

ÍNDICE

EXTRATO	vii
ABSTRACT	ix
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	4
2.1. Cacaueiro (<i>Theobroma cacao</i> L.).....	4
2.1.1. Abordagem Sócio-econômica.....	5
2.2. Banco Ativo de Germoplasma (BAG) Vegetal	7
2.3. Mapeamento Genético Utilizando Marcadores Moleculares	8
2.4. Populações para Mapeamento Genético	12
2.5. Análise Genômica e de Expressão gênica.....	14
2.6. Genética Clássica e Genética Reversa.....	16
2.7. Genes letais	18
2.7.1. Gene letal - seus efeitos biológicos e aplicações	18
2.7.2. Gene <i>Luteus-Pa</i>	18
3. CAPÍTULO I	21
Genetic mapping of <i>Theobroma cacao</i> (Malvaceae) seedlings of the Parinari series, carriers of the lethal gene <i>Luteus-Pa</i>	21
Abstract.....	22
3.1. Introduction	22
3.2. Material and Methods.....	24
3.2.1. Plant material and growth conditions	24
3.2.2. Extraction of DNA from <i>T. cacao</i> leaves	25
3.2.3. Amplification of DNA with PCR	26
3.2.4. Statistical analysis.....	28
3.3. Results and Discussion.....	28

3.4. References.....	32
4. CAPÍTULO II	41
Photosynthesis, chloroplast ultrastructure, chemical composition and oxidative stress in <i>Theobroma cacao</i> hybrids with the lethal gene <i>Luteus-Pa</i> mutant	41
Abstract.....	41
4.1. Introduction	42
4.2. Material and methods.....	44
4.2.1. Plant material and growth conditions	44
4.2.2. Chlorophyll Fluorescence	44
4.2.3. Leaf gas exchange.....	45
4.2.4. Growth Parameters.....	46
4.2.5. Peroxidases and polyphenol oxidases activities	47
4.2.6. Proteins and Rubisco quantification.....	47
4.2.7. Mass spectrometry analysis.....	48
4.2.8. Mineral macro and micronutrients determination	48
4.2.9. Total soluble sugars and starch	49
4.2.10. Photosynthetic pigments.....	49
4.2.11. TEM analysis	50
4.2.12. Statistical analysis.....	51
4.3. Results.....	51
4.3.1. Gas exchange and chlorophyll emission.....	51
4.3.2. Dry matter	52
4.3.3. Enzymes	52
4.3.4. Proteins.....	53
4.3.5. Mineral nutrients	54
4.3.6. Carbohydrates	54
4.3.7. Chlorophyll concentration	54
4.3.8. TEM analysis	55
4.4. Discussion.....	55
4.4.1. Gas exchange.....	56
4.4.2. Fluorescence emission	56
4.4.3. Biomass	56
4.4.4. Mineral nutrients	57
4.4.5. Peroxidases and polyphenol oxidases.....	57

4.4.6. Protein profiles	58
4.4.7. Sugars and starch.....	59
4.4.8. Chloroplasts.....	60
4.4.9. Pigments.....	60
Acknowledgements	61
4.5. References.....	62
5. CAPÍTULO III	81
Differential expression of the lethal mutant ' <i>Luteus-Pa</i> ' in cacao of the Parinari (Pa) series	81
Abstract.....	82
5.1. Introduction	82
5.2. Results	85
5.3. Discussion.....	86
5.4. Materials and Methods.....	90
5.4.1. Plant material and cultivation conditions	90
5.4.2. Identification of mutant and wild type seedlings	90
5.4.3. RNA extraction and construction of subtractive library.....	91
5.4.4 Plasmid minipreps and sequencing of the subtractive library	92
5.4.5 Real time (RT-PCR).....	93
5.4.6 Experimental design	94
Acknowledgments	95
5.5. References.....	95
Supplementary Material	106
6. CONCLUSÕES GERAIS.....	109
7. REFERÊNCIAS COMPLEMENTARES.....	111

1. INTRODUÇÃO

O sul da Bahia, por mais de 100 anos, foi considerado como uma das principais regiões produtoras de cultivo do cacauzeiro (*Theobroma cacao* L.) do país. No Brasil, esta cultura ganhou importância como alternativa para o desenvolvimento econômico de regiões como a Amazônia brasileira (DIAS, 2001). As lavouras de cacau no sul da Bahia ocupam uma área correspondente a 600 mil ha distribuídos em aproximadamente 30 mil propriedades rurais (DIAS, 2001).

