; LOUREIRO, M. E. Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. **Euphytica.** v. 173, p. 255-26, 2010.

CENCI A.; COMBES, M-C.; LASHERMES, P. Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. **Mol Genet Genomics**. v. 283, p. 493–501, 2010.

CENCI, A.; COMBES, M. C.; LASHERMES, P. Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. **Plant Mol Biol.** v. 78, p. 135-145, 2011.

CHARRIER, A., ESKES, A. Coffee. In: Charrier A, Jacquot, S.; Hamon, D.; Nicolas, C.; (eds.). Cirad, Repères. **Tropical plant breeding.** p. 128-152, 1997.

DIOLA, V.; LOUREIRO M. E.; BRITO, G. G.; CAIXETA, E. T.; ZAMBOLIN, E. M.; SAKIYAMA, N. S.; PEREIRA, L. F. P. Construção do mapa genético e físico para um gene de resistência a Hemileia vastatrix em café. **Anais XII Congresso Brasileiro de Fisiologia Vegetal**, 2009.

DIOLA, V. Resistência à ferrugem do cafeeiro: mapeamento genético, físico e análise da expressão gênica em resposta a infecção de *H. vastatrix*. Tese de doutorado, Fisiologia Vegetal. Universidade Federal de Viçosa, Viçosa-MG. 98 p. 2010.

DIOLA, V.; BRITO, G. G.; CAIXETA, E. T.; MACIEL-ZAMBOLIM, E.; SAKIYAMA, N. S.; LOUREIRO, M.E. High-density genetic mapping for coffee leaf rust resistance. **Tree Genetics and Genomes.** v. *7*, p. 1199-1208, 2011.

DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation method for small quantities of fresh tissues. **Phytochemical Bulletin.** v. 19, p. 11-15, 1987

HERRERA, P. J. C.; ALVARADO, A. G; CORTINA, G. H A.; COMBES, M-C; ROMERO, G. G; LASHERMES, P. Genetic analysis of partial resistance to coffee leaf rust (*Hemileia vastatrix* Berk & Br.) introgressed into the cultivated *Coffea arabica* L. from the diploid *C. canephora* species. **Euphytica.** v. 167, p. 57-67, 2009.

LASHERMES, P.; COUTURON, E.; MOREAU, N.; PAILLARD, M.; LOUARN, J. Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. **Theor. Appl. Genetics.** v. 93, p. 458-462, 1996.

LASHERMES, P.; COMBES, M.; RIBAS, A.; CENCI, A.; MAHE, L.; ETIENNE, H. Genetic and physical mapping of the SH3 region that confers resistance to leaf rust in coffee tree (*Coffea arabica* L.). **Tree Genetics Genomes.**v. 6, p. 973–980, 2010.

MAHÉ, L.; COMBES, M. C.; LASHERMES, P. 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 Molecular in Biology.** v. 64, p. 699-711, 2007.

MAHÉ, L.; COMBES, M. C.; VÁRZEA, V. M. P.; GUILHAUMON, C.; LASHERMES, P. Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica*). **Molecular Breeding.**v. 21, p. 105–113, 2008.

MARRA, M. A.; KUCABA, T. A.; DIETRICH, N. L.; GREEN, E. D.; BROWNSTEIN, B.; WILSON, R. K.; MCDONALD, K. M. **Genome Research**. High throughput fingerprint analysis of large-insert clones v. 7, p. 1072-1084, 1997.

OROZCO-CASTILLO, C.; CHALMERS, K. J.; WAUGH, R.; POWELL, W. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. **Theorical Applied Genetics.** v. 87, p. 934–940, 1994.

PENDERGRAST, M. Coffee second only to oil? Is coffee really the second largest commodity? Mark Pendergrast investigates and finds some startling results. **Journal Tea Coffee Trade.** v. 181, p. 38-41, 2009.

PRAKASH, N. S.; MARQUES, D. V.; VARZEA, V. M. P.; SILVA, M. C.; COMBES, M. C.; LASHERMES, P. Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *Coffea arabica* L. **Theoretical Applied Genetics.** v. 109, p. 1311-1317, 2004.

RIBAS, A.; CENCI, A.; COMBES, M-C.; ETIENNE, H.; LASHERMES. P. Organization and molecular evolution of a disease-resistance gene cluster in coffee trees, **BMC Genomics.** v. 12, p. 240, 2011.

