

UNIVERSIDADE ESTADUAL DE SANTA CRUZ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR



**Identificação de polimorfismo em ESTs da interação cacau-
*Moniliophthora perniciosa***

Lívia Santos Lima

ILHÉUS – BAHIA – BRASIL

Fevereiro de 2007

LÍVIA SANTOS LIMA

**Identificação de polimorfismo em ESTs da interação cacau-
*Moniliophthora perniciosa***

**Dissertação apresentada à
Universidade Estadual de Santa Cruz,
como parte das exigências para
obtenção do título de Mestre em
Genética e Biologia Molecular.**

ILHÉUS – BAHIA – BRASIL

Fevereiro de 2007

LÍVIA SANTOS LIMA

**Identificação de polimorfismo em ESTs da interação cacau-
*Moniliophthora perniciosa***

Dissertação apresentada à
Universidade Estadual de Santa Cruz,
como parte das exigências para obtenção
do título de Mestre em Genética e Biologia
Molecular.

APROVADA: 26 de fevereiro de 2007

Dr. Didier Clément
(Cirad/CEPLAC)

Prof. Dr. Júlio César de
Mattos Cascardo
(UESC)

Prof. Dr. Ronan Xavier
Corrêa
(UESC)

Dr^a Fabienne Micheli
(Cirad/UESC – Orientadora)

"Não confunda jamais conhecimento com sabedoria. Um o ajuda a ganhar a vida; o outro a construir uma vida."

(Sandra Carey)

Ao anjinho que acaba de chegar em minha família para trazer muita luz e alegria... minha sobrinha e afilhada Luiza,

DEDICO

Aos meus pais, Luciene e Arquibaldo, que abaixo de Deus são os principais responsáveis pela pessoa que sou hoje, e que muitas vezes se sacrificaram para que os meus sonhos se realizassem,

Às minhas irmãs, Leila e Lílian, e o meu cunhado Cláudio por todo incentivo e amor,

A Eloisio, por todo amor, incentivo e paciência e por estar ao meu lado, mesmo em pensamento, todas as vezes que precisei,

Com amor

OFEREÇO

AGRADECIMENTOS

A Deus, que me deu força e serenidade para persistir diante das dificuldades e para cumprir esta jornada.

À Dr^a Fabienne Micheli, pela orientação, confiança e constante disponibilidade, presteza e boa vontade.

À Dr^a Karina Peres Gramacho pela co-orientação, amizade e pela grande contribuição ao meu crescimento pessoal e profissional nesses quatro anos de convivência.

À Prof^a Dr^a Fernanda Gaiotto pela importante contribuição no desenvolvimento deste trabalho.

Ao Mestrado em Genética e Biologia Molecular da Universidade Estadual de Santa Cruz e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela oportunidade de realização do curso.

Aos professores do Mestrado em Genética e Biologia Molecular da UESC, pelos ensinamentos.

Ao Centro de Pesquisas do Cacau (CEPEC) da Comissão Executiva da Lavoura Cacaueira (CEPLAC), pela permissão da execução do trabalho em suas dependências.

Ao CNPq e CIRAD, pelo apoio financeiro indispensável na condução deste trabalho.

Aos doutores Abelmon Gesteira e Uilson Vanderlei Lopes, pelas sugestões e colaboração neste trabalho.

Ao Dr Nicolas Carels e Renato Novais, pela grande colaboração nos trabalhos de bioinformática.

Aos amigos do laboratório de Fitopatologia molecular da CEPLAC/CEPEC, José Reis, Joselito, Nara, Walnize, Flávia, Francisca, Fernanda, Débora, Dilton e Lurdinha, pelo convívio, pela amizade e pelos momentos compartilhados.

Aos colegas do Laboratório de Biotecnologia da CEPLAC/CEPEC, pelo apoio e companheirismo.

Aos colegas de turma, em especial a Nadja, Brena, Tarcilla e Heliana pelo apoio, pelo incentivo e pela sincera amizade.

Ao amigo Ricardo Franco, pelo grande incentivo e amizade.

A todos que, direta ou indiretamente, colaboraram na condução deste trabalho.

ÍNDICE

EXTRATO	iii
ABSTRACT	v
LISTA DE FIGURAS	vii
LISTA DE TABELAS	ix
1. INTRODUÇÃO	1
2. REVISÃO BIBLIOGRÁFICA	3
2.1. Origem, dispersão e importância econômica do cacaueteiro	3
2.2. Origem, propagação e sintomas da vassoura-de-bruxa	4
2.3. Melhoramento do cacaueteiro visando resistência durável à vassoura-de-bruxa.....	6
2.4. Resistência do cacaueteiro ao fungo <i>Moniliophthora perniciosa</i>	9
2.5. Seqüências expressas (ESTs) no cacaueteiro	12
2.6. Marcadores moleculares de alta resolução	14
2.6.1. Polimorfismos de nucleotídeos únicos (<i>Single Nucleotide Polymorphisms</i> - SNPs)	15
2.6.2. Marcadores microssatélites.....	17
3. METODOLOGIA.....	20
3.1. Seleção de ESTs	20
3.2. Re-sequenciamento dos cDNA selecionados.....	20
3.3. Processamento e análise de ESTs	21
3.4. Identificação de SNPs em ESTs potencialmente ligados a resistência a <i>M. perniciosa</i>	22
3.5. Construção de <i>primers</i> a partir dos SNPs.....	22
3.6. Identificação de SSR em ESTs potencialmente ligados a resistência <i>M. perniciosa</i>	23

3.7. Material Vegetal	25
3.8. Extração de DNA	25
4. RESULTADOS	27
4.1. Seleção de ESTs relacionados com resistência	27
4.2. Identificação de SNPs em ESTs	35
4.3. Identificação de SSR em ESTs.....	50
5. DISCUSSÃO	62
5.1. Genes relacionados com resistência das bibliotecas de interação cacau– <i>Moniliophthora perniciosa</i>	62
5.2. Identificação de SNPs em ESTs de cacau.....	64
5.3. Identificação de SSRs em ESTs de cacau.....	68
6. CONCLUSÕES	70
7. REFERÊNCIAS BIBLIOGRÁFICAS	71

EXTRATO

LIMA, Livia Santos. Universidade Estadual de Santa Cruz, Ilhéus, fevereiro de 2007. **Identificação de polimorfismo em ESTs de cacau associados à interação cacau-*Moniliophthora perniciosa***. Orientadora: Fabienne Micheli. Co-orientadora: Karina Peres Gramacho. Conselheira: Fernanda Amato Gaiotto.

A vassoura-de-bruxa causada pelo fungo *Moniliophthora perniciosa* é a doença de maior impacto na cacauicultura no Brasil. Sua principal forma de controle é a utilização de clones resistentes, porém é preciso aumentar a base genética nos plantios comerciais para que se tenha uma resistência mais duradoura. A identificação de genes expressos tem sido usada como uma importante abordagem experimental na seleção de novos genes de resistência, e a busca de polimorfismos pode ser uma ótima alternativa para contribuir com a seleção de cacauzeiros duravelmente resistentes à vassoura-de-bruxa. O objetivo deste trabalho foi identificar polimorfismos (SNPs e SSR) em bibliotecas de ESTs da interação cacau-*Moniliophthora perniciosa*, confeccionadas em trabalhos anteriores (GESTEIRA et al., 2004; GESTEIRA et al., 2007; LANAUD et al., 2006; ZAIDAN et al., 2005). Para a busca de polimorfismos do tipo SNPs, foram selecionados 153 ESTs possivelmente relacionados com resistência. Estas seqüências foram alinhadas, utilizando o programa Clustal W, com todos os genes das bibliotecas supracitadas, mais as seqüências oriundas de bibliotecas subtrativas cedidos pelo CIRAD e as seqüências de cacau existentes em banco de dados (JONES et al., 2002; VERICA et al., 2004). Os resultados dos alinhamentos múltiplos de seqüências foram imediatamente processados pelo programa CLOURE-D (<http://imtech.res.in/~anand/cloure.html>). Assim foi possível a identificação de 85 possíveis SNPs sendo mais freqüentes os do tipo transição. Grande parte dos SNPs foram encontrados em seqüências de

genes envolvidos em transdução de sinais, síntese de proteínas e defesa. Para a validação dos SNPs foi testada a técnica de AS-PCR, porém os resultados não foram conclusivos, necessitando de mais experimentos para a obtenção de resultados mais convincentes. Para a identificação de SSRs foram utilizadas 10 seqüências de meristemas de TSH 1188 e Catongo, e 675 seqüências de fruto de TSH 1188, obtendo-se 125 SSRs. Foram utilizados 32 pares de *primers* para amplificação em 21 genótipos de cacauero com diferentes níveis de resistência e dois *bulks* referentes à genótipos susceptíveis e resistentes a vassoura-de-bruxa, que permitiu a obtenção de quatro locos polimórficos envolvidos em transdução de sinais e em função ainda desconhecida e 16 alelos para resistência do cacauero à vassoura-de-bruxa. Os resultados mostraram uma separação entre os genótipos susceptíveis e resistentes, e uma separação entre os genótipos resistentes, necessitando porém de estudos mais aprofundados neste sentido.

ABSTRACT

LIMA, Livia Santos. Universidade Estadual de Santa Cruz, Ilhéus, February 2007. **Identification of polymorphisms in ESTs of cacao associated to the interaction cacao-*Moniliophthora perniciosa***. Advisor: Fabienne Micheli. Advisor Committee: Karina Peres Gramacho and Fernanda Amato Gaiotto.

Witches' broom caused by fungi *Moniliophthora perniciosa*, is the main disease occurring in the cacao plantations in Brazil. The main form of control of this disease is the utilization of resistant cacao clones, but, for this, it is necessary to increase the genetic base of commercial plantlets to obtain a durable resistance. The identification of expressed genes was used as an important experimental design to select new resistance genes and new polymorphisms, which may contribute to the selection of cacao trees durably resistant to witches' broom. The objective of the work present here was the identification of polymorphisms (SNPs and SSR) in cacao-*Moniliophthora perniciosa* interaction EST libraries previously obtained (GESTEIRA et al., 2004; GESTEIRA et al., 2007; LANAUD et al., 2006; ZAIDAN et al., 2005). The detection of SNP polymorphism was made in 153 ESTs related to resistance. Using the Clustal W program, these sequences were aligned with all the sequences of the libraries cited above, the sequences from subtractive libraries provided by CIRAD, and cacao sequences present in databanks (JONES et al., 2002; VERICA et al., 2004). The results of the sequence multiple alignments were processed by the CLOURE-D program. Thus, it was possible to identify 85 possible SNPs, with a high frequency of transition-type. Most of the SNPs were encountered in sequences of genes involved in signal transduction, protein synthesis and defence. To validate the SNPs, we tested the AS-PCR method, but results was not conclusive. More

experiments are planned to obtain more convincing results. On the other hand, 125 SSRs were obtained by the analysis of 10 sequences from TSH 1188 and Catongo meristems, and 675 sequences from TSH 1188 fruits. Thirty-two *primers* were designed and used to amplify 23 genotypes of cacao with different levels of resistance. Four polymorphic loci involved in signal transduction or in unknown function sequence classes were obtained, and 27 alleles related to resistance of cacao to witches' broom were identified. The results had shown a separation of the susceptible and resistant genotypes, and a certain separation between the resistant genotypes, needing however deepened studies in this direction.

LISTA DE FIGURAS

Figura 1. Ciclo de vida de <i>Moniliophthora perniciosa</i> e sintomas relacionados ocorrendo no cacaueteiro.....	6
Figura 2. Exemplo de polimorfismo baseado em variação pontual de um único polimorfismo (SNPs)	16
Figura 3. Variações na seqüência de um gene podem levar ou não a alterações na estrutura da proteína codificada por este	16
Figura 4. Base genética e detecção de polimorfismo de microssatélites	19
Figura 5. Diagrama mostrando a técnica AS-PCR para validação dos SNPs	23
Figura 6. Esquema mostrando a obtenção dos <i>primers</i> que flanqueiam as regiões dos SSR e os produtos de amplificação	24
Figura 7. Percentual de ESTs selecionados por classe funcional	35
Figura 8. Exemplo de resultado obtido pelo uso do programa CLUSTAL W	36
Figura 9. Exemplo de resultado obtido pelo após utilização do programa CLOURE-D.....	37
Figura 10. Porcentagem de SNPs identificados por função dos ESTs.....	45

Figura 11. Otimização do AS-PCR quanto a concentração de MgCl ₂ em gel de agarose 2%, com anelamento a 58,1°C, utilizando DNA de TSH 1188 (T) e Catongo (C) e os <i>triplets</i> 5, 7, 11, 14.....	47
Figura 12. AS-PCR em gel de agarose 2%, utilizando anelamento a 58,1°C, dois genótipos para cada conjunto de <i>primers</i> (T = TSH 1188 e C= Catongo) e 9 <i>triplets</i> de <i>primers</i>	48
Figura 13. AS-PCR em gel de poliacrilamida 6%, utilizando 58,1°C como temperatura de anelamento	49
Figura 14. Número de microssatélites (SSR) por repetição	54
Figura 15. Porcentagem de SSRs identificados por função dos ESTs.....	55
Figura 16. Distribuição de alelos por locos dos <i>primers</i> SSR obtido pelo programa Genetix, a partir de 31 <i>primers</i> SSR e clones de cacauero com diferentes níveis de resistência a vassoura-de-bruxa	59
Figura 17. Géis de SSR dos <i>primers</i> EST-SSR-27, EST-SSR-28, EST-SSR-29, com os genótipos de cacau com diferentes níveis de resistência	59
Figura 18. Análise de componente principal dos clones com diferentes níveis de resistência usados para análise de polimorfismo por SSR.....	61

LISTA DE TABELAS

Tabela 1. Similaridade genética entre clones de cacauzeiros estimada com base em marcadores RAPD (<i>Random Amplified Polymorphic DNA</i>).....	9
Tabela 2. Descrição dos dados das bibliotecas de cacau.....	14
Tabela 3. Genótipos de cacau com diferentes níveis de resistência a <i>M. pernicioso</i> obtidos por análise RAPD utilizados para avaliação de polimorfismo ...	26
Tabela 4. Genes relacionados à resistência, selecionados a partir das bibliotecas de interação cacau- <i>M. pernicioso</i> confeccionadas pela UESC (meristema suscetível e resistente) e CEPLAC (fruto resistente).....	29
Tabela 5. Numero de SNPs candidatos identificados através do programa CLOURE-D a partir dos 153 genes selecionados	39
Tabela 6. Descrição dos SNPs identificados entre genótipos.....	42
Tabela 7. Descrição e quantificação dos tipos de substituições identificados nos diferentes genótipos	44
Tabela 8. Seqüências dos <i>primers</i> desenhados para desenvolver o método de AS-PCR e respectivos genes, temperatura de desnaturação (T_m) e anelamento (T_A),e tamanhos dos produtos amplificados no controle e nos SNPs	46

Tabela 9. Descrição de SSR identificados em 561 genes da interação cacau- <i>M. pernicioso</i>	51
Tabela 10. Seqüências de <i>primers</i> SSR obtidos a partir dos genes de interação cacau- <i>M. pernicioso</i>	57
Tabela 11. <i>Primers</i> utilizados na amplificação dos clones, número total de alelos por loco e alelos obtidos em cada um, em pares de bases.....	60
Tabela 12. Freqüência de quatro locos microssatélites de ESTs da interação cacau- <i>M. pernicioso</i>	60

1. INTRODUÇÃO

A vassoura-de-bruxa do cacau (Theobroma cacao L.), causada por *Moniliophthora perniciosa* (Stahel) Singer (AIME; PHILLIPS-MORA, 2005) ocorre em plantações de cacau da América do Sul e Ilhas do Caribe. No Brasil, após permanecer confinada à região Amazônica, a doença foi constatada em maio de 1989 no sudeste da Bahia, região responsável por 84.5% da produção brasileira de cacau (PEREIRA, 1989). Desde então, a produção de cacau foi reduzida drasticamente observando-se perdas de até 100% em muitas plantações. Isto levou ao abandono de muitas das propriedades, resultando numa séria crise econômica e social na região sul da Bahia.

A principal forma de controle consiste no uso de variedades resistentes, e as principais fontes de resistências derivam de seleções de Scavinas. Entretanto já foi observado que a resistência de híbridos oriundos destas tem sido suplantada por *M. perniciosa* (ANDEBRHAN et al., 1998). A obtenção de novas fontes de resistência a *M. perniciosa* tornou-se prioridade do programa de melhoramento de países produtores de cacau. Com o objetivo de aumentar a base genética do cacau para resistência a este fungo, novos genótipos vêm sendo selecionados entre variedades melhoradas e acessos de cacau silvestres disponíveis nos bancos de germoplasma (ANDEBRHAN et al., 1998).

Recentemente, os laboratórios da CEPLAC/CEPEC, da UESC e do CIRAD desenvolveram trabalhos de genômica para identificar genes expressos envolvidos na resistência durável à vassoura-de-bruxa, a partir de híbridos de Sca 6 (CEPLAC/UESC) e do próprio Sca 6 (CIRAD) (GESTEIRA et al., 2004; GESTEIRA et al., 2007; LANAUD et al., 2006; ZAIDAN et al., 2005). Assim, bibliotecas de cDNA de meristemas e frutos de materiais resistentes e susceptíveis foram seqüenciadas, o que permitiu a identificação de genes expressos relacionados à resistência e a

susceptibilidade a vassoura-de-bruxa. A busca de polimorfismo nesses ESTs (SNPs e SSRs) na expectativa de correlacionar a sua função com a resistência observada em populações segregantes, vem a ser uma sequência natural de pesquisa.

Os SNPs são as variações de sequências mais abundantes em vários genomas (CHO et al., 1999; PICOULT-NEWBERG et al., 1999). Em plantas, SNPs foram desenvolvidos com sucesso em várias espécies como no milho (CHING et al., 2002; THORNSBERRY et al., 2001), *Arabidopsis* (CHO et al., 1999; NORDBORG et al., 2002), cana-de-açúcar (GRIVET et al., 2003) e trigo (SOMERS et al., 2003). Eles correspondem a marcadores bialélicos e têm se mostrados eficientes na obtenção de mapas genéticos de alta resolução, nos estudos de diversidade ou de recombinação, bem como em seleção assistida por marcadores (BUNDOCK; HENRY, 2004; GUIMARÃES; COSTA, 2002; GUPTA et al., 2004; RAFALSKI, 2002). Apesar do grande potencial desse marcador, a utilização deste em cacau é inexistente até o momento.

Os microssatélites, também conhecidos como SSR (*Simple Sequence Repeats*), são poderosos marcadores moleculares com herança mendeliana e consistem de trechos de DNA de unidades mono-, di-, tri-, tetra- ou penta-nucleotídica repetidas ao acaso, dispersas por todo o genoma (HAMADA et al., 1982). Apesar dos diversos estudos que mostraram a eficiência dos SSR em cacau (ALBUQUERQUE, 2006; FALEIRO et al., 2003; LANAUD et al., 1999; MOTAMAYOR et al., 2002; SERENO et al., 2006), ainda não se dispõem de dados de SSR obtidos a partir de seqüências de ESTs, o que é uma abordagem muito interessante, e já que já foi obtida para outras culturas com muito sucesso (BRONDANI et al., 1998).

Dentro deste contexto, esse trabalho objetivou a identificação de polimorfismo em ESTs da interação cacau-*Moniliophthora perniciosa* utilizando os SNPs e SSRs, com o intuito de serem posteriormente usados em trabalhos de mapeamento, co-localização com QTL relacionado com vassoura-de-bruxa, estudos de diversidade e também seleção assistida por marcadores (SAM).

2. REVISÃO BIBLIOGRÁFICA

2.1. Origem, dispersão e importância econômica do cacau

O cacau, pertence à ordem Malvales, família Malvaceae e gênero *Theobroma* (ALVERSON et al., 1999). É uma planta perene, arbórea e dicotiledônea, nativa das florestas tropicais úmidas das Américas Central e do Sul, cuja origem provavelmente situa-se nas nascentes dos rios Amazonas e Orinoco (FIGUEIRA et al., 1992). Esta espécie foi descrita por Lineu em 1737 como *Theobroma fructus*, sendo definitivamente designado como *Theobroma cacao* L em 1753 (ALVERSON et al., 1999).

Da Amazônia, o cacau dispersou em dois grandes grupos raciais distintos: o Crioulo e o Forasteiro. O Crioulo é cultivado da América do Sul até o México sendo considerado o primeiro cacau domesticado, que produz frutos grandes com casca enrugada, sementes com interior branco ou violeta pálido e que fornecem o chocolate de melhor qualidade (CUATRECASAS, 1964). O Forasteiro está limitado mais à oeste da Cordilheira dos Andes, do norte do Brasil até as Guianas, e é considerado o verdadeiro cacau brasileiro, caracterizando-se por possuir frutos ovóides, com casca lisa e sementes com o seu interior violeta escuro (ROSÁRIO et al., 1978). Nesse grupo encontram-se as mutações da variedade comum, o Catongo e o Amelonado e é considerado o grupo que apresenta maior diversidade genética e melhor desempenho agrônomico (MARITA et al., 2001). A partir da hibridização natural desses dois grupos em plantações devastadas por doenças em Trinidad surgiu um terceiro grupo, o Trinitário, com características bastante distintas entre os seus principais representantes. Suas sementes apresentam coloração interna que varia do amarelo pálido até o roxo escuro e o seu produto final é de qualidade

intermediária. Esse grupo foi introduzido na região cacauera baiana pela Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC), para utilização em programas de melhoramento genético (CHEESMAN, 1944).

Na Bahia, o cacauero foi introduzido em 1746 pelo colono francês Luís Frederico Warneaux, que trouxe as sementes do Pará. Estas foram plantadas na fazenda Cubículo, situada à margem direita do Rio Pardo, no atual município de Canavieiras (VELLO; GARCIA, 1971), Sul da Bahia, com muito sucesso. No final do século XIX, a Bahia já era um grande produtor mundial e o principal produtor nacional, o cacau chegando a representar 40 a 50% do total das exportações do estado da Bahia (PEREIRA et al., 1989).

A cultura do cacauero é de grande relevância por possuir valor industrial e alimentício. A sua importância econômica é traduzida pelo consumo de chocolate sob as mais variadas formas por todo o mundo, também pela disseminação da utilização da manteiga de cacau, polpas, dentre outros produtos (PURDY; SCHMIDT, 1996). Além disso, as plantações de cacauero exercem um importante papel na conservação e na preservação ambiental da Mata Atlântica e da Floresta Amazônica, visto que é cultivado sob a floresta nativa conservando assim a sua fauna e flora (ARAÚJO, 1997).

2.2. Origem, propagação e sintomas da vassoura-de-bruxa

Dentre as principais doenças que afetam a cacauicultura, a vassoura-de-bruxa é uma das mais importantes, podendo causar danos de até 100% em muitas plantações (PEREIRA et al., 1989). Originária da planície Amazônica foi descrita pela primeira vez em 1875 no Suriname, causando sérios prejuízos no município de Saramacca (HOLLIDAY, 1952). Hoje ocorre em todos os países da América Latina e Ilhas do Caribe e, em 1989 foi relatada na região Sul da Bahia, em focos nos municípios de Camacan e Urucuça (PEREIRA et al., 1989). Desde então essa doença propagou-se por toda região cacauera da Bahia onde encontrou condições propícias ao seu desenvolvimento, transformando-se em uma epidemia que atinge a maior parte das lavouras, provocando drástica redução na produção do cacau, o que causou grave crise econômica, social e ecológica nessa região, levando ao abandono de muitas propriedades. Atualmente a produção de cacau no Brasil está

em torno de 169.000 toneladas de amêndoas, situando o país em quinto lugar na produção mundial (FAO, 2005).

A vassoura-de-bruxa é causada pelo fungo *Moniliophthora (=Crinipellis) perniciosa* (Stahel) Singer (AIME; PHILLIPS-MORA, 2005), pertencente à classe dos Basidiomicetos, família Tricholomataceae e ordem Agaricales.

Os basidiósporos são os únicos propágulos comprovadamente infectivos de *M. perniciosa* e têm habilidade para infectar os tecidos meristemáticos de qualquer parte da planta (flores, frutos e ramos), causando inicialmente uma hipertrofia (fase biotrófica) seguida por uma necrose (fase necrotrófica). Em meristemas vegetativos, *M. perniciosa* causa sintomas como a ausência da dominância apical e hipertrofia dos brotos com aspecto de vassouras. Em torno de 60-90 dias após a infecção as partes doentes tornam-se secas. Estas, sob condições de alta umidade relativa ($\geq 80\%$), precipitação de 1.000 a 2.000 mm e temperaturas entre 22 e 28°C produzem os basidiocarpos (PURDY; SCHMIDT, 1996). Desta maneira, os basidiocarpos liberam esporos para o início de novas infecções (Figura 1).

A epidemiologia da vassoura-de-bruxa no Sudeste da Bahia é distinta das diferentes regiões cacaueiras do Brasil. Na região Amazônica a doença é menos severa pelo fato de os plantios serem mais diversificados. As variedades de cacauero plantadas são na maioria híbridos, nos quais alguns de seus genitores possuíam níveis aceitáveis de resistência a *M. perniciosa* (ANDEBRHAN et al., 1993). No Sul da Bahia, a doença tomou proporções devastadoras, devido a suscetibilidade do hospedeiro e condições climáticas altamente propícias à introdução do patógeno (LUZ et al., 1997).

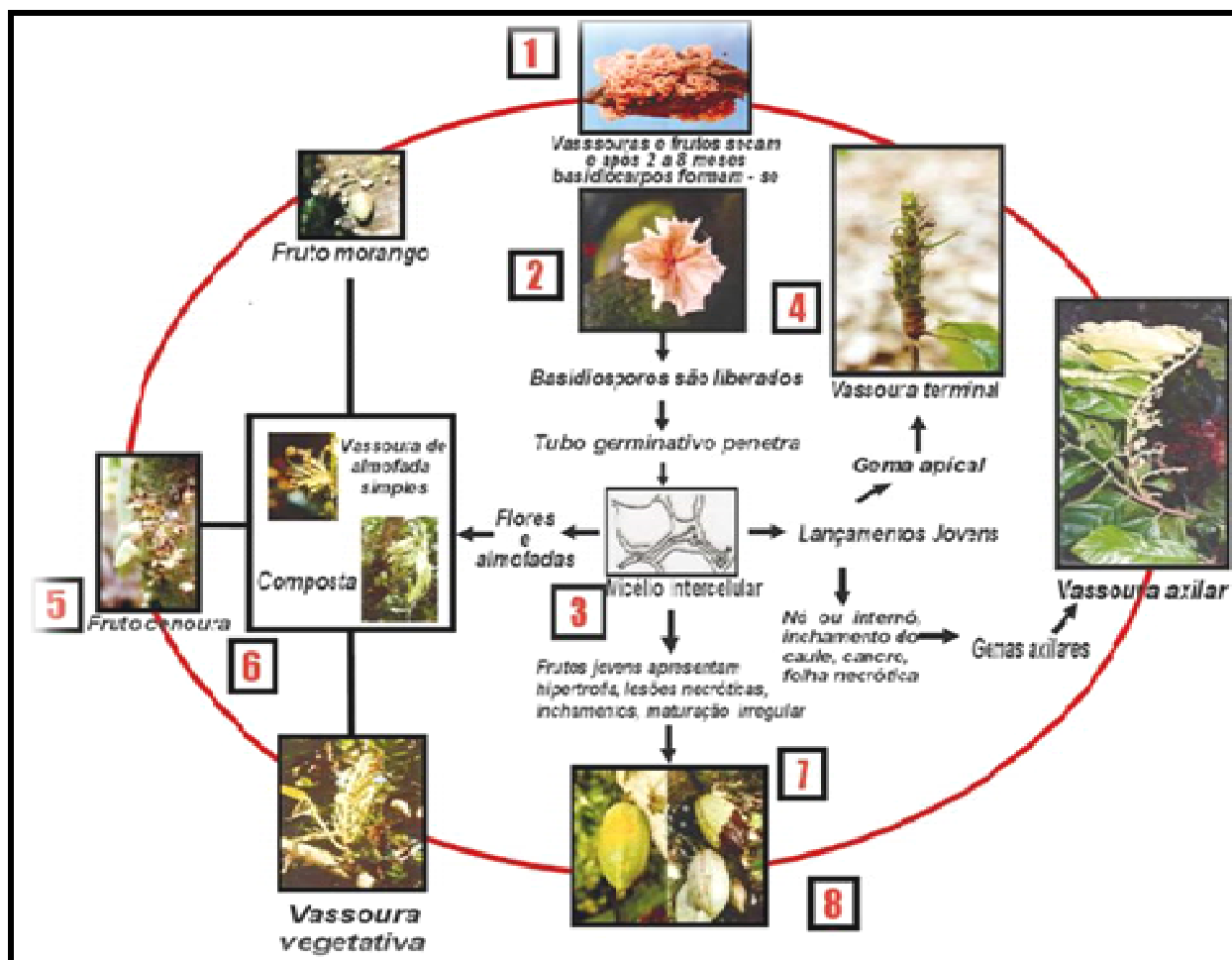


Figura 1. Ciclo de vida de *Moniliophthora perniciosa* (adaptado de WHEELER, 1989, fonte CEPLAC) e sintomas relacionados ocorrendo no cacauero.

2.3. Melhoramento do cacauero visando resistência durável à vassoura-de-bruxa

O manejo da vassoura-de-bruxa recebeu considerável atenção desde o início do século XX (EVANS; PRIOR, 1987; FULTON, 1989; PEREIRA, 2000; PURDY; SCHMIDT, 1996; WHEELER; SUÁREZ, 1993). O controle deve ser realizado através da integração de procedimentos que consiste no controle cultural, químico, biológico, e genético através do uso de variedades resistentes, naturalmente ou por indução (FRIAS, 1991; LUZ et al., 1997) de forma a reduzir a incidência da doença, aumentar a produtividade e recuperar a lavoura cacauera (WHEELER; SUÁREZ, 1993).

A resistência genética a doenças é um dos principais componentes do sistema integrado de proteção de plantas (FLORIS; ALVAREZ, 1996) e em longo prazo é a medida de controle mais eficiente no combate da vassoura-de-bruxa.

Entende-se por resistência, a capacidade da planta em atrasar ou evitar a entrada e, ou, a subsequente atividade de um patógeno em seus tecidos. Embora as plantas estejam normalmente expostas a um número incalculável de microorganismos na natureza, a resistência das mesmas mostra-se como regra, enquanto que a suscetibilidade aos agentes fitopatogênicos mostra-se como exceção (PASCHOLATI; LEITE, 1995).

O primeiro programa de melhoramento do cacauero visando a obtenção de genótipos resistentes ao *M. pernicioso* foi iniciado no Equador a partir de 1918, através da seleção massal de plântulas selecionadas de árvores sem sintomas, conhecidas como “refractarias” (BARTLEY, 1986). Cacaueros silvestres foram coletados na região Amazônica, centro de origem do cacauero, por via seminal em 1938, e por via clonal em 1942 (POUND 1938; 1943), e enviados para Trinidad onde foram avaliados para resistência à *M. pernicioso*. Em casa de vegetação, os clones Scavina 6 (Sca 6) e Scavina 12 (Sca 12) mostraram-se como genótipos livres de vassoura-de-bruxa, mas em condições de campo mostraram-se susceptíveis. Ainda assim os referidos genótipos continuam sendo usados em cruzamentos para a incorporação de genes, obtendo-se bons resultados práticos (ANDERBRHAN et al., 1998).