O banco de germoplasma é a infra-estrutura física onde se processa a conservação *ex situ* dos acessos na forma de coleções permanentes de pólen, sementes, culturas de tecidos ou de coleções de plantas mantidas em campo (BORÉM e MIRANDA, 2005). Os bancos de germoplasmas de cacau da CEPLAC/CEPEC (BA) e CEPLAC/ Estação Experimental José Haroldo (ERJOH), Marituba (PA), representados por diferentes acessos de cacauzeiro em condições de campo (DIAS, 2001), constituem a matéria-prima necessária para a criação de novas variedades mais produtivas, mais adaptadas às regiões de cultivo e mais resistentes a pragas e doenças de importância econômica.

Os genótipos da série Parinari (Pa) são oriundos de coleções encontradas por Pound (1938), nos arredores do lugar com mesmo nome, no Peru. Esses genótipos apresentam importância agrônômica devido ao aspecto de resistência à *Phytophthora* sp, além das características de precocidade e produtividade. Os genótipos Pa exibem variabilidades expressivas para diversos caracteres, dentre os quais encontra-se a diversidade de expressão fenotípica para incompatibilidade, indicando a existência de vários alelos associados a este fator (YAMADA et al., 1994). Os membros dessa série têm um papel importante nos programas de melhoramento genético, por possuírem características desejáveis, principalmente em relação às doenças (Yamada et al., 1982). O mutante '*Luteus-Pa*', designado para um fator fisiológico que causa a morte de plântulas de cacau, aproximadamente, aos 30 a 40 dias de idade, foi identificado pela primeira vez em alguns acessos da série Pa.

Objetivou-se no presente trabalho (i) avaliar as características fisiológicas do híbrido Pa 30 x Pa 169 e seu recíproco, com suas respectivas progênies, visando elucidar os possíveis mecanismos letais do gene '*Luteus-Pa*' nos indivíduos homocigotos recessivos; (ii) mapear o gene '*Luteus-Pa*' com base em marcadores moleculares; (iii) analisar sua expressão diferenciada em plantas selvagens e mutantes, em progênies segregantes para essa característica; (iv) verificar se ocorrem alterações ultraestruturais nas plantas mutantes (especialmente na morfologia dos cloroplastos em células do mesofilo foliar), e (v) avaliar os efeitos fisiológicos e bioquímicos desse gene em plantas de cacau da série Parinari.

Espera-se que os estudos com o mutante '*Luteus - Pa*', baseados em análises moleculares e parâmetros fisiológicos, possibilitem o aumento do conhecimento sobre a genética da espécie *Theobroma cacao* L. e dêem subsídios

aos programas de melhoramento; permitindo, dessa forma, o uso desse gene como marcador genético e ferramenta em teste de paternidade e nas seleções de cacauzeiros em fazendas.

2. REVISÃO DE LITERATURA

2.1. Cacaueiro (*Theobroma cacao* L.)

A espécie *Theobroma cacao* L. (Malvaceae), planta nativa das florestas quentes e úmidas das terras baixas do México e da América Central e das bacias do rio Amazonas e Orinoco (METCALFE e CHALK, 1979), se caracteriza por ser uma espécie perene, com grande importância para as regiões produtoras de cacau do Brasil (FALEIRO et al., 2001).

O cacaueiro caracteriza-se ainda por ser uma árvore de pequeno porte, podendo atingir seis metros de altura, de copa globosa e baixa, com pequenas flores inseridas no tronco, nos ramos principais e na axila das folhas caducas, de onde também surgem frutos de tamanho e formato variáveis (LORENZI e MATOS, 2002). A espécie *T. cacao* é comercialmente explorada para a produção de sementes destinadas ao preparo de derivados e subprodutos do cacau, sobretudo na sua forma mais popular, o chocolate, podendo também ser transformado em cosméticos, bebidas finas, geléias, sorvetes e sucos.

