**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Desenvolvimento de marcadores microssatélites e caracterização do
germoplasma de mamona (*Ricinus communis* L.)**

Miklos Maximiliano Bajay

**Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Genética e Melhoramento de Plantas**

**Piracicaba
2009**

Miklos Maximiliano Bajay
Biólogo

**Desenvolvimento de marcadores microssatélites e caracterização
do germoplasma de mamona (*Ricinus communis* L.)**

Orientador:
Prof. Dr. JOSÉ BALDIN PINHEIRO

**Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Genética e Melhoramento de Plantas**

Piracicaba
2009

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Bajay, Miklos Maximiliano

Desenvolvimento de marcadores microssatélites e caracterização do germoplasma de mamona (*Ricinus communis* L.) / Miklos Maximiliano Bajay. - - Piracicaba, 2009.
96 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2009.
Bibliografia.

1. Diversidade genética 2. Germoplasma vegetal 3. Mamona 4. Marcador molecular
Melhoramento genético vegetal I. Título

CDD 633.85
B165d

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

AGRADECIMENTOS

Agradeço a todos aqueles que colaboraram nas diversas etapas para meu crescimento científico, à todas as pessoas e instituições que direta ou indiretamente contribuíram com seu apoio para a realização deste trabalho.

Agradeço minha família pelo apoio, pelo exemplo e pelo amor, Terezinha, Sérgio, Árpád, Stephanie e também agradeço Mariana Cisotto por seu carinho, suporte e pela ajuda essencial para a conclusão dessa dissertação.

Agradeço todos os amigos que fizeram parte dessa fase de minha formação, ao Mário Sérgio Sigrist, Carlos Eduardo Araújo Batista, aos companheiros de IAC e UNICAMP, Thiago da Matta, Milene Möller, Bruno Mulato e todo o suporte e valiosa ajuda da Marisa Monteiro, Aluana Gonçalves Abreu e em especial ao Marcelo Cavallari que me ajudou muito em diversas etapas.

Agradeço aos membros da banca de defesa pela leitura e pela contribuição crítica no aperfeiçoamento de idéias.

Agradeço também Márcia Barreto de Medeiros Nóbrega da Empresa Brasileira de Pesquisa Agropecuária - Embrapa Algodão, Tammy Aparecida Manabe Kiihl do Instituto Agronômico de Campinas – IAC e Maurício Dutra Zanotto, da Unesp Botucatu, por terem cedido os genótipos de mamona que serviram como base para o desenvolvimento da pesquisa.

Agradeço à FAPESP, pela bolsa de mestrado concedida, sem a qual não poderia ter participado dos congressos brasileiro de genética.

Agradeço aos funcionários e a todo corpo docente do Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” ESALQ-USP pela valiosa contribuição na minha formação. Em especial aos funcionários do Laboratório de Diversidade Genética e Melhoramento Domingos de Salvio Amaral e Marcio Araújo Silva.

Sou imensamente grato à Maria Imaculada Zucchi, que me acompanha desde os tempos de IAC, pela confiança e pelo incentivo ao longo dessa trajetória e José Baldin Pinheiro, meu orientador, agradeço a oportunidade e por todo seu esforço em me auxiliar para a direção dessa pesquisa.

SUMÁRIO

RESUMO.....	7
ABSTRACT.....	9
LISTA DE FIGURAS.....	11
LISTA DE TABELAS.....	13
1 INTRODUÇÃO.....	15
2 DESENVOLVIMENTO.....	17
2.1 Revisão Bibliográfica.....	17
2.1.1 Descrição geral.....	17
2.1.2 Aspectos Socioeconômicos.....	19
2.1.3 Recursos genéticos e melhoramento.....	24
2.1.4 Marcadores Moleculares: microssatélites.....	29
2.2 Material e métodos.....	33
2.2.1 Material experimental.....	33
2.2.2 Extração e quantificação de DNA.....	36
2.2.3 Construção de biblioteca genômica enriquecidas com SSR.....	38
2.2.4 Sequenciamento da região de interesse dos plasmídeos.....	44
2.2.5 Desenho dos oligonucleotídeos de DNA.....	45
2.2.6 Caracterização e genotipagem dos microssatélites.....	46
2.3 Resultados e discussão.....	51
2.3.1 Marcadores SSR desenhados.....	51
2.3.2 Condições de amplificação dos locos microssatélites.....	54
2.3.3 Análises estatístico-genéticas.....	55
3. CONCLUSÕES.....	83
REFERÊNCIAS.....	85

ANEXOS.....	93
-------------	----

RESUMO

Desenvolvimento de marcadores microssatélites e caracterização do germoplasma de mamona (*Ricinus communis* L.)

A mamona (*Ricinus communis*) é uma cultura de clima tropical de importância social e econômica no Brasil e no mundo. As sementes contêm um óleo com excelentes propriedades para o uso industrial. Esse óleo é o mais indicado para a produção de biodiesel, pois é o único solúvel em álcool e requer menos calor que os outros óleos vegetais para a produção de combustível. A grande diversidade genética observada no germoplasma de mamoneira possibilita a seleção de genótipos superiores a partir do material base, no entanto não existem informações sobre a caracterização molecular desses acessos, representando um sério risco quanto a perda de diversidade genética. O sucesso dos programas de melhoramento depende do conhecimento sobre a diversidade genética do banco ativo de germoplasma. Os microssatélites são ferramentas moleculares muito úteis em estudos de diversidade. Desta forma uma biblioteca genômica foi construída para a seleção dos fragmentos contendo marcadores moleculares microssatélites (SSR), a serem utilizados para estimar a diversidade genética dos acessos do banco de germoplasma da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), do Instituto Agronômico de Campinas (IAC) e da UNESP - Botucatu. A partir dos 41 locos analisados, 26 apresentaram polimorfismo, revelando 111 alelos (4,27 por loco). Destes, 37 alelos foram considerados raros devido à frequência genética menor do que 5%. 41 alelos foram privados, sendo observados em somente uma das três populações (germoplasma) analisadas. O valor da riqueza alélica média foi 2,554. A heterozigosidade média esperada ($H_E=0,473$) foi muito maior do que a observada ($H_O = 0,054$), indicando taxa relativamente alta de endogamia nos genótipos estudados, confirmada através do alto valor médio de G_{IS} (0,845). O valor obtido de $G_{ST'}$ (0,364) foi superior a 0,25, indicando que existem fortes indícios de estruturação na amostra. Os 26 locos apresentaram desvios na proporção do Equilíbrio de Hardy-Weinberg ($P<5\%$) depois da correção de Bonferroni. Os resultados obtidos indicam a formação de dois grupos geneticamente distintos, um grupo é composto pelos acessos da EMBRAPA e o outro grupo é composto pelos acessos do IAC e pelos genótipos da UNESP - Botucatu. Esse fato indica que a diversidade genética da mamona está bem representada nesses bancos de germoplasma. Estas informações serão de extrema importância para os programas de melhoramento da mamoneira. O conhecimento gerado nesse estudo sobre a diversidade genética da mamona pode ser utilizado para uma melhor eficiência na conservação dessa diversidade, no manejo do germoplasma, guiando programas de melhoramento genético no desenvolvimento de novos genótipos.

Palavras-chave: Marcadores microssatélites; Mamona; Diversidade genética; Germoplasma

ABSTRACT

Development and characterization of microsatellite markers for castor (*Ricinus communis* L.) germplasm

The castor bean (*Ricinus communis* L.) is an oil plant with a high economic and social value in Brazil and over the world. The seeds contain oil with excellent properties for industrial uses. It is the only vegetable oil soluble in alcohol, presenting high viscosity and requiring less heating than other oils during the production of biodiesel. Castor diversity is well represented in Brazilian germplasm collections, but there is little information on the molecular characterization of the accessions, and diversity loss is a serious risk. The success in a breeding program depends on the knowledge about the genetic diversity of the available germplasm. Microsatellite is an important tool for estimating the Genetic diversity, This study aims to develop microsatellite markers from an enriched microsatellite DNA library for castor and characterize in genotypes accessions from the castor germplasm of the Brazilian Agricultural Research Company (EMBRAPA), castor germplasm of the Agronomic Institute at Campinas (IAC) and genotypes from State University of São Paulo (UNESP) - Botucatu. From the 41 loci microsatellites analyzed, 26 showed polymorphism revealing 111 alleles independent (4,27 alleles by loco). 37 alleles have presented an inferior frequency 5%, being considered as rare alleles. 41 alleles are particular, being observed in only one of the three analyzed populations. The average allelic richness was 2,554. The average $H_O=0,054$ was smaller than the average $H_E=0,473$, evidencing a heterozygote deficit (average $G_{IS}=0,845$). The samples has shown a strong structure with a significant differentiation ($G_{ST}'= 0.364$). The Twenty six loci differ significantly from Hardy–Weinberg Equilibrium ($P<5\%$) after Bonferroni correction. The Results of STRUCTURE graphic, main component analysis and dendrograms shows the formation of two genetically distinct groups, a group is composed by the accesses of the EMBRAPA and the other group is composed by the accesses of the IAC and the genotypes of UNESP - Botucatu. This fact indicates that the genetic diversity of castor is well represented. Knowledge on the genetic diversity of castor can be used to gain a better understanding on genetic diversity conservation, and germplasm management, guiding breeding programs and conservation strategies.