Na Amazônia Brasileira o programa foi iniciado em 1976 pela CEPLAC, utilizando os clones da série Scavina como principais fontes de resistência à *M. pernicioso*. Em 1990, o Centro de Pesquisas do Cacau (CEPEC), em Ilhéus, Bahia, iniciou um programa de melhoramento genético visando a substituição da variedade “comum”, estimada em pelo menos 60% dos plantios existentes na região, por variedades mais produtivas e resistentes a *M. pernicioso* (LUZ et al., 2003). Para tal, desenvolveu-se um sistema semi-automatizado de inoculação artificial com apoio do American Cocoa Research Institute (ACRI), o qual vem permitindo a avaliação de um grande número de progênies de cacau sob condições controladas. Com a utilização deste sistema, um total de 341 progênies de cacau foram avaliadas quanto à resistência à vassoura-de-bruxa. Com a chegada da doença na Bahia, dado que havia uma estrutura estabelecida e os conhecimentos sobre os Scavinas e seus descendentes, foi possível lançar variedades resistentes, todas descendentes do clone Sca 6. Para a situação emergencial da Bahia estes genótipos desempenharam o seu papel, baixando o potencial de inóculo na região. Recentemente foi possível conduzir estudos mais amplos, comparando as respostas de um número realmente

expressivo de genótipos de diferentes origens, quer através de inoculação artificial, quer pela incidência natural da doença em plantas adultas no campo (LUZ et al., 1999; PIRES et al., 1999a).

Duas linhas básicas têm sido utilizadas tradicionalmente no melhoramento genético do cacau: seleção de clones e seleção de variedades híbridas. Os primeiros trabalhos de seleção de clones foram conduzidos em Trinidad, na década de 1930, tendo como base áreas de Crioulos e Trinitários, resultando na série ICS (*Imperial College Selection*). Os clones foram multiplicados por enraizamento de estacas e distribuídos aos agricultores (PIRES, 2003). Alguns desses materiais foram cruzados com Sca 6 e Sca 12 e a seleção dentro das progênes deu origem as séries TSH (*Trinidad Selected Hybrids*) e TSA (*Trinidad Selected Amazonian*) que também foram distribuídas como sendo materiais resistentes (BARTLEY, 1994). Mais recentemente foi obtida a série EET (*Estacion Experimental Tropical*) e o clone CCN 51 (*Coleccion Castro Naranja*) de alta produtividade e resistente a vassoura-de-bruxa no Equador, que também foram distribuídos aos agricultores.

Dos principais genitores utilizados pela CEPLAC desde o início do programa de melhoramento, somente quatro clones, Sca 6, Sca 12, IMC 67 (*Mixed Collection*) e TSA 644 foram utilizados por apresentarem resistência a *M. pernicioso* (YAMADA et al., 2001). No período de 1995 a 2002, vinte clones resistentes foram lançados pela CEPLAC, sendo a maioria destes genitores descendentes de Scavina ou IMC 67. Observações recentes têm revelado a quebra de resistência de alguns destes clones em condições de campo na região Sul da Bahia (ALBUQUERQUE et al., 2005).

A obtenção de novas fontes de resistência a *M. pernicioso* tornou-se prioridade do programa de melhoramento de países produtores de cacau. Com o objetivo de aumentar a base genética do cacau para resistência a este fungo, novos genótipos vêm sendo selecionados tanto entre variedades já desenvolvidas como nos acessos de cacau silvestres disponíveis nos bancos de germoplasma (ANDEBRHAN et al., 1998).

Estudos desenvolvidos por Marita et al. (2001), constataram uma grande diversidade genética nos clones de cacau que apresentaram baixa incidência de vassoura-de-bruxa nos bancos de germoplasma da região Sul da Bahia. De 270 clones, 180 foram avaliados como resistentes, 67% pertenciam à séries brasileiras do Cruzeiro do Sul ou eram descendentes de Sca 6, ou Sca 12. Com base na

observação do caráter resistência à vassoura-de-bruxa, em campo e em casa-de-vegetação, e também com base na distância genética do clone Sca 6, cerca de 20 genótipos têm se mostrado altamente promissores para a obtenção de novas fontes de resistência, por conseguinte passaram a ser base do programa de melhoramento genético do CEPEC (PIRES, 2003; Tabela 1).

Clone	CCN-10	CCN-51	C-Sul-3	EET-392	EET-399	ICS-32	Ma-16	MO-20	Moq-216	Na-33	OC-67	P-4B	RB-36	Sca-6	Sca-12	SPA-12
CCN-10	100	56	39	33	46	49	50	45	41	46	41	53	37	31	29	47
CCN-51		100	33	34	44	46	49	50	39	41	45	45	35	37	35	49
C.Sul-3			100	36	45	39	34	42	31	40	37	35	40	39	38	44
EET-392				100	45	33	30	39	35	35	36	37	30	48	48	35
EET-399					100	39	45	47	40	54	40	46	30	33	34	55
ICS-32						100	43	53	32	49	63	38	34	35	33	40
Ma-16							100	45	37	48	47	50	34	30	30	59
MO-20								100	32	43	58	44	36	41	40	40
MOQ-216									100	41	30	37	26	32	32	47
Na-33										100	38	48	33	39	39	54
OC-67											100	39	30	32	31	37
P-4B												100	41	31	31	59
RB-36													100	32	32	41
Sca-6														100	98	28
Sca-12															100	28
SPA-12																100

Tabela 1. Similaridade genética entre clones de cacauzeiros estimada com base em marcadores RAPD (*Random Amplified Polymorphic DNA*). De acordo com Pires et al. (2003).

2.4. Resistência do cacauzeiro ao fungo *Moniliophthora perniciosa*

As alterações que constituem a resposta de defesa das plantas frente à infecção causada por patógenos têm sido intensamente estudadas com o objetivo do melhoramento vegetal visando resistência a doenças (KOMBRINK; SOMSSICH, 1997). A interação entre a planta e o patógeno pode ser dividida em dois tipos básicos: interação compatível e interação incompatível. Na interação compatível o patógeno invade o tecido vegetal, estabelece a relação parasítica e coloniza a planta, provocando a doença. Na interação incompatível, o patógeno, ao infectar o tecido vegetal, encontra as defesas da planta, que são rapidamente ativadas impedindo sua multiplicação e produzindo resistência (CAMARGO; BERGAMIN FILHO, 1995).

A interação incompatível, que produz o fenômeno de resistência, pode ser classificada em monogênica ou poligênica de acordo com o número de genes envolvidos (FLORIS; ALVAREZ, 1996). Segundo Vanderplank (1968), existem resistências que são efetivas contra algumas raças do patógeno e resistências que são efetivas contra todas as raças. No primeiro caso temos as resistências monogênicas, e no segundo caso temos as resistências poligênicas. A resistência monogênica age no sentido de reduzir a quantidade efetiva de inóculo inicial, fazendo com que o início da epidemia seja atrasado. A resistência poligênica diminui o tamanho das lesões produzidas pelo patógeno, aumenta seu período de incubação e diminui o número de esporos produzidos por lesões. Todos os seus efeitos são parciais, quantitativos e mais duráveis (AGRIOS, 1997; VALLE et al., 1999; VANDERPLANK, 1968).

Segundo Robinson (1978), provavelmente a resistência monogênica nunca ocorre desacompanhada da resistência poligênica, e esta ocorre em todas as plantas contra todos os patógenos, apesar de alguns cultivares não apresentarem controle a um nível agrônomico satisfatório. O valor da resistência monogênica pode ser aumentado, quando reforçado com níveis úteis de resistência poligênica (ROBINSON, 1976).

Apesar da grande importância do cacauzeiro, os progressos no melhoramento para resistência à vassoura-de-bruxa têm sido modestos, em comparação com os alcançados em muitas culturas. Entre as principais razões para este fato estão o desconhecimento do controle genético dos mecanismos de resistência e da interação planta-patógeno. O entendimento de como os alelos de resistência do hospedeiro interagem-se com os alelos de (a)virulência do patógeno é fundamentalmente importante para a definição de estratégias de melhoramento visando a resistência (VANDERPLANK, 1968).

Diferenças genéticas no patossistema *Theobroma cacao-Moniliophthora perniciosa* têm sido detectadas por marcadores genéticos, por avaliações fenotípicas de patogenicidade e por melhoramento genético visando resistência à doença (GOMES et al., 2000; NIELLA, 2000; PIRES et al., 2001; WHEELER; MEPSTED, 1988). Os resultados evidenciam que provavelmente o cacauzeiro apresenta resistência monogênica e oligogênica e poligênica (FALEIRO et al., 2002; NELSON, 1978; PARLEVLIET, 1993; PARLEVLIET; ZADOKS, 1977).

Reforçando a existência de resistência monogênica, estudos demonstraram que a base genética dos clones Sca 6 e 12 é dominante (BARTLEY, 1986), que seus genes possuem alta herdabilidade (PIRES et al., 1999; PURDY et al., 1997), e que quando estes clones são inoculados com diferentes isolados de *M. perniciosa* apresentam comportamento diferenciado (WHEELER, 1989). Niella (2000), estudando a patogenicidade de três isolados de *M. perniciosa* provenientes da Amazônia em quatro clones de cacau, demonstrou uma reação diferencial de Sca 6 a *M. perniciosa* sob pressão elevada de inóculo e sugeriu a existência de resistência monogênica. Apesar deste comportamento diferenciado, os clones Sca 6 e 12 ainda possuem um razoável nível de resistência.

Em favor da resistência poligênica, temos que estes clones, quando inoculados com isolados de *M. perniciosa*, apresentam menor incidência a doença, menor número e tamanho médio de vassouras por clone e menos esporulação do que os clones expressando os sintomas típicos da doença (PIRES et al., 1999; LUZ et al., 1999). Segundo Pires (1999), isto pode ser uma indicação de genes de efeitos menores, o que explica em parte a razoável resistência dos seus descendentes. A identificação de QTL (*Quantitative Trait Loci*) com genes de efeitos menores reforçam este argumento (CROUZILLAT et al., 1996; LANAUD et al., 1996; LANAUD et al., 2004; MOTILAL et al., 2000; PUGH et al., 2004; QUEIROZ et al., 2003; RISTERUCCI et al., 2000).

Evidências da resistência poligênica também foram apontadas por Robinson (1978) ao verificar que cacauzeiros resistentes submetidos a uma maior pressão de inóculo tornavam-se menos resistentes, e que aqueles moderadamente resistentes tornavam-se altamente resistentes quando submetidos a menor pressão de inóculo, afirmando a existência de resistência poligênica, que é uma resistência quantitativamente variável, e herdada poligeneticamente e sobretudo é uma resistência permanente.

Evidências da existência destes dois tipos de resistência foram observadas por Faleiro et al. (2002) através do estudo da capacidade específica de interação (CEI) de quatro clones de cacau inoculados com isolados de *M. perniciosa* de diferentes municípios da região cacauzeira, onde foi demonstrada a existência não muito pronunciada da resistência monogênica e a existência de resistência poligênica nas progênies de Sca 6 X ICS 1 e nas progênies de polinização aberta de

TSH 1188, e afirmaram que a resistência poligênica teve maior efeito no TSH 1188 do que no Sca 6 devido a presença de duas fontes de resistência.

A resistência monogênica é preferida em alguns programas de melhoramento, principalmente em culturas de ciclo curto. Como apontou Zadoks (1999), para culturas perenes, como é o caso do cacaueteiro, é muito importante a presença de diferentes genes de resistência para diminuir a pressão de inóculo do patógeno e aumentar a durabilidade da resistência (ZADOKS, 1999). Uma estratégia adotada para plantas perenes é a piramidização de diferentes genes de resistência, visando aumentar a eficiência e a durabilidade da resistência. Esse tipo de estratégia é atualmente o principal objetivo do Programa de Melhoramento do Cacaueteiro do CEPEC/CEPLAC (PIRES et al., 1996b).

2.5. Seqüências expressas (ESTs) no cacaueteiro

A análise de bibliotecas de cDNA através do seqüenciamento dos ESTs correspondentes (*Expressed Sequence Tags*) é uma abordagem potente no objetivo de selecionar apenas os genes que estão sendo expressos em determinado organismo ou em determinada condição fisiológica. Este método permite a identificação de genes de interesse (como genes envolvidos na resistência no caso de interações plantas-patógenos) e tem sido usado como uma importante abordagem experimental para elucidar não somente a função gênica, mas também para compreender como os mecanismos moleculares estão relacionados com processos biológicos (VEDOY et al., 1999). ESTs podem também ser utilizados como fontes de polimorfismo não neutro, sendo que este será de fato procurado em regiões codantes e expressas, aumentando o número de genes candidatos para estudo de QTLs (HATEY et al., 1998).

Nos últimos cinco anos, diferentes bibliotecas de ESTs foram obtidas em cacaueteiro. Em 2002, Jones et al. seqüenciaram 4.455 ESTs de folhas e sementes e obtiveram um *unigene* contendo 1.380 sequences. A análise desses genes permitiu a identificação de 46 seqüências similares a genes de defesa com função conhecida e outros de função ainda desconhecida. Em 2004, Verica e colaboradores fizeram um estudo de identificação de genes expressos através da confecção de bibliotecas subtrativas em folhas de cacau tratadas com indutores de resposta de defesa. Foi

identificado um *unigene* contendo 1.256 sequências, representando os genes expressos nas folhas elicítadas (Tabela 2). Entre essas seqüências foram identificados 330 genes induzidos durante a resposta de defesa.

Recentemente, os laboratórios da CEPLAC/CEPEC, da UESC e do CIRAD desenvolveram trabalhos de genômica para indentificar genes expressos envolvidos na resistência durável à vassoura-de-bruxa, a partir de híbridos de Sca 6 (CEPLAC/UESC) e do próprio Sca 6 (CIRAD) (GESTEIRA et al., 2004; GESTEIRA et al., 2007; LANAUD et al., 2006; ZAIDAN et al., 2005). Assim, bibliotecas de cDNA de meristemas e frutos de materiais resistentes e susceptíveis (inoculados e não inoculados) foram seqüenciadas, o que permitiu a identificação de genes expressos relacionados à resistência e a susceptibilidade. Na UESC foram geradas 6.884 seqüências em duas bibliotecas de meristemas de cacau de material suscetível (Catongo) e resistente (TSH 1188) a vassoura-de-bruxa. Deste total de ESTs, 2.926 correspondem a *unigenes*, sendo um *unigene* de 1719 sequências para TSH 1188 e um de 1207 sequências para Catongo (Tabela 2). Estes *unigenes* são consistidos de 199 contigs e 1520 sequências únicas no caso de TSH 1188, e 142 contigs e 1065 sequências únicas no caso de Catongo. As duas bibliotecas tiveram uma alta especificidade quando comparadas entre si, e quando comparadas com as duas outras bibliotecas publicadas de cacau (JONES et al., 2002; VERICA et al., 2004). A análise das seqüências permitiu atribuir uma categoria funcional em 54% das seqüências obtidas das duas bibliotecas. Na CEPLAC foram geradas 1.274 seqüências de frutos de TSH 1188, e formados 191 contigs e 675 sequências únicas, formando um *unigene* de 866 seqüências (Tabela 2). Dessas 866 seqüências, 552 apresentaram homologia com seqüências de banco de dados de proteínas não redundantes “nr” através do BLAST X. Das 1.274 seqüências geradas, 11 tiveram homologia com proteínas de fungos e destas nove apresentaram homologia com seqüências do banco de dados de *M. pernicioso* (ZAIDAN et al., 2005). No CIRAD foram construídas bibliotecas SSH (*Suppression Subtractive Hybridization*) de meristemas de Sca 6 para facilitar a identificação adicional de resistência ou genes de defesa expressos em cacau durante infecção por *Moniliophthora pernicioso*. Nas seqüências das bibliotecas de cDNA da UESC e CEPLAC foram encontradas homologias com proteínas de particular interesse como provável peroxidase putativa, superoxide dismutase, β -1-3-glucanases, proteínas relacionadas à patogênese, Avr9/Cf-9, entre outras que podem estar relacionadas

com a resistência à vassoura-de-bruxa. Estes genes, além de dar informações funcionais, poderão ser utilizados como fonte de polimorfismo (marcadores moleculares) para posterior mapeamento no mapa de referência e colocação com QTLs relacionados de resistência à *M. perniciosa* (FALEIRO et al., 2006).

Tabela 2. Descrição dos dados das bibliotecas de cacau. nd: não determinado; R: resistente a *M. perniciosa*; S: suscetível a *M. perniciosa*; SSH: *Supression Subtractive Hybridization*.

Cultivar	Fenótipo	Tecido	Biblioteca	Instituição	N° de seqüências geradas	Sequencia única	Contig	Unigene	Referencias
TSH 1188	R inoculado	Meristema	<i>Full-lenght</i>	UESC	3613	1520	199	1719	GESTEIRA et al., 2007
	R inoculado	Fruto	<i>Full-lenght</i>	CEPLAC	1274	675	191	866	Zaidan et al., 2005
Sca 6	R inoculado	Fruto	SSH	CIRAD	nd	nd	nd	nd	Lanaud et al., 2006
Amelonado P7B R10 Spec 54 UF221 Sic 5	R/S inoculado	Folha Semente	<i>Full-lenght</i>	Masterfood UK, University of Nottingham	5545	nd	nd	1380	JONES et al., 2002
Forastero	nd	Folha	SSH	Pennstate University	2114	nd	nd	1256	VERICA et al., 2004
Catongo	S/inoculado	Meristema	<i>Full-lenght</i>	UESC	3271	1065	142	1207	GESTEIRA et al., 2007

2.6. Marcadores moleculares de alta resolução

Com o advento das técnicas modernas de biologia molecular surgiram diversos métodos de detecção de polimorfismo ao nível de DNA. Aliadas às técnicas de clonagem e seqüenciamento, estas metodologias têm possibilitado um rápido acúmulo de informações sobre a estrutura de genomas. Os marcadores moleculares são ferramentas úteis na análise de variação genética e podem ser usados para avaliar níveis de diversidade genética e relações filogenéticas intra e inter específicas (LEAL-BERTIOLI, 1998). Além disso, os marcadores permitem em um curto espaço de tempo, analisar um número praticamente ilimitado de marcas polimórficas em nível de DNA, sem influência do ambiente (PIRES et al., 2000). Uma série de técnicas de biologia molecular têm sido desenvolvidas para estudos de polimorfismos, porém a escolha da técnica depende das inferências biológicas inerentes à pesquisa em questão, assim como, do recurso financeiro e pessoal

técnico disponível (MUELLER; WOLFENBARGER, 1999). Quando comparado com outras culturas de importância agrícola, o cacaueteiro ainda é pouco estudado com o emprego dessas técnicas; entretanto, nos últimos anos, esses estudos foram intensificados, permitindo a obtenção de mapas genéticos (FALEIRO et al., 2006; PUGH et al., 2004).

Entre as técnicas de detecção de polimorfismos estão os marcadores de polimorfismos de nucleotídeos únicos (SNPs) e os microssatélites (SSR), que são marcadores altamente polimórficos e bastante empregados em estudos de polimorfismo entre seqüências e em estudos de resistência a patógenos. Estas técnicas permitem a obtenção de um número virtualmente ilimitado de marcadores moleculares cobrindo todo o genoma do organismo (SANTOS, 2005).

2.6.1. Polimorfismos de nucleotídeos únicos (*Single Nucleotide Polymorphisms* - SNPs)

Os polimorfismos de nucleotídeos únicos (*Single Nucleotide Polymorphisms* – SNPs) correspondem a posição onde existe uma alternância dos nucleotídeos A, C, G, T em uma freqüência alélica mínima de 1% numa dada população (BROOKES, 1999) e são as variações de seqüências mais abundantes em vários genomas (CHO et al., 1999; PICOULT-NEWBERG et al., 1999). O SNP é individual quando ocorre uma mutação dentro de uma seqüência de DNA e é haplogrupo quando mutações ocorrem em conjunto. As mutações são de ponto, do tipo transição, quando há uma substituição de uma purina por outra purina ou de uma pirimidina por outra pirimidina; e do tipo transversão, quando a substituição é de uma purina por uma pirimidina e vice-versa (Figura 2). Em regiões codificadoras, quando resultam em uma substituição de aminoácido, são denominados não-sinônimos, a substituição podendo ser considerada conservativa ou não-conservativa em função das características dos aminoácidos envolvidos na troca. Nesses casos, podem ocorrer modificações estruturais e funcionais nas proteínas. Quando essas substituições não alteram o aminoácido formado elas são conhecidas como sinônimas, apesar das conseqüências danosas geradas pela alteração do mRNA, da tradução e da expressão gênica (KWOK; GU, 1999) (Figura 3).

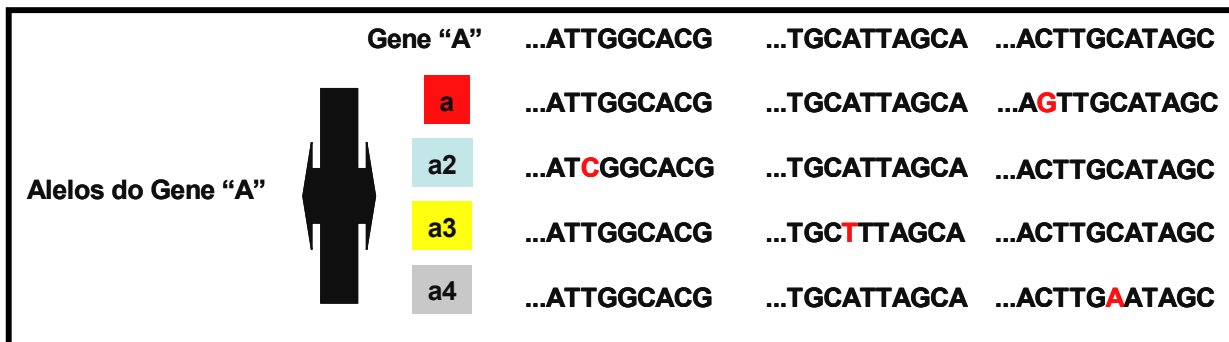


Figura 2. Exemplo de polimorfismo baseado em variação pontual de um único polimorfismo (SNPs). A forma do gene "A" está representada na linha superior. Logo abaixo está a representação da seqüência de nucleotídeos de quatro alelos de "A". Três regiões onde ocorre polimorfismo estão explicitadas e intercaladas por reticências, que representam a parte invariante do gene. No alelo "a", há substituição de C por G na terceira região. No alelo "a2", há substituição de T por C na primeira região. No alelo "a3", há substituição de A por T na segunda região. No alelo "a4", há substituição de C por A na terceira região.

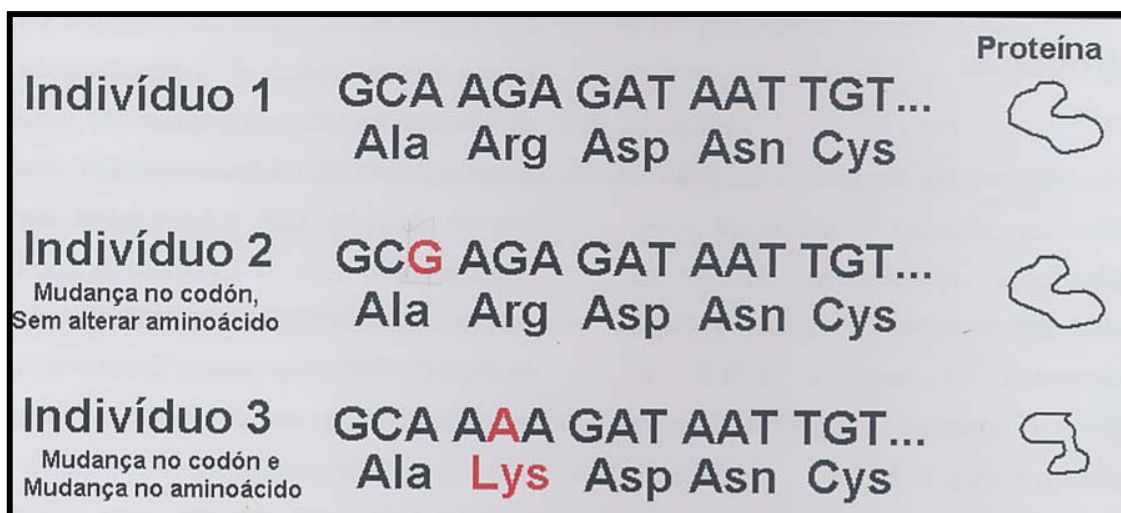


Figura 3. Variações na seqüência de um gene podem levar ou não a alterações na estrutura da proteína codificada por este. No indivíduo 1 é mostrado um SNP do tipo sinônimo, onde a substituição do nucleotídeo (A/G) não altera o aminoácido formado, não alterando a estrutura da proteína. No indivíduo 3 é mostrado um SNP do tipo não-sinônimo onde a substituição do nucleotídeo (G/A), altera o aminoácido formado, alterando a estrutura da proteína.

Os SNPs podem ser identificados através de sequenciamento direto do DNA, de informações disponíveis em banco de dados, ou através de métodos bioquímicos que investigam a presença de variantes (MUNERATO, 2005). Após a identificação dos SNPs é necessário validá-los, ou seja, determinar a frequência do SNP em um grupo de indivíduos, certificando-se de sua natureza polimórfica. Dependendo do número de amostras utilizadas nos estudos de seqüenciamento, se adotado um numero maior ou igual a 100 amostras, a validação dos SNPs é desnecessária (MUNERATO, 2005). A maior parte dos estudos visando a identificação de SNPs, tem explorado as seqüências presentes nos bancos de dados, incluindo seqüências

de cDNA ou ESTs (BUNDOCK; HENRY, 2004; GUIMARÃES; COSTA, 2002; KOTA et al., 2003). Nessa estratégia o processo de validação é extremamente importante antes de utilizar os SNPs identificados em estudos de genotipagem em larga escala.

A grande vantagem dos SNPs reside na abundância de polimorfismos entre alelos de um determinado gene. Em humanos, por exemplo, SNPs ocorrem em uma frequência de 1 a cada 1.000 nucleotídeos (COLLINS et al., 1999). Em vegetais, informações acerca desse tipo de marcadores são escassos e possivelmente uma frequência maior possa ser encontrada. Bhatramaki et al. (2000) apresentaram um trabalho referente à frequência de SNPs em milho e relataram a ocorrência de um polimorfismo a cada 70 bases em média, comprovando desta forma uma alta frequência de SNPs para a cultura do milho. Soori et al. (1999) estudaram a frequência de SNPs em batata. Um fragmento de DNA do hospedeiro ligado ao gene *RYadg* de resistência ao vírus PVY, foi amplificado por PCR e seqüenciado. Verificou-se que a substituição de um nucleotídeo aboliu o único local de reconhecimento da enzima de restrição (Bbv), resultando na substituição do aminoácido nessa posição somente para o genótipo de batata que carregava o gene *RYadg*. Os autores consideraram que o SNP pode ser determinante de um gene de resistência ou parte do gene *RYadg*.

Em plantas, SNPs foram desenvolvidos com sucesso em várias espécies como o milho (CHING et al., 2002; THORNSBERRY et al., 2001), *Arabidopsis* (CHO et al., 1999; NORDBORG et al., 2002), cana de açúcar (GRIVET et al., 2003) e trigo (SOMERS et al., 2003). Na literatura ainda não se dispõe de dados de SNPs desenvolvidos para o cacaueteiro, que podem vir a ser marcadores muito eficientes para obtenção de marcas associadas a resistência à vassoura-de-bruxa. Eles correspondem a marcadores bialélicos e têm-se mostrado eficientes na obtenção de mapas genéticos de alta resolução, nos estudos de diversidade ou de recombinação, bem como em seleção assistida por marcadores (BUNDOCK; HENRY, 2004; GUIMARÃES; COSTA, 2002; GUPTA et al., 2004; RAFALSKI, 2002).

2.6.2. Marcadores microssatélites

Os microssatélites também conhecidos como SSR (*Simple Sequence Repeats*) são poderosos marcadores moleculares com comportamento mendeliano

e consistem de trechos de DNA de unidades mono-, di-, tri-, tetra- ou penta-nucleotídica repetidas ao acaso, dispersas por todo o genoma (HAMADA et al., 1982). A técnica de SSR revela polimorfismo em um loco devido a diferenças no número de vezes (n) que uma seqüência nucleotídica (por exemplo um dinucleotídeo CA) se repete naquele loco ((CA)_n; Figura 4). Os marcadores SSR caracterizam-se por serem co-dominantes, abundantes, dispersos por todo o genoma, multialélicos e dependentes de pequena quantidade de DNA dos indivíduos analisados. O conteúdo informativo de um loco SSR é bastante alto, por se tratarem de seqüências de altas taxas evolutivas (LITT; LUTY, 1989). A análise dos SSR é realizada por meio da técnica de PCR (*Polymerase Chain Reaction*), utilizando-se *primers* com 18 a 25 bases complementares às regiões que flanqueiam as repetições. Essas variações no número de repetições constituem-se em variações no comprimento do segmento detectado pela reação de polimerase em cadeia e na separação de fragmentos amplificados em gel de eletroforese. O notável valor desta classe de marcadores é a facilidade de troca de informações entre laboratórios, visto que as seqüências dos primers proporcionam uma linguagem comum e a identificação alélica é direta e objetiva (POWELL et al., 1996). Os microssatélites são classificados em três famílias: de repetições puras, compostas ou interrompidas. Nas repetições puras, os locos de microssatélites são formados por um único motivo repetido (5'-TATTATTATTATTATTAT-3'); nas repetições compostas, mais de um motivo compõem o microssatélite (5'-TATATATATACACACACA-3'); e nas repetições interrompidas os motivos são intercalados por nucleotídeos que não fazem parte da unidade de repetição (5'-CACATTCACACACATTCA-3') (PINTO et al., 2001).

O advento desses marcadores e o avanço das técnicas de biologia molecular têm permitido a construção de mapas genéticos para diversas culturas agrícolas anuais e perenes, tais como soja (CORRÊA, 1995), café (PEARL et al., 2004), cevada (HORI et al., 2003), milho (BRUNELLI et al., 2002), eucalipto (GRATTAPAGLIA; SEDEROFF, 1994; THAMARUS et al., 2002) além do cacauzeiro (FALEIRO et al 2004; LANAUD et al., 1996; PUGH et al., 2004). Uma abordagem muito interessante e bastante útil é o desenvolvimento deste tipo de marcador a partir de ESTs. Estes banco de dados de seqüências representa uma fonte extremamente útil e importante para o descobrimento de novos microssatélites (BRONDANI et al., 1998). SSR desenvolvidos a partir de ESTs já foram

desenvolvidos em *Eucalyptus* obtendo-se ótimos resultados (BRONDANI et al., 1998), mostrando-se como uma excelente abordagem a ser adotada para a cultura do cacau, onde esse tipo de trabalho é ainda inexistente.