RIJO, L.; RODRIGUES, C.J. Processo de infecção da *Hemileia vastatrix* Berk. Et Br. em cultivares susceptíveis e resistentes de *Coffea arabica* L. **Garcia de Orta.** v. 5, p. 23-24, 1978.

SAMBROOK, J.; FRITSCH, E.F.E.; MANIATIS, T. Molecular Cloning - A Laboratory Manual, New-York. **Cold Spring Habour Laboratory Press.** 2nd Edition, 1989.

VAN DER VOSSEN, H. Coffee breeding practices. In: Clarke RJ, Vitzthum OG (eds) Coffee Recent Developments. **Agronomy**. v. 1, p. 184-201, 2001.

VÁRZEA, V.M.P.; MARQUES, D.V. Population variability of *Hemileia vastatrix* vs. coffee durable resistance. IN: ZAMBOLIM, L., ZAMBOLIM, E.M; VÁRZEA, V.M.P. (Eds.) Durable Resistance to Coffee Leaf Rust. Viçosa: Universidade Federal de Viçosa. p. 53-74, 2005.

WICKER, T.; YAHIAOUI, N.; KELLER, B. Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. **Plant Journal**. v. 51, p. 631-641 2007.

6 CONCLUSÃO GERAL

No presente trabalho foi construída e disponibilizada uma biblioteca BAC de Hibrido de Timor acesso CIF 832/2, demonstrando sua utilização para análises genômicas em *Coffea.* A biblioteca BAC apresentou insertos com uma média superior a 100kb, oferecendo a vantagem de reduzir o número de clones necessários para uma cobertura adequada do genoma, simplificando a seleção e geração de contigs BAC. Isso pode facilitar trabalhos de sequenciamento e de mapeamento físico do genoma reduzindo o número de *gaps* desses esforços.

Combinações das técnicas de PCR de *pools* da biblioteca e hibridização de membranas propiciou uma forma rápida e econômica de seleção de clones BAC para três diferentes finalidades: 1) isolamento de M6PR, um gene envolvido em diversas respostas de tolerância a estresses bióticos e abióticos; 2) seleção de clones BAC representativos de quatro grupos distintos de ligação em *C. arabica*; 3) identificação de clones para as marcas relacionadas ao locus SH3 de resistência a ferrugem.

Futuros estudos de sintenia e comparação de sequências baseados nesta e em outras bibliotecas de BACs e em dados de projetos de sequenciamento permitirão uma análise comparativa em espécies vegetais, visando uma melhor compreensão dos processos relacionados à evolução dos genomas.

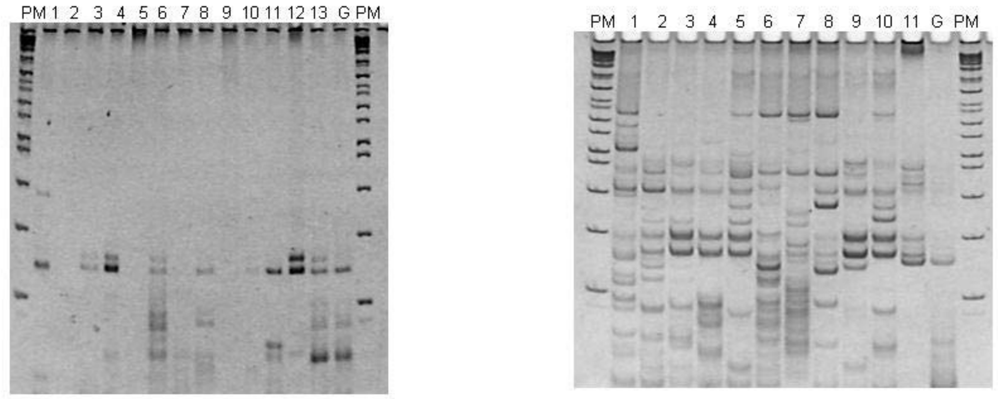
Uma vez que *C. arabica* apresenta um genoma poliplóide, esta biblioteca BAC oferece a oportunidade para reconstruir a filogenia de lócus homeólogos neste genoma, a fim de melhor investigar os mecanismos de especiação no gênero *Coffea.*

A identificação de clones da biblioteca BAC com marcas SSR e de resistência a ferrugem poderão ser de grande relevância para seleção assistida por marcadores (MAS) em cafeeiro. Adicionalmente fornecem as bases para estudos de mapeamento voltados a clonagem de genes de importância para o melhoramento vegetal e envolvidos na resistência à ferrugem.