Keywords: Microsatellite markers; Castor Bean; Genetic diversity; Germplasm

LISTA DE FIGURAS

Figura 1 - Evolução das exportações e importações de óleo de mamona no Brasil (2006-2008).....	21
Figura 2 - Perfil de gel de agarose 1% corados com Syber Safe, mostrando a quantificação de 63 acessos do banco de germoplasma do IAC.....	37
Figura 3 - Proporção dos motivos dos SSR encontrados na biblioteca genômica construída para <i>Ricinus communis</i>	51
Figura 4A - Perfil de gel de agarose, mostrando a amplificação de 15 locos SSR (Rco1 a Rco15).....	55
Figura 4B - Perfil de gel de agarose, mostrando a amplificação de 15 locos SSR (Rco15 a Rco30).....	55
Figura 4C - Perfil de gel de agarose, mostrando a amplificação de 11 locos SSR (Rco30 a Rco41).....	55
Figura 5 - Perfil de dois géis de acrilamida 7%, correspondentes aos loco Rco06 (sete alelos) e Rco29 (três alelos) respectivamente.....	57
Figura 6 - Histogramas das frequências alélicas dos 26 locos de SSR, estimados para cada uma das três populações de <i>Ricinus communis</i> utilizadas no estudo.....	58,59,60
Figura 7 - Número de alelos por loco polimórfico das três populações presentes no estudo.....	60
Figura 8 - Número total de alelos nas três populações presentes no estudo.....	61
Figura 9 - Números de alelos privados das três populações de <i>Ricinus communis</i> utilizadas no estudo.....	63
Figura 10 - Números de alelos raros das três populações de <i>Ricinus communis</i> utilizadas no estudo.....	63
Figura 11 - Diversidade genética por loco polimórfico das três populações presentes no estudo.....	64
Figura 12 - Dendrograma dos acessos que correspondem ao banco de germoplasma da Embrapa.....	71

Figura 13 - Dendrograma dos acessos que correspondem ao banco de germoplasma do IAC.....	72
Figura 14 - Dendrograma dos genótipos de Botucatu.....	73
Figura 15 - Dendrograma de todas as amostras agrupadas nas três populações.....	74
Figura 16 - Dendrograma de todas as amostras.....	75
Figura 17 - Análise dos componentes principais envolvendo todos os genótipos.....	76
Figura 18 - Valores de ΔK para cada valor de K, calculado nos acessos do bando de germoplasma da EMBRAPA.....	78
Figura 19 - Teste de atribuição para os acessos do banco de germoplasma da EMBRAPA de <i>Ricinus communis</i> avaliados (K=3).....	78
Figura 20 - Valores de ΔK para cada valor de K, calculado nos acessos do banco de germoplasma do IAC.....	79
Figura 21 - Teste de atribuição para os acessos do banco de germoplasma do IAC de <i>Ricinus communis</i> avaliados (K=2).....	79
Figura 22 - Valores de $\ln P(D)$ para cada valor de K, calculado nos acessos genótipos de BOTUCATU.....	80
Figura 23 - Valores de ΔK para cada valor de K, calculado em todos os genótipos.....	81
Figura 24 - Teste de atribuição para todos os genótipos de <i>Ricinus communis</i> avaliados.....	81

LISTA DE TABELAS

Tabela 1 - Identificação dos acessos provenientes do banco de germoplasma de mamona da Embrapa Algodão (CNPA)	34
Tabela 2 - Identificação dos acessos provenientes do banco de germoplasma de mamona do IAC.....	35
Tabela 3 - Identificação dos genótipos cedidos pela UNESP – Botucatu.....	36
Tabela 4 - Marcadores SSR sintetizados, com os respectivos motivos, temperaturas de anelamento, tamanhos do produto de PCR formado e as porcentagem de CG.....	52,53
Tabela 5 - Descrição dos 26 <i>primers</i> utilizados no estudo.....	56
Tabela 6 - Riqueza alélica por loco e por população.....	62
Tabela 7 - Estimativas por loco de índices relacionados à variabilidade genética dos 121 genótipos analisados distribuídos em três populações.....	66
Tabela 8 - Teste exato de aderência ao Equilíbrio de Hardy-Weinberg para cada loco dentro de cada população.....	68

1 INTRODUÇÃO

A cultura da mamona vem ganhando destaque no cenário agrícola brasileiro, proporcionando estímulos ao seu plantio e ao desenvolvimento tecnológico. O Brasil é o terceiro maior produtor de mamona que é uma cultura de clima tropical, tendo grande importância para a economia do semiárido nordestino. Entretanto, o pólo de produção de mamona tem a tendência de se deslocar para os Estados de Mato Grosso, Goiás e norte de Minas Gerais. A produção brasileira desta oleaginosa declinou nas últimas décadas e gerou a necessidade da importação de óleo bruto proveniente da China e Índia.

A agroindústria da mamona envolve uma cadeia produtiva, processando a matéria-prima e transformando-a em óleo e torta. A torta ou farelo de mamona, obtida no processamento industrial, é um produto muito importante para a aplicação direta na agricultura como adubo orgânico (SAVY FILHO, 2005).

A utilização de óleo de mamona para produção de biodiesel é uma das alternativas brasileiras para redução da importação de petróleo. O óleo de rícino é a melhor substância para produzir biodiesel porque é único que é solúvel no álcool, e requer menos calor que os outros óleos vegetais para a produção de combustível. O aumento da demanda para o óleo de mamona proporcionará o aumento das áreas agrícolas exploradas pela cultura, gerando milhares de postos de trabalho diretos e indiretos (BELTRÃO, 2003).

Tendo em vista a importância econômica da mamona para o país, ainda existem muitas características agronômicas desfavoráveis para a cultura, como baixa produtividade e suscetibilidade às principais doenças e insetos que ocorrem devido ao cultivo utilizando-se sementes dos próprios produtores, o que resulta em um alto nível de heterogeneidade (AZEVEDO ; BELTRÃO, 2007). Portanto, o fomento à ricinocultura deve ser acompanhado de investimentos em pesquisas agronômicas (FREITAS *et. al* 2005).

O sistema de reprodução da mamoneira é caracterizado pela ocorrência simultânea da autofecundação e do cruzamento natural. A autofecundação não acarreta em perda de vigor da mamoneira. Assim, no melhoramento geralmente são utilizados métodos preconizados para as plantas autógamas, mas estudos recentes estão sendo desenvolvidos com o intuito de comparar os métodos de melhoramento para alógamas e autógamas para essa espécie (AZEVEDO ; BELTRÃO, 2007).

A grande diversidade genética observada no germoplasma de mamoneira gera possibilidade para a seleção a partir do material de base, para isso é muito importante a caracterização do germoplasma. Assim, buscou-se o desenvolvimento de uma biblioteca genômica para a seleção dos fragmentos contendo marcadores moleculares microssatélites (SSR), e foram utilizados para estimar a diversidade genética entre genótipos do banco de germoplasma do Instituto Agronômico de Campinas (IAC), a Empresa Brasileira de Pesquisa Agropecuária e genótipos do programa de melhoramento da UNESP Botucatu.

A mamona foi escolhida por ser uma espécie de grande importância no Brasil para produção de biodiesel, pois seu óleo é apropriado para um amplo uso na indústria e medicina. A caracterização molecular de coleções do banco de germoplasma utilizando marcadores microssatélites é inédita no país e as informações geradas são de suma importância para a identificação, exploração e conservação da variabilidade genética dessa espécie.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

2.1.1 Descrição geral

Ricinus communis L. é uma planta oleaginosa que pertence à família das euforbiáceas e é única no gênero *Ricinus* (MOSHKIN, 1986; WEISS, 1983). Alguns autores consideram a ocorrência de subespécies, que se cruzam e produzem descendentes férteis (SAVY FILHO, 1999 a e b). A mamoneira tem ampla distribuição geográfica e ocorre espontaneamente por todo território nacional, sendo uma fornecedora de matéria-prima para a fabricação de diversos produtos. O óleo da mamoneira é utilizado em vários setores da indústria, se apresentando como um dos óleos vegetais mais versáteis encontrados na natureza.

A mamoneira é uma planta monóica, heliófila e xerófila. A maioria dos autores consideram que seu centro de origem está localizado na Etiópia e no leste da África Gurgel (1945) e Moshkin (1986), porém outros autores afirmam que a planta é originária da Índia (FONTQUER, 1979; PEREIRA, 1929; e PERES, 1997). Segundo Purseglove (1968), a mamoneira foi cultivada no Egito a mais de 4000 anos a.c.

A espécie é de clima tropical, necessitando de temperatura média de 25°C, variando entre 20 °C e 30 e altitude de pelo menos 300 m, sendo seu ótimo de 650 m. Resiste muito bem à seca, porém é não tolerante à salinidade. Necessita pelo menos de 500 mm de precipitação por ciclo, porém para atingir alta produtividade (acima de 2.000 kg/ha), precisa de uma média de 900 mm de chuva/ciclo (BELTRÃO, 2003).

O sistema radicular da planta é pivotante e profundo. Ao longo das raízes ocorre o desenvolvimento de radículas, o que aumenta a eficiência na absorção de água. As folhas da mamoneira são simples, alternas, longo pecioladas, glabras e palmatilobadas (WEISS, 1983).

A mamoneira apresenta protoginia, portanto, as anteras das flores masculinas atingem a maturação em aproximadamente cinco a dez dias depois da maturação do pistilo nas flores femininas. O pólen é expelido das flores masculinas e transportado pelo vento para inflorescências de outras plantas (cruzamento) ou para flores femininas da mesma planta (autofecundação) (BERTOZZO, 2009).

O fruto é geralmente uma cápsula deiscente contendo três sementes, a cápsula rompe espontaneamente com a maturação do fruto, liberando as sementes. Porém, atualmente já existe no mercado cultivares indeiscentes que possibilitam a colheita mecânica. As principais cultivares brasileiras que apresentam indeiscência são: Guarani, IAC-226, IAC-2028 e Al-Guarany 2002.

A inflorescência é terminal, denominada racemo (cacho), composta por uma ráquis, onde são distribuídas cimas dicásicas. O padrão de desenvolvimento da parte aérea da mamoneira é caracterizado pela emissão de ramos laterais logo após o crescimento da inflorescência primária na haste principal. Cada ramo termina em uma inflorescência (BERTOZZO, 2009).

O desenvolvimento das ramificações é o principal fator de produção, uma vez que cada ramo vai formar um racemo de mamona. O número de racemos, de frutos por racemo, o peso da semente, peso de cem sementes e o teor de óleo da semente são os componentes relacionados à produção de mamona. O florescimento da mamoneira é chamado botanicamente de simpodial, com determinado intervalo entre a emissão das inflorescências primárias e secundárias (TÁVORA, 1982).

As plantas de porte anão apresentam taxa de cruzamento de 25%, já nas demais, com ramificação aberta, essa taxa pode chegar a 40% (GURGEL, 1945; MOREIRA *et al.*, 1996).

2.1.1 Aspectos Socioeconômicos

Entre as espécies cultivadas economicamente no Brasil a mamoneira é uma das menos exigentes em termos de clima, solo e manejo cultural. Sua produção é tradicional no Nordeste brasileiro e mais de 63% da área cultivada com mamona se encontra no estado da Bahia, principalmente na região de Irecê (responsável por 80% da produção) onde os sistemas de produção utilizados ainda são precários (VIEIRA e LIMA, 1999). A cultura da mamona apresenta resistência ao clima adverso e representa uma das poucas alternativas de trabalho e renda para o agricultor do semiárido baiano. A região Nordeste brasileira dispõe de mais de 45 milhões de hectares de terras com aptidão para a exploração econômica desta cultura (NÓBREGA, 2008).