Com base nas características morfológicas e na distribuição geográfica, a espécie *T. cacao* foi subdividida em três grupos genéticos denominados Forasteiro, Criolo e Trinitário (N'GORAN et al., 1994; ALMEIDA e VALLE, 2007; ALMEIDA e VALLE, 2009). O tipo Criolo tem sido cultivado há longo tempo na América Central e no norte da América do Sul. O primeiro Forasteiro cultivado foi do Baixo Amazonas, sendo cultivado principalmente no Brasil e na Venezuela (SOUNIGO et al., 2003).

A espécie *T. cacao* é uma planta preferencialmente alógama, sendo definida como um grupo de indivíduos que possuem um conjunto de genes que são mantidos por meio da fecundação cruzada em um mesmo local e época (BORÉM e MIRANDA, 2005; ALMEIDA e VALLE, 2007; ALMEIDA e VALLE, 2009). Segundo Borém e Miranda (2005), as plantas alógamas possuem um conjunto de genes organizados, com propriedades complexas e integradas e, devido a isso, são tratadas como população. O melhoramento de espécies alógamas baseia-se no modo de reprodução e na história evolutiva da espécie, avaliada por meio das freqüências genéticas e alélicas (Dias, 2001).

2.1.1. Abordagem Sócio-econômica

A importância sócio-econômica do cultivo de *Theobroma cacao* é muito grande para os países produtores, em particular para o Brasil (LEAL, 2004), onde cerca de três milhões de pessoas dependem direta ou indiretamente dessa cultura (FALEIRO et al., 2001). O país é o oitavo produtor mundial de *T. cacao* e tem o quinto maior parque industrial chocolateiro do mundo. A cadeia produtiva do cacau envolve, investimentos da ordem de 2 bilhões de reais, sendo 1,7 bilhão no setor primário (LEAL, 2004), se constituindo assim numa importante fonte de receitas

públicas, renda e emprego (FALEIRO et al., 2001). A região cacauera do Sul da Bahia, a maior área de produção do Brasil, é composta por 71 municípios, abrangendo 1.372.342 habitantes (IBGE, 2009).

O êxodo rural da região cacauera do Sul da Bahia para os grandes centros urbanos no Brasil deveu-se, principalmente, ao surgimento da vassoura-de-bruxa na região por volta de 1989, que contribuiu grandemente com a queda na produção de *T. cacao*, associados aos baixos preços da amêndoa praticados no mercado internacional.

Conforme dados apresentados por Zugaib et al. (2006), no Brasil, o cacau é produzido em mais de 40 mil propriedades rurais distribuídas em mais de 150 municípios, existindo cerca de cinco indústrias de processamento de cacau (Cargill, Joanes, Barry Calebaut, Nestlé e Indeca) que diferenciam seus produtos em líquido, torta, manteiga e pó, e 57 indústrias de fabricação de chocolate, sendo 19 delas de grande porte com mais de 500 empregos gerados e 38 de pequeno porte. O Sul da Bahia é a principal região produtora de cacau do Brasil, sendo responsável por 72,8% da produção Nacional, safra 2007/08 (MANDARINO e SENA GOMES, 2009). Conforme estes autores, até o ano de 1989, esta região era livre da enfermidade conhecida como “vassoura-de-bruxa”, causada pelo fungo *Moniliophthora perniciosa*, responsável por elevadas perdas em produção. Em decorrência dessa doença a produtividade média da região cacauera da Bahia passou de 650 kg de cacau seco por ha⁻¹ ano⁻¹, para cerca de 200 kg ha⁻¹ ano⁻¹, nos últimos 19 anos.