ANEXOS

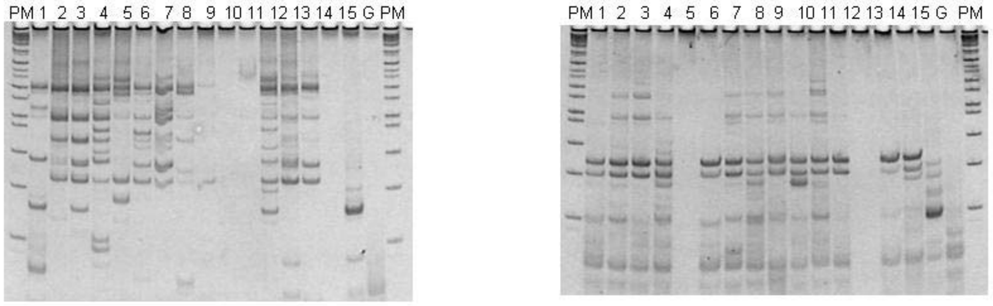
ANEXO A

Figura 1 – Seleção por dos superpools da biblioteca BAC HT CIFC 832/2. PM: DNA ladder 1 Kb plus, G: DNA genomico HT



SP SSR-16

SP SSR-18

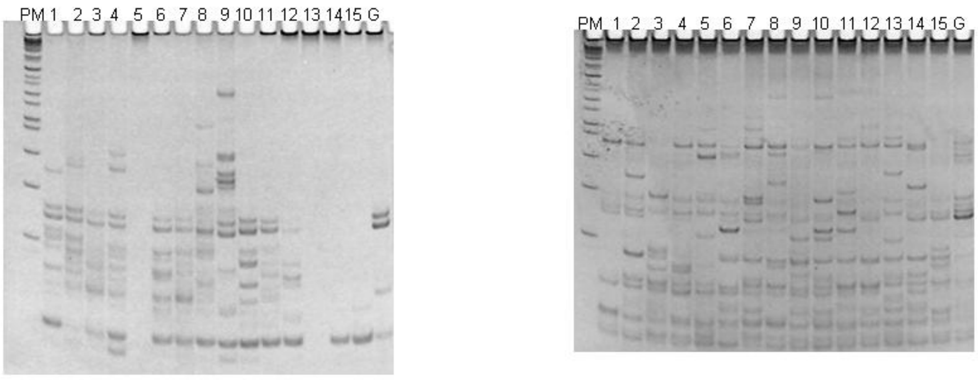


SP ACCG-1

SP CCG-3

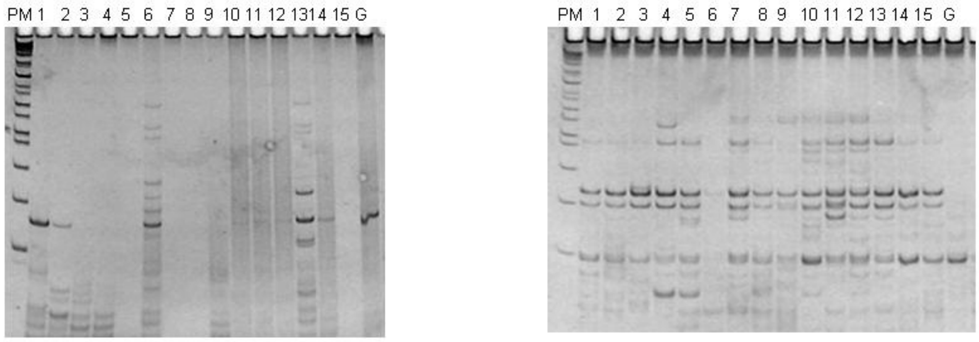
ANEXO B

Figura 2 – Seleção por dos superpools da biblioteca BAC HT CIFC 832/2. PM: DNA ladder 1 Kb plus, G: DNA genomico HT



Pool de Placa SSR-16

Pool de Placa SSR-18



Pool de Placa ACCG-1

Pool de Placa CCG-3

ANEXO C

Figura 3 – Perfil de restrição dos clones BAC positivos para as marcas dos grupos de ligação (SSR-16, SSR-18, ACGG-1, CCG-3) digeridos com *Hind* III.