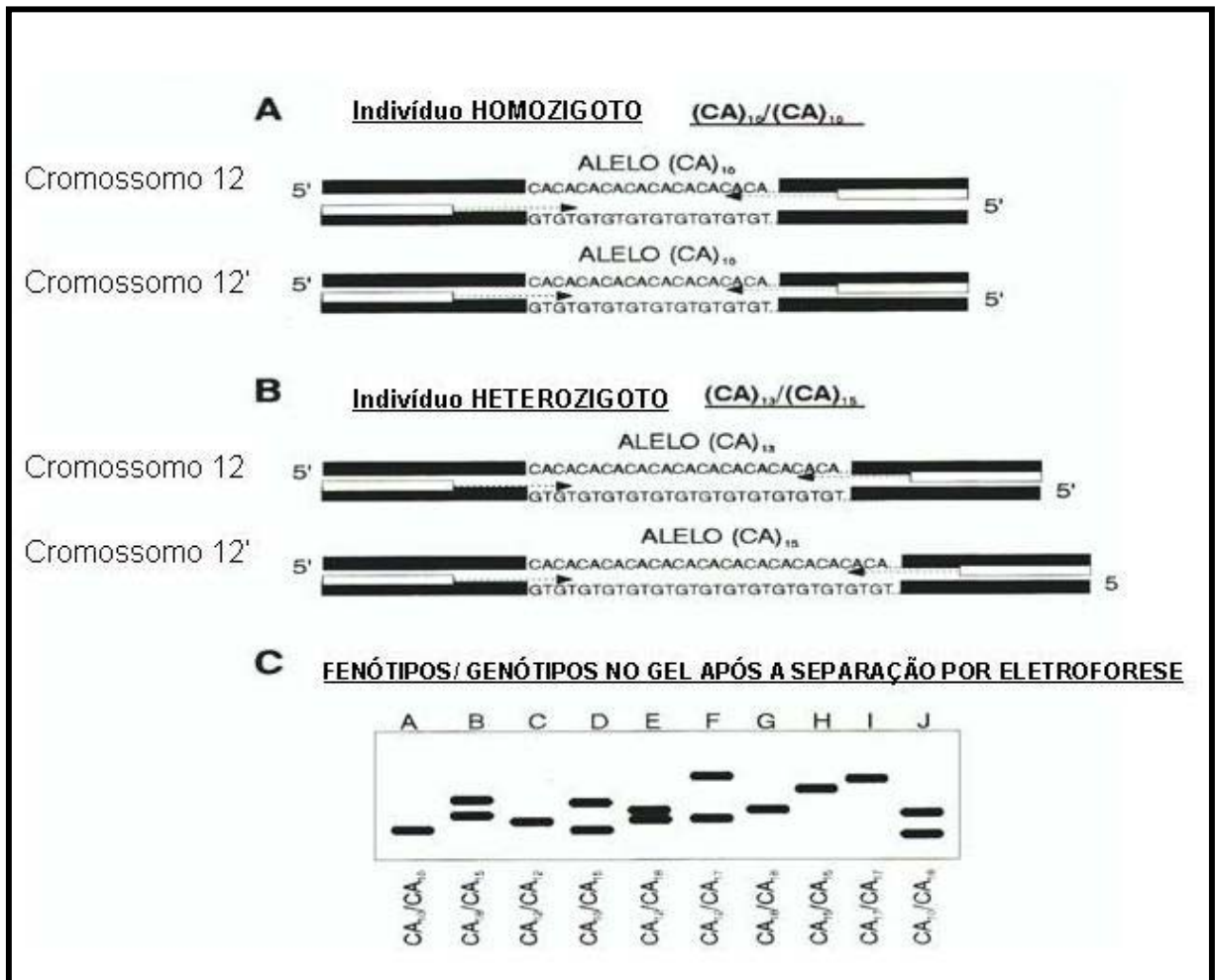


Figura 4. Base genética e detecção de polimorfismo de microssatélites. Painéis A e B ilustram genótipos homocigoto e heterocigoto, respectivamente, para uma região genômica que compreende um microssatélite de elementos (CA). Painel C ilustra um gel de eletroforese com diferentes genótipos homocigotos (banda única) e heterocigotos (duas bandas) em indivíduos diplóides. Fonte: Ferreira; Gratapaglia (1995).

3. METODOLOGIA

3.1. Seleção de ESTs

Para a análise de SNPs foram selecionados 153 genes de resistência a partir de bibliotecas da interação cacau-*M. perniciosa* de meristemas e de frutos previamente confeccionadas (GESTEIRA et al., 2004; GESTEIRA et al., 2007; LANAUD et al. 2006; ZAIDAN et al., 2005) a seguir: (i) TSH 1188 (planta resistente), bibliotecas de meristema (UESC) e de fruto (CEPLAC); (ii) Catongo (planta suscetível), biblioteca de meristemas (UESC); (iii) Sca 6 (planta resistente), biblioteca de frutos do SCA 6 (CIRAD/CEPEC). Para a seleção dos genes de resistência tomou-se como critério os genes comuns as três bibliotecas (TSH 1188 /Catongo UESC; TSH 1188 CEPLAC), alguns genes comuns às bibliotecas de cacau existentes (JONES et al, 2002, variedades Amelonado, P 7B, R 10, Spec 54, UF 221 e Sic 5; VERICA et al., 2004, Forastero), e outros genes descritos em literatura (CORAM; PANG, 2005; SANDERS et al., 2002; URBAN et al., 1999). A análise de SSR foi feita a partir do total de genes de resistência da biblioteca de TSH 1188 de frutos (CEPLAC) e de genes selecionados das bibliotecas de meristemas (UESC).

3.2. Re-seqüenciamento dos cDNA selecionados

Os genes provindos das bibliotecas fisicamente disponíveis (UESC, CEPLAC) foram re-seqüenciados no intuito de minimizar as possibilidades de encontrar falsos SNPs durante a detecção de polimorfismo. Os 153 clones referentes aos genes de resistência selecionados foram organizados em placa (96 poços) e submetidos à extração do DNA plasmidial usando o método de lise alcalina (SAMBROOK et al.,

1989) adaptado para placa. A qualidade e a quantidade dos plasmídios foram verificadas em gel de agarose 1%-TBE-BET.

Foi feito o seqüenciamento com kit Big DyEnamic Terminator V. 3.0 segundo especificações da Applied Biosystems (ABI) usando o *primer forward* M13-F 5'-TAAACGACGGCCAGT-3'. A reação foi efetuada nas seguintes condições: 95°C por 2 minutos; 40 ciclos de 95°C por 15 segundos, 50°C por 30 segundos e 60°C por 4 minutos. Em seguida o DNA foi purificado adicionando-se 20 µL de isopropanol 65%, mantido à temperatura ambiente por 15 minutos e centrifugado por 40 minutos a 4.000 rpm. O isopropanol foi removido invertendo-se a placa em papel toalha e imediatamente adicionado-se 100 µl de etanol 60% e centrifugado por 5 minutos a 4000 rpm. O etanol foi removido e a placa colocada em termociclador a 60°C por 2 minutos para remoção completa do álcool. O DNA foi ressuspendido em 10 µL de formamida HiDI, denaturado por 3 minutos a 95°C, e colocado no gelo por cerca de 5 minutos. O seqüenciamento foi feito no ABI3100 disponível do laboratório de biotecnologia da CEPLAC.

3.3. Processamento e análise de ESTs

As seqüências obtidas a partir do re-seqüenciamento foram processadas (eliminação da seqüência do vetor, dos adaptadores, do polyA) e analisadas (qualidade Phred > 10, tamanho da seqüência > 90 pb, EWING et al., 1998). Em seguida foi feito um BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) para comparação com seqüências já depositadas nos bancos de dados no intuito de verificar a similaridade das seqüências. Posteriormente as seqüências com similaridade menor que 5,E-05 e tamanho maior que 167 pares de bases foram analisadas por meio de bioinformática para detecção de polimorfismo do tipo SNPs e SSRs, utilizando os programas CLUSTAL W e CLOURE-D para detecção de SNPs e um programa em linguagem PEARL e em SAS (comunicação pessoal, Dr Uilson Lopes) para detecção de SSRs.

3.4. Identificação de SNPs em ESTs potencialmente ligados a resistência a *M. perniciosus*

Para a detecção de SNPs as seqüências obtidas a partir do re-sequenciamento foram alinhadas às seqüências de um mesmo cDNA presente em diferentes bibliotecas: Catongo, TSH 1188, Sca 6 e bibliotecas de cacau disponíveis nos bancos de dados, descritas no item 3.1. Dos resultados obtidos, excluiu-se das análises subseqüentes todas as seqüências com 100% de homologia e as que blastaram com tamanho pequeno, inferior a 60 pb. As demais seqüências foram submetidas ao programa Clustal W (<http://align.genome.jp/>) que tem por finalidade o alinhamento múltiplo de seqüências e os resultados foram diretamente processados pelo programa CLOURE-D (<http://imtech.res.in/~anand/cloure.html>), para identificação dos SNPs. O programa CLOURE-D utiliza os alinhamentos do Clustal W e realça somente os nucleotídeos diferentes em relação à seqüência de referência, que no caso são os 153 genes relacionados à resistência que foram previamente selecionados.

3.5. Construção de *primers* a partir dos SNPs

Das substituições identificadas foram selecionadas as que aconteceram entre genótipos por possuírem maiores chances de serem validados como SNPs. Para a validação foram construídos *primers* AS-PCR (PCR alelo-específico) segundo a metodologia de Soleimani et al. (2003). Essa técnica consiste na amplificação direta no DNA genômico usando 3 *primers* simultaneamente na reação. Dois *primers* foram desenhados para amplificar a seqüência do DNA genômico (seqüência controle). O *primer* específico foi desenhado para cada possível SNP de forma que o nucleotídeo da posição 3' do *primer* detecte somente o alelo do possível SNP (Figura 5). Foi utilizado o programa *Primer Design Report* para o desenho dos *primers*.

O volume utilizado para reação foi de 20 μL , contendo 30 ng de DNA, 0,2 mmol.L^{-1} do *primer* AS-PCR, 0,2 mmol.L^{-1} do *primer forward*, 0,4 mmol.L^{-1} do *primer reverse*, 2,0 mmol.L^{-1} de MgCl_2 , 0,2 mmol.L^{-1} de cada dNTP (Fermentas), 1X de

tampão de PCR $(\text{NH}_4)_2\text{So}_4$, 1 unidade de *Taq* DNA polimerase (Fermentas). As ampliações foram conduzidas em um *Eppendorf Mastercycler Gradient*, nas seguintes condições: um passo inicial de 94°C por 2 minutos; 35 ciclos de 94°C por 30 segundos, 58,1°C por 30 segundos, extensão a 72°C por 30 segundos; e um passo final de extensão a 72°C por 10 minutos. Os produtos de PCR foram verificados em géis de agarose a 1,2% e 3% corados com brometo de etídeo e em gel de poliacrilamida a 6%, tendo como tampão de corrida TBE 1X (89 mmol.L⁻¹ Tris, 89 mmol.L⁻¹ de ácido bórico e 2 mmol.L⁻¹ EDTA). O gel de poliacrilamida foi corado com nitrato de prata de acordo com Creste et al. (2001) e Gramacho et al. (2003). Utilizou-se como marcador de peso molecular o *ladder* 10 pb (Fermentas).

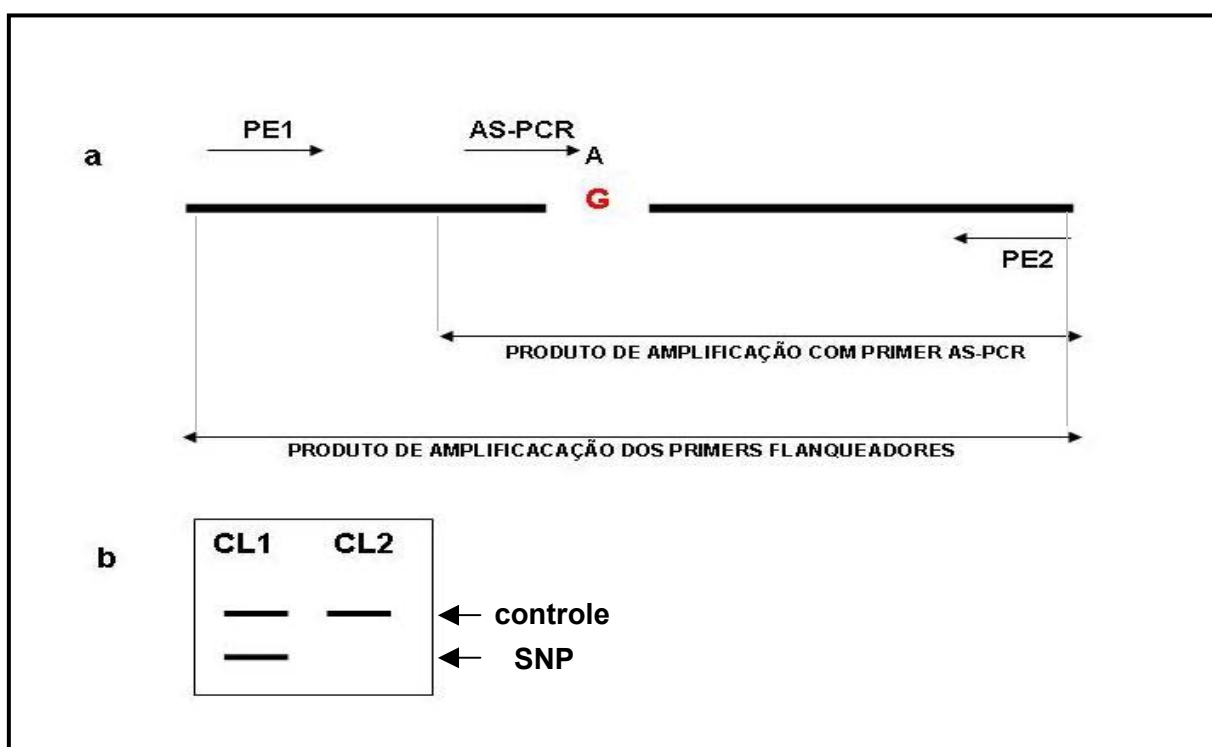


Figura 5. Diagrama mostrando a técnica AS-PCR para validação dos SNPs. (a) Alinhamento das seqüências com uma transição (SNP) G/A. Os *primers* PE1 e PE2 foram desenhados para amplificar a seqüência do controle. O *primer* AS-PCR foi desenhado para detectar o polimorfismo. (b) Banda de DNA resultando da amplificação por AS-PCR. CL1 tem duas bandas indicando a presença do SNP, entretanto o CL2 tem somente uma banda mostrando a ausência do alelo G para este loco.

3.6. Identificação de SSR em ESTs potencialmente ligados a resistência *M. perniciosa*

A identificação de SSR foi feita a partir do total de genes da biblioteca de interação cacau-*M. perniciosa* de fruto da CEPLAC e das seqüências selecionadas

da UESC, utilizando um programa desenvolvido por Uilson V. Lopes (CEPLAC) em linguagem PEARL e em SAS (comunicação pessoal). Em seguida, foram desenhados *primers* que anelam nas regiões que flanqueiam os SSR, utilizando os programas *Primer Design Report* e o *Primer 3.0* (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Figura 6).

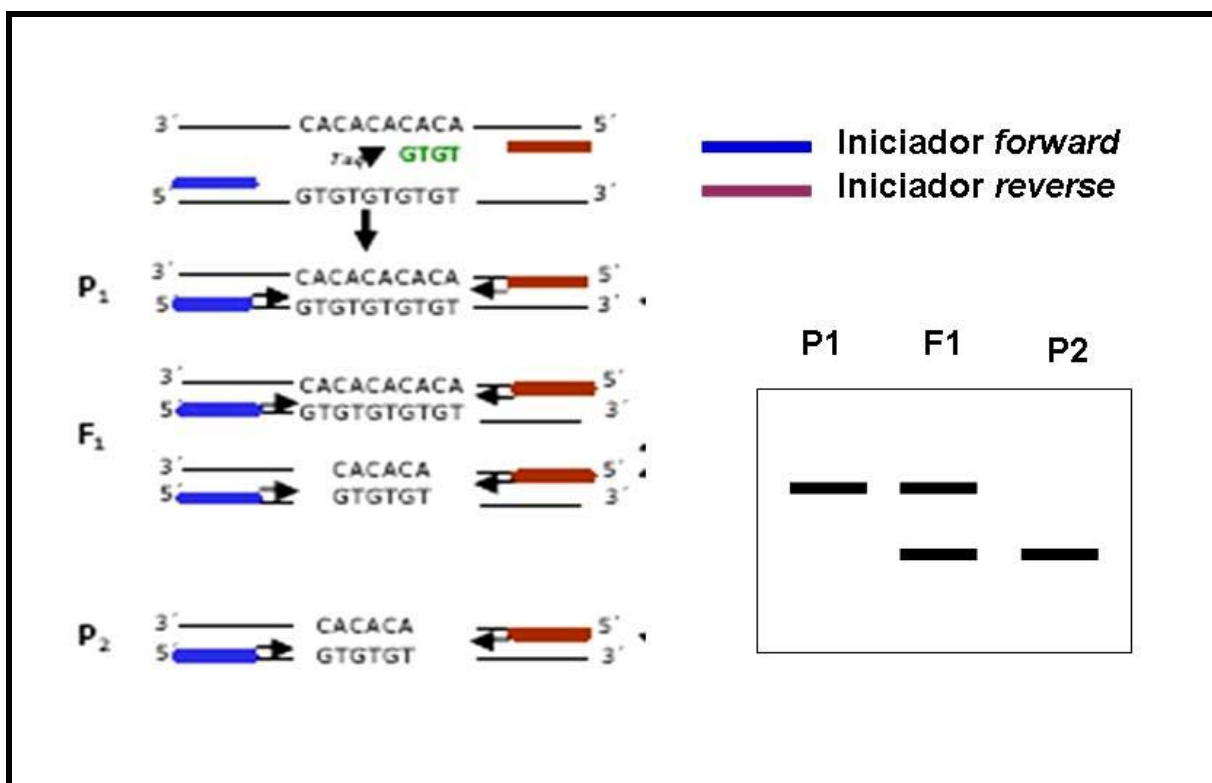


Figura 6. Esquema mostrando a obtenção dos *primers* que flanqueiam as regiões dos SSR e os produtos de amplificação. P1 e P2 representam uma fita do cromossomo de indivíduos homocigotos com elementos de SSR GT(5) e GT(3), respectivamente, e F1 mostra um indivíduo heterocigoto com os dois alelos.

O volume utilizado para reação foi de 20 μL , contendo 30 ng de DNA, 0,2 mmol.L^{-1} de cada *primer*, 2,0 mmol.L^{-1} de MgCl_2 , 0,2 mmol.L^{-1} de cada dNTP (Ludwig Biotecnologia Ltda), 1X de tampão e 1 unidade de *Taq* DNA polimerase (Ludwig Biotecnologia Ltda). Foi feito um gradiente a 56°C para obter a temperatura de anelamento dos *primers* e também foram testados com o protocolo *touch down* (TD). Na reação de *touch down* houve 10 ciclos para denaturação a 94°C por 4 min, anelamento do *primer* a 60-48°C usando 1°C de decréscimo, e extensão a 72°C por 1 minuto e 30 segundos, seguido de 30 ciclos a 94°C por 30 segundos, 48°C por 1 minuto, e 72°C por 1 minuto e 30 segundos, e uma extensão final de 4 minutos a 72°C. Para a reação com as temperaturas definidas através do gradiente foi utilizado

94°C por 2 minutos; 40 ciclos de 94°C por 15 segundos, T_A °C por 1 minuto, extensão a 72°C por 1 minuto e 30 segundos e extensão final a 72°C por 7 minutos. Os fragmentos obtidos após amplificação foram separados por eletroforese em gel de poliacrilamida a 6%, tendo como tampão de corrida TBE 1X (89 mmol.L⁻¹ Tris, 89 mmol.L⁻¹ de ácido bórico e 2 mmol.L⁻¹ EDTA). O gel de poliacrilamida foi corado com nitrato de prata de acordo com Creste et al. (2001) e Gramacho et al. (2003). Utilizou-se como marcador de peso molecular o *ladder* 10 pb (Invitrogen). A detecção de polimorfismo foi baseada no comprimento dos alelos amplificados em cada loco, registrados em géis de poliacrilamida, tomando por base os pesos moleculares do *ladder* 10 pb colocados nas laterais de cada gel.

3.7. Material vegetal

A avaliação do polimorfismo (SNPs, SSR) foi feita no DNA de parentais de clones de cacau possuindo diferentes níveis de resistência à vassoura-de-bruxa (Tabela 3). Foram adicionados também à análise de polimorfismo alguns genótipos com níveis de resistência, que não foram avaliados por RAPD mas que mostram-se como bons materiais para estudos de avaliação de resistência (Tabela 3), o DNA do Catongo (genótipo suscetível) e os DNAs de dois *bulks* obtidos de seis genótipos resistentes e seis genótipos susceptíveis respectivamente, selecionados a partir da contagem de vassouras em campo (dados não publicados).

3.8. Extração de DNA

Amostras de folhas de cada genótipo foram coletadas e armazenadas a -80°C até o momento da extração do DNA. O DNA genômico foi extraído utilizando-se o método do CTAB (DOYLE; DOYLE, 1990). Após a extração, a concentração do DNA foi estimada por espectrofotometria a 260 nm (SAMBROOK et al., 1989). Integridade e pureza do DNA foi verificado por eletroforese em gel de agarose 0,8%. Após a quantificação, as amostras de DNA de boa qualidade foram diluídas para a concentração de 10 ng/uL.

Tabela 3. Genótipos de cacau com diferentes níveis de resistência a *M. perniciosa* obtidos por análise RAPD, utilizados para avaliação de polimorfismo. nd: não determinado.

Clone	Origem	Quantidade de vassouras	Marcas de Scavina (RAPD)
R1	Peru	3	0
R2	Peru	1	2
R3	Trinidad	11	1,25
R4	Peru	15	1,4
R5	Equador	2	4
R6	Brasil	nd	nd
R7	Equador	18	1
R8	Guiana Francesa	nd	nd
R9	Equador	4	0
R10	Brasil	8	1
R11	Equador	14	1
R12	Brasil	7	1
R13	Venezuela	23	0
R14	Equador	57	2,2
R15	Trinidad	nd	nd
R16	nd	nd	nd
R17	Trinidad	nd	nd
R18	Colombia	30	1
R19	Brasil	35	1
R20	Peru	nd	nd

4. Resultados

4.1. Seleção de ESTs relacionados com resistência

Embora os escores dos BLASTs tenham sido considerados, a avaliação se uma dada similaridade poderá ou não ser envolvida no processo da infecção do cacau à *M. perniciosa* foi determinada pela anotação dos genes com similaridades com outros ESTs e informações de literatura. Os 153 genes com similaridade com genes de função conhecida representam genes envolvidos direta ou indiretamente no processo de resistência. O agrupamento destes genes de acordo com suas funções potenciais está descrito na Tabela 4. Destes 153 genes, 143 correspondem a genes da biblioteca de cDNA de fruto de TSH 1188 (material resistente, CEPLAC), 6 correspondem a genes da biblioteca de cDNA de meristema de TSH 1188 (material resistente, UESC) e 4 correspondem a genes da biblioteca de cDNA de meristema de Catongo (material suscetível, UESC).

A seleção dos genes levou em conta a presença dos mesmos nas diferentes bibliotecas (meristema e fruto), nas bibliotecas de cacau existentes nos bancos de dados (JONES et al., 2002; VERICA et al., 2004) e através de dados existentes na literatura (AUSUBEL et al., 1995; CHEONG et al., 2002; ECKER; DAVIS, 1987; HAHN, 1996).

Os genes selecionados foram agrupados em 16 classes funcionais segundo Journet et al. (2002) e Gesteira et al. (2007). A classe referente aos genes não homólogos não foi contemplada visto que foi feita uma seleção prévia na base da homologia (Figura 7). Houve apenas 1 gene selecionado das classes “tráfego de vesícula e secreção de proteína”, “citoesqueleto”, “cromatina e metabolismo de DNA”

e “ciclo de divisão celular”, totalizando 4% dos genes. Além disso, 32% dos genes estão envolvidos na classes “síntese de proteínas e processamento” e “metabolismo primário”; 28% estão envolvidos em “defesa e resgate de célula” e “transdução de sinal”, sendo 14% de cada; 12% estão envolvidos nas classes de “estrutura celular e metabolismo” e “metabolismo secundário e hormonal”; 8% estão envolvidos na classe “expressão de genes e metabolismo de RNA”; 10% estão relacionados ao “estímulo abiótico e desenvolvimento”; e 1% não tiveram função conhecida (Figura 7).

Entre os genes selecionados a partir de dados da literatura estão: *leucine-rich-repeat-transmembrane protein kinase* que possui um papel importante como sinalizador celular (JALALI et al., 2006); AVR9 que é um gene de resistência bastante estudado em tomate, que codifica uma proteína com repetições ricas em leucina e confere resistência ao fungo *Cladosporium fulvum* (ROMEIS et al., 1999); β -1,3-glucanase que é uma proteína relacionada a patogênese que pode atuar na degradação da parede celular de fungos fitopatogênicos (MAUCH et al., 1988). Foram selecionados alguns genes já descritos em envolvimento com patogenicidade como os genes *ABC transporter* (URBAN et al., 1999), *protein kinase* (KIM et al., 2000) e *serine/threonine protein kinase* (DUFRESNE et al., 1998)

Foram encontrados genes em comum nas bibliotecas de fruto, de meristema e nas bibliotecas do banco de dados, como o gene superoxide dismutase cuja importância no mecanismo de proteção de plantas (JABS et al, 1997) e cujo envolvimento em resposta hipersensível em batatas infectadas com *Phytophthora infestans* foram demonstrados (DUKE, 1983). Foram encontrados também os genes *non cell autonomous* e *heat shock protein 80* que estão envolvidos em síntese de proteínas (SORGER; PELHAM, 1987).

Tabela 4. Genes relacionados à resistência, selecionados a partir das bibliotecas de interação cacau-*M. pernicioso* confeccionadas pela UESC (meristemas suscetível e resistente) e CEPLAC (fruto resistente). As classes foram definidas como descrito por Journet et al. (2002) e Gesteira et al. (2007).

Classe	Anotação funcional	Espécie	E value	Tamanho (bp)	Biblioteca	Referência
I	Xyloglucan endotransglycosylase	<i>Gossypium hirsutum</i>	1E-38	361	CEPLAC/ FRUTO/ TSH 11 88	-
I	Arabinogalactan-protein (AGP20)	<i>Arabidopsis thaliana</i>	2E-17	502	CEPLAC/ FRUTO/ TSH 11 88	SCHULTZ et al., 2002
I	Cellulose synthase catalytic subunit	<i>Gossypium hirsutum</i>	1E-16	416	CEPLAC/ FRUTO/ TSH 11 88	RICHMOND; SOMERVILLE, 2000
I	Cellulose synthase, catalytic subunit, putative	<i>Arabidopsis thaliana</i>	7E-08	217	CEPLAC/ FRUTO/ TSH 11 88	RICHMOND; SOMERVILLE, 2000
I	Cinnamoyl coa reductase	<i>Fragaria x ananassa</i>	6E-46	324	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
I	Extensin-like protein	<i>Cucumis sativus</i>	8E-05	457	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
I	Polygalacturonase-like protein	<i>Fragaria x ananassa</i>	2E-24	373	CEPLAC/ FRUTO/ TSH 11 88	FAVARON et al., 2004
I	Probable cellulase (EC 3.2.1.4)	<i>Gossypium hirsutum</i>	8E-44	491	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
I	Putative cinnamoyl-coa reductase	<i>Oryza sativa</i>	5E-26	489	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
II	Ankyrin repeat family protein	<i>Arabidopsis thaliana</i>	9E-10	406	CEPLAC/ FRUTO/ TSH 11 88	DONG, 2004
III	ABC transporter	<i>Arabidopsis thaliana</i>	3E-43	420	UESC/MERISTEMA/ TSH 1188	SMART; FLEMING, 1996
III	Copper transporter family protein	<i>Arabidopsis thaliana</i>	2E-20	481	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
IV	Calnexin 1 (CNX1)	<i>Arabidopsis thaliana</i>	1E-54	464	CEPLAC/ FRUTO/ TSH 11 88	WANG et al., 2005
IX	10 kda chaperonin (CPN10)	<i>Arabidopsis thaliana</i>	9E-40	479	CEPLAC/ FRUTO/ TSH 11 88	BAKER et al., 1994
IX	26S proteasome non-ATPase regulatory subunit	<i>Arabidopsis thaliana</i>	1E-53	445	CEPLAC/ FRUTO/ TSH 11 88	REINHECKEL et al., 1998
IX	26S proteasome regulatory particle non-ATPase subunit8	<i>Oryza sativa</i>	1E-25	486	CEPLAC/ FRUTO/ TSH 11 88	REINHECKEL et al., 1998
IX	26S Proteasome regulatory subunit S5A	<i>Mesembryanthemum crystallinum</i>	2E-50	438	CEPLAC/ FRUTO/ TSH 11 88	REINHECKEL et al., 1998
IX	Aspartic proteinase	<i>Theobroma cacao</i>	5E-37	458	CEPLAC/ FRUTO/ TSH 11 88	SIMÕES; FARO, 2004
IX	Aspartic proteinase	<i>Theobroma cacao</i>	5E-37	401	CEPLAC/ FRUTO/ TSH 11 88	SIMÕES; FARO, 2004
IX	ATP-dependent clp protease ATP-binding subunit clpa	<i>garden pea</i>	4E-33	461	CEPLAC/ FRUTO/ TSH 11 88	HAHN, 1996
IX	ATP-dependent Clp protease proteolytic subunit	<i>Nicotiana tabacum</i>	1E-83	489	CEPLAC/ FRUTO/ TSH 11 88	HAHN, 1996
IX	Cathepsin B-like cysteine proteinase (EC 3.4.22.-)	<i>Triticum aestivum</i>	5E-55	492	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
IX	Copper chaperone (CCH)-related	<i>Arabidopsis thaliana</i>	2E-13	427	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
IX	Copper homeostasis factor putative (CCH)	<i>Arabidopsis thaliana</i>	1E-28	494	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
IX	Cyclophilin	<i>Arabidopsis thaliana</i>	5E-67	445	UESC/MERISTEMA/ TSH 1188	LEE et al., 2007
IX	Heat-shock protein 80	<i>Euphorbia esula</i>	2E-56	587	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002; VERICA et al., 2004
IX	Methionyl aminopeptidase, putative	<i>Arabidopsis thaliana</i>	3E-35	449	CEPLAC/ FRUTO/ TSH 11 88	-