A mamona é uma cultura fortemente relacionada ao componente social, uma vez que é cultivada principalmente por pequenos produtores, sendo muito importante para a agricultura familiar e podendo apresentar economicidade elevada. Os agricultores do semiárido baiano mantêm o cultivo de mamona como um seguro em anos de seca, uma vez que as lavouras tradicionais de milho e feijão têm perdas irreversíveis em períodos de veranico superiores a 30 dias, enquanto que a mamona sofre uma menor perda (CARVALHO *et al.*, 2005). A mamoneira é cultivada em quase todos os municípios do Semi-Árido baiano (Carfanaum e Ouro-lândia são atualmente os maiores produtores), garantindo o sustento de milhares de famílias (NÓBREGA, 2008).

A cultura da mamona tem grande importância para a economia do semiárido nordestino, por sua resistência à seca, como fator fixador de mão de obra, gerador de emprego e de matéria-prima indispensável ao desenvolvimento da região e do país (AZEVEDO *et al.*, 1997). Entretanto, o polo de produção de mamona tende a se deslocar para o norte de Minas Gerais e para os Estados de Mato Grosso e Goiás.

No Brasil a ricinocultura vem há algum tempo sofrendo sérios problemas econômicos, uma vez que o preço pago ao produtor depende fortemente das oscilações do mercado internacional (VIEIRA; LIMA, 1999).

O Brasil era o primeiro produtor mundial de mamona até a metade da década de 1980, porém uma série de fatores internos e externos provocou uma redução significativa na área plantada com a cultura, razão pela qual o Brasil, atualmente, ocupa a terceira posição, atrás da Índia e da China (SANTOS *et al.*, 2001).

Na safra de 1974 foi produzido aproximadamente 573.000 t de mamona; já em 1996, a produção nacional foi de apenas 122.000 t, o que representa uma redução de 79% (ALMEIDA *et al.*, 2002; SANTOS *et al.*, 2001; VIEIRA; LIMA, 1999).

Atualmente a Índia é o principal país produtor, que contribui com 68,2%, da produção mundial de óleo. A China encontra-se em segundo lugar, sendo responsável por 14,6% da produção mundial que é totalmente destinada para o próprio consumo, e o Brasil vem em terceiro com 9,2%. Estes três países produzem 92% de toda mamona comercializada no mundo (CONAB, 2009).

A produção de mamona em bagas no Brasil na safra de 2008/2009 foi de 130,9 mil toneladas, representando acréscimo de 6,1% em relação à safra de 2007/08. Foram plantados 163,9 mil hectares de mamona no país, atingindo uma produtividade de 799 kg/ha. Em Roterdã/Holanda, o preço FOB do óleo bruto de mamona é cotado em US\$ 1.450,00/tonelada. Conforme apresentado na Figura 1, a exportação brasileira do óleo de mamona em 2008 totalizou 7.431 toneladas, representando redução de 46,02%, quando comparada com aquela efetuada em 2007 que foi 13.773,2 toneladas, enquanto a importação aumentou em 16,82%. Observa-se também que as importações estão superando as exportações desde 2008. Isso é decorrente do maior consumo de óleo no mercado interno e da baixa produção que não foi suficiente para atender a demanda nacional (CONAB 2009). Tornando necessária a importação do óleo proveniente da Índia e da China.

Figura 1 - Evolução das exportações e importações de óleo de mamona no Brasil (2006-2008)
Fonte: CONAB, 2009

A introdução da mamona no Brasil ocorreu no século XVI, durante a colonização portuguesa, por ocasião da vinda dos escravos africanos. Na era colonial, dela se extraía o óleo para a lubrificação das engrenagens e dos mancais dos engenhos de cana. Atualmente, a agroindústria da mamona envolve uma cadeia produtiva, processando a matéria-prima e transformando-a em óleo e torta. O óleo, em suas propriedades químicas originais é apropriado para grande variedade de usos de importância industrial (ATSMON *et al.*, 1989).

A torta de mamona é utilizada como adubo orgânico de alto valor econômico, apresentando propriedades nematocidas (AKTAR e MAHMOOD, 1996; SAVY FILHO, 1999b); inseticidas (CARLINI ; SÁ, 2002) encontrando aplicações na fruticultura, horticultura e floricultura (LIMA *et al.*, 2005). Quando a torta é desintoxicada, pode ser utilizada na alimentação animal, porém ainda é um processo oneroso. A haste também pode fornecer celulose para a fabricação de papel (AZZINI *et al.*, 1984).

Segundo Bertozzo (2009) o óleo de mamona, também chamado óleo de rícino, era utilizado inicialmente para geração de luz e para fins medicinais (inflamações em geral, dor de ouvido, assaduras, purgativo e unguento para as moléstias das articulações).

Dentre as aplicações modernas do óleo de mamona estão: revestimentos protetores (tintas e vernizes), plásticos, nylon, lubrificantes para aviões e naves espaciais, impermeabilizantes de superfície, fluídos hidráulicos, produtos farmacêuticos, vidros à prova de bala, cabos de fibra óptica, lentes de contato e plastificantes. Recentemente uma nova aplicação do óleo de mamona envolve a biomedicina, substituindo próteses de silicone com a vantagem de não sofrer rejeição pela biocompatibilidade do material (NOBREGA, 2008).

A demanda mundial por combustíveis de origem renovável é crescente e o Brasil tem potencial para ser um grande exportador mundial, principalmente no contexto atual de mudanças climáticas. Comparado ao óleo diesel derivado de petróleo, o biodiesel pode reduzir significativamente as emissões líquidas de gás carbônico - CO₂, um dos grandes causadores do agravamento do efeito estufa, visto que o mesmo é reabsorvido quando do crescimento das plantas que fornecerão a matéria-prima. Além disso, o biodiesel permite a redução das emissões de fumaça e praticamente elimina as emissões de óxido de enxofre (LIMA *et al.*, 2005).

Conceitualmente, o biodiesel é um combustível renovável, biodegradável e ambientalmente correto, sucedâneo ao óleo diesel mineral, constituído de uma mistura de ésteres metílicos ou etílicos de ácidos graxos obtidos da reação de transesterificação de qualquer triglicerídeo com um álcool de cadeia curta.

A produção de biodiesel a partir da mamona pode reduzir a importação de combustível fóssil, além de promover inclusão social através da agricultura familiar. O Brasil importa 16% do diesel que consome, a custos superiores a um bilhão de dólares por ano (EMBRAPA, 2009).

A mamona é considerada pelo governo como uma planta de excelente potencial e incentiva seu plantio, principalmente nas regiões mais carentes do Brasil.

O governo brasileiro tornou-se um dos maiores divulgadores e promotores dessa cultura, ao sinalizar que essa deve ser a principal oleaginosa no processo de substituição do diesel brasileiro. O objetivo do governo é realizar um programa de grande benefício social, garantindo uma fonte contínua de renda para as famílias das regiões que estejam à margem do processo de desenvolvimento econômico do país (MYCZKOWSKI, 2006).

O Programa Nacional de Biodiesel está incentivando a produção de mamona e deve promover a expansão da área de plantio, não só na região nordeste, mas também nas regiões centro-oeste, sudeste e sul do Brasil.

A utilização de óleo de mamona para produção de biodiesel é uma das alternativas brasileiras para redução da importação de petróleo, além de proporcionar o decréscimo da emissão de poluentes e gases nocivos na atmosfera. O aumento da demanda pelo óleo de mamona proporcionará o aumento das áreas agrícolas exploradas pela cultura, gerando milhares de postos de trabalho diretos e indiretos (BELTRÃO, 2003). Estudos têm demonstrado que o teor de óleo das sementes é proporcional a soma do calor recebido pela planta em todo o seu ciclo vegetativo (AZEVEDO; BELTRÃO, 2007).

O óleo de rícino é a melhor substância para a produção de biodiesel, pois é o único solúvel em álcool, além de não requerer o calor e a energia de outros óleos vegetais para transformá-los em combustível. As pesquisas para utilização do óleo de mamona como biodiesel no Brasil estão em pleno desenvolvimento, o que ampliará as possibilidades de utilização da mamona no agronegócio brasileiro (EMBRAPA, 2008). Estudos multidisciplinares recentes concluíram que a mamona constitui atualmente a cultura de sequeiro mais rentável de certas áreas do Semi-Árido nordestino (HOLANDA *et al.*, 2004).

Segundo a Resolução nº 7/2008 da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), o biodiesel derivado somente do óleo de mamona apresenta viscosidade muito alta e pode prejudicar o funcionamento de motores. No entanto, a mamona ainda permanece como opção energética, contudo o óleo deve ser misturado a outros óleos em misturas com até 30% de óleo de mamona para garantir a viscosidade determinada pelos órgãos regulatórios.

As sementes de mamona contêm aproximadamente 46% de óleo denominado óleo de rícino e conhecido internacionalmente como castor oil, sua composição é formada de 90,2% de ácidos ricinoléicos, 4,4% de ácido linoleicos e 2,2% de ácidos oleicos. A presença de grandes quantidades de ácidos graxos hidroxilados no óleo de mamona reflete em suas propriedades coligativas, resultando em alta viscosidade e alta densidade (FREITAS, 2005).

Segundo Beltrão (2003), é o único glicerídeo feito pela natureza que é solúvel em álcool e sua viscosidade é dez vezes maior que o óleo de girassol. O Biodiesel obtido de óleo de mamona apresenta um custo reduzido comparado ao biodiesel obtido de outros óleos vegetais, pois é solúvel em álcool e sua reação de transesterificação pode ocorrer em temperatura ambiente. Por não conter enxofre em sua composição, esse óleo contém um ótimo índice de cetano, que indica uma melhor qualidade de ignição, e ele tem mais oxigênio, realizando uma combustão mais completa (CONCEIÇÃO *et al.* 2007).