2.2. Banco Ativo de Germoplasma (BAG) Vegetal

Os recursos genéticos vegetais compreendem plantas cultivadas e espécies silvestres com valor comprovado ou mesmo potencial (BORÉM e MIRANDA, 2005). A manutenção desses recursos realiza-se por meio do estabelecimento de áreas de proteção ambientais e pela coleta e manutenção desses materiais *ex situ*, os quais passam a ser denominados bancos ativos de germoplasma (BAGs). O BAG consiste em um repositório onde se armazena a variabilidade genética de uma ou de várias espécies. Geralmente, a base física do germoplasma é mantida como centros ou instituições públicas e privadas que conservam as coleções sob a forma de sementes, explantes ou plantas em condições de campo. Além disso, existem várias formas de conservação de germoplasma vegetal. A conservação de germoplasma de espécies frutíferas, por exemplo, é quase que exclusivamente realizada em condições de campo, com poucas duplicatas de acessos de algumas espécies na forma de sementes ortodoxas e, ou cultura de meristema *in vitro*. A formação de BAGs contribui para prevenção de possíveis perdas genéticas e possibilita o estabelecimento de sítios ou áreas de coletas que contenham maior variabilidade.

A caracterização de BAGs de cacaueteiro têm sido basicamente morfoagronômica ou baseada em características econômicas. Embora essa caracterização permita uma distinção razoável entre os acessos, as comparações entre grupamentos gênicos ficam impossibilitadas. Assim, torna-se necessário à utilização de técnicas mais rápidas, eficazes e confiáveis para auxiliar na caracterização, avaliação e triagem dos genótipos nas coleções como, por exemplo, a aplicação das técnicas de marcadores moleculares (HOSBINO et al., 2003).

A avaliação e caracterização dos acessos que compõem as coleções no BAG visam ao conhecimento das potencialidades dos mesmos, portanto, são de grande

importância, e permitem, assim, a identificação da variabilidade entre e dentro das populações. Por causa da grande importância dos BAGs, um esforço internacional é necessário para a caracterização de todas as importantes coleções de germoplasma de cacaueteiro do mundo (FALEIRO et al., 2002).

O melhoramento de plantas tem sido conduzido com alguns objetivos específicos, porém sua missão é a elevação do valor econômico das espécies (BORÉM e MIRANDA, 2005). A população da série Parinari (Pa), originária do Peru e pertencente ao grupo Forastero (WILDE et al., 1992), é constituída de um grande número de genótipos amplamente aplicados em programas de melhoramento devido ao vigor, precocidade e resistência a algumas doenças (RISTERUCCI et al., 2000).

2.3. Mapeamento Genético Utilizando Marcadores Moleculares

O desenvolvimento de mapas genéticos é considerado uma das aplicações de maior impacto da tecnologia de marcadores moleculares (CARNEIRO e VIEIRA, 2002). Diversos mapas de ligação em plantas podem ser obtidos por meio de populações segregantes derivadas de cruzamentos entre linhagens puras (GRATTAPAGLIA e SEDEROFF, 1994). O mapa de ligação reproduz a ordem linear dos genes e, ou marcadores no cromossomo e quantifica o grau de ligação entre locos, além de ter uma elevada resolução, devido a uma estimativa precisa da frequência de recombinação (FALKE et al., 2006). A clonagem posicional é um método para clonagem de genes de interesse, e depende da avaliação de marcadores ligados ao mesmo (KAYE et al., 2003). Conforme Carneiro e Vieira (2002), nessa abordagem o efeito fenotípico do gene é verificado avaliando-se inicialmente a característica de interesse na progênie, e sua posição cromossômica

aproximada é determinada pelo mapeamento. Uma das aplicações mais importantes dos mapas genéticos é a localização de genes que controlam características de importância econômica (CARNEIRO e VIEIRA, 2002) ou marcadores a eles associados.

A caracterização de populações, por meio de marcadores moleculares, têm sido utilizadas principalmente na análise de coleções de germoplasma (HOSBINO et al., 2003). Nos programas de melhoramento de plantas, a identificação e a discriminação de genótipos, bem como os estudos da diversidade, constituem-se em aplicação, em curto prazo, de marcadores moleculares (FERREIRA e GRATTAPAGLIA, 1998). Toda caracterização também pode ser útil para obter informações que subsidiem a seleção de genótipos para populações de mapa. Os mais variados tipos de marcadores moleculares têm sido utilizados para caracterização de recursos genéticos de cacaueteiro, tanto quanto à variabilidade genética e filogenia de *Theobroma*, e quanto à extensão dessa variabilidade genética entre acessos selecionados (MARITA et al., 2001). Além disso, possibilita a caracterização de genes homólogos que conferem resistência a doenças (LANAUD et al., 1999; KUHN et al., 2003).