Tabela 4. Continuação

Clas	Anotação Funcional	Espécie	E value	Size (bp)	Biblioteca	Referências
IX	Non-cell-autonomous heat shock cognate protein 70	<i>Cucurbita maxima</i>	1E-25	379	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002; VERICA et al., 2004
IX	Pentameric polyubiquitin	<i>Arabidopsis thaliana</i>	1E-167	1149	CEPLAC/ FRUTO/ TSH 11 88	ZENG et al., 2004
IX	Polyubiquitin	<i>Aster tripolium</i>	1E-87	506	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
IX	Protease inhibitor	<i>Vitis vinifera</i>	4E-19	496	CEPLAC/ FRUTO/ TSH 11 88	DANGL; JONES, 2001
IX	Protease inhibitor/seed storage	<i>Arabidopsis thaliana</i>	7E-21	451	CEPLAC/ FRUTO/ TSH 11 88	DANGL; JONES, 2001
IX	Protease inhibitor/seed storage/ (LTP) family protein	<i>Arabidopsis thaliana</i>	7E-21	413	CEPLAC/ FRUTO/ TSH 11 88	DANGL; JONES, 2001
IX	Proteasome subunit alpha type 5	<i>Glycine max</i>	1E-41	490	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
IX	Type I proteinase inhibitor-like protein	<i>Citrus x paradisi</i>	2E-09	251	CEPLAC/ FRUTO/ TSH 11 88	TIAN et al., 2007
IX	Ubiquitin conjugating protein	<i>Avicennia marina</i>	4E-86	769	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
IX	U-box domain-containing protein	<i>Arabidopsis thaliana</i>	5E-38	386	UESC/MERISTEMA/ TSH 1188	ZENG et al., 2004
V	6-phosphogluconate dehydrogenase family protein	<i>Arabidopsis thaliana</i>	1E-19	329	CEPLAC/ FRUTO/ TSH 11 88	DANSON et al., 2004
V	Aconitate hydratase, cytoplasmic	<i>Cucurbita cv. Kurokawa amakuri</i>	6E-13	351	CEPLAC/ FRUTO/ TSH 11 88	PINHEIRO DE CARVALHO et al., 2003
V	Cytochrome c oxidase subunit 6b, putative	<i>Arabidopsis thaliana</i>	3E-38	416	CEPLAC/ FRUTO/ TSH 11 88	DANGL; JONES, 2001
V	Cytosolic phosphoglycerate kinase 1	<i>Populus nigra</i>	6E-37	481	CEPLAC/ FRUTO/ TSH 11 88	BOWEN et al., 1988
V	Ferredoxin [2fe-2s]	<i>Citrus sinensis</i>	3E-61	597	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	Ferredoxin [2Fe-2S] fd1 precursor	<i>Citrus sinensis</i>	3E-61	597	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	Ferredoxin family protein	<i>Arabidopsis thaliana</i>	6E-34	486	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	Ferredoxin family protein	<i>Arabidopsis thaliana</i>	6E-34	486	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	GDSL-motif lipase/hydrolase family protein	<i>Arabidopsis thaliana</i>	3E-55	468	CEPLAC/ FRUTO/ TSH 11 88	OH et al., 2005
V	Glutamate-1-semialdehyde 2,1-aminomutase (GSA)	<i>common tobacco</i>	6E-70	433	CEPLAC/ FRUTO/ TSH 11 88	-
V	Glutaredoxin family protein	<i>Arabidopsis thaliana</i>	2E-19	427	CEPLAC/ FRUTO/ TSH 11 88	ROUHIER et al., 2004
V	Haloacid dehalogenase-like hydrolase family protein	<i>Arabidopsis thaliana</i>	2E-27	481	CEPLAC/ FRUTO/ TSH 11 88	-
V	Lipase class 3 family protein	<i>Arabidopsis thaliana</i>	2E-15	456	CEPLAC/ FRUTO/ TSH 11 88	OH et al., 2005
V	Monodehydroascorbate reductase (NADH)	<i>Arabidopsis thaliana</i>	3E-25	453	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	NADH-ubiquinone oxidoreductase-related	<i>Arabidopsis thaliana</i>	2E-23	324	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	Nadp-dependent malic enzyme (nadp-me)	<i>Populus balsamifera</i>	8E-09	242	CEPLAC/ FRUTO/ TSH 11 88	PINHEIRO DE CARVALHO et al., 2003
V	NADP-dependent malic protein	<i>Ricinus communis</i>	2E-18	433	CEPLAC/ FRUTO/ TSH 11 88	PINHEIRO DE CARVALHO et al., 2003
V	Peroxioredoxin	<i>Populus tremula x Populus tremuloides</i>	2E-68	575	CEPLAC/ FRUTO/ TSH 11 88	AUSUBEL et al., 1995

Tabela 4. Continuação

Clas	Anotação Funcional	Espécie	E value	Size (bp)	Biblioteca	Referências
V	Phenylalanine-ammonia lyase	<i>Citrus clementina</i> x <i>Citrus reticulata</i>	1E-155	1053	CEPLAC/ FRUTO/ TSH 11 88	AUSUBEL et al., 1995
V	Putative ATP/GTP-binding protein	<i>Oryza sativa</i>	5E-24	261	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
V	Putative blue copper-binding protein	<i>Oryza sativa</i>	4E-08	475	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	S-adenosyl-L-methionine synthetase	<i>Elaeagnus umbellata</i>	1E-37	435	CEPLAC/ FRUTO/ TSH 11 88	KAWALLECK et al. 1992
V	Small blue copper protein Bcp1	<i>Boea crassifolia</i>	3E-24	526	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
V	Sterol carrier protein 2 (SCP-2) family protein	<i>Arabidopsis thaliana</i>	3E-50	481	CEPLAC/ FRUTO/ TSH 11 88	-
V	Thioredoxin H-type (TRX-H)	<i>Ricinus communis</i>	7E-56	485	CEPLAC/ FRUTO/ TSH 11 88	RIVAS et al., 2004
V	UDP-glucuronosyl/UDP-glucosyl transferase	<i>Arabidopsis thaliana</i>	2E-33	454	CEPLAC/ FRUTO/ TSH 11 88	-
VI	Caffeine synthase	<i>Camellia sinensis</i>	1E-16	415	UESC/MERISTEMA/ TSH 1188	JONES et al., 2002
VI	Caffeine synthase	<i>Camellia sinensis</i>	3E-45	417	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
VI	Chalcone synthase	<i>Hypericum androsaemum</i>	1E-74	390	CEPLAC/ FRUTO/ TSH 11 88	JAKOBEK; LINDGREN, 2002
VI	Chalcone synthase	<i>Rubus idaeus</i>	1E-74	439	CEPLAC/ FRUTO/ TSH 11 88	JAKOBEK; LINDGREN, 2002
VI	Flavonol synthase	<i>Ginkgo biloba</i>	2E-48	485	CEPLAC/ FRUTO/ TSH 11 88	CORAM; PANG, 2005
VI	Leucoanthocyanidin dioxygenase	<i>Vitis labrusca</i> x <i>Vitis vinifera</i>	3E-51	455	CEPLAC/ FRUTO/ TSH 11 88	-
VI	Polyphenol oxidase, chloroplast precursor (PPO)	<i>Catechol oxidase</i>	8E-90	696	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
VI	Pseudo-hevein	<i>Hevea brasiliensis</i>	1E-57	498	CEPLAC/ FRUTO/ TSH 11 88	KOO et al., 2002
VI	Putative calmodulin-binding protein	<i>Arabidopsis thaliana</i>	3E-39	501	CEPLAC/ FRUTO/ TSH 11 88	HAHN, 1996
VII	Putative Tam3-transposase	<i>Oryza sativa</i>	2E-08	432	CEPLAC/ FRUTO/ TSH 11 88	-
VIII	Acinus protein	<i>Oryza sativa</i>	2E-13	442	UESC/MERISTEMA/ CATONGO	EPRINTSEV, 2002
VIII	AP2 domain containing protein	<i>Prunus armeniaca</i>	4E-25	534	CEPLAC/ FRUTO/ TSH 11 88	STOCKINGER et al., 1997
VIII	BZIP transcription factor ATB2	<i>Glycine max</i>	7E-43	803	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
VIII	Gag-pol polyprotein	<i>Vitis vinifera</i>	1E-14	293	UESC/MERISTEMA/ TSH 1188	-
VIII	Myb family transcription factor	<i>Arabidopsis thaliana</i>	6E-17	467	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
VIII	PHD finger family protein	<i>Arabidopsis thaliana</i>	3E-24	501	CEPLAC/ FRUTO/ TSH 11 88	-
VIII	Putative leucine zipper protein	<i>Gossypium hirsutum</i>	1E-94	492	CEPLAC/ FRUTO/ TSH 11 88	ZHANG et al., 1999
VIII	Putative zinc finger protein zmzf	<i>Zea mays</i>	5E-16	317	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
VIII	Zinc finger (C2H2 type) family protein	<i>Arabidopsis thaliana</i>	6E-40	507	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
VIII	Zinc finger (C3HC4-type RING finger) family protein	<i>Arabidopsis thaliana</i>	1E-47	459	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
VIII	Zinc finger (C3HC4-type RING finger) family protein	<i>Arabidopsis thaliana</i>	1E-47	459	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
VIII	Zinc knuckle (CCHC-type) family protein	<i>Arabidopsis thaliana</i>	1E-48	534	CEPLAC/ FRUTO/ TSH 11 88	-
VIII	Ring-box protein-related	<i>Arabidopsis thaliana</i>	1E-36	352	CEPLAC/ FRUTO/ TSH 11 88	-
X	C2 Domain-containing protein	<i>Arabidopsis thaliana</i>	5E-21	331	CEPLAC/ FRUTO/ TSH 11 88	PLANT et al., 1997

Tabela 4. Continuação

Clas	Anotação Funcional	Espécie	E value	Size (bp)	Biblioteca	Referências
X	Calcineurin-like phosphoesterase family protein	<i>Arabidopsis thaliana</i>	2E-22	280	CEPLAC/ FRUTO/ TSH 11 88	COWEN; LINDQUIST, 2005
X	Calcium-binding EF	<i>Arabidopsis thaliana</i>	5E-35	297	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
X	Calcium-binding EF hand family protein	<i>Arabidopsis thaliana</i>	5E-35	297	CEPLAC/ FRUTO/ TSH 11 88	VERICA et al., 2004
X	Calmodulin	<i>Pyrus communis</i>	1E-65	489	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
X	Calmodulin	<i>Brassica napus</i>	3E-08	333	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
X	CBL-interacting protein kinase 10 (CIPK10)	<i>Arabidopsis thaliana</i>	5E-05	420	CEPLAC/ FRUTO/ TSH 11 88	MAHALINGAM et al., 2003
X	Double-stranded DNA-binding family protein	<i>Arabidopsis thaliana</i>	1E-30	449	CEPLAC/ FRUTO/ TSH 11 88	HAHN, 1996
X	Homoserine kinase	<i>Arabidopsis thaliana</i>	1E-25	497	CEPLAC/ FRUTO/ TSH 11 88	MAHALINGAM et al., 2003
X	Nucleoside diphosphate kinase B	<i>Arabidopsis thaliana</i>	2E-65	469	CEPLAC/ FRUTO/ TSH 11 88	MAHALINGAM et al., 2003
X	Phloem calmodulin-like-domain protein kinase PCPK1	<i>Cucurbita maxima</i>	6E-37	553	CEPLAC/ FRUTO/ TSH 11 88	MAHALINGAM et al., 2003
X	Phloem calmodulin-like-domain protein kinase PCPK1	<i>Cucurbita maxima</i>	6E-37	553	CEPLAC/ FRUTO/ TSH 11 88	MAHALINGAM et al., 2003
X	Phospholipase A2 family protein	<i>Arabidopsis thaliana</i>	3E-46	451	CEPLAC/ FRUTO/ TSH 11 88	-
X	Putative phosphatidylinositol-4-phosphate 5-kinase	<i>Oryza sativa</i>	2E-56	441	CEPLAC/ FRUTO/ TSH 11 88	-
X	Putative transformer serine/arginine-rich ribonucleoprotein	<i>Oryza sativa</i>	4E-18	378	CEPLAC/ FRUTO/ TSH 11 88	MAHALINGAM et al., 2003
X	Rac GDP-dissociation inhibitor 1	<i>Oryza sativa</i>	5E-10	323	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
X	Ras-related GTP-binding family protein	<i>Arabidopsis thaliana</i>	1E-77	476	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
X	Ras-related GTP-binding protein, putative	<i>Arabidopsis thaliana</i>	1E-38	391	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
X	Ras-related protein ARA-5	<i>Arabidopsis thaliana</i>	3E-46	438	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2003
X	Serine/threonine protein kinase, putative	<i>Arabidopsis thaliana</i>	4E-25	410	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
X	Signal recognition particle 9 kda protein	<i>Arabidopsis thaliana</i>	2E-50	371	CEPLAC/ FRUTO/ TSH 11 88	KIM et al., 2004
XI	Fasciclin-like arabinogalactan family protein	<i>Arabidopsis thaliana</i>	1E-10	453	CEPLAC/ FRUTO/ TSH 11 88	GILSON et al. 2001
XI	Kinesin-like calmodulin binding protein	<i>Gossypium hirsutum</i>	2E-69	482	CEPLAC/ FRUTO/ TSH 11 88	HAHN, 1996
XII	Beta-1,3-glucanase	<i>Salix gilgiana</i>	9E-39	317	CEPLAC/ FRUTO/ TSH 11 88	AUSUBEL et al., 1995
XII	Peroxidase, putative	<i>Arabidopsis thaliana</i>	4E-08	244	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII	Proline-rich protein	<i>Glycine max</i>	5E-32	790	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII	Proline-rich protein	<i>Glycine max</i>	5E-32	790	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII	Proline-rich protein	<i>Glycine max</i>	3E-17	940	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII	Selenoprotein	<i>Arabidopsis thaliana</i>	9E-16	206	UESC/MERISTEMA/ CATONGO	FU et al., 2002
XII.A	Superoxide dismutase	<i>Fagus sylvatica</i>	2E-58	417	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002; VERICA et al., 2004
XII.A	Apoptosis inhibitor	<i>Arabidopsis thaliana</i>	1E-10	275	UESC/MERISTEMA/ TSH 1188	DANGL; JONES, 2001

Tabela 4. Continuação

Clas	Anotação Funcional	Espécie	E value	Size (bp)	Biblioteca	Referências
XII.A	Avr9/Cf-9 rapidly elicited protein 141	<i>Nicotiana tabacum</i>	1E-17	270	CEPLAC/ FRUTO/ TSH 11 88	ROMEIS et al., 1999
XII.A	Disease resistance-responsive protein	<i>Arabidopsis thaliana</i>	2E-56	509	UESC/MERISTEMA/ CATONGO	HAMOND-KOSAK; JONES, 1997
XII.A	Ests AU082419(E61744),AU 031498(E61744)	<i>Arabidopsis thaliana</i>	5E-31	455	CEPLAC/ FRUTO/ TSH 11 88	-
XII.A	Glutathione S-transferase	<i>Gossypium hirsutum</i>	8E-56	453	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
XII.A	Glutathione S-transferase GST 18	<i>Glycine max</i>	2E-19	363	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
XII.A	Glutathione S-transferase GST 18	<i>Glycine max</i>	2E-19	363	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
XII.A	Glutathione S-transferase GST 22	<i>Glycine max</i>	4E-85	737	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
XII.A	Glutathione S-transferase U1	<i>Malva pusilla</i>	6E-37	459	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
XII.A	Glutathione S-transferase U2	<i>Malva pusilla</i>	9E-53	404	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
XII.A	Harpin inducing protein	<i>Nicotiana tabacum</i>	1E-26	423	CEPLAC/ FRUTO/ TSH 11 88	MATTINEN et al., 2004
XII.A	Lectin-related	<i>Arabidopsis thaliana</i>	1E-23	409	CEPLAC/ FRUTO/ TSH 11 88	ROOPASHREE et al., 2006
XII.A	Leucine-rich repeat transmembrane protein kinase	<i>Arabidopsis thaliana</i>	1E-10	406	CEPLAC/ FRUTO/ TSH 11 88	CORAM; PANG, 2005
XII.A	Metallothionein-like protein	<i>Gossypium hirsutum</i>	7E-24	482	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII.A	Metallothionein-like protein	<i>Gossypium hirsutum</i>	7E-24	482	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII.A	Pathogenesis-related protein	<i>Pyrus pyrifolia</i>	1E-57	480	CEPLAC/ FRUTO/ TSH 11 88	CORAM; PANG, 2005
XII.A	Pathogenesis-related protein 4B	<i>Nicotiana tabacum</i>	2E-57	479	CEPLAC/ FRUTO/ TSH 11 88	CORAM; PANG, 2005
XII.A	Putative cuzn-superoxide dismutase	<i>Populus tremula x Populus tremuloides</i>	3E-18	186	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002; VERICA et al., 2004
XII.A	Putative pathogenesis related protein 1 precursor	<i>Vitis vinifera</i>	6E-30	410	CEPLAC/ FRUTO/ TSH 11 88	LINTHORST, 1991
XII.A	Resistance protein	<i>Lactuca sativa</i>	3E-19	488	UESC/MERISTEMA/ CATONGO	HAMOND-KOSAK; JONES, 1997
XII.B	ABA stress ripening protein - common ice plant	<i>Mesembryanthemum crystallinum</i>	3E-7	1186	CEPLAC/ FRUTO/ TSH 11 88	PFLIEGER et al., 2001
XII.B	Auxin-binding protein ABP19a precursor	<i>Prunus persica</i>	7E-25	461	CEPLAC/ FRUTO/ TSH 11 88	SHIMOMURA et al., 1993
XII.B	Drought-responsive family protein	<i>Arabidopsis thaliana</i>	4E-11	428	CEPLAC/ FRUTO/ TSH 11 88	-
XII.B	Far-red impaired responsive protein, putative	<i>Arabidopsis thaliana</i>	2E-20	494	CEPLAC/ FRUTO/ TSH 11 88	-
XII.B	Floral homeotic protein DEFICIENS	<i>Antirrhinum majus</i>	4E-50	525	CEPLAC/ FRUTO/ TSH 11 88	DE BODT et al., 2003
XII.B	Nodulin mtn3 family protein	<i>Arabidopsis thaliana</i>	8E-43	466	CEPLAC/ FRUTO/ TSH 11 88	GAMAS et al., 1998
XII.B	Nodulin-related	<i>Arabidopsis thaliana</i>	6E-10	491	CEPLAC/ FRUTO/ TSH 11 88	GAMAS et al., 1998
XII.B	Putative senescence-associated protein	<i>Pisum sativum</i>	7E-22	167	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII.B	Putative wound-induced protein	<i>Medicago sativa subsp. x varia</i>	3E-22	454	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII.B	Stress-induced protein sti1-like protein	<i>Arabidopsis thaliana</i>	5E-19	466	CEPLAC/ FRUTO/ TSH 11 88	-
XII.B	Thaumatococin	<i>Vitis riparia</i>	2E-59	445	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002

Tabela 4. Continuação

Clas	Anotação Funcional	Espécie	E value	Size (bp)	Biblioteca	Referências
XII.B	Thaumatococcus-like protein	<i>Sambucus nigra</i>	2E-43	427	CEPLAC/ FRUTO/ TSH 11 88	CHEONG et al., 2002
XII.B	Translationally controlled tumor protein homolog (TCTP)	<i>Glycine max</i>	7E-49	467	CEPLAC/ FRUTO/ TSH 11 88	LOPEZ; FRANCO, 2006
XII.B	Translationally controlled tumor protein homolog (TCTP)	<i>Medicago sativa</i>	9E-13	430	CEPLAC/ FRUTO/ TSH 11 88	LOPEZ; FRANCO, 2006
XII.B	Universal stress protein (USP) family protein	<i>Arabidopsis thaliana</i>	1E-64	544	CEPLAC/ FRUTO/ TSH 11 88	-
XII.B	Wound- induced basic protein	<i>Phaseolus vulgaris</i>	7E-18	466	CEPLAC/ FRUTO/ TSH 11 88	JONES et al., 2002
XII.C	PRLI-interacting factor, putative	<i>Arabidopsis thaliana</i>	5E-13	498	CEPLAC/ FRUTO/ TSH 11 88	LINTHORST, 1991

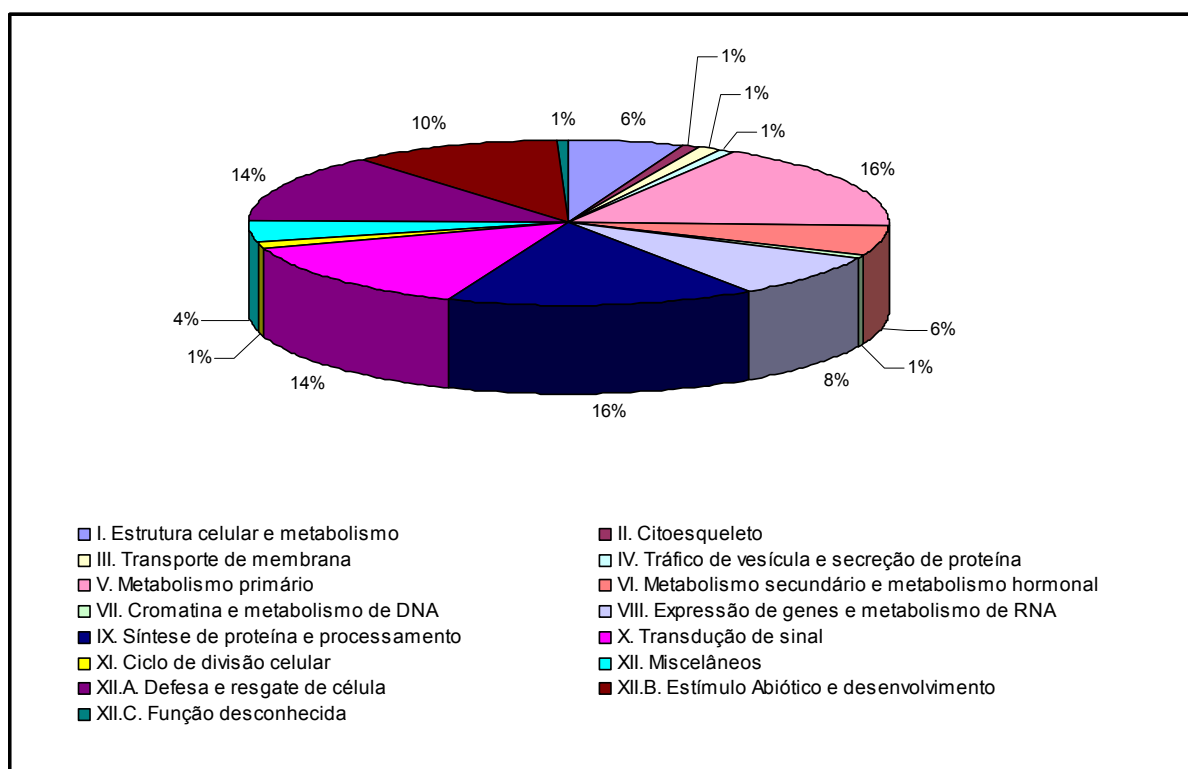


Figura 7. Percentual de ESTs selecionados por classe funcional. A classificação dos ESTs relacionados à resistência foi efetuada de acordo com Journet et al. (2002) e Gesteira et al. (2007).

4.2. Identificação de SNPs em ESTs

Os 153 genes que apresentaram similaridade significativa (item 4.1) foram considerados como seqüências de referência, a partir das quais foi identificada homologia em 54 genes através do alinhamento múltiplo de seqüências, como exemplificado na Figura 8.

A seqüência de referência (gene *ABA stress ripening protein* da biblioteca de fruto de TSH 1188) foi alinhada com os mesmos genes presentes nas bibliotecas supracitadas, tendo no exemplo uma seqüência do banco de dados de cacau (CF974722), uma seqüência de Catongo (UEBINTSP001025.F.fasta.seq.A01), uma seqüência de TSH 1188 de meristema (UEBINTRT001050.F.G10.esd) e uma seqüência da biblioteca subtrativa de Sca 6 (KZ0ACF1YC09FM1).

```

Query= ABA_stress_ripening_protein.ab1
CLUSTAL W (1.83) multiple sequence alignment

ABA_stress_ripening_protein.ab      GGGGATCGCATCCCAATCCACCTTAATTAGCATCACATTGCCATATTTCC
CF974722                             -----CTTAATTAGCATCACATTGCCATATTTCC
UEBINTSP001025.F.fasta.seq.A01      GGGGATCGCATCCCAATCCACCTTAATTAGCATCACATTGCCATATTTCC
UEBINTRT001050.F.G10.esd           GGGGATCGCATCCCAATCCACCTTAATTAGCATCACATTGCCATATTTCC
KZ0ACFLYC09FM1                      GGGGATCGCATCCCAATCCACCTTAATTAGCATCACATTGCCATATTTCC
                                         *****

ABA_stress_ripening_protein.ab      TTCTGCTCTCTTTTTTAAAGCATACTCAACCAACCACCACTCCAAAATGGC
CF974722                             TTCTGCTCTCTTTTTTAAAGCATACTCAACCAACCACCACTCCAAAATGGC
UEBINTSP001025.F.fasta.seq.A01      TTCTGCTCTCTTTTTTAAAGCATACTCAACCAACCACCACTCTAAAATGGC
UEBINTRT001050.F.G10.esd           TTCTGCTCTCTTTTTTAAAGCATACTCAACCAACCACCACTCCAAAATGGC
KZ0ACFLYC09FM1                      TTTTGC-----
                                         ** ***

ABA_stress_ripening_protein.ab      TGAGGAG AAGCAACACCACCACCTCTTCCACCACCACAAGGAGGATGAAA
CF974722                             TGAGGAG AAGCAACACCACCACCTCTTCCACCACCACAAGGAGGATGAAA
UEBINTSP001025.F.fasta.seq.A01      TGAGGAG AAGCAACACCACCACCTCTTCCACCACCACAAGGAGGATGAAA
UEBINTRT001050.F.G10.esd           TGAGGAG AAGCAACACCACCACCTCTTCCACCACCACAAGGAGGATGAAA
KZ0ACFLYC09FM1                      -----

ABA_stress_ripening_protein.ab      AGCCTGTT
CF974722                             AGCCTGTT
UEBINTSP001025.F.fasta.seq.A01      AGCCTGTT
UEBINTRT001050.F.G10.esd           AGCCTGTT
KZ0ACFLYC09FM1                      -----

```

Figura 8. Exemplo de resultado obtido pelo uso do programa CLUSTAL W. Refere-se a seqüência *ABA stress ripening protein* (TSH 1188/ Fruto), alinhada com o mesmo gene presente em diferentes bibliotecas: **CF974722** (banco de dados de cacau); **UEBINTSP001025** (Catongo/meristema); **UEBINTRT001050** (TSH 1188/meristema); **KZ0ACFLYC09FM1** (Sca 6/ meristema). Asterisco (*) refere-se a nucleotídeos iguais quando todas seqüências são completas e traço (-) refere-se a ausência de nucleotídeo na seqüência.

Para identificação dos SNPs o resultado do CLUSTAL W dos 54 genes foi processado pelo programa CLOURE-D. O programa CLOURE-D permitiu identificar apenas os nucleotídeos diferentes em relação a seqüência de referência. A Figura 9 mostra o resultado do processamento feito pelo CLOURE-D a partir do alinhamento múltiplo feito pelo CLUSTAL W tomando-se o gene *ABA stress ripening protein*, como referência. O programa fez o processamento desses alinhamentos e colocou um ponto nos nucleotídeos iguais a seqüência de referência e um traço onde ocorreu a ausência do nucleotídeo. No caso da seqüência *ABA stress ripening protein* utilizada como exemplo na figura 9, foram identificados cinco possíveis SNPs, sendo dois na seqüência do banco de dados de cacau do tipo transição (T/C), dois na seqüência de Sca 6 do tipo transição (C/T e T/C), um na seqüência de Catongo do tipo transição (C/T) e um na seqüência de TSH 1188 do tipo transição (T/C).



Figura 9. Exemplo de resultado obtido após utilização do programa CLOURE-D. Refere-se a seqüência ABA stress ripening protein (TSH 1188/fruto) alinhada com o mesmo gene presente em diferentes bibliotecas: CF974722 (banco de dados de cacau); UEBINTSP001025 (Catongo/meristema); UEBINTRT001050 (TSH 1188/meristema); KZ0ACF1YC09FM1 (Sca 6/ meristema). São visualizados em azul cinco possíveis SNPs do tipo transição acontecendo nas seqüências CF974722, UEBINTSP001025, KZ0ACF1YC09FM1 e UEBINTRT001050. O traço (-) refere-se a ausência do nucleotídeo e ponto (.) refere-se à nucleotídeos iguais a seqüência de referência.

Dentre os 54 genes que apresentaram homologia, 31 apresentaram possíveis SNPs. Foi identificado um total de 133 possíveis SNPs, dois quais 85 correspondem a substituições entre genótipos e ocorreram nos genes superoxide dismutase, *calcium binding protein*, *heat shock protein*, entre outros (Tabela 5).

Para as seqüências que apresentaram possíveis SNPs entre genótipos, 10 tiveram mais de um possível SNP. O gene *Ras-related GTP-binding family protein* teve 22 SNPs em sua seqüência, o gene superoxide dismutase teve cinco SNPs, o gene *calmodulin* teve quatro SNPs e o gene *calcium binding protein* teve três SNPs e todos estes genes estão envolvidos em transdução de sinais. No gene *ubiquitin conjugating protein* foram identificados 18 SNPs, no gene *protease inhibitor* foram identificados quatro SNPs, no gene *pentameric polyubiquitin* foram identificados três SNPs, no gene *26 S proteasome* foram identificados dois SNPs e no gene *heat shock* foram identificados dois SNPs, todos esses genes estão envolvidos em síntese de proteínas. Por fim, em *S-adenosyl-L-methionine synthetase* foram identificados cinco SNPs e este gene está envolvido em metabolismo primário. As

demais seqüências de genes tiveram apenas um SNP identificado em suas seqüências (Tabela 5).