2.1.2 Recursos genéticos e melhoramento

Ricinus communis ($2n=20$) é explorada comercialmente como uma cultura anual que floresce de 90 a 120 dias após o plantio (BALDANZI *et al.*, 2003), entretanto é classificada pela botânica como uma planta perene, que pode chegar até cinco anos de idade (SAVY FILHO, 1999a). O tipo de inflorescência da mamoneira é panícula (cacho), com um comprimento de 15 a 50 cm, as flores estão dispostas nos racemos terminais, as flores femininas ocupam a parte superior e as masculinas a parte inferior do eixo da inflorescência. Essa arquitetura da planta possibilita a polinização controlada (AZEVEDO *et al.*, 1997; SOUSA *et al.*, 2004).

A mamoneira é explorada em dois sistemas distintos de cultivo. O cultivo isolado é o mais empregado por grandes produtores, utilizando cultivares de porte anão ou materiais híbridos; enquanto o sistema consorciado é típico do semiárido nordestino, onde predomina a utilização de cultivares de porte médio a alto (VIEIRA; LIMA, 1999).

A mamoneira apresenta baixa produtividade em solos muito pobres, porém, quando cultivada em solos extremamente férteis, favorecem o crescimento vegetativo excessivo, prolongando o início do período reprodutivo no ciclo de vida da planta. O pH do solo ótimo para o desenvolvimento da planta deve estar entre 5.5 e 6.5 (AZEVEDO *et al.*, 1997).

O plantio realizado em épocas inadequadas está entre as principais causas do baixo desempenho da mamoneira no Brasil (HEMERLY *et al.*, 1981). A época de plantio está relacionada com a distribuição e quantidade de precipitação. Em regiões de alta pluviosidade, a época de plantio deve ser ajustada de forma que não ocorram grandes volumes de precipitação nas fases de amadurecimento e secagem dos frutos (TÁVORA *et al.*, 1982). No entanto o excesso de umidade é prejudicial em qualquer período do ciclo da lavoura, sendo crítico nos estágios de plântula, maturação e colheita (AZEVEDO *et al.*, 1997).

De acordo com Carvalho (2005), para alcançar produtividades economicamente viáveis com a cultura da mamona é preciso suprir adequadamente as necessidades nutricionais da planta e realizar melhorias no sistema de produção, principalmente utilizando sementes de cultivares desenvolvidas através do melhoramento genético. Somente com esses progressos a mamona poderá se consolidar como uma das principais fontes de matéria-prima para o Programa de Biodiesel, produzindo em quantidade e qualidade suficientes para a oferta de produtos a preços que viabilizem a produção do mesmo.

No Brasil, até o momento, não foram desenvolvidas máquinas destinadas especificamente para a colheita de mamona, sendo utilizadas colheitadeiras de cereais adaptadas. Para que a colheita mecânica seja aprimorada, são necessárias informações adicionais sobre as características agrônômicas da cultura, tendo em vista que o florescimento da mamoneira ocorre sequencialmente, resultando na desuniformidade entre os diferentes racemos da planta. O período de enchimento das sementes de cada ordem de racemo ocorre sob diferentes condições ambientais, que provoca variações na contribuição dos racemos de cada ordem na produtividade total (FANAN *et al.*, 2009; KOUTROUBAS *et al.*, 2000).

Os quatro híbridos distribuídos comercialmente no Brasil apresentam as seguintes características: porte baixo, frutos indeiscentes, precocidade, maior proporção de plantas femininas e altura que pode chegar até 1,50 m quando cultivados sob irrigação (TÁVORA, 1982). Eles atualmente vêm sendo cultivados como cultura de “safrinha”, em sucessão ao milho ou à soja nos estados de Goiás e Mato Grosso (SAVY FILHO, 2005).

O primeiro programa de melhoramento genético da mamoneira desenvolvido no Brasil foi iniciado em São Paulo, pelo Instituto Agrônomo de Campinas – IAC, em 1936 (KRUG *et al.*, 1943). Naquele ano, a secção de genética deste Instituto lançou as bases de um plano de melhoramento, com o objetivo de desenvolver cultivares de mamoneiras mais produtivas, com melhores níveis de resistência às doenças, resistência a insetos e com outras características de interesse agrônomo. A partir de 1937 foram instalados vários ensaios de competição de genótipos de mamoneira de porte alto e anã, visando à identificação de cultivares mais produtivas bem como a realização de trabalhos de melhoramento (VIEIRA; LIMA, 1999).

No Brasil não se observa depressão por endogamia na mamona, possibilitando a aplicação de métodos de melhoramento de espécies autógamas.

A exploração da heterose é normalmente aplicada a plantas alógamas, no entanto também pode ser aplicado no desenvolvimento de cultivares híbridas de mamona, resultando em um modo eficaz de aumentar a produtividade (FREIRE *et al.*, 2007). Esse processo é possível devido à utilização de linhagens portadoras do alelo recessivo responsável pela diócia para flores femininas, dessa maneira o desenvolvimento de híbridos é possível e é a estratégia de melhoramento mais viável para o desenvolvimento de cultivares adaptadas a colheita mecânica (MOSHKIN; WEISS, 1983). Na Índia, os esforços em melhoramento genético das três últimas décadas resultaram no desenvolvimento de variedades *high-yielding* e de híbridos adaptados às situações agroclimáticas diversas (SUJATHA *et al.*, 1998).

O melhoramento genético já solucionou diversos problemas inerentes à cultura da mamoneira: aumento de produtividade, aumento do teor de óleo na semente, diminuição do porte da planta para facilitar a colheita mecânica, diminuição do grau de deiscência do fruto a fim de evitar o desperdício no campo e proporcionar um menor número de colheitas, além do aumento do nível de resistência a algumas das principais doenças que ocorrem no país (VIEIRA ; LIMA, 1999). Dentre as principais pragas que atingem a mamoneira podemos destacar: a podridão da raiz e do caule causadas, respectivamente, pelos fungos *Macrophomina phaseolina* e *Botryodiplodia theobromae*, a murcha-de-fusarium (*Fusarium oxysporum f. sp. ricini*) e o mofo-cinzento (*Botrytis*

ricini), a cigarrinha (*Agallia* sp. e *Empoasca* sp.), o percevejo-verde (*Nezara viridula*) e as lagartas (*Agrotis* spp. e *Thalesa citrina*) (FREIRE *et al.* 2001).

No entanto, falta de cultivares melhoradas causa várias dificuldades para a cultura da mamoneira, portanto, o melhoramento genético da cultura devem atender as necessidades da indústria, buscando o aumento da produtividade e a adequação do teor de óleo da semente a fim de satisfazer as necessidades da indústria. Também devem ser objetivos desses programas outras características como precocidade e aumento na porcentagem de flores femininas (BERTOZZO, 2009).

Atualmente temos apenas seis principais cultivares que estão sendo utilizadas no Brasil: BRS 149 Nordestina e BRS Paraguaçu, desenvolvidas pela Embrapa / EBDA para plantio em região semiárida e para uso na agricultura familiar, com plantio e colheita manual (parcelada), ciclo longo e boa tolerância à seca, sendo cultivares de porte médio, frutos semi-deiscentes e semente contendo 49% de óleo; BRS Energia foi desenvolvida em rede pela Embrapa, EBDA e Emparn e lançada em 2007 para a região Nordeste, com porte baixo e as sementes apresentam 48% de óleo; AL Guarany 2002 foi desenvolvida pela Coordenadoria de Assistência Técnica Integral do Estado de São Paulo (CATI) para as condições climáticas do Estado de São Paulo e pode ser plantada em diferentes regiões do país com características climáticas similares, com porte médio e sementes contendo 48% de óleo na semente; IAC 80 foi desenvolvida pelo Instituto Agrônomo de Campinas (IAC) para as condições climáticas do Estado de São Paulo, mas é muito utilizada em diversas regiões do país, lançada em 1982, apresenta porte entre médio e alto, com sementes contendo 47% de óleo; IAC 2028 foi lançada pelo IAC em 2007 e desenvolvida para as condições edafoclimáticas do estado de São Paulo, porte baixo, frutos indeiscentes e sementes com 47% de óleo (EMBRAPA, 2009).

Os bancos de germoplasma funcionam como um método de conservação *ex situ*, em que amostras da variabilidade genética de determinada espécie é conservada em condições controladas, fora do habitat da espécie (NASS, 2001). Estes têm como objetivo evitar a perda de alelos que diminui a variabilidade genética disponível (erosão genética) de determinada espécie, que diminui a eficiência do o processo de seleção e aumenta a vulnerabilidade a pragas e condições abióticas adversas (YORINORI *et al.*,

2001). Em geral o armazenamento das sementes é feito em condições de temperatura e umidade baixas para que a viabilidade seja preservada por décadas (BORÉM; MIRANDA, 2005).

A variabilidade genética disponível nos acessos preservados em bancos de germoplasma ainda é pouco explorada pelos melhoristas, principalmente devido à carência na caracterização desses recursos genéticos (NASS, 2001; VALLS, 2007).

A caracterização de bancos de germoplasma fornece informações fundamentais para a utilização da variabilidade genética de modo eficiente em programas de melhoramento genético e é muito importante para a conservação da diversidade genética.

Observa-se grande variabilidade no germoplasma de mamona no que diz respeito à produtividade, teor de óleo na semente, precocidade e outras características agronômicas, mas baixa ou nenhuma variabilidade quanto à resistência às principais doenças que ocorrem no país. Há, portanto, necessidade de introdução de variabilidade genética externa a fim de se conseguir fontes de resistência às principais doenças da cultura (VIEIRA; LIMA, 1999).

As principais instituições envolvidas no melhoramento genético da mamoneira atualmente são o IAC e a EMBRAPA – Algodão. A Embrapa dispõe de mais de 500 acessos em seu banco de germoplasma, que fazem parte da coleção de base da Embrapa Recursos Genéticos e Biotecnologia e representam grande parte da variabilidade genética usada atualmente nos programas de melhoramento genético da mamoneira (FREIRE *et al.*, 2007).

O IAC iniciou os projetos envolvendo a cultura da mamona em 1936 e seu banco de germoplasma constitui-se de uma coleção de trabalho formada por aproximadamente 500 progênies e linhagens, obtidas através de diversas hibridações entre variedades locais e introduzidas, conservadas em sua pureza e variabilidade genética ao longo dos últimos 70 anos (SAVY FILHO, 2007).