Os marcadores moleculares são necessários para dar suporte ao mapeamento físico (KAYE et al., 2003). O uso de marcadores moleculares permite o dissecamento dos componentes gênicos das características quantitativas e, conseqüentemente, o acompanhamento da segregação dos genes envolvidos na característica (BORÉM e MIRANDA, 2005). O emprego de diferentes tipos de marcadores moleculares vem sendo a estratégia mais utilizada atualmente para análises de mapeamento genético de plantas. Os marcadores têm facilitado tais estudos porque permitem, em curto espaço de tempo, analisar um número

praticamente ilimitado de marcas polimórficas em nível de DNA, sem influência do ambiente (LERCETEAU et al., 1997). O sucesso da utilização da tecnologia de marcadores depende da avaliação de um grande número de marcadores altamente polimórficos, cujos locos podem estar ligados a genes de interesse (YU et al., 2000).

Marcadores microssatélites ou SSR (*simple sequence repeat*) tem sido uma ferramenta valiosa para realização de estudos de mapeamento gênico, bem como na integração de outros mapas e para análise da variabilidade genética presente em germoplasma de espécies vegetais (KAYE et al., 2003). Este tipo de marcador molecular faz parte de uma classe de DNA repetitivo presente em todos os organismos vivos (FERREIRA e GRATTAPAGLIA, 1998). Marcadores SSR são codominantes, permitindo a distinção entre genótipos homozigóticos e heterozigóticos, e ainda parecem sofrer menor pressão de seleção, por serem localizados, em geral, em regiões não codificantes (HOSBINO et al., 2003; OLIVEIRA et al., 2006). Contudo, em determinadas espécies, a freqüência de SSR é idêntica entre regiões codificantes e não-codificantes. Devido ao alto nível de especificidade dos locos e de polimorfismo, os SSR permitem a identificação de grupos de ligação e a integração ou comparação de mapas entre várias progênes (RISTERUCCI et al., 2000). Os SSR têm muitas vantagens, entre as quais estão a grande abundância no genoma (PUGH et al., 2004), o alto polimorfismo e o multi-alelismo (KAYE et al., 2003), que são importantes no mapeamento. Estas características fazem de SSR uma opção atrativa para o aumento da densidade de mapas de ligação de cacauero, que, quando baseados em marcadores codominantes, podem ser bastante informativos (PUGH et al., 2004).

A técnica de RAPD é sensível, rápida, relativamente simples e tem sido amplamente utilizada em programas de melhoramento genético de plantas,

constituindo-se em um instrumento eficiente para estudos de mapeamento de genes de importância agrônômica (LAMES et al., 2004). Marcadores RAPD foram usados em estudos de mapeamento genético (QUEIRÓS, 1999), de fluxo gênico, na identificação de combinações de parentais heterozigóticos e na caracterização de clones de cacauzeiros para representar os três tipos de subpopulações: crioulo, forastero e trinitário (WILDE et al., 1992; DIAS, 2001). Os marcadores RAPD tiveram avanços significativos, na capacidade de geração de mapas de ligação (GRATTAPAGLIA e SEDEROFF, 1994). Esses *primers* únicos caracterizam-se por possuírem uma seqüência nucleotídica arbitrária (WILLIAMS et al., 1990). Este método tem como desvantagem ser pouco informativo por causa de sua dominância, que é marcada pelo fato de não permitir a distinção entre genótipos homozigóticos e heterozigóticos. Além disso, discute-se que há baixa repetibilidade entre locos neste procedimento (PUGH, et al., 2004). No entanto, essa limitação pode ser suplantada por transformar as marcas, que se mantém ligadas a genes de interesse, em marcadores SCAR (*Sequence-Characterized Amplified Regions*). Os marcadores do tipo SCAR são altamente específicos e não apresentam os problemas típicos dos RAPD (BORÉM e MIRANDA, 2005). Os marcadores SCAR identificam locos específicos para amplificação com pares de *primers* de oligonucleotídeos definidos, e ainda se caracterizam por apresentar polimorfismo de amplificação idêntico ao revelado por marcadores RAPD (CORRÊA et al., 2000). A principal vantagem do SCAR é a grande especificidade de pareamento com o DNA e, portanto, reprodutibilidade (BORÉM e MIRANDA, 2005).