Tabela 5. Número de SNPs candidatos identificados através do programa CLOURE-D a partir dos 153 genes selecionados. Os genes que possuem 0 são os que apresentaram 100% de homologia não apresentando portanto nenhuma substituição

Anotação funcional	Espécie	Tamanho (bp)	N° SNPs total	N° SNPs entre genótipos
26S proteasome non-ATPase regulatory subunit	<i>Arabidopsis thaliana</i>	445	2	2
26S proteasome regulatory particle non-ATPase subunit8	<i>Oryza sativa</i>	486	0	0
ABA stress ripening protein - common ice plant	<i>Mesembryanthemum crystallinum</i>	1186	1	1
ABC transporter	<i>Arabidopsis thaliana</i>	420	1	1
Acinus protein	<i>Oryza sativa</i>	442	0	0
Ankyrin repeat family protein	<i>Arabidopsis thaliana</i>	406	1	1
AP2 domain containing protein	<i>Prunus armeniaca</i>	534	0	0
Apoptosis inhibitor	<i>Arabidopsis thaliana</i>	275	0	0
Aspartic proteinase	<i>Theobroma cacao</i>	401	1	1
Auxin-binding protein ABP19a precursor	<i>Prunus persica</i>	461	0	0
C2 Domain-containing protein	<i>Arabidopsis thaliana</i>	331	0	0
Calcium-binding EF hand family protein	<i>Arabidopsis thaliana</i>	297	6	3
Calmodulin	<i>Pyrus communis</i>	489	5	4
Calmodulin	<i>Brassica napus</i>	333	0	0
Cathepsin B-like cysteine proteinase (EC 3.4.22.-)	<i>Triticum aestivum</i>	492	1	1
Chalcone synthase	<i>Rubus idaeus</i>	439	0	0
Cinnamoyl CoA reductase	<i>Fragaria x ananassa</i>	324	0	0
Cytochrome c oxidase subunit 6b, putative	<i>Arabidopsis thaliana</i>	416	1	1
Disease resistance-responsive protein	<i>Arabidopsis thaliana</i>	509	0	0
Ferredoxin family protein	<i>Arabidopsis thaliana</i>	486	1	1
Flavonol synthase	<i>Ginkgo biloba</i>	485	1	1
Glutathione S-transferase	<i>Gossypium hirsutum</i>	453	0	0
Heat-shock protein 80	<i>Euphorbia esula</i>	587	4	2
Leucine-rich repeat transmembrane protein kinase	<i>Arabidopsis thaliana</i>	406	3	0
Leucoanthocyanidin dioxygenase	<i>Vitis labrusca x Vitis vinifera</i>	455	1	1
Methionyl aminopeptidase, putative	<i>Arabidopsis thaliana</i>	449	0	0
Monodehydroascorbate reductase (NADH)	<i>Arabidopsis thaliana</i>	453	1	0
Non-cell-autonomous heat shock cognate protein 70	<i>Cucurbita maxima</i>	379	6	2
Nucleoside diphosphate kinase B	<i>Arabidopsis thaliana</i>	469	0	0
Pathogenesis-related protein	<i>Pyrus pyrifolia</i>	480	0	0
Pathogenesis-related protein 4B	<i>Nicotiana tabacum</i>	479	5	0
Pentameric polyubiquitin	<i>Arabidopsis thaliana</i>	1149	3	3
Peroxidase putative	<i>Arabidopsis thaliana</i>	244	1	1
Protease inhibitor	<i>Vitis vinifera</i>	496	4	4
Protease inhibitor/seed storage	<i>Arabidopsis thaliana</i>	451	7	0
Pseudo-hevein	<i>Hevea brasiliensis</i>	498	15	0
Putative cinnamoyl-CoA reductase	<i>Oryza sativa</i>	489	0	0
Putative CuZn-superoxide dismutase	<i>Populus tremula x Populus tremuloides</i>	186	5	1
Putative pathogenesis related protein 1 precursor	<i>Vitis vinifera</i>	410	2	1
Putative transformer serine/arginine-rich ribonucleoprotein	<i>Oryza sativa</i>	378	0	0
Putative zinc finger protein ZmZf	<i>Zea mays</i>	317	0	0
Ras-related GTP-binding family protein	<i>Arabidopsis thaliana</i>	476	22	22

Tabela 5. Continuação

Anotação funcional	Espécie	Tamanho (bp)	N° SNPs total	N° SNPs entre genótipos
Resistance protein	<i>Lactuca sativa</i>	488	0	0
S-adenosyl-L-methionine synthetase	<i>Elaeagnus umbellata</i>	435	4	4
Superoxide dismutase	<i>Fagus sylvatica</i>	417	7	5
Thaumatococin	<i>Vitis riparia</i>	445	0	0
Thaumatococin-like protein	<i>Sambucus nigra</i>	427	0	0
Translationally controlled tumor protein homolog (TCTP)	<i>Glycine max</i>	467	0	0
Translationally controlled tumor protein homolog (TCTP)	<i>Medicago sativa</i>	430	2	2
Type I proteinase inhibitor-like protein	<i>Citrus x paradisi</i>	251	0	0
Ubiquitin conjugating protein	<i>Avicennia marina</i>	769	18	18
Universal stress protein (USP) family protein	<i>Arabidopsis thaliana</i>	544	1	1
Zinc finger (C2H2 type) family protein	<i>Arabidopsis thaliana</i>	507	0	0
Zinc finger (C3HC4-type RING finger) family protein	<i>Arabidopsis thaliana</i>	459	1	1

Os 85 SNPs identificados entre genótipos foram mantidos para as demais análises por serem considerados “possíveis SNPs verdadeiros”. As 48 demais substituições não foram adotadas como “SNPs verdadeiros” por se tratarem do mesmo genótipo, sendo possivelmente algum artefato de seqüenciamento ou diferentes alelos do mesmo gene, o que pode ser também um resultado interessante, porém, optou-se por se utilizar somente os 85 genótipos com a finalidade de minimizar ao máximo as possibilidades de falsos resultados.

A descrição dos possíveis SNPs identificados entre genótipos (posição na seqüência, tipo, seqüência onde foi encontrado) é apresentada na Tabela 6. Por exemplo, na sequencia correspondente ao gene *calmodulin*, quando comparado com o mesmo gene em diferentes genótipos, apresentou uma substituição na posição 59 do alinhamento múltiplo de seqüências do tipo transição de uma citosina para uma timina, e dentre as seqüências usadas no alinhamento a substituição ocorreu na seqüência CA_794461 do banco de dados de cacau. Outro exemplo foi a seqüência *protease inhibitor* que apresentou uma substituição do tipo transversão de uma timina para uma adenina na posição 97 e na seqüência de Catongo. Dos possíveis SNPs identificados, 2 ocorreram em seqüências da biblioteca subtrativa de Sca 6, 23 ocorreram em seqüências do banco de dados de cacau, 17 em seqüências de Catongo e 43 aconteceram em seqüências de TSH 1188 (Tabela 7).

Tabela 6. Descrição dos SNPs identificados entre genótipos. * Refere-se aos SNPs usados para a construção dos *primers* AS-PCR. ** Refere-se se a posição a partir da extremidade 5' da sequência. *** Refere-se a sequência em que ocorreu a substituição.

Gene	Posição**	SNP	Gene SNP***
* Calcium binding	59	C/T	Calcium binding
* Heat shock proten 80	74	G/A	heat shock proten 80
* Heat shock proten 80	85	G/T	heat shock proten 80
* Peroxidase putative	70	C/A	Peroxidase putative
* Protease inhibithor	98	C/A	protease inhibithor.ab1
* Protease inhibithor	97	T/A	UEBINTSP001012.F.fasta.seq.D07
* Putative CuZn superoxide dismutase	99	A/G	Superoxide dismutase
* Superoxide dismutase	60	G/C	Superoxide dismutase
* Superoxide dismutase	91	G/T	Superoxide dismutase
26S proteasome regulatory	33	G/A	26S proteasome regulatory.ab1
26S proteasome regulatory	140	T/C	26S proteasome regulatory.ab1
ABA stress ripening protein	53	C/T	KZ0ACF1YC09FM1
ABC transporter	71	G/A	UEBINTRT001027.F.H01.esd
Ankyrin repeat family protein	20	T/C	UEBINTSP001002.F.fasta.seq.H02
Aspartic proteinase	31	T/C	CA797632
Calcium binding	92	G/A	CEPLAC 40.seq.G09 013 G09.ab1
Calcium binding	139	T/A	UEBINTRT001027.F.G05.esd
Calmodulin	59	C/T	CA794461
Calmodulin	16	G/A	CA794461
Calmodulin	66	G/A	CA794461
Calmodulin	61	T/C	CA794461
Cathepsin B-like cysteine proteinase	99	G/A	CF972716
Cytocrome c oxidase subuint 6b	114	T/G	CF972836
Ferredoxin	118	T/C	KZ0ACD1YJ05FM1
Flavonol synthase subunit	14	A/G	flavonol synthase .ab1
Leucoanthocyanidin dioxgenase	75	C/A	leucoanthocyanidin.ab1
Non cell autonomous	69	A/G	UEBINTSP001014.F.fasta.seq.G05
Non cell autonomous	48	G/A	UEBINTSP001016.F.fasta.seq.H03
Pentameric polyubiquitin	126	C/G	UEBINTRT001025.F.H03.esd
Pentameric polyubiquitin	121	G/A	CF973009
Pentameric polyubiquitin	125	T/A	UEBINTRT001025.F.H03.esd
Protease inhibithor	84	T/C	UEBINTSP001012.F.fasta.seq.D07
Protease inhibithor	120	T/G	UEBINTRT001007.F.F02.esd
Putative pathogenesis related protein 1 precursor	177	T/A	UEBINTSP001051.F.fasta.seq.D03
Ras related GTP binding	68	C/A	CA798591
Ras related GTP binding	118	C/A	CA798591
Ras related GTP binding	69	G/A	CA798591
Ras related GTP binding	80	G/A	CA798591
Ras related GTP binding	133	G/A	CA798591
Ras related GTP binding	52	G/C	Ras related GTP binding .ab1
Ras related GTP binding	145	T/A	Ras related GTP binding .ab1
Ras related GTP binding	148	T/A	Ras related GTP binding .ab1
Ras related GTP binding	49	T/C	CA798591
Ras related GTP binding	136	T/C	CA798591
Ras related GTP binding	43	T/C	Ras related GTP binding .ab1

Tabela 6. Continuação

Gene	Posição**	SNP	Gene SNP***
Ras related GTP binding	46	T/C	Ras related GTP binding.ab1
Ras related GTP binding	55	T/C	Ras related GTP binding.ab1
Ras related GTP binding	61	T/C	Ras related GTP binding.ab1
Ras related GTP binding	73	T/C	Ras related GTP binding.ab1
Ras related GTP binding	103	T/C	Ras related GTP binding.ab1
Ras related GTP binding	154	T/C	Ras related GTP binding.ab1
Ras related GTP binding	157	T/C	Ras related GTP binding.ab1
Ras related GTP binding	160	T/C	Ras related GTP binding.ab1
Ras related GTP binding	70	T/G	CA798591
Ras related GTP binding	124	T/G	CA798591
Ras related GTP binding	139	T/G	CA798591
S-adenosyl-L-methionine synthetase	148	A/T	CF974275
S-adenosyl-L-methionine synthetase	134	C/A	CF974275
S-adenosyl-L-methionine synthetase	154	C/A	CF974275
S-adenosyl-L-methionine synthetase	140	T/C	CF974275
Superoxide dismutase	113	G/T	B06 004 B06.ab1
Superoxide dismutase	106	G/T	CEPLAC 14.seq.B06 004 B06.ab1
Superoxide dismutase	108	T/C	Superoxide dismutase
Translationally controlled1	181	C/T	UEBINTRT001016.F.F06.esd
Translationally controlled1.1	231	C/T	UEBINTRT001016.F.F06.esd
Ubiquitin conjugation protein	129	C/A	ubiquitin conjugation protein
Ubiquitin conjugation protein	56	G/A	ubiquitin conjugation protein
Ubiquitin conjugation protein	80	G/A	ubiquitin conjugation protein
Ubiquitin conjugation protein	83	G/A	ubiquitin conjugation protein
Ubiquitin conjugation protein	108	G/A	ubiquitin conjugation protein
Ubiquitin conjugation protein	77	G/C	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	89	T/A	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	95	T/C	ubiquitin conjugation protein
Ubiquitin conjugation protein	102	T/C	ubiquitin conjugation protein
Ubiquitin conjugation protein	120	T/C	ubiquitin conjugation protein
Ubiquitin conjugation protein	123	T/C	ubiquitin conjugation protein
Ubiquitin conjugation protein	65	T/C	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	68	T/C	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	75	T/C	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	90	T/C	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	91	T/C	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	94	T/C	UEBINTSP001047.F.fasta.seq.D07
Ubiquitin conjugation protein	111	T/G	ubiquitin conjugation protein
Universal stress protein	14	G/A	universal stress protein.ab1
Zinc Finger	44	G/A	Zinc Finger.ab1

Dos 85 SNPs candidatos, 54 correspondem a substituição do tipo transição e 31 correspondem a substituições do tipo transversão. Dos possíveis SNPs tipo transição, 27 ocorreram em seqüências de TSH 1188 alinhadas com diferentes genótipos, 12 em seqüências de Catongo, 13 em seqüências dos bancos de dados de cacau e apenas duas nas seqüências de Sca 6. Dos SNPs tipo transversão 16 ocorreram nas seqüências de TSH 1188, 10 nas seqüências dos bancos de dados de cacau, cinco nas seqüências de Catongo, e nas seqüências de Sca 6 não houve nenhuma substituição deste tipo (Tabela 7).

Tabela 7. Descrição e quantificação dos tipos de substituições identificados nos diferentes genótipos. A tabela mostra o número de SNPs do tipo transição e transversão por genótipo e o total de SNPs de cada tipo.

Biblioteca	Numero de SNPs	
	Transições	Transversões
Catongo	12	5
TSH 1188	27	16
Banco de dados	13	10
Sca 6	2	0
Total	54	31

Dentre os SNPs candidatos, 74% estão em seqüências de genes envolvidos em “defesa e resgate de célula”, “transdução de sinais” e “síntese de proteínas”. Trinta e três por cento correspondem a SNPs em seqüências de genes envolvidos em “síntese de proteínas e processamento”, 8% em “metabolismo primário” e “estímulo abiótico e desenvolvimento”; 4% em “metabolismo secundário e metabolismo hormonal”; e 4% estão envolvidos em “transporte de membrana” e “expressão de genes e metabolismo de RNA”. Não foi identificada nenhuma substituição nos genes das classes compreendidas em “estrutura celular e metabolismo”, “divisão celular”, “metabolismo de DNA e cromatina”, “miscelâneos”, “tráfego de vesícula”, “função desconhecida” e “não homólogos” (Figura 10).

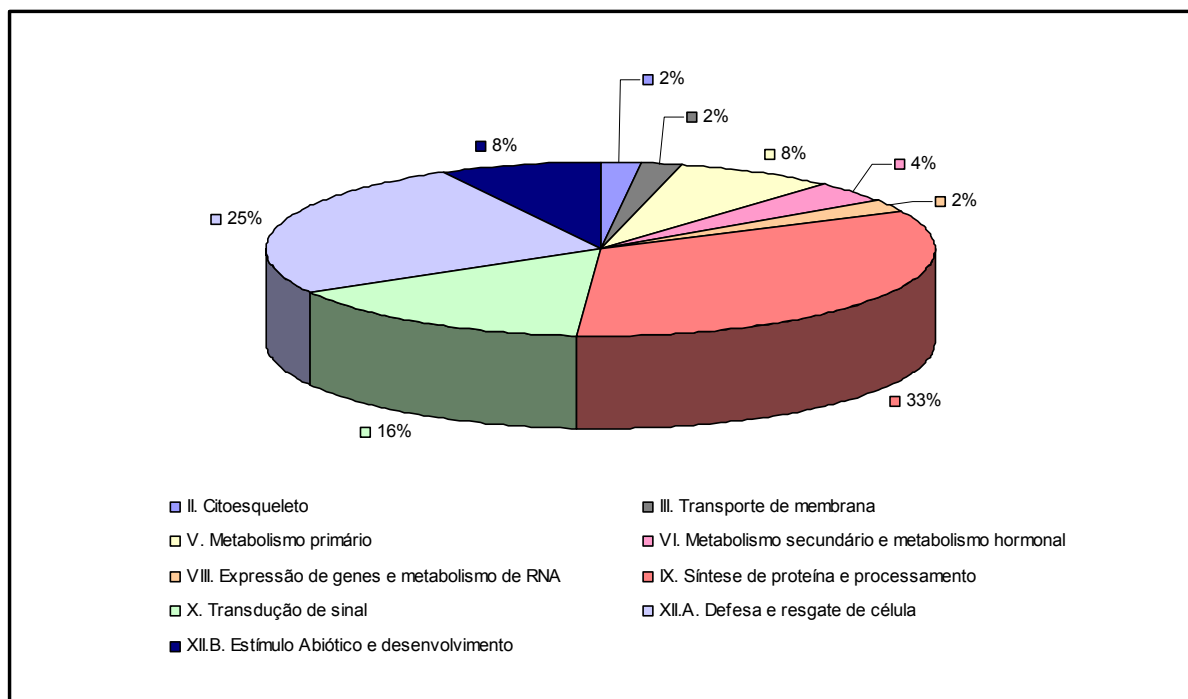


Figura 10. Porcentagem de SNPs identificados por função dos ESTs. As classes funcionais foram determinadas de acordo com Journet et al. (2002) e Gesteira et al., 2007

Dos SNPs identificados entre genótipos foram desenhados nove conjuntos de *primers* segundo a técnica AS-PCR (Tabela 8). Foram utilizados os DNAs do Catongo e TSH 1188 para a otimização da amplificação PCR. Nos experimentos de otimização de temperatura, os *primers* foram testados à temperatura de 56°C, gradiente e *touch-down* (TD 60-48), não obtendo amplificação do DNA genômico com nenhum dos *primers* controle e AS-PCR (dados não mostrados). Amplificações só foram obtidas quando: (i) os *primers* foram submetidos a condições de amplificação de maior estringência, ou seja, sofreram aumento em sua temperatura de anelamento (58°C, 60°C e 62°C) visando aumentar a especificidade do produto amplificado, (ii) utilizando a concentração do *primer forward* e do *primer AS-PCR* a 2 pmol, do *primer reverse* a 4 pmol, e (iii) mudando o tampão de KCl para (NH₄)₂SO₄. As temperaturas otimizadas estão apresentadas na Tabela 8.

Tabela 8. Seqüências dos *primers* desenhados para desenvolver o método de AS-PCR e respectivos genes, temperatura de desnaturação (T_m) e anelamento (T_A), e tamanhos dos produtos amplificados no controle e nos SNPs

Gene	Biblioteca	Primer forward			Primer reverse			T _A [°C]	Tamanho do produto de amplificação controle (pb)	Tamanho do produto de amplificação SNPs (pb)
		Nome*	Seqüência	T _m [°C]	Nome*	Seqüência	T _m [°C]			
Superoxide dismutase	TSH 1188/CEPLAC	Tc_SNP1Ff	CGGGGACTCTTTTCTTTC	65	Tc_SNP1R	ACAGCCATCATCACCAAC	63	58,1	337	303
		Tc_SNP1F	GGTGCTCTGAGATCACATAC	62						
Calcium binding	TSH 1188/CEPLAC	Tc_SNP5Ff	TTCACGACATGGGAGTG	62	Tc_SNP5R	TGTTGAGGTCAAGTGCTG	62	58,1	114	103
		Tc_SNP5F	GGAGTGGTGATCATCGAT	62						
Protease inhibitor	Catongo/UESC	Tc_SNP7Ff	TGGGAGCAAAAGGGAGG	69	Tc_SNP7R	CCCAAACGCGAACCCTA	70	58,1	125	113
		Tc_SNP7F	GGAGGAGGCAGCAGCAA	68						
Peroxidase putative	TSH 1188/CEPLAC	Tc_SNP8Ff	GGCCCTCGAGAAGCTTT	68	Tc_SNP8R	TTTACCTACTGGGCCCG	67	58,1	96	54
		Tc_SNP8F	CGGGCCCCGATAGGTAAA	66						
Putative CuZn superoxide dismutase	TSH 1188/CEPLAC	Tc_SNP9Ff	CATCGTTGGAGACAACA	59	Tc_SNP9R	CAGCCATTGGTGGTATC	61	58,1	147	83
		Tc_SNP9F	CATGTGAAAGGAAAGATG	59						
Heat-shock protein 80	TSH 1188/CEPLAC	Tc_SNP11Ff	CTTCAGCCTTGATGACC	60	Tc_SNP11R	GTCGACTTCCTCCATCT	58	58,1	135	109
		Tc_SNP11F	GAAACTTGGATTGAGCA	60						
Heat-shock protein 80	TSH 1188/CEPLAC	Tc_SNP12Ff	GCTTCAGCCTTGATGAC	61	Tc_SNP12R	GCAAGATGGAGGAAGTC	59	58,1	160	70
		Tc_SNP12F	TGAGCATTGATGAGGAT	60						
Protease inhibitor	TSH 1188/CEPLAC	Tc_SNP13Ff	TCCTGGGAGCAAAAGCG	70	Tc_SNP13R	GCACAAGAAATCCGGGG	69	58,1	110	94
		Tc_SNP13F	GGAGGAGGCAGCAGCAA	70						
Superoxide dismutase	TSH 1188/CEPLAC	Tc_SNP14F	TTCTTTCTTTTGTACTGC	55	Tc_SNP14R	CACATTCACTGCTAATAA	53	58,1	143	87
					Tc_SNP14RR	CTCAGTAGTTGGACCATC	55			

* Os *primers* F são os *primers* AS-PCR para amplificação do SNP. Os *primers* Ff e R são os *primers* flanqueadores

Os resultados da otimização da técnica quanto à concentração de magnésio foram obtidos após as otimizações acima descritas. A concentração ótima de MgCl_2 foi obtida nas reações contendo $4,3 \text{ mmol. L}^{-1}$ de MgCl_2 (Figura 11), podendo visualizar bandas mais nítidas.

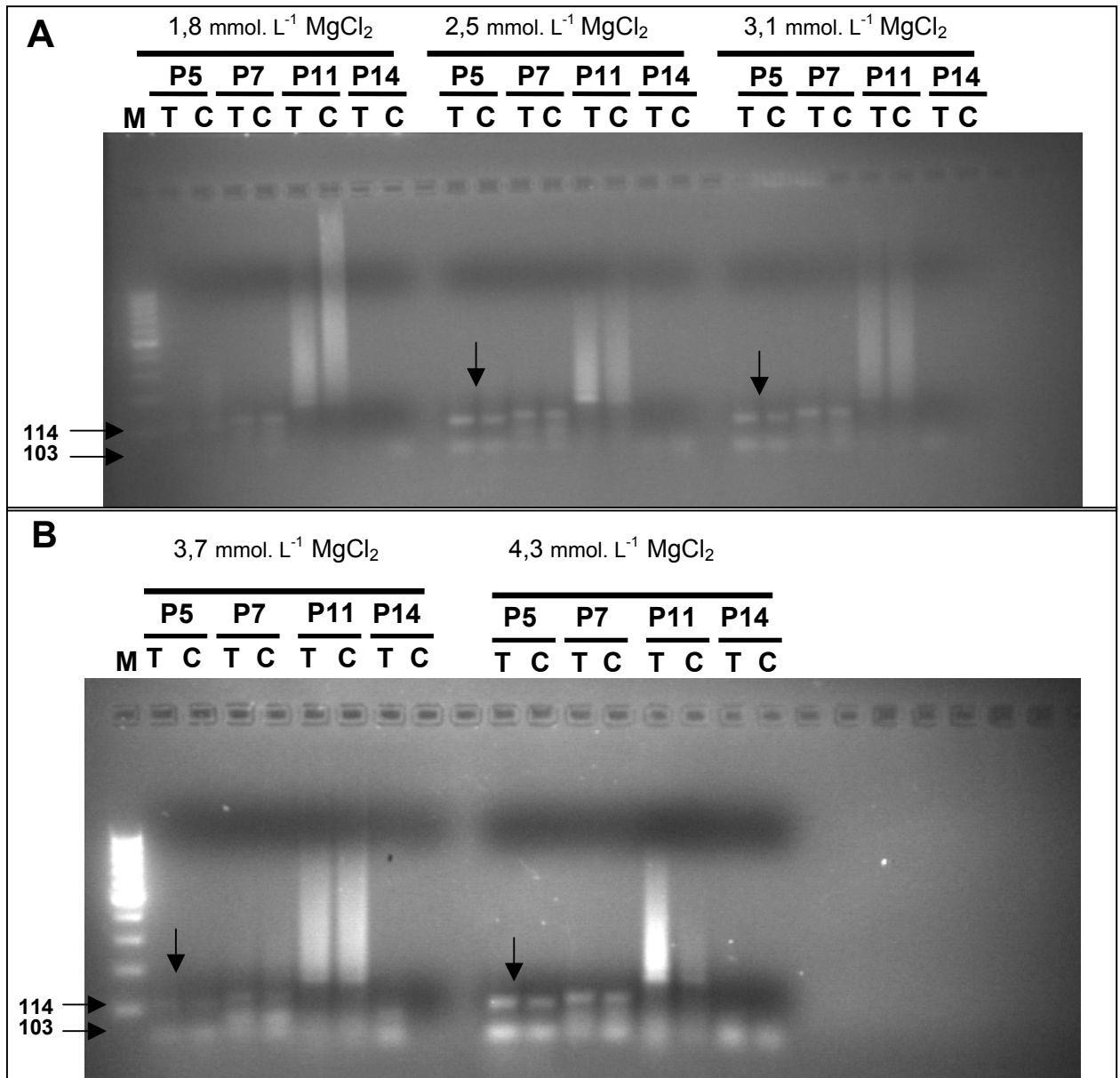


Figura 11. Otimização do AS-PCR quanto a concentração de MgCl_2 em gel de agarose 2%, com anelamento a $58,1^\circ\text{C}$, utilizando DNA de TSH 1188 (T) e Catongo (C) e os *triplets* 5, 7, 11, 14. (A) Concentrações à $1,8 \text{ mmol. L}^{-1}$, $2,5 \text{ mmol. L}^{-1}$ e $3,1 \text{ mmol. L}^{-1}$ de MgCl_2 ; (B) Concentrações à $3,7 \text{ mmol. L}^{-1}$ e $4,3 \text{ mmol. L}^{-1}$ de MgCl_2 . A seta mostra a consistência dos resultados com o *triplet* 5 amplificando o controle e o *primer* AS-PCR para os DNAs do TSH 1188 e do Catongo. M = marcador de peso molecular (100 pb, Invitrogen).

Os fragmentos gerados pelas ampliações foram analisados por meio da eletroforese em gel de agarose a 1,2%, 2%, e 3% corado em solução contendo

brometo de etídio (0,5 mg/mL) e visualizados sob luz UV, e em gel de poliacrilamida 6% não-desnaturante. Melhores resultados foram obtidos em gel de agarose 2%, garantindo repetibilidade dos fragmentos amplificados com o *triplet* de *primers* 5; 114 pb para os *primers* controle e 103 pb para o *primer* AS-PCR. Por outro lado, apenas fragmentos referentes ao controle foram amplificados com os *primers* 9 e 11, respectivamente 147 pb e 135 pb (Figura 12). Os fragmentos amplificaram no tamanho esperado (Tabela 8). Foi possível, em gel de poliacrilamida corado com prata, a visualização dos produtos de amplificação controle e AS-PCR para o *primer* 5 e do produto de amplificação de 143 pb referente aos *primers* controle do *primer* 14, porém os resultados em gel de poliacrilamida mostraram-se sem repetibilidade, gerando muitos fragmentos inespecíficos (Figura 13).

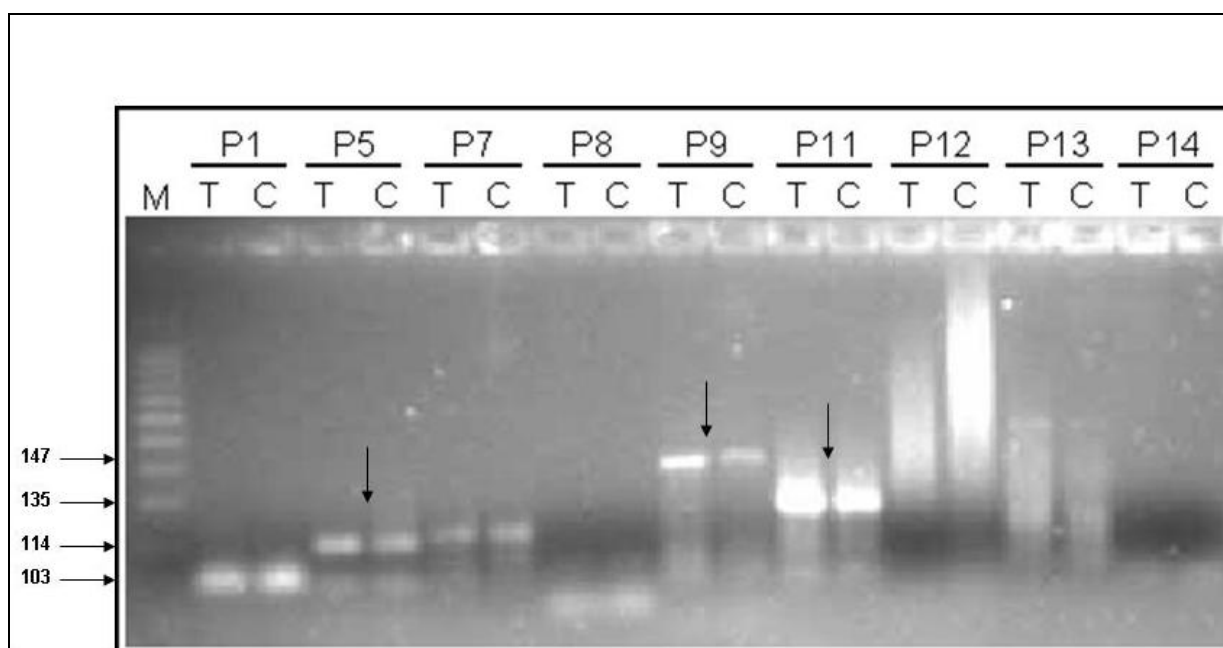


Figura 12. AS-PCR em gel de agarose 2%, utilizando anelamento a 58,1°C, dois genótipos para cada conjunto de *primers* (T = TSH 1188 e C= Catongo) e 9 *triplets* de *primers*. As setas mostram a visualização do produto de amplificação do controle e do *primer* AS-PCR para o *primer* 5 e o produto de amplificação do controle para os *primers* 9 e 11. M = marcador de peso molecular (100 pb, Invitrogen).