A grande variabilidade apresentada pela mamona é observada em características botânicas e agronômicas, podendo ser avaliada através de polimorfismo de DNA. O grupo TIGR sequenciou e reuniu aproximadamente 400Mpb do genoma da mamona através do sequenciamento por shotgun, gerando 50000 ESTs de diferentes tecidos que ajudam a descoberta e anotação de genes (CASTOR BEAN GENOME DATABASE, 2009).

Em coleções de germoplasma, a caracterização molecular é muito utilizada para complementar dados fenotípicos, pois fornece um estudo direto do genótipo e fornecem parâmetros para a identificação de genitores favoráveis à obtenção de populações segregantes, em programas de hibridação. Técnicas multivariadas envolvendo componentes principais e agrupamentos quantificam e visualizam a diversidade, identificam genótipos desejáveis e grupos de similaridade que possam se constituir duplicatas e, também aperfeiçoam seus manejos pela identificação dos caracteres mais informativos para serem empregados na caracterização e melhoramento genético (CRUZ *et al.*, 2004).

2.1.3 Marcadores Moleculares: Microssatélites

As tecnologias de análise molecular da variabilidade do DNA permitem determinar pontos de referência nos cromossomos, tecnicamente denominados “marcadores moleculares”. Marcador molecular é todo e qualquer fenótipo molecular proveniente de um gene expresso (isoenzimas), ou de um segmento específico de DNA, correspondendo a regiões expressas ou não do genoma (FERREIRA ; GRATTAPAGLIA, 1998).

Os marcadores moleculares podem ser utilizados em análise genética com diversas finalidades. No melhoramento clássico, as aplicações em curto prazo incluem a identificação e a discriminação de genótipos, a médio e longo prazo, os marcadores permitem quantificar a variabilidade genética existente ao nível de sequência de DNA e correlacioná-la com a expressão fenotípica em procedimentos de mapeamento genético. Nas aplicações sintéticas, a informação gerada na fase analítica, é integrada às metodologias de seleção e recombinação de genótipos, como uma ferramenta adicional para promover o avanço genético (FERREIRA ; GRATTAPAGLIA, 1998).

Microssatélites, ou sequências simples repetidas (SSR), são compostos por sequências de um a seis pares de base de comprimento repetidos em tandem, abundantes e dispersos por todo o genoma dos organismos eucarióticos. Como marcadores codominantes, os locos específicos são usados frequentemente para o *fingerprinting* de DNA, teste de paternidade, para a construção de mapas de ligação e os estudos genéticos da população (HAYDEN *et al.*, 2001).

O atributo principal dos locos microssatélites é seu nível elevado de diversidade alélica (MCCOUCH *et al.*, 1997). Devido a erros que podem ocorrer durante a replicação do DNA, diferentes indivíduos de uma mesma espécie podem apresentar um número variado de repetições dentro de um mesmo microssatélite, os quais constituem diferentes alelos (ALZATE-MARIN *et al.*, 2005).

Como outras classes de DNA repetitivo, os microssatélites possuem altas taxas de mutação. Elas variam entre 10^{-3} e 10^{-4} por loco por gameta por geração e o principal mecanismo que origina os alelos nesses marcadores é o deslize (“slippage”) da DNA polimerase durante a replicação. Entretanto essas sequências podem ser geradas também por recombinação, através de *crossing-over* desigual ou conversão gênica quando se trata de sequências repetidas em tandem mais longas (GOLDSTEIN, 1999). Durante a replicação de uma região repetitiva, as fitas de DNA separam-se e associam-se novamente de forma incorreta. Assim, seriam geradas cópias de trechos de DNA (alelos) com diferentes tamanhos ou números de repetições de um determinado motivo no próximo ciclo de replicação, por meio da inserção ou deleção de uma unidade de repetição (SCHLÖTTERER ; TAUTZ *et al.*, 1992).

A distribuição, a frequência e as classes de microssatélites variam entre as diferentes classes de organismos, entretanto em plantas, os dinucleotídeos mais comuns são AT/TA (SCHLÖTTERER ; TAUTZ *et al.*, 1992). As repetições dos SSR são classificadas em:

- (a) perfeitas – ATATATATATAT- quando não apresentam interrupção
- (b) imperfeitas – ATATATATGGATAT- quando a sequência é interrompida por algumas bases que não pertencem ao motivo.
- (c) compostas –ATATATATCGCGCG- quando dois ou mais tipos de repetição se apresentam contíguos (BORÉM ;CAIXETA, 2006)

Os microssatélites são marcadores ideais para análises de diversidade genética em genética de conservação e de populações (POWELL *et al.*, 1995, RAJORA *et al.*, 2000), construção de mapas genéticos de alta resolução, identificação de locos controladores de características de interesse (DEVEY *et al.*, 1996) e seleção assistida por marcadores (WELSING *et al.*, 1998; BORÉM; MIRANDA, 2005).

Esses marcadores também são utilizados na identificação de espécies e híbridos, determinação da paternidade e no monitoramento de cruzamentos controlados (VAN DE VEN; MCNLCOL, 1996).

Os marcadores SSR têm sido muito utilizados em estudos de diversidade genética em bancos de germoplasma, devido à robustez, confiabilidade, praticidade operacional e por serem marcadores mais informativos geneticamente. Para se amplificar uma região de microssatélites utiliza-se um par de *primers*, que são sequências únicas de oligonucleotídeos, homólogos às regiões franqueadoras dos microssatélites. Assim é possível obter uma amplificação específica da região que contém a sequência repetida, e o polimorfismo entre as bandas foi decorrente dos números diferentes de elementos simples repetidos. Cada segmento diferente de DNA representa um alelo daquele loco específico (ALVES *et al.*, 2002).

Os marcadores moleculares SSR apresentam características altamente desejáveis: são marcadores codominantes, ou seja, ambos os alelos de um indivíduo heterozigoto são visualizados no gel; estão ampla e uniformemente distribuídos pelo genoma dos eucariotos; são altamente multialélicos; são amplificados via PCR, o que facilita sua obtenção mesmo com poucas quantidades de DNA; uma vez desenvolvidos os *primers* que amplificam tais regiões do genoma, podem ser facilmente compartilhados entre laboratórios (SUGANUMA ; CIAMPI, 2001).

Os SSR são os marcadores que possuem o mais elevado conteúdo de informação de polimorfismo, denominado "P I C" (*Polymorphism Information Content*) (FERREIRA; GRATTAPAGLIA, 1998).

Esses marcadores são encontrados preferencialmente em regiões não codificadoras, mas, em menor frequência, também são encontrados nas regiões codificadoras (SCHLÖTTERER; TAUTZ *et al.*, 1992).

A principal limitação para o emprego da tecnologia dos microssatélites é a grande quantidade de tempo e recursos necessários para o desenvolvimento de *primers* específicos para os locos de microssatélites de cada espécie.

Entretanto, a conservação de sítios microssatélites entre espécies relacionadas ocorre frequentemente, possibilitando aproveitar *primers* (heterólogos), que foram desenvolvidos para uma espécie e que podem ser aplicados em espécies filogeneticamente próximas (FERREIRA; GRATTAPAGLIA, 1998).

A construção de bibliotecas genômicas pode ser realizada por meio de vários protocolos diferentes de enriquecimento. Dentre os quais, o protocolo desenvolvido por Billotte *et al.*, (1999) se destaca por sua eficiência e relativa baixa demanda de recursos.

Outros marcadores moleculares, no caso os dominantes: RAPD (ANTHONISEN, 2006), DAF (OLIVEIRA, 2004) e AFLP (ALLAN *et al.* 2007) também foram utilizados para caracterizar a variabilidade genética em mamona. No entanto, Allan *et al.* (2007) concluiu que os locos SSR utilizados em seu estudo foram mais polimórficos e portanto, mais informativos do que os locos AFLP utilizados no mesmo estudo, apresentando uma heterozigosidade muito superior.

2.2 Material e métodos

2.2.1 Material experimental

A Embrapa Algodão (CNPA) cedeu genótipos, citados na Tabela 1, provenientes de 38 acessos representativos de seu banco de germoplasma de mamona, que foram semeadas e numeradas de 1 a 38.

Também foram utilizados no presente estudo, 76 acessos provenientes do banco de germoplasma de mamona do Instituto Agrônômico de Campinas (IAC), (Tabela 2).

Além destes acessos também foram utilizados sete genótipos cedidos pela UNESP – Botucatu, (Tabela 3), que são utilizados no desenvolvimento de híbridos no programa de melhoramento da instituição coordenado pelo professor Mauricio Dutra Zanutto.

Todos os acessos foram plantados em triplicata procurando-se garantir a germinação de todos os genótipos e após 30 dias da semeadura foram retiradas amostras de tecido foliar jovem para a extração de DNA genômico.

No presente estudo, o termo “população” será utilizado para fazer referência a cada um dos três conjuntos de genótipos (EMBRAPA, IAC e UNESP-BOTUCATU), não tendo o sentido de população natural ou panmítica.