2.4. Populações para Mapeamento Genético

A seleção de população para mapeamento envolve a escolha de genitores e a determinação do tipo de cruzamento, sendo considerada uma etapa crucial para o sucesso da construção de mapas. As características desejáveis dos genitores escolhidos devem ser contrastantes para gerar a população de mapeamento (CARNEIRO e VIEIRA, 2002). Quando o polimorfismo é baixo, podem-se realizar cruzamentos interespecíficos para obter maior número de locos informativos (CARNEIRO e VIEIRA, 2002). Os híbridos interespecíficos vêm sendo utilizados como estratégia para gerar população segregante para mapeamento, tanto em plantas como animais (OHARA et al., 2005).

As populações mais utilizadas para o mapeamento genético são provenientes de retrocruzamentos, gerações F_2 , conjuntos de linhagens puras recombinantes, e conjuntos de linhagens duplo-haplóides obtidas de gametas F_1 , de quem os genitores são linhas puras. Dessa forma, o desequilíbrio de ligação gerado pode ser visualizado nas populações segregantes, permitindo a análise de ligação (MORAES, 2005). A maioria dos experimentos de mapeamento utiliza populações oriundas do cruzamento entre linhagens endogâmicas. No entanto, em espécies frutíferas, a obtenção de tais populações é impraticável, seja pelo tempo necessário para sua obtenção ou, ainda, pela grande depressão por endogamia, que ocorre nessas espécies quando autofecundadas. Assim, a abordagem “pseudocruzamento teste” é uma alternativa para a construção de mapas genéticos em tais espécies. Tal abordagem se baseia na análise de uma população resultante do cruzamento de dois indivíduos altamente heterozigóticos e na busca de marcas informativas de locos revelados por marcadores moleculares (MORAES, 2005).

As análises de mapeamento são ferramentas indispensáveis ao programa de melhoramento do cacaueteiro, que, por meio da variabilidade genética, promoverá a criação de novos cultivares (LEAL, 2004). O cacaueteiro possui alta variabilidade genética que permite a manipulação ampla da espécie pelo melhoramento (DIAS, 2001). A caracterização e avaliação dos acessos que compõem as coleções são de fundamental importância, para se conhecer as potencialidades das mesmas e identificar a variabilidade entre e dentro as populações (LEAL, 2004).

A população de linhagens isogênicas próximas (NILs) seria como uma “biblioteca genômica com insertos”, que são fragmentos cromossômicos de genomas de parentais doadores e um vetor de um genoma parental recorrente. A qualidade da população NIL pode ser determinada por uma série de fatores, como: cobertura máxima do genoma doador; número mínimo de introgressões homozigóticas por NIL; tipo de marcadores usados, preferencialmente marcadores codominantes transferíveis como os RFLPs e os SSRs (EDUARDO et al., 2005).

A análise de *bulks* segregantes (BSA – *bulked segregant analysis*) caracteriza-se por ser um procedimento rápido para identificação efetiva de marcadores em regiões específicas do genoma. Para um ou poucos genes (oligogenes) que controla(m) um caráter, tal como resistência a doenças de natureza monogênica, os marcadores moleculares ligados podem ser identificados rapidamente pela metodologia de análise de grupos segregantes. Este método envolve a comparação de dois “*pools*” de amostras de DNA de populações segregantes originadas de cruzamentos simples e pode ser usado no preenchimento de regiões pouco saturadas (*gaps*), em mapas genéticos, e na localização de marcadores adicionais no estudo de mapeamento localizado (MICHELMORE et al., 1991). Esse método tem sido amplamente utilizado para saturar regiões genômicas