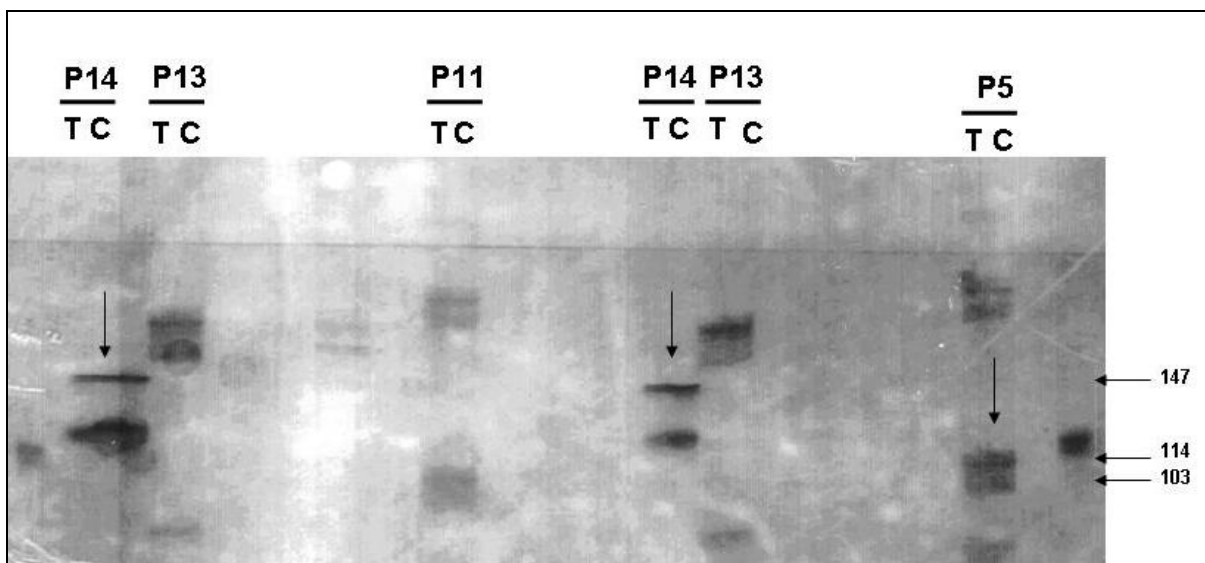


Figura 13. AS-PCR em gel de poliacrilamida 6%, utilizando 58,1°C como temperatura de anelamento. Foram utilizados 2 genótipos para cada conjunto de *primers* (T = TSH 1188 e C= Catongo). As setas mostram a visualização do produto de amplificação do controle e do *primer* AS-PCR para o *primer* 5 e o produto de amplificação do controle para o *primers* 14. M = marcador de peso molecular (100 pb, Invitrogen)

Apesar dos resultados obtidos descritos acima com os *triplets* 5, 9, 11 e o 14, com relação aos experimentos de validação dos SNPs não foi possível a reprodutibilidade esperada com a técnica AS-PCR. Segundo a metodologia adotada os *primers* deveriam amplificar os dois fragmentos (controle e AS-PCR) no genótipo do TSH 1188 e amplificar só o fragmento dos *primers* controle no genótipo do Catongo, para SNPs desenhados a partir do TSH 1188 e vice-versa. Os fragmentos amplificados foram do tamanho esperado para a detecção dos SNP, 114 pb (controle) e 103 pb (AS-PCR), mas a amplificação dos dois fragmentos aconteceu nos dois genótipos, TSH 1188 e Catongo, o que não era esperado. Além disso os fragmentos obtidos com os *primers* 9, 11 e 14 não amplificaram o *primer* AS-PCR. Por isso a validação desses SNPs por AS-PCR não foi satisfatória.

4.3. Identificação de SSR em ESTs

A Identificação de SSR foi feita nos 561 genes das bibliotecas de interação cacau-*M. pernicioso* de fruto de TSH 1188 da CEPLAC, e nos 10 genes das bibliotecas de meristema de TSH 1188 e Catongo da UESC. Um total de 125 SSR foi identificado. A Tabela 9 mostra os genes onde foram identificados os 125 SSR com o tamanho da seqüência, o número de SSR por seqüência e o *motif* de repetição. Os genes *ATGP3*, *bZIP transcription factor ATB2*, *caffeine synthase*, *constans-like protein 2*, *double-stranded DNA-binding family protein*, *drought responsive family protein*, *expressed protein*, *polyphenol oxidase*, *putative photosystem I reaction centre subunit IV*, *small nuclear ribonucleoprotein G*, *putative small nuclear ribonucleoprotein G* tiveram mais de um microssatélite identificados em suas seqüências, sendo que o gene *ATGP3* teve três microssatélites identificados em sua seqüência.

Dos 125 SSR, 63,7% correspondem a *motifs* com apenas quatro repetições, 6,5% a *motifs* com seis repetições e 14,5% a *motifs* com cinco repetições. Foi identificado um *motif* com 14 repetições, três *motifs* com nove repetições e quatro *motifs* com sete repetições. Dos microssatélites identificados, 58,4% correspondem aos *motifs* TC, GA, CT e AG, sendo que 28 correspondem ao *motif* TC, 16 correspondem ao *motif* GA, 15 correspondem ao *motif* CT e 14 correspondem ao *motif* GA (Figura 14).

Dos SSR identificados 23% foram identificados em genes correspondentes a função de “síntese de proteínas e processamento”, 14% correspondente a “expressão de genes e metabolismo de RNA”, 9% correspondente a “metabolismo primário”, 8% correspondente a “estímulo abiótico e desenvolvimento”, 5% correspondente a “transdução de sinais e regulação pos-transcricional”, 3% correspondente a “cromatina e metabolismo de DNA” e “tráfego de vesícula”, 2% correspondente a “citoesqueleto”, “miscelânea”, “defesa e resgate de célula” e “metabolismo secundário e metabolismo hormonal”, e 26% foram identificados em genes de função desconhecida. Não foi encontrado SSR em genes envolvidos em “transporte de membrana” e “ciclo de divisão celular” (Figura 15)

Tabela 9. Descrição de SSR identificados em 561 genes da interação cacau-*M. perniciosa*. * Refere-se aos SSR selecionados para análise de polimorfismo.

Anotação funcional	Espécie	Tamanho (pb)	Nº SSR	Motif
100 pct identical to gp:ECOTNPA 3	<i>Hypothetical protein tnpX - Escherichia coli</i>	567	1	AT(4)
20S proteasome alpha 6 subunit	<i>Nicotiana benthamiana</i>	498	1	GAA(4)
*26S proteasome regulatory subunit S5A	<i>Mesembryanthemum crystallinum</i>	438	1	CT(4)
2'-hydroxy isoflavone/dihydroflavonol reductase homolog	<i>Glycine max</i>	429	1	GT(4)
40s ribosomal protein s21	<i>Oryza sativa</i>	426	1	CT(4)
40s ribosomal protein s23	<i>Oryza sativa</i>	731	1	AGA(4)
40s ribosomal protein s23	<i>Oryza sativa</i>	365	1	AGA(4)
40s ribosomal protein s25	<i>Glycine max</i>	483	1	AGA(4)
40s ribosomal protein s25	<i>Glycine max</i>	529	1	GAA(4)
60S ribosomal protein L12	<i>Capsicum annuum</i>	428	1	TC(4)
60S ribosomal protein L12	<i>Capsicum annuum</i>	511	1	TC(4)
60S ribosomal protein L13A (RPL13aD)	<i>Arabidopsis thaliana</i>	517	1	AGC(4)
60S ribosomal protein L21	<i>Oryza sativa</i>	466	1	GA(5)
60S ribosomal protein L34 (RPL34A)	<i>Arabidopsis thaliana</i>	394	1	AG(4)
60S ribosomal protein L35	<i>Euphorbia esula</i>	525	1	AG(4)
60S ribosomal protein L35	<i>Euphorbia esula</i>	348	1	AG(4)
60s ribosomal protein l44	<i>Gossypium hirsutum</i>	509	1	CT(4)
ABA stress ripening protein	<i>Mesembryanthemum crystallinum</i>	1186	1	TA(4)
* ABC transporter	<i>Arabidopsis thaliana</i>	420	1	CT(4)
Antiquitin 1	<i>Homo sapiens</i>	461	1	TC(4)
ATGP3	<i>Arabidopsis thaliana</i>	552	3	TC(4) TC(4) TC (6)
* ATP-dependent clp protease ATP	<i>Nicotiana tabacum</i>	489	1	AG(4)
B12D protein	<i>Ipomoea batatas</i>	477	1	CT(4)
Bet v I allergen family protein	<i>Arabidopsis thaliana</i>	507	1	AG(4)
Beta-tubulin R2242	<i>Oryza sativa</i>	501	1	GAG(5)
* bzip transcription factor ATB2	<i>Glycine max</i>	803	2	CT(4)/ CT(4)
C-8,7 sterol isomerase	<i>Arabidopsis thaliana</i>	525	1	GA(4)
* Caffeine synthase	<i>Camellia sinensis</i>	415	3	GA(4)/ AT(7)/ CA(5)
* Calmodulin-binding family protein	<i>Arabidopsis thaliana</i>	507	1	CTG(4)
Calnexin 1 (CNX1)	<i>Arabidopsis thaliana</i>	464	1	GA(4)
Calreticulin precursor	<i>Ricinus communis</i>	377	1	TC(4)
Cold-regulated LTCOR12	<i>Lavatera thuringiaca</i>	432	1	CT(4)
Constans-like protein 2	<i>Malus x domestica</i>	271	2	AG(4) AG(4)
Copper chaperone (CCH)-related	<i>Arabidopsis thaliana</i>	427	1	AG(6)
*Copper chaperone (CCH)-related	<i>Arabidopsis thaliana</i>	494	1	GAAA(4)
Cytochrome b5, putative	<i>Arabidopsis thaliana</i>	305	1	AG(4)
* DNA polymerase-related	<i>Arabidopsis thaliana</i>	388	1	AT(9)
* Double-stranded DNA-binding family protein	<i>Arabidopsis thaliana</i>	449	2	CT(4)/ TGA (4)
* Drought responsive family protein	<i>Arabidopsis thaliana</i>	428	2	CT(4)/ TC(4)
Drought-responsive family protein	<i>Arabidopsis thaliana</i>	428	1	CT(4)
Eukaryotic translation initiation factor 5A isoform I	<i>Hevea brasiliensis</i>	745	1	GA(6)
Eukaryotic translation initiation factor 5A isoform I	<i>Hevea brasiliensis</i>	570	1	AG(4)
Expressed protein	<i>Arabidopsis thaliana</i>	539	1	GA(4)
Expressed protein	<i>Arabidopsis thaliana</i>	444	1	TC(4)
* Expressed protein	<i>Arabidopsis thaliana</i>	410	2	TA(5) AC(4)
Expressed protein	<i>Arabidopsis thaliana</i>	325	1	GA(4)
Expressed protein	<i>Arabidopsis thaliana</i>	296	1	TCC(5)
Expressed protein	<i>Arabidopsis thaliana</i>	459	2	TC(4) ATC(4)
* Expressed protein	<i>Arabidopsis thaliana</i>	373	1	CT(9)

Tabela 9. Continuação

Anotação funcional do gene	Espécie	Tamanho (pb)	Nº SSR	Motif
Expressed protein	<i>Arabidopsis thaliana</i>	276	1	TC(4)
Expressed protein	<i>Arabidopsis thaliana</i>	561	1	TC(4)
Expressed protein	<i>Arabidopsis thaliana</i>	394	1	AG(4)
Expressed protein	<i>Arabidopsis thaliana</i>	510	1	TC(4)
Expressed protein	<i>Arabidopsis thaliana</i>	277	1	GT(4)
Expressed protein	<i>Arabidopsis thaliana</i>	443	1	AG(4)
Expressed protein	<i>Arabidopsis thaliana</i>	408	1	AGG(4)
* Expressed protein	<i>Arabidopsis thaliana</i>	494	2	CTG(4) CTG(7)
Expressed protein	<i>Arabidopsis thaliana</i>	516	1	CT(5)
Glutamine synthetase	<i>Zea mays</i>	351	1	CT(4)
* Glutaredoxin family protein	<i>Arabidopsis thaliana</i>	427	1	TCT(4)
* GTP-binding nuclear protein RAN1	<i>Arabidopsis thaliana</i>	521	1	CT(4)
Histone H2B-3 - tomato	<i>Lycopersicon esculentum</i>	481	1	AGA(5)
Hypothetical protein	<i>Arabidopsis thaliana</i>	417	1	CT(5)
Hypothetical protein	<i>Cicer arietinum</i>	472	1	CG(4)
Hypothetical protein	<i>Oryza sativa</i>	797	1	GA(6)
Hypothetical protein	<i>Oryza sativa</i>	476	1	TCT(4)
KH domain-containing quaking protein	<i>Arabidopsis thaliana</i>	516	1	TC(4)
Lectin related	<i>Arabidopsis thaliana</i>	409	1	CT(4)
* Methionyl Aminopeptidase	<i>Arabidopsis thaliana</i>	449	1	AT(5)
Monodehydroascorbate reductase (NADH)	<i>Arabidopsis thaliana</i>	324	1	ATA(6)
Nascent polypeptide-associated complex (NAC)	<i>Arabidopsis thaliana</i>	520	1	GAT(4)
Peptidyl-prolyl cis-trans isomerase 1	<i>Malus x domestica</i>	498	1	ACC(4)
* Peroxiredoxin	<i>Populus tremula x Populus tremuloides</i>	575	1	TC(4)
* Polyphenol oxidase	<i>Catechol oxidase</i>	696	2	CT(4)/ CT(4)
PQ-loop repeat family protein	<i>Arabidopsis thaliana</i>	471	1	TC(6)
* Profilin 1	<i>Ricinus communis</i>	560	1	AGA(7)
* Proline rich protein	<i>Glycine max</i>	790	1	TGG(5)
Proline-rich protein	<i>Glycine max</i>	790	1	CTC(5)
Proline-rich protein	<i>Glycine max</i>	940	1	TA(4)
Protease inhibitor	<i>Vitis vinifera</i>	496	1	TC(4)
Protease inhibitor/ Seed storage	<i>Arabidopsis thaliana</i>	413	1	TC(4)
Protease inhibitor/seed storage/lipid transfer protein (LTP)	<i>Arabidopsis thaliana</i>	451	1	TC(4)
* Protein F2D10.18	<i>Arabidopsis thaliana</i>	786	1	CT(9)
Putative 60S ribosomal protein L31	<i>Oryza sativa</i>	455	1	TGA(4)
Putative pathogenesis related protein 1 precursor	<i>Vitis vinifera</i>	410	1	AT(5)
Putative photosystem I reaction centre subunit IV	<i>Populus euramericana</i>	514	2	AG(6) ACA(4)
* Putative transformer serine/arginine-rich ribonucleoprotein	<i>Oryza sativa</i>	378	1	AG(4)
Squamosa promoter binding protein-homolog 4	<i>Antirrhinum majus</i>	528	1	CT(5)
RAB8E	<i>Lotus corniculatus</i>	435	1	TG(4)
Ribosomal protein L33	<i>Castanea sativa</i>	476	1	CT(4)
Ribosomal protein L33	<i>Castanea sativa</i>	530	1	CT(4)
* Ring box protein related	<i>Arabidopsis thaliana</i>	352	1	AG(4)
Seed specific protein Bn15D1B	<i>Brassica napus</i>	466	1	CTG(6)
Small nuclear ribonucleoprotein G, putative	<i>Arabidopsis thaliana</i>	387	2	CT(4) CT(4)
Small nuclear ribonucleoprotein G, putative	<i>Arabidopsis thaliana</i>	537	2	CT(4) TC(5)
Synaptobrevin-like protein	<i>Pyrus pyrifolia</i>	471	1	TAT(4)
Transaldolase, putative	<i>Arabidopsis thaliana</i>	456	1	TC(4)
* Translationally controlled tumor protein homolog (TCTP)	<i>Hevea brasiliensis</i>	467	1	CT(4)

Tabela 9. Continuação

Anotação funcional do gene	Espécie	Tamanho (pb)	N° SSR	Motif
Translationally controlled tumor protein homolog (TCTP)	<i>Glycine max</i>	570	1	AG(4)
* Translationally controlled tumor protein homolog (TCTP)	<i>Medicago sativa</i>	430	1	AT(4)
* Universal stress protein (USP) family protein	<i>Arabidopsis thaliana</i>	544	1	AG(14)
Unknown	<i>Arabidopsis thaliana</i>	452	1	TCT(5)
Unknown protein	<i>Arabidopsis thaliana</i>	371	1	AG(4)
Unknown protein	<i>Arabidopsis thaliana</i>	347	1	AG(4)
Unknown protein	<i>Arabidopsis thaliana</i>	290	1	AT(4)
Xyloglucan endotransglycosylase	<i>Gossypium hirsutum</i>	361	1	TC(4)
* zinc finger (AN1-like) family protein	<i>Arabidopsis thaliana</i>	325	1	TC(7)
* Zinc Finger (C3HC4 type ring finger)	<i>Arabidopsis thaliana</i>	459	1	TC(4)
* Zinc finger (C3HC4)	<i>Arabidopsis thaliana</i>	459	1	CAT(5)
* Rac GDP dissociation Inhibitor	<i>Oryza sativa</i>	323	1	TC(4)

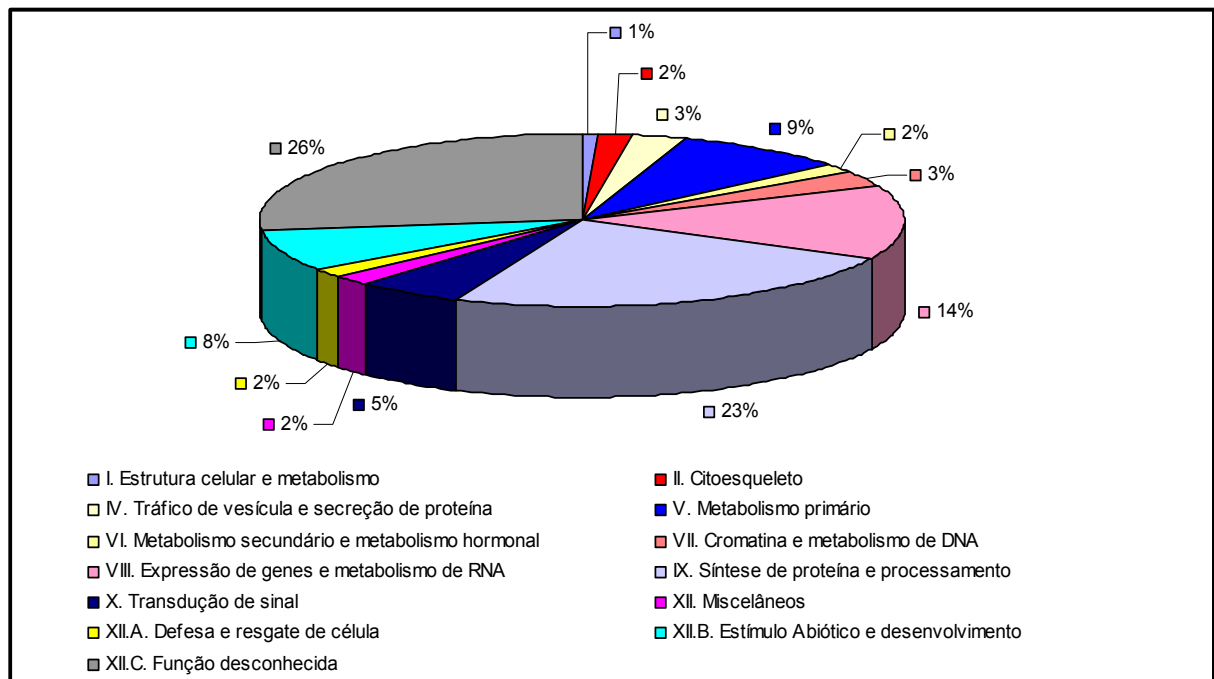


Figura 15 - Porcentagem de SSRs identificados por função dos ESTs. As classes funcionais foram determinadas de acordo com Journet et al. (2002) e Gesteira et al. (2007)

A partir dos 125 SSR identificados foram construídos 32 pares de *primers* flanqueando os microssatélites para análise de polimorfismo. Para 47% destes, foi possível obter a temperatura de anelamento, sendo que 7 pares de *primers* tiveram sua temperatura otimizada utilizando o programa TD 60-48, 5 deles tiveram a temperatura de anelamento em 56°C e 62°C e 59,5°C e dois pares de *primers* anelaram a 56,8°C (Tabela 10). Inicialmente foi verificada a temperatura dos *primers* em gel de agarose a 3%, para uma posterior análise de polimorfismo em gel de poli-acrilamida 6% corado com prata. Dos 15 *primers* testados quanto ao polimorfismo em 20 clones de cacauero descritos na Tabela 3, no Catongo e nos *bulks* resistentes e susceptíveis, 26,7% foram polimórficos (EST-SSR-26, EST-SSR-27, EST-SSR-28 e EST-SSR-29) e 73,3% foram monomórficos (EST-SSR-1, EST-SSR-2, EST-SSR-6, EST-SSR-7, EST-SSR-9, EST-SSR-10, EST-SSR-12, EST-SSR-13, EST-SSR-20, EST-SSR-24, EST-SSR-25, EST-SSR-31). Um total de 27 alelos foi identificado, sendo 16 alelos referentes aos locos polimórficos (Figura 16). Os *primers* 26 e 29 permitiram identificar ambos 3 alelos, o *primer* 27 indentificou 4 alelos e o *primer* 28 apresentou-se como o mais polimórfico indentificando 6 alelos (Figura 16 e 17). Foram encontrados indivíduos heterozigotos para os locos 27 e 28 (Figura 16). Os alelos foram designados pelo tamanho do(s) fragmento(s) em pares de bases, identificados em cada locus (Tabela 11). A média de diversidade genética

total (H_T) para todos os clones foi de 0,4694 e a média da diversidade genética esperada (H_e) entre os locos polimórficos foi de 0,4462 (Tabela 11).

A frequência alélica em todos os locos variou de 0,0313 a 0,913, sendo que as frequências menores foram obtidas com o *primer* 28 para três alelos e a frequência maior obtida com o *primer* 29 em um alelo (Tabela 12).

Tabela 10. Sequências de *primers* SSR obtidos a partir dos genes de interação cacau-*M. perniciosa*. TD (*touch-down*), T_m (temperatura de desnaturação) e T_A (temperatura de anelamento).

Gene	Primers forward e reverse*	T _m (°C)	T _A (°C)	Tamanho do produto de amplificação (pb)
ABC Transporter	EST-SSR 1-F	55,75	56,0	160
	EST-SSR1-R	55,87		
BZIP Transcription factor ATB2	EST-SSR-2-F	56,08	TD 60-48	211
	EST-SSR-2-R	55,91		
Caffeine synthase	EST-SSR-3-F	56,61		209
	EST-SSR-3-R	55,89		
Cooper chaperone	EST-SSR-4-F	55,94		161
	EST-SSR-4-R	55,50		
Dissociation inhibitor	EST-SSR-5-F	56,12		209
	EST-SSR-5-R	55,98		
Double stranded dna binding	EST-SSR-6-F	56,04	TD 60-48	194
	EST-SSR-6-R	56,00		
Glutaredoxin family protein	EST-SSR-7-F	56,92	56,0	155
	EST-SSR-7-R	56,36		
Zinc finger	EST-SSR-8-F	56,61		157
	EST-SSR-8-R	55,83		
Polyphenol oxidase	EST-SSR-9-F	55,90	TD 60-48	259
	EST-SSR-9-R	55,59		
Putative transformer	EST-SSR-10-F	55,89	TD 60-48	179
	EST-SSR-10-R	55,84		
Translationally controled tumor protein	EST-SSR-11-F	56,61		238
	EST-SSR-11-R	55,70		
Zinc finger	EST-SSR-12-F	56,59	TD 60-48	152
	EST-SSR-12-R	56,09		
ATP dependent clp protease	EST-SSR-13-F	55,45	TD 60-48	196
	EST-SSR-13-R	56,04		
Caffeine synthase	EST-SSR-14-F	55,97		245
	EST-SSR-14-R	55,97		
Calmodulin binding family protein	EST-SSR-15-F	56,24		235
	EST-SSR-15-R	56,08		
Copper chaperone	EST-SSR-16-F	55,40		178
	EST-SSR-16-R	55,77		
Double stranded dna binding	EST-SSR-17-F	55,59		177
	EST-SSR-17-R	55,68		
Drought reponsive family protein	EST-SSR-18-F	56,61		233
	EST-SSR-18-R	56,14		
GTP Binding nuclear protein	EST-SSR-19-F	56,05		233
	EST-SSR-19-R	56,03		
Peroxiredoxin	EST-SSR-20-F	56,36	56,0	171
	EST-SSR-20-R	56,07		
Proline rich protein	EST-SSR-21-F	55,82		160
	EST-SSR-21-R	55,08		
Ring box protein related	EST-SSR-22-F	55,77		163
	EST-SSR-22-R	56,04		
Universal stress protein	EST-SSR-23-F	52,04		249
	EST-SSR-23-R	52,68		
Proteasome	EST-SSR-24-F	55,81	TD 60-48	178
	EST-SSR-24-R	55,97		

Tabela 10. Continuação

Gene	Primers forward e reverse*	T _m (°C)	T _A (°C)	Tamanho do produto de amplificação (pb)
Profilin 1	EST-SSR-25-F	71,0	62,0	173
	EST-SSR-25-R	70,0		
DNA polymerase-related	EST-SSR-26-F	63,0	56,8	192
	EST-SSR-26-R	65,0		
Expressed protein	EST-SSR-27-F	68,0	62,0	92
	EST-SSR-27-R	69,0		
Protein F2D10.18	EST-SSR-28-F	64,0	56,8	104
	EST-SSR-28-R	63,0		
Expressed protein	EST-SSR-29-F	60,0	59,5	179
	EST-SSR-29-R	62,0		
Zinc finGer (AN1-like) family protein	EST-SSR-30-F	67,0	59,5	282
	EST-SSR-30-R	66,0		
Expressed protein	EST-SSR-31-F	65,0	59,5	197
	EST-SSR-31-R	65,0		
Universal stress protein	EST-SSR-32-F	63,0		188
	EST-SSR-32-R	64,0		
Universal stress protein	EST-SSR-33-F	56,61	62,0	168
	EST-SSR-33-R	56,10		

* As seqüências dos *primers* não estão registrados devido ao fato de estarem em fase de publicação.

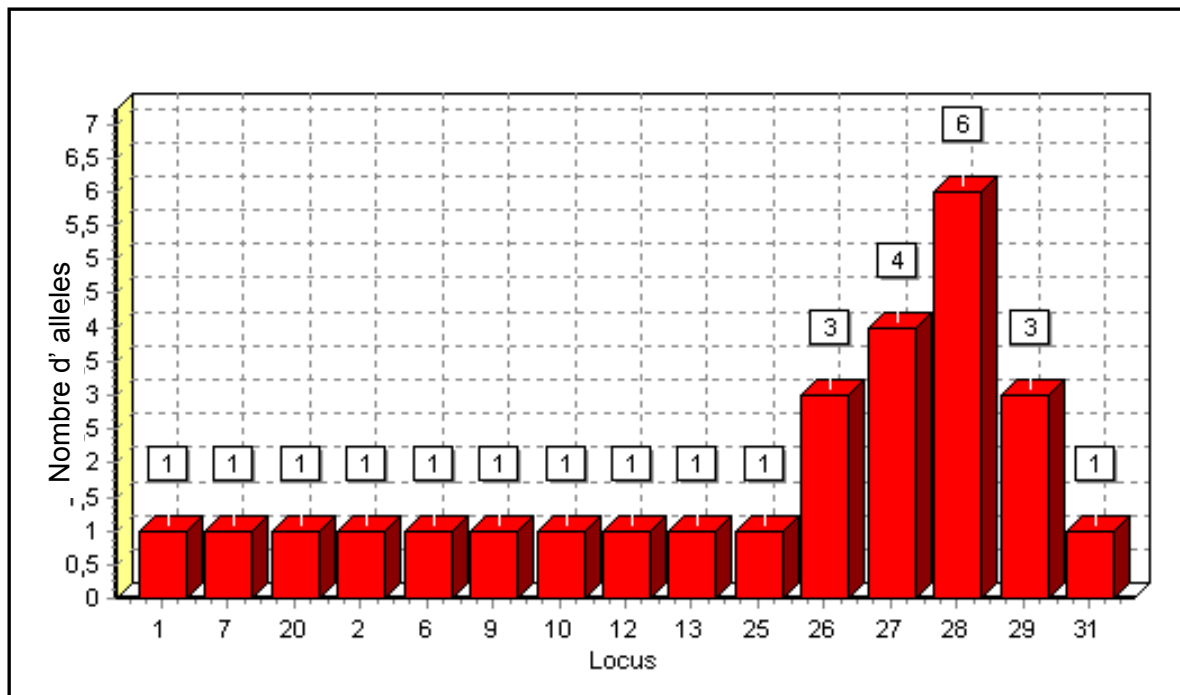


Figura 16. Distribuição de alelos por locos dos *primers* SSR obtido pelo programa Genetix, a partir de 31 *primers* SSR e clones de cacauero com diferentes níveis de resistência a vassoura-de-bruxa.

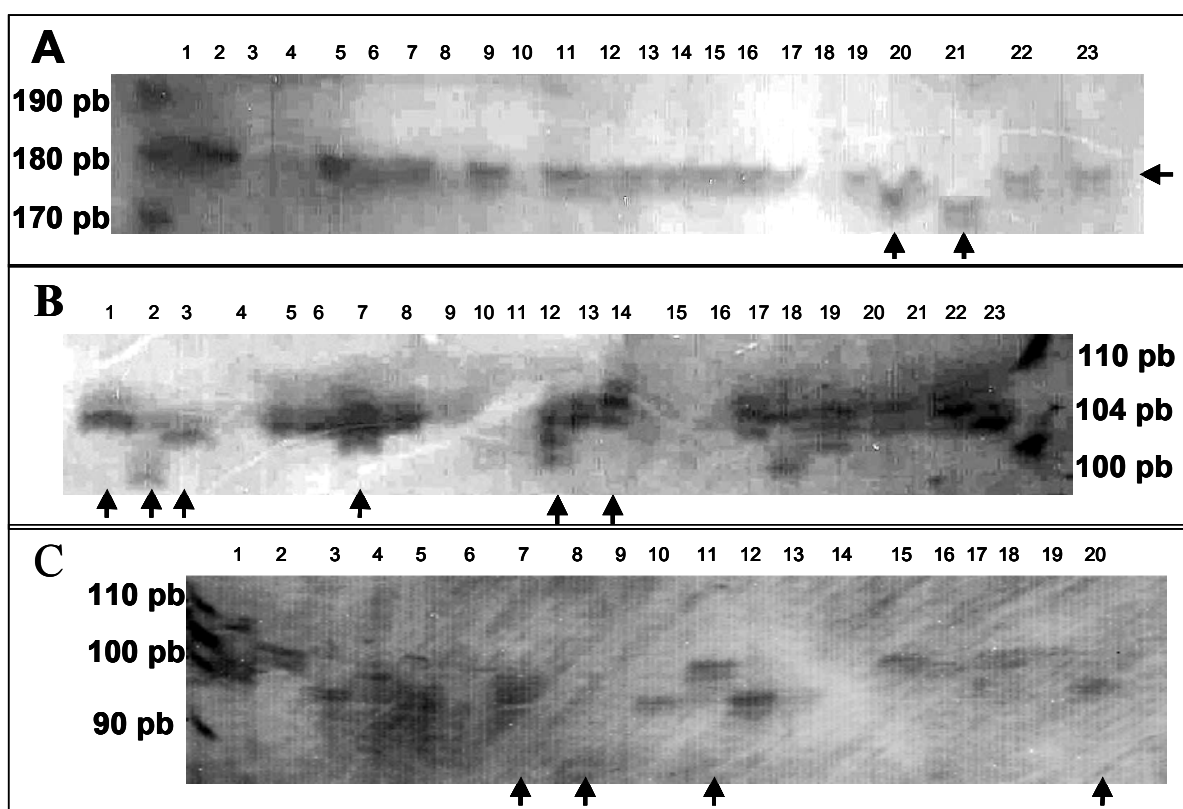


Figura 17. Géis de SSR dos *primers* EST-SSR-27, EST-SSR-28, EST-SSR-29, com os genótipos de cacau com diferentes níveis de resistência. (A) *primer* 29 referente ao gene *Expressed protein*, (B) *primer* 28 referente ao gene *Protein F2D10.18*, (C) *primer* 27 referente ao gene *Expressed protein*. Os genótipos estão numerados de 1 a 23, sendo 1 (R1), 2 (R2), 3 (R3), 4 (R4), 5(R5), 6(R6), 7(R7), 8 (R8), 9 (R9), 10 (R10), 11(R11), 12(R12), 13(R13), 14(R14), 15 (R15), 16(R16), 17(R17), 18(R18), 19 (R19), 20(R20), 21(R21), 22(R22), 23(R23), conforme Tabela 3. As setas indicam os alelos encontrados para cada *primer*. O tamanho dos fragmentos obtidos foi estimado com o marcador 10 pb (Invitrogen).