Tabela 1 - Identificação dos acessos provenientes do banco de germoplasma de mamona da Embrapa

Número	Identificação
1	CNPAM 2001-50
2	CSRD 20
3	CNPAM 2000-72
4	CNPAM 2000-79
5	PERNAMBUCANA 02
6	BRS NORDESTINA
7	CSRD 2
8	PAJÉ
9	CSRN 393
10	CNPAM 20001-89
11	CNPAM 20001-57
12	CNPAM 99-5
13	CSRN 379
14	CNPAM 2000-87
15	CNPAM 2000-73
16	CNPAM 2000-48
17	CNPAM 99-44
18	CNPAM 99-32
19	CNPAM 99-14
20	CNPAM 2003
21	BRA 3908
22	CNPAM 99-13
23	CSRN 05
24	BRA 10111
25	CSRN 142
26	CNPAM 99-30
27	CNPAM 99-31
28	CNPAM 2000-9
29	CNPAM 2000-20
30	CNPAM 2000-3
31	BRA 10235
32	BRA 10100
33	BRA 8869
34	CSRM 5
35	BRA 8877
36	BRA 5916
37	CSRD 268
38	CSRD 10P

Tabela 2 - Identificação dos acessos provenientes do banco de germoplasma de mamona do IAC

Número	Identificação	Número	Identificação
1	PB49II	39	PB30II
2	PB22II	40	TS10
3	PB77II	41	TS38
4	PB73II	42	PB29
5	PB09II	43	PB32II
6	PB56II	44	PB06II
7	PB44	45	IAC80
8	PB46II	46	PB36
9	PB68II	47	TS13
10	PB33II	48	PB52
11	PB75II	49	PB56
12	TS02	50	TS23
13	PB55II	51	TS30
14	PB44II	52	PB45II
15	PB28	53	PB76II
16	PB70II	54	TS09
17	PB29II	55	PB42II
18	PB04II	56	PB15II
19	TS12	57	PB34II
20	PB35	58	PB47II
21	TS15	59	IAC2028
22	IAC226	60	PB48
23	TS17	61	PB67II
24	PB26II	62	PB65II
25	PB19II	63	PB72II
26	PB07II	64	TS07
27	PB23	65	PB74II
28	TS16	66	PB53
29	TS27	67	PB43II
30	TS06	68	Guarani
31	PB32	69	PB43
32	PB59II	70	PB52II
33	TS03	71	PB20II
34	PB69II	72	PB05II
35	PB46	73	PB39
36	PB12II	74	TS04
37	PB10II	75	PB08II
38	PB38	76	PB28II

Tabela 3 - Identificação dos genótipos cedidos pela UNESP – Botucatu

Número	Identificação
1	MRZ. 03
2	Parcela 13XParcela 13 (cv. Guarani)
3	Parcela 17.1 (Cv. Guarani)
4	MRZ.02 X Parcela 49
5	China
6	Israel
7	Sara (hibrido)

2.2.2. Extração e quantificação de DNA

O DNA dos acessos foi extraído seguindo o protocolo CTAB (DOYLE ; DOYLE, 1990). Foram coletadas folhas jovens recém-expandidas, as quais foram desidratadas em estufa a 50°C por 12 horas. As folhas secas foram maceradas com a utilização de nitrogênio líquido até a obtenção de um pó muito fino, que facilita a atividade do tampão (quebra da parede celular). 30mg do material macerado foi transferido para um eppendorf de 2mL, identificados para cada amostra e adicionou-se imediatamente 800µl de tampão de extração CTAB 3% (brometo de cetiltrimetilamonio), contendo: TRIS (tris- (hidroximetil-aminometano)) - HCl 0,1 M pH 8,0; NaCl 1,2 M; CTAB 3%; EDTA (ácido etileno diamino tetracético) 30 mM pH 8,0 e mercaptoetanol a 3%. Os tubos foram agitados e incubados em banho maria durante 60 minutos a 65°C. A cada 15 minutos os tubos foram invertidos lentamente para que houvesse uma homogeneização. Durante este período o EDTA inibe a atividade enzimática das nucleases, quelando os seus cofatores, os íons de magnésio. O CTAB é o detergente que solubiliza as membranas. Após o banho maria foram adicionados 500µl de clorofórmio/álcool isoamílico (24/1), os tubos foram agitados levemente por 5 minutos e centrifugados durante 5 minutos a 14000 rpm. O sobrenadante foi transferido para novos tubos. O clorofórmio desnatura as proteínas que se precipitam pois são insolúveis na fase aquosa aonde se encontra o DNA da amostra. Foram adicionados 700µl do volume transferido de sobrenadante de isopropanol gelado, misturando levemente até formar um precipitado. Os tubos foram levados ao congelador (-18°C)

por 30 minutos e depois centrifugados a 14000 rpm por 10 minutos. O sobrenadante foi descartado e o precipitado ficou secando em temperatura ambiente por 10 minutos. Adicionou-se 500µl de etanol 70% para remover as impurezas do precipitado, os tubos foram centrifugados a 14000 rpm por 5 minutos e o etanol foi descartado. Depois o procedimento foi repetido utilizando 500µl de etanol 95%. O precipitado (*pellet*) foi deixado secar em temperatura ambiente por 2 horas para evaporar os resíduos de etanol e dissolvido em 100µl de tampão TE (10 mM Tris-HCL pH 8,0; 1 mM EDTA pH 8,0), acrescidos de 1µl de RNase (10 mg/mL) em cada tubo. Os tubos foram finalmente armazenados a temperatura de - 20°C no freezer.

Para avaliar a qualidade e concentração do DNA extraído foram utilizadas eletroforeses em géis de agarose 1% corados com Syber safe, comparando-se com o DNA-padrão do fago lambda (Figura 2). Para todas as amostras foi realizada a extração de DNA em duplicata, buscando assegurar que a quantidade do mesmo seja suficiente para todas as etapas seguintes da pesquisa. O DNA das amostras foi diluído para uma concentração final de 10ng/µl a fim de se padronizar as reações de PCR.

Figura 2 – Perfil de gel de agarose 1% corados com Syber Safe, mostrando a quantificação de 63 acessos do banco de germoplasma do IAC

2.2.3 Construção de biblioteca genômica enriquecidas com SSR

Para a caracterização da variabilidade genética na espécie foi construída uma biblioteca genômica para a seleção dos fragmentos contendo marcadores moleculares microssatélites (SSR). O protocolo utilizado para a construção da biblioteca foi descrito por Billotte et. al. (1999), com modificações, e otimizado pelo pesquisador Angie-Marie Risterruci, do CIRAD/França.

Inicialmente amostrou-se, ao acaso, um dos genótipos cujo DNA foi extraído para ser modelo da espécie, sendo esse acesso: PERNABUCANA 02. O DNA foi digerido para gerar fragmentos de tamanhos adequados. Um total de 6,5 µg de DNA foi digerido pela enzima *Rsa* I (*Invitrogen* cat. 15424-013), na concentração de 0,5 U/µL, em presença de 10% (v/v) de tampão (fornecido pelo fabricante da enzima) em água ultrapura autoclavada, em banho-maria a 37°C por 15 horas. A digestão foi verificada após eletroforese em gel de agarose 1,2% (p/v) por 1,5 h a 60 V.

O produto da digestão foi ligado aos adaptadores *Rsa*21 (5' CTCTTgCTTACgCgTggACTA 3') (IDT, *Integrated DNA Tchnologies, Inc.* ref. 15350710) e *Rsa*25 (5' pTAgTCCACgCgTAAgCAAgAgCACA 3'), a fim de garantir que todos os fragmentos tivessem uma terminação comum e conhecida. A reação foi realizada com 20% (v/v) do produto da digestão, 0,4µM de *Rsa*21, 0,4µM de *Rsa*25, 8% (v/v) de enzima T4 DNA Ligase (*Invitrogen*, cat.15224-025) e 20% (v/v) do tampão fornecido pelo fabricante da enzima, em água ultrapura autoclavada, por 20°C por 2h em termociclador (BIO-RAD My Cycler).

Após a ligação aos adaptadores, foi realizada uma etapa de pré-amplificação, visando um aumentar a quantidade de fragmentos que foram ligados aos adaptadores. A reação foi feita utilizando-se 3,0 µl de fragmentos ligados aos adaptadores, aos quais adicionaram-se 47 µl de um coquetel contendo 0,2 mM de dNTPs , 0,4 µM de *Rsa*21, 5 unidades de *Taq* DNA polimerase (*Fermentas Life Sciences*), 1,5 mM de MgCl₂ (*Fermentas Life Sciences*) e 10% (v/v) de tampão (*Fermentas Life Sciences*) contendo 750 mM de Tris-HCl (pH 8,8) e 200 mM de (NH₄)₂SO₄ em água ultrapura autoclavada.

As reações de PCR foram feitas em termociclador (BIO-RAD My Cycler) no qual se submeteu as amostras inicialmente a 95°C durante 4 minutos, depois a 20 ciclos de 94°C por 30 segundos, então a 60°C por 1 minuto e por fim a 72°C por 2 minutos. Após os 20 ciclos, as amostras foram mantidas por 8 minutos a uma temperatura de 72°C.

Os produtos da PCR foram submetidos à eletroforese em gel de agarose 1,0% (p/v) em tampão TBE 1X, contendo Syber Safe, utilizando uma voltagem de 110 V por 1,5 hora. O *Gene Ruler 100 bp DNA Ladder Plus* (Fermentas Life Sciences) foi utilizado como marcador de peso molecular. Decorrido o tempo de eletroforese, o gel foi retirado da cuba e levado para o aparelho de fotodocumentação (Major Science UVDI). Revelando-se as bandas de DNA marcado com o Syber Safe, sobre fonte de luz ultravioleta.

O produto de PCR foi purificado através do *Kit Concert Rapid PCR Purification System* (Life Technologies cat. 11458), conforme instruções do fabricante, preparando o DNA para a etapa seguinte. A purificação foi verificada em gel de agarose 1,0% (p/v) em tampão TBE 1X, contendo Syber Safe, utilizando uma voltagem de 110 V por 1 hora.

Após a purificação, foi realizada uma etapa de seleção de fragmentos contendo microssatélites. A seleção foi obtida através de *beads* magnéticas (Promega, cat. Z5481) e de sondas biotiniladas contendo repetições em *tandem* de dinucleotídeos CT (5' biotOOOOOTCTCTCTCTCTCTCTCT 3' – Nome: BiotCT) e GT (biotOOOOOGTGTGTGTGTGTGTGTGTGT 3' – Nome: BiotGT).

A solução de 600 µL de *beads* foi lavada com 300 µL de tampão SSC 0,5 X (obtido através de diluição, em água ultrapura, de tampão SSC 20X, composto por Cloreto de Sódio 3M e Citrato de Sódio 0,3M em água ultrapura, com pH igual a 7,0) e o procedimento foi repetido mais três vezes.

Paralelamente, o DNA foi preparado através de diluição de 40 µL em 460 µL de água ultrapura e posterior incubação em banho-maria a 95°C por 15 minutos. Após a incubação, a amostra recebeu 13µL de SSC 20x, 3µL do oligonucleotídeo CT e 3µL do oligonucleotídeo GT, e foi deixada em temperatura ambiente, sendo agitada suavemente por 1 minuto a cada 2 minutos, durante um tempo total de 20 minutos.

Posteriormente adicionou-se a mistura de DNA e oligonucleotídeos ao tubo das *beads*, incubado por 10 minutos a temperatura ambiente, sempre agitando gentilmente. O tubo foi colocado em uma *rack* magnética e o sobrenadante foi aspirado e descartado. As *beads* foram ressuspensas em SSC 0,1X e lavadas por quatro vezes. Finalmente foram ressuspensas em 250µL de água ultrapura, e o sobrenadante após magnetização foi transferido para um novo tubo de 1,5ml.

Os fragmentos selecionados foram então amplificados via PCR, utilizando-se 20µl desses fragmentos, aos quais adicionaram-se 80 µl de um coquetel contendo 0,2 mM de dNTPs, 0,4 µM do *primers* Rsa21, três unidades de *Taq* DNA polimerase (*Fermentas Life Sciences*), 1,5 mM de MgCl₂ (*Fermentas Life Sciences*) e 10% (v/v) de tampão (*Fermentas Life Sciences*) composto de 750 mM de Tris-HCl (pH 8,8) e 200 mM de (NH₄)₂SO₄ em água ultrapura autoclavada.

As reações de PCR foram realizadas em termociclador (BIO-RAD My Cycler), no qual foram submetidas as amostras inicialmente a 95°C por 1 minuto, e depois a 25 ciclos de amplificação. Cada ciclo submeteu as amostras a 94°C por 40 segundos, 60°C por 1 minuto e finalmente a 72°C por 2 minutos. Após os 25 ciclos, as amostras foram mantidas por 5 minutos a uma temperatura de 72°C.

Os produtos da PCR foram submetidos à eletroforese em gel de agarose 1,0% (p/v) em tampão TBE 1X, contendo Syber Safe, utilizando uma voltagem de 70 V por 1,5 hora. O *Gene Ruler 100 bp DNA Ladder Plus* (*Fermentas Life Sciences*) foi utilizado como marcador de peso molecular. Decorrido o tempo de eletroforese, o gel foi retirado da cuba e levado para o aparelho de fotodocumentação (Major Science UVDI). O gel foi fotografado sobre fonte de luz ultravioleta, revelando as bandas de DNA marcado com o Syber Safe.

Os fragmentos amplificados foram então ligados a um vetor de clonagem *pGem* – *T-Vector* (*kit pGem-T Vector System I -Promega* cat. A3600). A ligação foi feita através de reação a 4°C por 15 horas, contendo 10 µL de tampão, 1 µL de plasmídeo, 1 µL de enzima ligase (todos componentes fornecidos pelo *kit*) e 8 µL dos fragmentos.

Células competentes da cepa DH5 α de *Escherichia coli*, preparadas através do protocolo de Cloreto de Rubídio, foram utilizadas. Cem microlitros de células competentes foram adicionadas a 4 μ L de fragmentos ligados ao vetor e deixadas sobre gelo durante 20 minutos. Assim as células sofreram choque térmico, sendo mergulhadas em banho-maria a 42°C por dois minutos, sendo retiradas e colocadas no gelo por 5 minutos. Após o choque térmico, em câmara-fluxo, adicionaram-se 450 μ L de meio LB (0,1% p/v de triptona, 0,05% p/v de extrato de levedura, 0,1 % p/v de NaCl em água ultrapura, com pH igual a 7,0) autoclavado e pré-aquecido a 37°C. Então as células foram deixadas em estufa a 37 °C por 50 minutos.

Após a incubação, o tubo contendo as células foi centrifugado por 30 segundos a 13000 rpm, e 350 μ L do sobrenadante foram descartados. O volume restante de 60 μ L, incluindo o *pellet*, foram aliquotados sobre placa de Petri contendo meio LB sólido (0,1% p/v de triptona, 0,05% p/v de extrato de levedura, 0,1 % p/v de NaCl e 0,15% p/v de *Select Agar* em água ultrapura, com pH igual a 7,0) contendo 100 μ g/ml de ampicilina. Também aliquotou-se na placa de Petri 60 μ L de 20% p/v IPTG e 60 μ L de 2% p/v X-GAL. As placas de Petri foram deixadas em estufa a 37°C por 15 horas.

Depois do tempo de crescimento das colônias, as de coloração branca foram repicadas em poços de microplaca do tipo ELISA (*Costar*, cat. 3795) contendo 140 μ L de meio 2YT-HMFM (1,6% p/v de triptona, 1% p/v de extrato de levedura, 0,5% p/v de NaCl, 0,0076% p/v de MgSO₄ (7H₂O), 0,045% p/v de Citrato de Sódio, 0,09% p/v de (NH₄)₂SO₄, 4,4% p/v de glicerol, 0,18% p/v de KH₂PO₄ e 0,47% p/v de K₂HPO₄ em água ultrapura) acrescido de 100 μ g/ml de ampicilina. As placas foram mantidas em estufa a 37°C por 15 horas, fechadas com adesivo selador (AB-Gene, cat. AB0558) e transferidas para freezer -20°C.

Realizou-se então uma reação de PCR para verificar se os clones transformados apresentavam os insertos. A reação foi feita com 1 μ L do conteúdo de cada poço da placa ELISA, aos quais adicionaram-se 24 μ L de um coquetel contendo 0,4 mM de dNTPs, 0,25 μ M de Rsa21, 1 unidades de *Taq* DNA polimerase (*Fermentas Life Sciences*), 0,4 mM de MgCl₂ (*Fermentas Life Sciences*) e 10% (v/v) de tampão (*Fermentas Life Sciences*) composto de 750 mM de Tris-HCl (pH 8,8) e 200 mM de (NH₄)₂SO₄ em água ultrapura autoclavada.

A reação de PCR foi feita em termociclador, (BIO-RAD My Cycler) inicialmente a 95°C por 4 minutos, com 30 ciclos de amplificação. Cada ciclo foi de 94°C por 30 segundos, 52°C por 45 segundos e finalmente a 72°C por 1,5 minutos. Após os 30 ciclos, as amostras foram mantidas por 8 minutos a uma temperatura de 72°C.

Os produtos da PCR foram submetidos à eletroforese em gel de agarose 1,5% (p/v) em tampão TBE 1X, contendo Syber Safe, utilizando uma voltagem de 110 V por 1 hora. Decorrido o tempo de eletroforese, o gel foi retirado da cuba e levado para o aparelho de fotodocumentação (Major Science UVDI). O gel foi fotografado sobre fonte de luz ultravioleta, revelando as bandas de DNA marcado com o Syber Safe. Assim, foi pode-se identificar quais os clones que apresentavam os insertos. A presença de duas bandas no mesmo poço do gel indica que existem dois clones com insertos diferentes no mesmo poço da placa, amostras desse tipo foram descartadas.

Os clones transformados foram mantidos a -80 °C até o início da etapa seguinte, na qual os plasmídios foram extraídos das células. Para esta etapa, foram utilizadas placas de 96 poços fundos (AB-GENE cat. AB0564), foi adicionado a eles 1ml de meio *Circle Grow* (QBIO-GENE cat. 3000-142) contendo 100 µg/ml de ampicilina. Em seguida 2 µl do conteúdo de cada poço da placa ELISA contendo as colônias foram adicionados a cada poço das placas de poços fundos, onde já estava o meio *Circle Grow*. As placas foram seladas com adesivo, que foi perfurado com uma agulha sobre cada poço da placa. As placas foram incubadas em *shaker* a 370 rpm por 22 horas.

Após o período de incubação, o adesivo das placas foi substituído por um novo, e as placas foram centrifugadas a 3000 rpm por 6 minutos, para que as células aderissem ao fundo da placa. O sobrenadante foi descartado por inversão das placas. A cada poço adicionou-se 240 µl de meio GTE (2,6% (v/v) de Tris 1M pH 7,4, 2,0% (v/v) de EDTA 0,5M pH 8,0 e 4,6% (v/v) de solução de glicose a 20%, em água ultrapura autoclavada). As placas foram novamente seladas e as células foram lavadas por agitação sobre vórtex por 2 minutos. As placas foram novamente centrifugadas a 4000 rpm por 6 minutos, e o sobrenadante foi descartado por inversão. As células sofreram outra lavagem com 80 µl de meio GTE, por agitação em vórtex por 5 minutos.

Cada poço das novas placas (do tipo ELISA) recebeu 5 µl de RNase (*Ribonuclease A*, *Sigma R6513*) a uma concentração de 10 mg/ml, preparada em solução de Tris 10 mM pH 7,5 e NaCl 15 mM em água ultrapura autoclavada. Logo após, 60 µl de suspensão de células foram aliquotados de cada poço das placas de poços fundos e transferidos para cada poço das placas ELISA. A cada poço adicionaram-se 60 µl de solução NaOH 0,2 M – SDS 1%. As placas foram seladas e agitadas por inversão 10 vezes e incubadas a temperatura ambiente por 10 minutos. Nesta etapa, a parede celular e a membrana das células foram rompidas.

As placas sofreram um pulso na centrífuga para limpar o adesivo selador. O adesivo foi retirado e cada poço recebeu 60 µl de solução de KOAc 3M a 4 °C. As placas novamente foram seladas e agitadas 10 vezes por inversão e sofreram um pulso na centrífuga para limpar o adesivo selador, que foi retirado. As placas foram incubadas a 90 °C por 50 minutos. Após a incubação, as placas foram resfriadas em gelo por 10 minutos e centrifugada a 4000 rpm por 4 minutos.

Uma placa-filtro (*Millipore* cat. MAGNV 2250) foi fixada sobre uma placa ELISA, com o auxílio de elásticos, de modo que os poços das duas placas ficassem alinhados. Então todo o volume de cada poço das placas anteriores foi aliquotado e transferido para cada poço de uma placa filtro deixando apenas um precipitado branco nas placas, que foram descartadas.

Os conjuntos placa-filtro + placa ELISA foram centrifugados a 4000 rpm por 4 minutos, a 20 °C. A placa-filtro foi removida e descartada. A cada poço das placas ELISA contendo o filtrado adicionou-se 110 µl de isopropanol. As placas foram seladas e invertidas por 20 vezes, sendo então centrifugadas a 4000 rpm por 45 minutos a 20 °C.

O sobrenadante foi descartado por inversão, e 200 µl de etanol 70 % foram adicionados. As placas foram centrifugadas a 4000 rpm por 5 minutos, e o sobrenadante foi novamente descartado. As placas foram colocadas invertidas sobre papel absorvente, e nesta posição, foram centrifugadas a 600 rpm por 1 minuto. Então foram deixadas para secar sobre a bancada por 30 minutos. Decorrido o tempo de secagem, adicionou-se 60 µl de água ultrapura autoclavada. As placas foram mantidas sobre a bancada, ao abrigo da luz, por 15 horas.