Tabela 11. *Primers* utilizados na amplificação dos clones, número total de alelos por loco e alelos obtidos em cada um, em pares de bases. H_e : heterozigosidade esperada para cada loco; H_o : heterozigosidade observada para cada loco; H_T : heterozigosidade total de cada loco.

<i>Primers</i>	Gene	Nº alelos	Alelos (pb)				H_e	H_o	H_T		
EST-SSR-26	DNA polimerase related	3	192	198	200		0.4753	0	0.5033		
EST-SSR-29	Expressed protein	3	173	175	179		0.1626	0	0.6857		
EST-SSR-27	Expressed protein	4	90	92	94	96	0.6511	0.3333	0.5188		
EST-SSR-28	Protein F2D10.18	6	102	104	106	108	110	112	0.4961	0.3125	0.1700
Total		16									
Média dos locos polimórficos		4					0.4462	0.16145	0.4694		

Tabela 12. Frequência de quatro locos microssatélites de ESTs da interação cacau-*M. pernicioso*

EST-SSR-26		EST-SSR-27		EST-SSR-28		EST-SSR-29	
Tamanho	Frequência	Tamanho	Frequência	Tamanho	Frequência	Tamanho	Frequência
192	0,6667	90	0,0667	102	0,0313	173	0,0435
198	0,0556	92	0,5000	104	0,6875	175	0,0435
200	0,2778	94	0,2000	106	0,1563	179	0,913
		96	0,2333	108	0,0313		
				110	0,0313		
				112	0,0625		

A Figura 18 mostra os resultados da análise de componentes principais (PCA). A distribuição dos clones de cacauzeiro utilizados para a análise encontram-se no plano com as respectivas porcentagens da variação genética por eixo, mostrando uma tendência de separação entre materiais resistentes e susceptíveis, e uma separação do clone R11 em relação ao Sca 6 e aos demais clones.

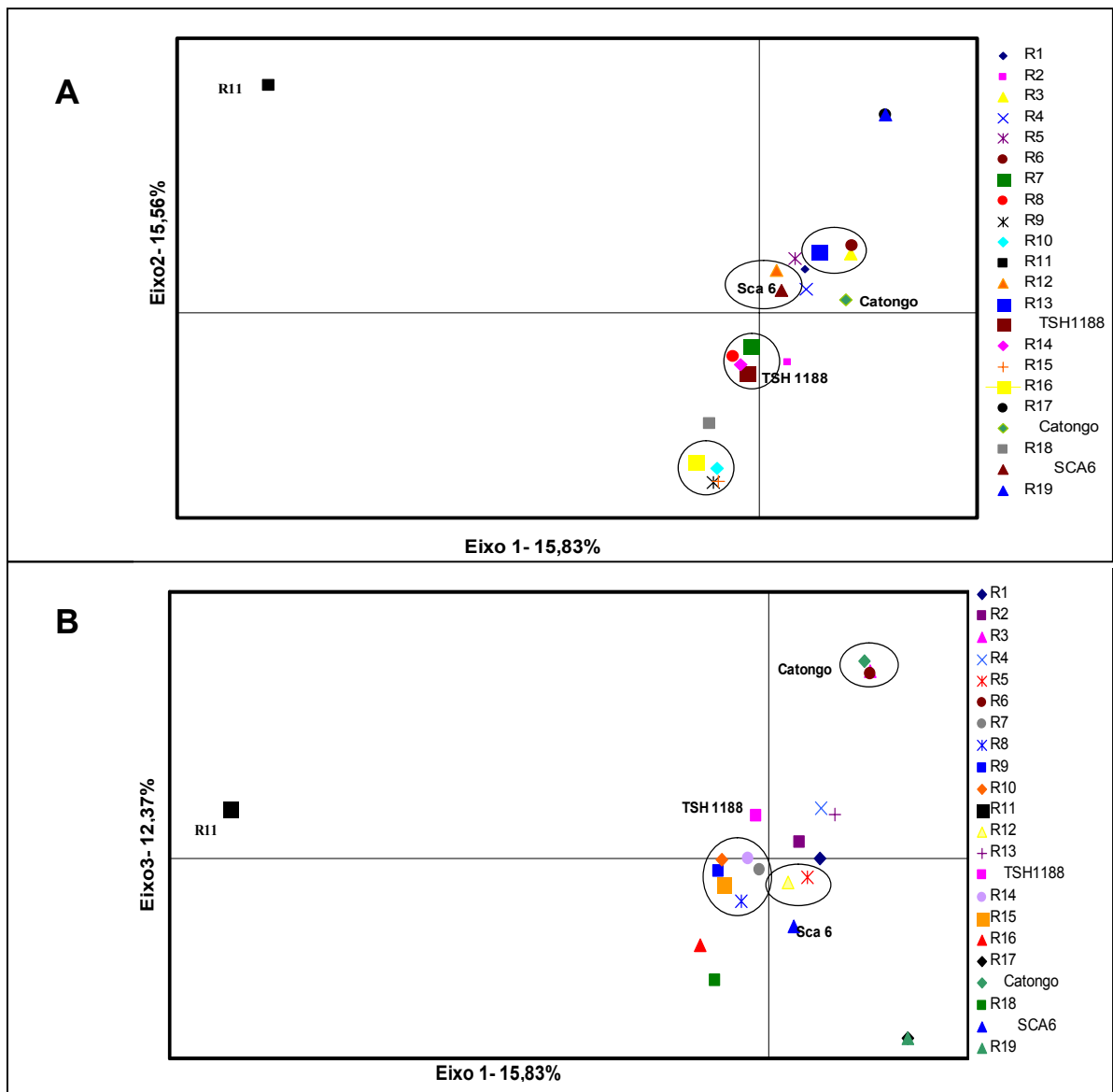


Figura 18. Análise de componente principal dos clones com diferentes níveis de resistência usados para análise de polimorfismo por SSR. (A) Distribuição dos clones no plano formado pelos eixos fatoriais 1 e 2; (B) Distribuição dos clones nos eixos fatoriais 1 e 3.

5. DISCUSSÃO

5.1. Genes relacionados com resistência das bibliotecas de interação cacau *Moniliophthora perniciosa*

As bibliotecas de ESTs mostram-se como uma importante abordagem experimental na seleção de novos genes de resistência, e pôde dar uma visão global da resposta do cacau a presença do fungo *Moniliophthora perniciosa*, possibilitando assim um melhor entendimento dos mecanismos de defesa envolvidos no processo de resistência à doença.

Os genes obtidos em bibliotecas de ESTs dão subsídios para diversos estudos, como por exemplo a possibilidade de inferir sobre fatores biológicos vitais ao desenvolvimento da planta, como vias de transdução de sinais, genes responsáveis pelo desenvolvimento da planta, estresse biótico e abiótico, regulação gênica, análise de polimorfismo e vários outros assuntos relacionados ao metabolismo da planta (ARRUDA, 2001). No presente estudo foram utilizadas as seqüências dos ESTs envolvidos na interação cacau-*M. perniciosa* para a identificação de polimorfismo, visando identificar marcas relacionadas a resistência de forma a contribuir com o programa de melhoramento genético do cacau.

Para a classificação dos genes, a metodologia descrita por Journet et al. (2002) permitiu um alcance de alta qualidade da anotação e a seleção de genes de resistência de todas as classes funcionais. Foi observado que muitos genes iguais foram obtidos nas bibliotecas de meristema e fruto e nas bibliotecas de cacau existentes em banco de dados (JONES et al., 2002; VERICA et al., 2004), por isso a maior parte dos genes foi selecionada da biblioteca de fruto de TSH1188, visto que de qualquer forma estaria abrangendo todos os possíveis genes envolvidos na

interação cacau-*Moniliophthora perniciosa*. Pôde-se observar a existência de genes de resistência diretamente relacionados a teoria gene-a-gene e genes de efeitos menores envolvidos na resistência. Com relação a genes relacionados com a resistência monogênica foi selecionado o gene AVR9 que é um gene de efeito maior, muito estudado em tomate conferindo resistência a *Cladosporium fulvum* (ROMEIS et al., 1999).

Entre os genes de efeitos menores têm-se vários genes envolvidos no mecanismo de resistência. Os genes superóxido dismutase estão envolvidos nos mecanismos celulares de inibição das espécies ativas de oxigênio (EAO). Estes genes normalmente estão ligados a um metal (Cu/Zn, Mn, Fe). As plantas normalmente têm Cu/Zn-SOD no citosol, Cu/Zn e ou Fé-SOD no cloroplasto e Mn-SOD na mitocôndria. Entre os genes selecionados estão representadas as formas Cu/ZN e Fe-SOD e os mesmos estavam presentes nas três bibliotecas utilizadas para a seleção e nas bibliotecas de cacau existentes em bancos de dados (JONES et al., 2002; VERICA et al., 2004).

Entre os genes selecionados, os genes fenilalanina amônia liase (PAL) e chalcona sintase (CHS) estão envolvidos no mecanismo de defesa. Estes genes participam da síntese de compostos secundários, como as fitoalexinas, que conduzem ao processo de defesa através da inibição do patógeno (AUSUBEL et al., 1995). Foram selecionados também genes que participam do fenômeno de resistência sistêmica adquirida (SAR), que está associada com o aumento da expressão de vários genes que codificam para as proteínas de defesa, as PR-proteínas ou proteínas relacionadas à patogênese (DURRANT; DONG, 2004). Estas proteínas são representadas aqui pelas PR4b, PR1, β -1,3-glucanase e thaumatine (CORAM; PANG, 2005). Acredita-se que as β -1,3-glucanase protegem as plantas contra a infecção por fungos, visto que os glucanos são um dos principais componentes de paredes de fungos e são decompostas por essas proteínas (LINTHORST, 1991). As PR1, sem atividade biológica conhecida, não induzem resistência vírus, mas sim ao fungo *Peronospora tabacina* quando superexpressas em plantas transgênicas de tabaco (LINTHORST, 1991).

Em suma, os ESTs obtidos da interação cacau-*Moniliophthora perniciosa* demonstram várias respostas bioquímicas e moleculares de defesa do cacau. Estas respostas devem ser detalhadamente estudadas para que melhor se compreenda as estratégias de defesa empregadas pela planta. A identificação de

polimorfismo a partir desses ESTs pode ser uma importante ferramenta para a obtenção de marcas associadas a respostas de defesa específicas.

5.2. Identificação de SNPs em ESTs de cacau

SNPs são marcadores genéticos altamente polimórficos e que vêm ganhando muito espaço em estudos de análises genéticas em plantas. Como exemplo têm-se os trabalhos realizados com milho (CHING et al., 2002; THORNSBERRY et al., 2001), *Arabidopsis* (CHO et al., 1999; NORDBORG et al., 2002), cana-de-açúcar (GRIVET et al., 2003) e trigo (SOMERS et al., 2003). Vários métodos têm sido utilizados para a identificação de SNPs, entre eles a descoberta através de dados de seqüenciamento é um método muito promissor, por ser uma barata e abundante fonte de SNPs (GU et al., 1998). Nesse estudo foi iniciado um trabalho pioneiro de identificação de SNPs a partir de dados de ESTs envolvidos na interação cacau-*Moniliophthora perniciosa*. A identificação de SNPs feita por alinhamento múltiplo de seqüências, mostrou-se muito interessante por permitir a comparação de seqüências de genes de resistência expressos e oriundos de diferentes genótipos. Por outro lado o método de alinhamento de seqüências é limitado pela exigência de seqüências em banco de dados e não distingue erros de incorporação durante a transcrição inversa de mRNA no caso de bibliotecas de cDNA. Porém essas dificuldades foram sanadas pela existência dos dados de sequenciamento de bibliotecas de cacau confeccionadas em trabalhos anteriores (GESTEIRA et al., 2004; JONES et al., 2002; LANAUD et al., 2006; VERICA et al., 2004; ZAIDAN et al., 2005) e pela estratégia adotada de re-sequenciar os genes selecionados para a identificação de SNPs.

Uma dificuldade na identificação dos SNPs foi a de detectar manualmente as substituições que ocorriam no alinhamento múltiplo de seqüências (Figura 9). Para resolver esse problema foi adotada a utilização do programa CLOURE-D que utilizou esses resultados do CLUSTAL W e realçou os nucleotídeos diferentes em relação a seqüência de referência, facilitando dessa forma a identificação dos SNPs. Assim foi possível a identificação de 85 SNPs candidatos de estarem envolvidos no processo de resistência à vassoura-de-bruxa. Esses dados podem ser importantes

informações para futuros trabalhos de piramidação de genes pelo programa de melhoramento genético do cacau.

Grande parte dos possíveis SNPs foram encontrados em genes envolvidos em transdução de sinais, síntese de proteínas e defesa e resgate de célula. Como exemplo tem-se o gene *Ras-related GTP-binding family protein* que teve um total de 22 SNPs candidatos, o qual está envolvido em transdução de sinais (CHEONG et al., 2002). Esses SNPs identificados em diferentes genótipos possuíram uma frequência relativamente alta de SNPs do tipo transição, que são substituições entre nucleotídeos do mesmo grupo químico. Esse tipo de resultado foi observado em programas prévios de identificação de SNPs envolvendo outras culturas (GUPTA et al., 2001; PICOULT-NEWBERG et al., 1999). Isso acontece porque uma transversão gera uma distorção de largura na molécula de DNA e a probabilidade do sistema de reparo da molécula atuar neste tipo de mutação é muito mais rara do que no caso de uma transição.

Para a validação dos SNPs foi utilizado o método AS-PCR, desenvolvido com sucesso em outros trabalhos, em arroz (HAYASHI et al., 2006) e cevada (BUNDOCK; HENRY, 2004; SOLEIMANI e al., 2003). É um método relativamente barato e rápido, possibilitando a detecção dos SNPs em gel de agarose (SOLEIMANI et al., 2003). Considerando que a probabilidade de obtenção de um SNP verdadeiro é maior entre genótipos diferentes, na obtenção de *primers* para detecção de SNPs utilizou-se os SNPs detectados via bioinformática que levou em consideração apenas a ocorrência de SNPs entre os genótipos diferentes, como por exemplo entre o genótipo resistente (TSH 1188) e o suscetível (Catongo), ou entre o forastero e o TSH 1188. Ao adotar esses critérios, mais os critérios exigidos para o desenho de *primers* AS-PCR, foi possível a obtenção de *triplets* de *primers* para os genes *superoxide dismutase*, *calcium binding*, *peroxidase putative*, *protease inhibitor* e *heat shock protein*. Infelizmente, no caso deste estudo com cacau não foram obtidos os resultados esperados: esperava-se a amplificação de dois fragmentos, um referente ao gene controle e o segundo referente a presença do SNP e apenas um fragmento (o controle) para ausência do SNP. Uma primeira explicação para este resultado seria a existência de erros de sequenciamento durante a identificação dos SNPs porém essa hipótese é muito pouco provável visto que os genes selecionados para o estudo foram re-sequenciados e analisados justamente para eliminar a ocorrência deste artefato.

Outro resultado não previsto que pode estar aliado ao seqüenciamento foi a presença dos fragmentos referente ao *primer* controle e ao *primer* AS-PCR para o *triplet* de *primers* 5 nos genótipos TSH1188 e Catongo (Figura 12). Esperava-se obter para esse *triplet* de *primers* a amplificação dos fragmentos de 114 pb (controle) e 103 pb (AS-PCR) no genótipo do TSH 1188, e a amplificação somente do fragmento de 103 pb (controle) no genótipo do Catongo, já que o possível SNP foi obtido em uma seqüência de TSH 1188. Embora especulativo, e apesar de dados anteriores mostrarem que o Catongo é um material que com alto índice de homozigose (PIRES, 2003), existe a possibilidade do Catongo apresentar-se heterozigoto para esse gene, que pode não ter sido identificada nos dados de seqüência, vez que, no seqüenciamento apenas um dos alelos tenha por acaso sido sequenciado. Uma segunda hipótese para o resultado não-esperado baseia-se na baixa especificidade da técnica utilizada e nas condições dos *primers*. Uma alternativa para ambas hipóteses seria a utilização da técnica chamada *tetra-primer ARMS-PCR*, uma modificação da AS-PCR feita por Ye et al. (2001). Nesta publicação, foram obtidos resultados mais consistentes, pela utilização de um conjunto de quatro *primers* na reação de PCR e pela visualização dos fragmentos em gel de poliacrilamida não-desnaturante, possibilitando a identificação de indivíduos heterozigotos. Outra técnica alternativa desenvolvida a partir do AS-PCR foi o PCR-CTPP que utiliza *primers* para os dois alelos de um SNP onde no primeiro par de *primers*, a base 3' do *primer* reverse é complementar ao alelo "x" e amplifica somente o alelo "x", e no segundo par de *primers* a base 3' do *primer forward* corresponde a base "y" e amplifica somente o alelo "y" (HAMAJIMA, 2003).

Uma dificuldade existente para a utilização da técnica AS-PCR está em obter bons *primers* para amplificação exatamente na posição do SNP, ou seja, no desenho dos *primers*, que deve considerar não somente a posição do SNP na posição 3' como também precisa considerar a competição dos *primers* entre si, daí a necessidade de testar os três *primers* uns com os outros. Conseqüentemente, neste conjunto de dados de 85 SNPs, foi possível o desenho de apenas nove conjuntos de *primers* AS-PCR para validação dos SNPs. Essa artefato foi demonstrado pelos autores Bundok et al. (2006) em um trabalho com cevada, onde foram desenhados um total de 230 *primers* AS-PCR desenvolvendo apenas 36 marcadores. Comparativamente, o resultado obtido aqui está nas proporções observadas. Além disso os *primers* poderiam não ter sido loco específicos e amplificaram alelos de

outros locos similares aos dos ESTs estudados. Esta hipótese decorre do fato de que genes de resistência podem ocorrer na forma de famílias gênicas duplicadas (HAMOND-KOSAK; JONES, 1997). Os *primers* 9, 11 e 14 amplificaram apenas os *primers* flanqueadores. Nesses casos o produto de amplificação maior pode ter agido como um competidor suficientemente efetivo para prevenir a amplificação do AS-PCR. É provável que alguns *primers* internos projetados para o AS-PCR tiveram qualidade sub-ótimas de PCR, resultando no fracasso da amplificação. Para superar este problema, modificações no método para o desenho dos *primers* alelo-específicos com aumento da especificidade e discriminação segura entre alelos foi sugerida em outros trabalhos (DRENKARD et al., 2000; KWOK et al., 1990)

Para a validação desses 85 SNPs candidatos identificados outros métodos alternativos podem ser realizados, como o redesenhamento dos *primers* segundo a técnica ARMS-PCR, já tendo hoje disponível um site específico para o desenho de *primers* para SNP segundo esta técnica http://cedar.genetics.soton.ac.uk/public_html/primer1.html (YE et al., 2001). Outra alternativa é a validação desses SNPs por seqüenciamento com a incorporação de *dideoxy* nucleotídeos usando o kit SNaPshot™ Multiplex (Applied Biosystems) de acordo com instruções do fabricante e analisados com o *software Gene Mapper* (SOLEIMANI et al., 2003). Por fim outra possibilidade é a utilização do PCR em tempo real em ensaio de discriminação alélica com a utilização do sistema Taq Man que utiliza sondas específicas para cada SNP a ser validado e marcados duplamente com um corante “repórter” e um corante “silenciador” (GIULIETTI et al., 2001).

Apesar dos resultados não esperados com a técnica AS-PCR, alternativas foram propostas e as análises de bioinformática puderam mostrar que a identificação de SNPs a partir de seqüências expressas podem mostrar resultados muito promissores a serem utilizados em trabalhos futuros de validação e genotipagem dessas potenciais marcas para assim serem utilizados no processo de piramidação de genes de resistência, um dos atuais objetivos do programa de melhoramento genético do cacauero.

5.3. Identificação de SSRs em ESTs de cacau

O uso de marcadores SSRs tornou-se uma rotina em análise de polimorfismo e genética de populações em plantas (GOLDSTEIN; SCHLOTTERER, 1999). Em cacau já se dispõe de alguns trabalhos utilizando SSR para seleção de marcas de resistência (ALBUQUERQUE, 2006; FALEIRO et al., 2003, LANAUD, 1999). O objetivo principal desse estudo foi o de identificar, otimizar e validar SSR a partir da interação cacau-*Moniliophthora perniciosa*, e este foi o primeiro estudo feito em cacau que descreve SSR em ESTs.

A identificação de polimorfismo a partir de dados de seqüências de ESTs da interação cacau-*Moniliophthora perniciosa* mostrou-se muito promissora. De 32 SSR selecionados 15 geraram fragmentos esperados, e foram gerados 16 alelos referentes aos locos polimórficos, com uma média de 4,0 alelos por loco, confirmando o potencial de utilização de ESTs para detecção de SSR. Um resultado bastante interessante foi uma certa separação entre genótipos resistentes e susceptíveis e entre genótipos resistentes para os genes específicos, assim como a separação entre o Sca 6 e os outros genótipos resistentes, demonstrando o envolvimento de diferentes genes de resistência à vassoura-de-bruxa. De certa forma esse trabalho confirma o estudo feito por Marita et al. (2001) e Pires (2003) no qual eles conseguiram separar por RAPD e SSR os clones utilizados nesse estudo em diferentes níveis de resistência à vassoura-de-bruxa. Vale salientar que neste estudo temos a vantagem de estar lidando com genes expressos da interação cacau-*Moniliophthora perniciosa*, assim sendo, marcas identificadas a partir destes genes nos clones resistentes podem nos dar informações mais específicas acerca da resistência.

Comparando-se nossos resultados com outros resultados obtidos para diferentes acessos de cacaueiro verifica-se que outros autores têm encontrado maior número de alelos por locos. Motamayor et al. (2002) encontraram uma média de 8,7 alelos/locos utilizando 16 SSR para comparar acessos de cacaueiro de diferentes países do grupo forasteiro. Sereno et al. (2006) avaliaram 11 locos de SSR em 94 acessos "CAB" e encontraram uma média de 4,45 alelos/locos. O menor número de alelos/locos encontrados aqui podem advir: (i) da proximidade genética entre os clones de cacaueiros resistentes, o que reforça a necessidade de se aumentar a base genética entre os acessos de cacaueiros resistentes usados no programa de

melhoramento; (ii) do tamanho dos *motifs* dos SSRs analisados, pois a maior parte dos SSRs identificados possuía um baixo número de repetições por *motif*, dificultando assim a detecção de polimorfismo. As frequências globais de SSRs em regiões expressas variam significativamente em estudos diferentes que dependem do comprimento mínimo do motivo repetido e dos critérios usados para identificá-los (VARSHNEY et al., 2005). Em um número de conjunto de dados de EST de plantas, SSRs maiores que 18 pb estão presentes em aproximadamente 2-7% em seqüências de ESTs (<http://www.genome.clemson.edu/projects/ssr>). Em nosso conjunto de dados *motifs* maiores ou iguais a 18 bp estavam presentes em 6,4 % das seqüências dos ESTs mostrando que os resultados estão na faixa esperada, apesar da grande quantidade de SSRs com tamanho abaixo de 18 pb.

Ainda assim os resultados foram satisfatórios levando-se em conta que foi possível obter quatro locos polimórficos, mesmo com um número baixo de repetições oriundos de genes expressos, e também pelo fato de que o mapeamento de SSR derivados desses ESTs permitem a localização de marcas muito mais próximas dos genes de resistência no mapa genético porque se trata de seqüências expressas. A utilização e análise de SSR em ESTs já foram realizadas com sucesso em outras espécies, como por exemplo na obtenção de marcadores SSR derivados de ESTs usados para traçar um mapeamento comparativo entre milho e arroz (YU et al., 2004) e no desenvolvimento e mapeamento de SSR derivados de ESTs em *Eucalyptus* (FALCÃO et al., 2004).

Com relação aos alelos encontrados, os alelos mais raros foram obtidos para o loco 28 referente ao gene F2D10.18 que é um gene membro da super família da proteína *DNA-binding protein* envolvida na transdução de sinais e regulação pós-transcricional, que é citado como um dos mecanismos-chave para o desencadeamento do processo de resistência na planta (HAHN, 1996). Outras marcas foram obtidas em genes expressos de função ainda desconhecida, que pode vir a ser um gene muito interessante no processo de resistência, e poder se tratar de um gene específico à interação cacau-*Moniliophthora perniciosa*.

Finalmente, o banco de dados de ESTs pode ser de grande utilidade para a obtenção de marcas de resistência a serem usadas pelo programa de melhoramento genético do cacau, e os resultados obtidos com a identificação de SSR mostraram-se muito interessantes para a utilização dessas marcas em trabalhos adicionais de mapeamento genético.

6. CONCLUSÕES

Os ESTs obtidos da interação cacau-*Moniliophthora perniciosa* pertencem a classes funcionais envolvidas em várias respostas moleculares de defesa do cacau, e são importantes ferramentas para a obtenção de marcas associadas a respostas de defesa específicas.

As análises de bioinformática puderam mostrar que a identificação de SNPs a partir de seqüências expressas podem mostrar resultados muito promissores a serem utilizados em trabalhos futuros de validação e genotipagem dessas potenciais marcas para assim serem utilizados pelo programa de melhoramento genético do cacau. Porém, neste estudo, a validação dos SNPs por AS-PCR não mostrou-se satisfatória para os estudos com os ESTs de cacau.

Cento e vinte e cinco SSRs foram identificados na biblioteca de TSH 1188 de fruto, dos quais os 15 locos que obtiveram amplificação foram testados quanto ao polimorfismo, revelando quatro locos polimórficos com dois locos em heterozigose e 11 locos monomórficos, entre os 20 genótipos, o Catongo e os *bulks* resistente e suscetível testados.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AGRIOS, G. N. **Plant pathology**, San Diego: Academic Press, 635p, 1997.

AIME, M. C.; PHILLIPS - MORA, W. The causal agents of witches' broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. **Mycologia**, v. 97, p. 1012 - 1022, 2005.

ALBUQUERQUE, P. S. B. **Mapas de ligação e Identificação de locos controladores de características quantitativas (QTLs) associados à resistência a *Crinipellis pernicioso* em acessos de cacauero (*Theobroma cacao*) originários da Amazônia Brasileira**. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura Luis de Queiroz, Universidade de São Paulo, Piracicaba, 133p, 2006.

ALBUQUERQUE, P. et al. Doenças do cacauero. In: KAMATI, H. et al. **Manual de Fitopatologia**, 4. ed, Piracicaba: Ceres, cap.18, v. 2, p. 151 - 164, 2005.

ALVERSON, W. S. et al. Phylogeny of the core Malvales: evidence from ndhF sequence data. **American Journal of Botany**, v. 86, p. 1474 - 1486, 1999.

ANDEBRHAN, T.; FURTEK, D. The use of Random Amplified Polymorphic DNA (RAPD) markers to distinguish isolates of *Crinipellis pernicioso*, causative agent of witches' broom disease of *Theobroma* spp. In: **International Cocoa Research Conference**, Yamoussoukro, Proceedings, Ivory Coast: Producers Alliance, p. 133 - 134, 1993.

ANDERBRHAN, T.; ALMEIDA, L. C.; NAKAYAMA, L. H. I. Resistência de *Theobroma cacao* L. a *Crinipellis pernicioso* (Stahel) Singer: a experiência da Amazônia Brasileira. **Agrotropica**, n. 2, v. 10, p. 49 - 60, maio 1998.

ARAÚJO, M. Conservação da Mata Atlântica na região cacauera da Bahia. In: **V Seminário da Reserva da Biosfera da Mata Atlântica**, Anais, União dos Palmares: Conselho Nacional da Reserva da Biosfera da Mata Atlântica, p. 11 - 16, 1997.

ARRUDA, P. Sugarcane Transcriptome. A land mark in plant genomics in the tropics. **Genetics and Molecular Biology**, v. 24, p. 3, 2001.

AUSUBEL, F. M. et al. Use of *Arabidopsis thaliana* defense - related mutants to dissect the plant response to pathogens. **Proceedings of the National Academy of Sciences of USA**, v. 92, p. 4189 - 4196, 1995.

BAKER, E. K.; COLLEY, N. J.; ZUKER, C. S. The cyclophilin homolog Nina a functions as a chaperone, forming a stable complex *in vivo* with its protein target rhodopsin. **EMBO Journal**, v. 13, p. 4886 - 4895, 1994.

BARTLEY, B. G. D. Cacao, *Theobroma cacao*. In: Breeding for Durable Resistance in Perennial Crops, Food Agric. Org. **Plant Production and Protection**, n. 70, p. 25 - 48, 1986.

BARTLEY, B. G. A review of cacao improvement. Fundamental methods and results. In: **Proceedings of the Internation Workshop on Cocoa Breeding Strategies**, Proceedings, Kuala Lumpur: Malaysia, p. 16, 1994.

BHATTRAMAKI, D. et al. Conserved single nucleotide polymorphism (SNP) haplotypes in maize. In: **Plant and Animal Genome Conference**, p. 188, Proceedings, San Diego: Scherago International, 2000.

BOWEN, D. et al. Nucleotide sequence of the phosphoglycerate kinase gene from the extreme thermophile *Thermus thermophilus*. Comparison of the deduced amino acid sequence with that of the mesophilic yeast phosphoglycerate kinase. **Biochemical Journal**, v. 2, p. 509 - 517, 1988.

BROOKES, A. J. The essence of SNPs. **Gene**, v. 234, p. 177 - 186, 1999.

BRONDANI, R. P. V. et al. Development, characterization an mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics**, New York, v. 97, p. 816 - 827, 1998.

BRUNELLI, K. R.; SILVA, H. P.; CAMARGO, L. E. A. Mapeamento de genes de resistência quantitativa a *Puccinia polysora* em milho. **Fitopatologia Brasileira**, v. 27, p. 134 - 140. 2002.

BUNDOCK, P. C.; HENRY, R. J. Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. **Theoretical and Applied Genetics**, v. 109, p. 543 - 551, 2004.

CAMARGO, L. E. A.; BERGAMIN FILHO, A. Controle genético. In: BERGAMIN FILHO, A.; KIMATI, H.; AMORIM, L. (Eds.). **Manual de Fitopatologia: princípios e conceitos.**, 3. ed. São Paulo: Agronômica Ceres, v.1, p. 729 - 760, 1995.

CHEESMAN, E. E. Notes on the nomenclature, classification and possible relationships of cacao populations. **Tropical Agriculture**, v. 21, p. 143 - 159, 1944.

CHEONG, Y. H. et al. Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*1. **Plant Physiology**, v. 129, p. 661 - 677, 2002.

CHING, A. et al. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetics**, v. 3, p. 1 - 19, 2002.

CHO, R. J. et al. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. **Nature Genetics**, v. 23, p. 203 - 207, 1999.

COLLINS, N. et al. Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. **Plant Cell**, v. 11, p. 1365 - 1376, 1999.

CORAM T. E.; PANG E. C. K. Isolation and analysis of candidate ascochyta blight defence genes in chickpea. Part II. Microarray expression analysis of putative defence-related ESTs. **Physiological and Molecular Plant Pathology**, v. 66, p. 201 - 210, 2005.

CORRÊA, R. X. **Mapeamento genético da soja (*Glycine max* (L.) Merrill) utilizando marcadores RAPD**. 1995. 97 f. Dissertação (Mestrado em Genética e Melhoramento), Universidade Federal de Viçosa, Viçosa, 1995.

COWEN, L. E.; LINDQUIST, S. Hsp90 Potentiates the Rapid Evolution of New Traits: Drug Resistance in Diverse Fungi. **Science**, v. 309, n. 5744, p. 2185 - 2189, 2005.

CRESTE, S.; TULMANN NETO, A.; FIGUEIRA, F. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, v. 19, p. 299 - 306, 2001.

CROUZILLAT D. et al. *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. **Theoretical and Applied Genetics**, v. 93, p. 205 - 214, 1996.

CUATRECASAS, J. Cacao and its allies a taxonomic revision of the genus *Theobroma*. **Contributions from the United States National herbarium**, Wahington, n. 6, v. 35, p. 379 - 614, 1964.

DANGL, J. L.; JONES, J. D. G. Plant pathogens and integrated defence responses to infection. **Nature**, v. 411, 2001.

DANSON, J.; WASANO, K.; NOSE, A. Infection of rice plants with the sheath blight fungus causes an activation of pentose phosphate and glycolytic pathways. **European Journal of Plant Pathology**, v.3, p. 555 - 561, 2004.

DE BODT, S. et al. Genome-wide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. **Journal Molecular Evolution**, v. 56, p. 573 - 586, 2003.

DONG, X. The role of membrane-bound ankyrin-repeat protein ACD6 in programmed cell death and plant defense. **Science STKE**, v. 24, 2004.

DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, p. 13 - 15, 1990.

DRENKARD, E. et al. A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. **Plant Physiology**, v. 124, p. 1483 - 1492, 2000.

DUFRESNE, M. et al. Clk1, a serine/threonine protein kinase-encoding gene, is involved in pathogenicity of *Colletotrichum lindemuthianum* on common bean. **Molecular Plant Microbe Interactions**, v. 11, p. 99 - 108, 1998.

DUKE, N. Involvement of superoxide anion generation in hypersensitive response of potato tuber tissues to infection with an incompatible race of *Phytophthora infestans*. **Physiology and Plant Pathology**, v. 23, p. 345 - 347, 1983.

DURRANT, W. E.; DONG, X. Systemic Acquired Resistance. **Annual Review of Phytopathology**, v. 42, p.185 - 209, 2004.

ECKER, J. R.; DAVIS, R. W. Plant defense genes are regulated by ethylene. **Proceedings of the National Academy of Sciences of USA**, v. 84, p. 5202 - 5206, 1987.

EPRINTSEV, A. T.; SEMENOVA, E. V.; POPOV, V. N. Induction of aconitate hydratase in hepatocytes of starving rats. **Biochemistry**, v. 67, p. 795 - 801, 2002.

EVANS, H. C.; PRIOR, C. Cocoa pod diseases: Causal agents and control. **Outlook on Agriculture**, v. 16, p 35 - 41, 1987.

EWING, B. et al. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. **Genome Research**, v. 8, p. 175 - 185, 1998.

FAVARON, F.; SELLA, L.; D'OVIDIO, R. Relationships among endopolygalacturonase, oxalate, pH, and plant polygalacturonase-inhibiting protein (PGIP) in the interaction between *Sclerotinia sclerotiorum* and soybean. **Molecular Plant Microbe Interaction**, v. 17, p. 1402 - 1049, 2004.

FALCÃO, C. L. et al. Desenvolvimento e mapeamento de microssatélites derivados de ESTs em *Eucalyptus*. **Circular Técnica**, Brasília, DF, 2004.

FALEIRO, F. G. et al. Mapping QTLs for witches' broom (*Crinipellis Perniciosa*) resistance in cacao (*Theobroma Cacao* L.). **Euphytica**, v.149, p. 227 - 235, 2006.

FALEIRO, F. G.; PIRES, J. L.; LOPES, U. V. Uso de marcadores moleculares RAPD e microssatélites visando a confirmação da fecundação cruzada entre *Theobroma cacao* e *Theobroma grandiflorum*. **Agrotrópica**, v. 15, p. 41 - 46, 2003.

FALEIRO, F. G. et al. Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando a obtenção de marcadores RAPD. **Agrotrópica**, v. 2, p. 31 - 34. 2002.

FAO. **FAOSTAT Agriculture Data**. Disponível em <http://faostat.fao.org/>. Acesso em 17 de julho de 2005.

FERREIRA M. E.; GRATTAPAGLIA D. Introdução ao uso de marcadores RAPD e RFLP em análise genética. **EMBRAPA/CENARGEN**, Brasília (EMBRAPA-CENARGEN Documento 20), 1995.

FIGUEIRA, A.; JANICK, J.; GOLDSBROUG, H. P. Genome size and DNA polymorphism in *Theobroma cacao*. **Journal of the American Society for Horticultural Science**, v. 117, p. 673 - 677, 1992.

FLORIS, E.; ALVAREZ, J. M. Nature of resistance of seven melon lines to *Sphaerotheca fuliginea*. **Plant Pathology**, v. 45, p. 155 - 160, 1996.

FRIAS, G. A. Infection biology of *Crinipellis pernicioso* on vegetative flushes of cacao. **Plant Disease**, v. 75, p. 552 - 556, 1991.

FU, LIAN-HAI. et al. A Selenoprotein in the Plant Kingdom : Mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in *Chlamydomonas reinhardtii* glutathione peroxidase. **The Journal of Biological Chemistry**, n. 29, v. 277, p. 25983 - 25991, 2002.

FULTON, R. H. The cacao disease trilogy: Black pod, Monilia pod rot, and witches' broom. **Plant Disease**, v. 73, p. 601 - 603, 1989.

GAMAS, P.; DE BILLY, F.; TRUCHET, G. Symbiosis-specific expression of two *Medicago truncatula* nodulin genes, MtN1 and MtN13, encoding products homologous to plant defense proteins. **Molecular Plant Microbe Interaction**, v.11, p. 393 - 403, 1998.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. **Genetics**, v. 137, p. 1121 - 1137, 1994.

GESTEIRA, A. et al. Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora pernicioso*. **Annals of Botany**, no prelo, 2007.

GESTEIRA A. et al. Understanding the *Theobroma cacao*-*Crinipellis pernicioso* interaction using ESTs and proteomic analyses. **Ingenic Newsletter**, v. 9, p 27 - 32, 2004.

GILSON, P. et al. AGP4 is an arabinogalactan protein whose expression is suppressed by wounding and fungal infection in *Nicotiana glauca*. **Plant Cell Biology**, v. 215, p.128 - 139, 2001.

GIULIETTI, A. et al. An overview of Real-Time quantitative PCR: applications to quantify cytokine gene expression. **Methods**, v. 25, p. 386 - 401, 2001.

GOLDSTEIN D. B.; SCHLOTTERER C. Microsatellites: Evolution and applications. Oxford: **Oxford University Press**, Oxford, 352p, 1999.

GOMES, L. M. C. et al. Diversidade genética de *Crinipellis perniciososa* na região sul da Bahia utilizando marcadores moleculares RAPD. **Fitopatologia Brasileira**, v. 25, p. 377, 2000.

GRAMACHO, K. P. et al. Development of microsatellite markers for the genetic analysis of *Crinipellis perniciososa*. In: **International Cocoa Research Conference**, Proceedings, Gana: Cambery, v. 14, 2003.

GRIVET, L. et al. ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. **Theoretical and Applied Genetics**, v. 106, p. 190 - 197, 2003.

GUIMARÃES, P. E. M.; COSTA, M. C. R. SNPs: Sutis diferenças em um código. **Biotecnologia Ciência e Desenvolvimento**, n.26, maio/junho 2002.

GUPTA, V. et al. Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. **Theoretical and Applied Genetics**, v.108, p. 743 - 749, 2004.

GUPTA, P. K.; ROY, J. K.; PRASAD, M. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science**, v. 80, p. 524 - 535, 2001.

HAYASHI, K.; YOSHIDA, H.; ASHIKAWA, L. Development of PCR-based allele-specific and InDel marker sets for nine blast resistance genes. **Theoretical and Applied Genetics**, v. 113, p. 251 - 260, 2006.

HAHN, M. G. Microbial elicitors and their receptors in plants. **Annual Review of Phytopathology**, v. 34, p. 387 - 412, 1996.

HAMADA, H.; PETRINO, M. G.; KAKUNAGA, T. A. A novel repeat element with Z-DNA-forming potential is widely found in evolutionary diverse eucariotic genomes. **Proceedings National Academy Science United States America**, v. 79, p. 6465-6469, 1982.

HAMAJIMA, A. T. N. et al. Duplex polymerase chain reaction with confronting two-pair primers (pcr-ctpp) for genotyping alcohol dehydrogenase β subunit (*adh2*) and aldehyde dehydrogenase 2 (*aldh2*). **Alcohol and Alcoholism**, v. 38, p. 407 - 410, 2003.

HAMMOND-KOSACK, K. E.; JONES, J. D. G. Inducible plant defense mechanisms and resistance gene function. **Plant Cell**, v. 8, p. 1773 -1791, 1996.

HATEY, F. et al. Expressed Sequence Tags for genes : a review. **Genetic Selection and Evolution**, v. 30, p. 521 - 554, 1998.

HOLLIDAY, P. Witches' broom disease of cacao (*Marasmius perniciosus* Stahel). London: **Her Majesty's Stationery Office**, 8 p, 1952.

HORI, K. et al. Efficient construction of high-density linkage map and its application to QTL analysis in barley. **Theoretical Applied Genetics**, v. 107, p. 806 - 813. 2003.

JABS, T. et al. Elicitor-stimulated ion fluxes and reactive oxygen species from the oxidative burst signal defense gene activation and phytoalexin synthesis in parsley. **Proceedings National Academy Science United States America**, v. 94, p. 4800-4805, 1997.

JAKOBEK, J. L.; LINDGREN, P.B. Expression of a bean acid phosphatase cDNA is correlated with disease resistance. **Journal of Experimental Botany**. n. 367, v. 53, p 387 - 399, 2002.

JONES, P. G. et al. Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. **Planta**, v. 216, p. 255 - 264, 2002.

JOURNET, E. P. D. et al. Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. **Nucleic Acids Research**, v.30, p. 5579 - 5592, 2002.

JALALI, B. L.; BHARGAVA, S.; KAMBLE, A. Signal transduction and transcriptional regulation of plant defence responses. **Journal of Phytopathology**, v.154, p. 65 - 74, 2006.

KAWALLECK, P. et al. Induction by fungal elicitor of S-adenosyl-L-methionine synthetase and S-adenosyl-L-homocysteine hydrolase mRNAs in cultured cells and leaves of *Petroselinum crispum*. **Proceedings National Academy Science United States America**, v. 89, p. 4713 - 4717, 1992.

KIM, Y. K. et al. A mitogenactivated protein kinase kinase required for induction of cytokinesis and appressorium formation by host signals in conidia of *Colletotrichum gloeosporioides*. **Plant Cell**, v.12, p. 1331 - 1343, 2000.

KOMBRINK, E.; SOMSSICH, I. E. Pathogenesis-related proteins and plant defense. In: CARROLL, G.; TUDZYNSKI, P. **The Mycota**, v. 5, p. 107 - 128, Part A: Plant Relationships (Carroll G. and Tudzynski P. eds.), Springer Verlag, , 1997.

KOO, J. C. et al. Over expression of a seed specific hevein like antimicrobial peptide from *Pharbitis nil* enhances resistance to a fungal pathogen in transgenic tobacco plants. **Plant Molecular Biology**, v.50, p. 441 - 452, 2002.

KOTA, R. et al. Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). **Molecular Genetics and Genomics**, v. 270, p. 24 - 33, 2003.

KWOK, P. Y.; GU Z. Single nucleotide polymorphism libraries: why and how are we building them?. **Molecular Medicine Today**, v. 5, p. 538 - 543, 1999.

KWOK, S. et al. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. **Nucleic Acids Research**, v.18, p. 999 - 1005, 1990.

LANAUD, C. et al. A large EST resource for *Theobroma cacao* including cDNAs isolated from various organs and under various biotic and abiotic stresses. In: **15TH International Cocoa Research Conference**, p.23, Proceedings, San Jose: Costa Rica, 2006.

LANAUD, C. et al. Characterization and genetic mapping of resistance and defense gene analogs in cocoa (*Theobroma cacao* L.). **Molecular Breeding**, v. 13, p. 211 - 227, 2004.

LANAUD, C. et al. Isolation and characterization of microsatellites in *Theobroma cacao* L. **Molecular Ecology**, v. 8, p. 2141 - 2143, 1999.

LANAUD, C. et al. Mapping Quantitative Trait Loci (QTL) for Resistance to *Phytophthora palmivora* in *Theobroma cacao* L. In: **International Cocoa Research Conference**, Trinidad, p. 99 - 105, 1996.

LEAL-BERTIOLI, S. C. M. O enfoque molecular na sistemática de fungos. **Revisão Anual de Patologia de Plantas**, v. 6, p. 197 - 230, 1998.

LEE, J. R. Molecular and functional characterization of a cyclophilin with antifungal activity from Chinese cabbage. **Biochemical and Biophysical Research Communications**, v. 16, p. 672 - 678, 2007.

LINTHORST, H. J. M. Pathogenesis-related proteins of plants. **Critical Reviews in Plant Science**, v. 10, p. 123 - 150, 1991.

LITT, M.; LUTY, J. A. A hipervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetics**, v. 44, p. 398 - 401, 1989.

LOPEZ, A. P.; FRANCO, A.R. Cloning and expression of cDNA encoding translationally controlled tumor protein from strawberry fruits. **Biomedical and Life Sciences**, v. 50, p. 447 - 449, 2006.

LUZ, E. D. M. N. et al. Evaluation of F1 cacao progênies for resistance to witches' broom disease. In: **International Congress of Plant Pathology**, Christchurch, Proceedings, p. 8, 2003.

LUZ, E. D. M. N. et al. Evaluation of cocoa in Brasil, for resistance to *Crinipellis pernicioso*. In: **12th International Cocoa Research Conference**. Proceedings. Lagos, Nigéria, Cocoa Producers' Alliance, p. 219 - 226, 1999.

LUZ, E. D. M. N. et al. Doenças do cacauero. In: ZAMBOLIM, L. ; VALE, F. X. R. **Controle de doenças de plantas: Grandes culturas**. Viçosa: UFV, v. 2, 131p, 1997.

MAHALINGAM, R. et al. Characterizing the stress/defense transcriptome of *Arabidopsis*. **Genome Biology**, v. 4, 14 p, 2003.

MARITA, J. M. et al. Analysis of genetic diversity in *Theobroma cacao* with emphasis on witches' broom disease resistance. **Crop Science**, v. 41, p. 1305 - 1316, 2001.

MATTINEN, L. et al. Identification and characterization of Nip necrosis inducing virulence protein of *Erwinia carotovora* subsp *carotovora*. **Molecular Plant Microbe Interaction**, v.17, p. 1366 - 1375, 2004.

MAUCH, F.; MAUCHMANI, B.; BOLLER, T. Antifungal hydrolases in pea tissue. 2. Inhibition on fungal growth by combination of chitinase and beta-1,3-glucanase. **Plant Physiology**, v. 88, p. 936 - 942, 1988.

MOTAMAYOR, J. C. et al. Cacao domestication I: the origin of the cacao cultivated by the Mayas. **Heredity**, v. 89, p. 380 - 386, 2002.

MOTILAL, L. A. et al. *Theobroma cacao* L.: Genome map and QTLs for *Phytophthora palmivora* resistance. In: **International Cocoa Research Conference**, Lagos. Proceedings, v.1, p. 111 - 118, 2000.

MUELLER, U. G.; WOLFENBARGER, L. L. AFLP genotyping and fingerprinting. **Trends Ecology Evolution**, v.14, p. 389 - 394, 1999.

MUNERATO, P. Curso: **Identificação de SNPs e genotipagem através de eletroforese capilar e sistema de detecção de seqüência**. ABI EXPERT Training Center-Applied Biosystems, São Paulo, 51p, 2005.

NELSON, R. R. Genetics of horizontal resistance to planta diseases. **Annual Review of Phytopathology**, v.16, p. 359 - 378, 1978.

NIELLA, G. R. **Frutificação *in vitro*, caracterização molecular e patogênica de *Crinipellis pernicios* (Stahel) Singer em *Theobroma cacao* L.** 2000. Tese (Doutorado em Agronomia). Universidade Federal de Lavras, Lavras, 75p, 2000.

NORDBORG, M. et al. The extent of linkage disequilibrium in *Arabidopsis thaliana*. **Nature Genetics**, v.30, p. 190 - 193, 2002.

OH, I. L. S. et al. Secretome Analysis Reveals an Arabidopsis Lipase Involved in Defense against *Alternaria brassicicola*. **The Plant Cell**, v.17, p. 2832 - 2847, october 2005.

PARLEVLIET, J. E. What is durable resistance, a general outline. In: JACOBS, T.H.; PARLEVLIET, J. E. **Durability of Disease Resistance**, Kluwer Academic Publishers, Dordrecht, p. 23 - 29, 1993.

PARLEVLIET, J. E.; ZADOKS, J. C. The integrated concept of disease resistance; a new view including horizontal and vertical resistance in plants. **Euphytica**, v. 26, p. 5 - 21,1977.

PASCHOLATI, S. F.; LEITE, B. Hospedeiro: Mecanismos de resistência. **Manual de Fitopatologia**, v. 1, p. 417- 452, 1995.

PEARL, H. M. et al. Construction of a genetic map for arabica coffee. **Theoretical Applied Genetics**, v. 108, p. 829 - 835. 2004.

PEREIRA, J. L. M. Management of witches' broom disease of cocoa: a contemporary retrospective. **Lagos: Cocoa Producers' alliance**, 36 p, 2000.

PEREIRA, J. L. et al. Primeira ocorrência da vassoura-de-bruxa na principal região produtora de cacau do Brasil. **Agrotropica**, v. 1, p. 70 - 81, 1989.

PFLIEGER, S.; LEFEBVRE, V.; CAUSSE, M. The candidate gene approach in plant genetics: a review. **Molecular Breeding**, v. 7, p. 275 - 291, 2001.

PICOULT-NEWBERG, L. et al. Mining SNPs from EST databases. **Genome Research**, v. 9, p. 167 - 174, 1999.

PINHEIRO DE CARVALHO, M. A. A. et al. Effect of metabolites of γ -aminobutyric shunt on activities of NAD- and NADP-isocitrate dehydrogenases and aconitate hydratase from higher plants. **Biology Bulletin**, v. 30, p. 236 - 242, 2003.

PINTO, L. et al. Isoenzimas e microssatélites em plantas. **Biotecnologia Ciência e Desenvolvimento**, n. 20, maio/junho 2001.

PIRES, J. L. **Avaliação quantitativa e molecular de germoplasma para o melhoramento do cacau com ênfase na produtividade, qualidade de frutos e resistência a doenças**. 2003. Tese (Doutorado em Genética e Melhoramento). Universidade Federal de Viçosa, Viçosa, 328p, 2003.

PIRES, J. L. et al. Variabilidade genética de fontes de resistência de *Theobroma cacao* a *Crinipellis pernicioso* com base em marcadores microssatélites. **Fitopatologia Brasileira**, v. 26, p. 347, 2001.

PIRES, J. L. et al. Diversity for phenotypic traits and molecular markers in CEPEC's germoplasm collection in Bahia, Brasil. In: **International Workshop on New Technologies and cocoa Breeding**, Kota Kinabalu, Sabah, Malaysia. Proceedings, S.1. INGENIC, p. 72 - 88, 2000.

PIRES, J. L. et al. Resistance to witches' broom - evaluation of genotypes from different origins. In: **12th International Cocoa Research Conference**. Proceedings. Lagos, Nigeria, Cocoa Producers' Alliance, p. 389 - 397, 1999.

PIRES J. L. et al. Cocoa breeding for witches' broom resistance at CEPEC, Bahia , Brazil. In: **International Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement**. Proceedings, Salvador, Brasil, p.910 - 1001, 1999a.

PIRES, J. L. et al. Cocoa breeding for witches's broom resistance at CEPEC, Bahia, Brasil. In: **Conferência Internacional de Pesquisas em Cacau**, Proceedings,. Lagos, Nigéria, Cocoa Producer's Alliance, p. 24 - 25, 1996b.

PLANT, P. J. et al. The C2 Domain of the Ubiquitin Protein Ligase Nedd4 Mediates Ca^{2+} -dependent Plasma Membrane Localization. **Journal of Biological Chemistry**, v. 272, p. 32329 - 32336, december 1997.

POWELL, W.; MACHRAY, G. C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends in Plant Science**, Oxford, n. 7, v. 1, p. 215 - 222, 1996.

POUND, F. J. Cocoa and witches' broom disease (*Maramius perniciosus*) of South America with notes on other species of *Theobroma*. **Cocoa Research**, v.1, p. 20 - 72, 1938.

POUND, F. J. Cocoa and witches' broom disease: report on a recent visit to the Amazon territory of Peru. In: Toxopeus H. **The archive of cocoa research**, London, United Kingdom, ACRI- IOCC, v. 1, 1943.

PURDY, L. H.; SCHMIDT, R.A. Status of cocoa witches broom: biology, epidemiology, and management. **Annual Review of Phytopathology**, v. 34, p. 573 - 594, 1996.

PURDY et al. An automated system for screening *Theobroma cacao* for resistance to witches' broom. **Agrotropica**, v. 9, p. 119 - 126, 1997.

PUGH, T. et al. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. **Theoretical and Applied Genetics**, v. 108, p. 1151 - 1161, 2004.

QUEIROZ, V. T. et al. Identification of a major QTL in cocoa (*Theobroma cacao* L.) associated with resistance to witches' broom disease. **Plant Breeding**, v. 122, p. 268 - 272, 2003.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, v. 5, p. 94 - 100, 2002.

REINHECKEL, T. et al. Comparative resistance of the 20S and 26S proteasome to oxidative stress. **Biochemical Journal**, v. 335, p. 637 - 642, 1998.

RICHMOND, T. A.; SOMERVILLE, C.R. The Cellulose Synthase Superfamily. **Plant Physiology**, v.124, p. 495 - 498, 2000.

RISTERUCCI, A. M. et al. Mapping of quantitative trait loci (QTLs) for resistance to *Phytophthora* in *Theobroma cacao* L. **INGENIC Newsletter**, v. 5, p. 9 - 10, 2000.

RIVAS, S. et al. CITRX thioredoxin interacts with the tomato Cf-9 resistance protein and negatively regulates defence **The EMBO Journal**, v. 23, p. 2156 - 2165, 2004.

ROBINSON, R. A. La "Escoba de Bruja", Enfermedad del cacao en Ecuador. In: **Informe FAOSTAT Agriculture Data**, 1978.

ROBINSON, R. A. **Plant pathossystems**. Springer Verlag, Berlin, Heidelberg, New York, 1976.

ROMEIS, T. et al. Rapid Avr9- and Cf-9-dependent activation of MAP kinases in tobacco cell cultures and leaves: convergence of resistance gene, elicitor, wound, and salicylate responses. **The Plant Cell**, v. 11, p. 273 - 287, February 1999.

ROOPASHREE, S. et al. Dual-function protein in plant defence: seed lectin from *Dolichos biflorus* (horse gram) exhibits lipoxygenase activity. **Biochemical Journal**, v. 395, p. 629-639, 2006.

ROSÁRIO, M. et al. Cacau: história e evolução no mundo. Ilhéus: **CEPLAC**, 46 p. 1978.

ROUHIER, N. et al. Poplar peroxiredoxin Q. A thioredoxin linked chloroplast antioxidant functional in pathogen defense. **Plant Physiology**, v. 134, p. 1027 - 1038, 2004.

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. Molecular Cloning: a laboratory manual. **Cold Spring Harbor Laboratory**, 2 ed, New York, 1989.

SANDERS, M. et al. An active role for endogenous b-1,3-glucanase genes in transgene-mediated co-suppression in tobacco. **The EMBO Journal**, v. 21, p. 5824 - 5832, 2002.

SANTOS, R. C. et al. Assesment of genetic diversity on a sample of cocoa accesions resistant to witches' broom disease based on RAPD and pedigree data. **Bragantia**, v. 64, p. 361 - 368, 2005.

SCHULTZ, C. J. et al. Using Genomic Resources to Guide Research Directions. The Arabinogalactan Protein Gene Family as a Test Case. **Plant Physiology**, v. 129, p. 1448 - 1463, 2002.

SERENO, M. L. et al. Genetic diversity and natural population structure of cação (*Theobroma cacao*) from the Brazilian Amazon evaluated by microsatellite markers. **Conservation Genetics**, v. 7, p. 13 - 24, 2006.

SHIMOMURA, S. et al. Structure of the Gene for an Auxin-Binding Protein and a Gene for 7SL RNA from *Arabidopsis thaliana*. **Plant and Cell Physiology**, v. 34, p. 633 - 637, 1993.

SIMÕES, I.; FARO, C. Structure and function of plant aspartic proteinases. **European Journal of Biochemistry**, v. 271, p. 2067 - 2075, 2004.

SMART, C. C.; FLEMING, A. J. Hormonal and Environmental Regulation of a Plant PDR5-like ABC Transporter. **The American Society for Biochemistry and Molecular Biology**, v. 271, p. 19351 - 19357, 1996.

SOLEIMANI, V. D.; BAUM B. R.; JOHNSON D. A. Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley. **Plant Molecular Biology Reporter**, v. 21, p. 281 - 288, 2003.

SOMERS, D. L. et al. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. **Genome**, v. 49, p. 431 - 437, 2003.

SORGER, P. K.; PELHAM, H. R. Purification and characterization of a heat-shock element binding protein from yeast. **The EMBO Journal**, v. 6, p. 3035 - 3041, 1987.

SOORI, V. A.; WATANABE, K. N. ; VALKONEN, J. P. T. Predicted kinase-3a motif of a resistance gene analogue as a unique marker for virus resistance. **Theoretical and Applied Genetics**, v. 99, p. 164 - 170, 1999.

STOCKINGER, E. J.; GILMOUR, S. J.; THOMASHOW, M. F. *Arabidopsis thaliana* *CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. **Proceedings National Academy Science United States America**, v. 94, p. 1035 - 1040, 1997.

THAMARUS, K. A. et al. A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre, and floral traits. **Theoretical and Applied Genetics**, v. 104, p. 379 - 387, 2002.

THORNSBERRY, J. M. et al. Dwarf polymorphisms associate with variation in flowering time. **Nature Genetics**, v. 28, p. 286 - 289, 2001.

TIAN, M. et al. A *Phytophthora infestans* cystatin like protein targets a novel tomato papain like apoplastic protease. **Plant Physiology**, v. 143, p. 364 - 77, 2007

URBAN, M., BHARGAVA, T.; HAMER, J. E. An ATP-driven efflux pump is a novel pathogenicity factor in rice blast disease. **The EMBO Journal**, v. 18, p. 512 - 521, 1999.

VANDERPLANK, J. E. **Plant Disease: Epidemics and control**. New York, Academic Press, 1968.

VALLE, L. A. C.; ALFENAS, A. C.; BROMMONSSCHENKEL, S. H. Resistência genética no controle de doenças de plantas. **Ação Ambiental**, v. 5, p. 20 - 23, 1999.

VARSHNEY, R. K.; RANER, A.; SORRELLS, M. E. Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**, v. 23, p. 48 - 55, 2005.

VEDOY, C. G.; BENGTON, M. H.; SOGAYAR, M. C. Hunting for differentially expressed genes. **Brazilian Journal of Medical and Biological Research**, v. 32, p. 877 - 884, 1999.

VELLO, F.; GARCIA, J. R. Características das principais variedades de cacau cultivadas na Bahia. **Revista Theobroma**, v.1, p. 3 - 10, 1971.

VERICA, J. A. et al. Isolation of ESTs from cacao (*Theobroma cacao* L.) leaves treated with inducers of the defense response. **Plant Cell Report**, v. 23, p. 404 - 413, 2004.

WANG, D.; WEAVER, N. D.; KESARWANI, M.; DONG, X. Induction of protein secretory pathway is required for systemic acquired resistance. **Science**, v. 13, p. 1036 - 1040, 2005.

WHEELER, B. E. J.; SUÁREZ, C. The pathosystem. In: RUDGARD, S. A.; MADDISON, A. C.; ANDEBRHAN, T. **Disease Management in Cocoa: Comparative Epidemiology of Witches' Broom**, eds. Chapman Hall, London, p. 9 - 19, 1993.

WHEELER, B. E. J. The growth of *Crinipellis pernicioso* in living and dead cocoa tissue. In: MOORE, D. et al. eds. **Developmental Biology of Higher Fungi**. Cambridge University Press, p.103 - 16, 1989.

WHEELER, B. E. J.; MEPSTED, R. Pathogenic variability among isolates of *Crinipellis pernicioso* from cocoa (*Theobroma cacao*). **Plant Pathology**, v. 30, p. 347 - 488, 1988.

YAMADA, M. M. et al. Genetic variability in cultivated cacao populations in Bahia, Brazil, detected by isozymes and RAPD markers. **Crop Breeding and Applied Biotechnology**, v. 1, p 377 - 384, 2001.

YE, S. et al. An efficient procedure for genotyping single nucleotide polymorphisms. **Nucleic Acids Research**. n. 17, v. 19, 2001.

YU, J. K. et al. EST derived SSR markers for comparative mapping in wheat and rice. **Molecular Genetics and Genomics**.v. 271, p. 742 - 751, 2004.

ZADOKS, J. C. Phytopathological aspects of disease resistance and resistance breeding in cocoa: An external review. In: **International Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement**, Salvador, Readings INGENIC, p. 17 - 22, 1999.

Z Aidan, H. A. et al. Caracterização molecular da resistencia do cacauero (*Theobroma cacao* L.) à vassoura de bruxa. In: **XXXVIII Congresso Brasileiro de Fitopatologia**, Brasilia, Brasil, 2005.

ZENG, L.. et al. Spotted leaf11, a Negative Regulator of Plant Cell Death and Defense, Encodes a U-Box/Armadillo Repeat Protein Endowed with E3 Ubiquitin Ligase Activity. **The Plant Cell**, v. 16, p. 2795 - 2802, 2004.

ZHANG, Y. et al. Interaction of NPR1 with basic leucine zipper protein transcription factors that bind sequences required for salicylic acid induction of the *PR-1* gene. **Proceedings National Academy Science United States America**, v. 96, p. 6523 - 6528, 1999.