A qualidade e concentração de DNA de uma amostra aleatória dos plasmídios foram verificadas em gel de agarose 1%. Utilizou-se 3 µl de plasmídios em comparação com um plasmídeo comercial (P-Gem 3Z, parte integrante do kit *Big Dye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems*). As placas contendo os plasmídeos extraídos foram congeladas a -20 °C.

2.2.4 Sequenciamento da região de interesse dos plasmídeos

Antes do sequenciamento foi realizada ainda uma etapa de purificação. A cada poço da placa foram adicionados 80µl de 65% (v/v) de isopropanol. A placa foi deixada em repouso, ao abrigo da luz por 15 minutos e centrifugada a 4000 rpm por 45 minutos. O sobrenadante foi descartado e o excesso absorvido com papel toalha. Adicionou-se 200 µl de 70% (v/v) de Etanol a cada poço. A placa foi centrifugada a 4000 rpm por 5 minutos e o sobrenadante descartado. Posteriormente a placa foi deixada à temperatura ambiente, em local protegido da luz, por uma hora, para que os poços ficassem totalmente secos. As amostras foram ressuspensas em 10 µl de Formamida ultrapura (*Hi-Di™, Applied Biosystems*). Imediatamente antes do sequenciamento as amostras foram desnaturadas em termociclador (BIO-RAD My Cycler), a 90°C por três minutos, e depois imediatamente colocadas sobre gelo por cinco minutos.

A reação de sequenciamento foi preparada em água ultrapura autoclavada com 4 µl de plasmídeo, 0,5 µM de *primer* SP6, 2 µl de *Big Dye Terminator v3.1 (Applied Biosystems)* e 2 µl de tampão *Save Money* (preparado com 10 % (v/v) de MgCl₂ 50 mM e 20 % (v/v) de Tris HCl pH 9,0 em água ultrapura autoclavada).

A reação de sequenciamento foi feita em termociclador (BIO-RAD My Cycler), no qual submeteram-se as amostras inicialmente a 96°C por 2 minutos, e, a seguir, a 25 ciclos de amplificação. Cada ciclo submeteu as amostras a 96°C por 45 segundos, 50°C por 30 segundos e por fim a 60°C por 4 minutos. A reação foi congelada a -20 °C.

O sequenciamento foi realizado em sequenciador *3730 DNA Analyzer (Applied Biosystems)* e as sequências obtidas foram processadas pelo *3730/3730xl Data Collection Software v3.0 (Applied Biosystems)*

2.2.5 Desenho dos oligonucleotídeos de DNA

As sequências obtidas foram transformadas em formato “FASTA” pelo programa BioEdit Sequence Alignment Editor e depois foram retirados os trechos relativos ao vetor e ao adaptador.

O programa Webtroll (disponível em <http://wsmartins.net/webtroll/troll.html>) permitiu a identificação de microssatélites nas seqüências obtidas. Os parâmetros utilizados para a identificação de microssatélites foram os seguintes: microssatélites compostos por motivos dinucleotídeos devem ter no mínimo sete repetições do motivo, microssatélites compostos por motivos trinucleotídeos devem ter no mínimo cinco repetições do motivo e microssatélites compostos por motivos tetranucleotídeos (ou mais) deveriam ter no mínimo três repetições do motivo.

O programa *Primer3* (disponível em http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) foi utilizado para o desenho dos *primers* que flanqueiam os microssatélites. Os parâmetros foram configurados para obtenção de *primers* cujo produto final da amplificação esteja no intervalo de 150 a 250 Pb, que é um tamanho compatível com a eletroforese em géis de poliacrilamida 7% corados com nitrato de prata; o tamanho dos *primers* deve estar entre 18 e 22 pb e a porcentagem de GC deve estar no intervalo de 40 a 60%.

O software Gene Runner v. 3.1 além de confirmar os valores para cada um dos parâmetros acima citados, o *software* também acusa a formação de estruturas secundárias, que são indesejáveis se ocorrerem em temperaturas próximas das temperaturas da reação de PCR. Desta maneira, foi possível avaliar a qualidade de cada par de *primers* obtido, selecionando os mais adequados para as reações de PCR.

O software Chromas 2 foi utilizado para avaliar a qualidade do sequenciamento nas regiões dos *primers* e dos microssatélites.

2.2.6 Caracterização e genotipagem dos microssatélites

A caracterização dos locos foi feita com 26 locos SSR polimórficos para a espécie estudada. Estes locos selecionados foram utilizados para a genotipagem dos acessos visando à caracterização do germoplasma de mamona.

A maioria dos *primers* utilizados no estudo é constituída de repetições de dinucleotídeos, como pode ser observado na Tabela 4, que quando detectados em géis de poliacrilamida de alta resolução podem apresentar “*shadow bands*” que dificultam a genotipagem, especialmente em indivíduos heterozigotos para alelos que diferem em poucos pares de bases. Com o objetivo de minimizar esses artefatos, os fragmentos foram misturados com 30% do volume de formamida e separados em gel desnaturante de poliacrilamida 7 %, contendo 5,6 M de ureia, em TBE 1X, conforme protocolo proposto por Litt *et al.* (1993).

Foram utilizadas cubas verticais de eletroforese com placas de capacidade para 100 amostras em gel de 0,7 mm de espessura. Antes da aplicação das amostras, o gel foi submetido a uma pré corrida de 60 minutos, visando obter uma temperatura razoável para a corrida dos fragmentos, que duraram em média 3 horas e meia, a 70 Watts de potência, utilizando o *ladder* de 10pb da invitrogen como marcador de peso molecular. Os géis obtidos foram corados em solução 0,2% de nitrato de prata.

A diversidade genética de um determinado loco marcador corresponde à probabilidade de ocorrência de polimorfismo entre indivíduos tomados ao acaso (RAFALSKY *et al.*, 1996) e é caracterizada por diversos índices como: heterozigosidade esperada, heterozigosidade observada e diversidade alélica.

O parâmetro H_E corresponde a probabilidade de dois gametas, tomados ao acaso de uma população, terem alelos diferentes em um determinado loco, sendo definida como a diversidade genética de Nei (1973) esperada segundo o princípio do Equilíbrio de Hardy-Weinberg (EHW). O parâmetro H_O representa a taxa real de indivíduos heterozigotos na população estudada. Assim tem-se que:

$$H_E = 1 - \sum (p_i^2)$$

$$H_O = 1 - \sum p_{ii}$$

Nas expressões descritas acima, p_i = frequência alélica estimada do i-ésimo alelo; p_{ii} = frequência estimada do homozigoto ii. A estimativa média desses valores é calculada pela média aritmética entre todos os locos analisados.

A relação entre essas estatísticas define o índice de fixação de Wright (f), que é um desvio da proporção dos indivíduos heterozigotos observados do esperado segundo o equilíbrio de Hardy-Weinberg.

O índice f é o valor individualizado do F_{IS} das estatísticas-F de Wright (WRIGHT, 1978) que é o valor médio do índice de fixação de alelos dentro de populações, e caracteriza o nível médio de endogamia ao nível populacional. As estatísticas-F permitem ainda a caracterização da distribuição da diversidade genética entre as populações (F_{ST}).

Índice de Fixação é uma estimativa do F_{IS} como uma medida do déficit ou excesso de heterozigosidade segundo WRIGHT (1978).

$$F_{IS} = (H_E - H_O) / H_E$$

$$F_{ST} = (H_T - H_E) / H_T$$

$$F_{IT} = (H_T - H_O) / H_T$$

$$1 - F_{IT} = (1 - F_{IS})(1 - F_{ST})$$

H_T é a heterozigosidade esperada na população total, assumindo que não ocorra subdivisão.

Os índices de fixação, os testes de aderência ao equilíbrio de Hardy-Weinberg e equilíbrio de ligação de locos dois a dois, foram realizados utilizando-se o programa GDA Genetic Data Analysis versão 1.0 (Lewis ; Zaykin, 2001).

H_S é a heterozigosidade esperada (não viesada) dentro da amostra, frequentemente denominada de diversidade genética.

$$H_S = \bar{n} [1 - \sum p_i^2 - H_O/2\bar{n}] / (\bar{n}-1)$$

$\bar{n}$ é a média harmônica do número de amostras.

D_{ST}' é a diversidade entre as amostras independente do número de amostras.

$$D_{ST}' = (n/np-1) H_T' - H_S$$

np é o número de amostras

H_T' é a heterozigosidade total independente do número de amostras.

$$H_T' = H_S + D_{ST}'$$

G_{ST}' é o coeficiente corrigido de diversidade relativa entre as populações (estimador de F_{ST}).

$$G_{ST}' = (H_T' - H_S) / H_T' = D_{ST}' / H_T'$$

G_{IS} é um estimador do índice de fixação F_{IS} .

$$G_{IS} = (H_S - H_O) / H_S = 1 - H_O / H_S$$

A diversidade genética e as estatísticas Nei foram estimadas sob modelo fixo. As frequências alélicas, o número de alelos por loco (A), a heterozigosidade observada (H_O) e esperada (H_E) e as estatísticas D_{ST} , H_S , H_T e G_{ST} desenvolvidas por Nei (1978), assim como a riqueza alélica foram determinadas utilizando o programa FSTAT 2.9.3 (GOUDET *et al.*, 1995).

O conteúdo de informação polimórfica (PIC – polymorphism information content), proposto por Anderson *et al.* (1993), foi calculado em função do número de alelos detectados, da sua distribuição e frequência na população estudada. A fim de se determinar os valores de cada marcador na detecção de polimorfismo entre os 121 acessos mamona analisados. Os valores de PIC por loco são determinados pela expressão:

$$PIC = 1 - \sum p_i^2$$

Nesta expressão, p_i é a frequência do alelo i na população. O cálculo é baseado no número de alelos detectados por determinado loco e a frequência relativa de cada alelo no conjunto dos 121 acessos de mamona analisados. Desta maneira o PIC está relacionado com o número de alelos, que, por sua vez, está diretamente associado à divergência genética e ao número de genótipos em estudo.

O teste exato com 10.000 permutações foi executado pelo Programa FSTAT 2.9.3 (GOUDET, 1995), para verificar significância da associação entre genótipos de cada par de locos, através da razão de verossimilhança. Foi aplicada a correção de Bonferroni ajustando-se o nível nominal de significância (5%) ao número de testes aplicados. O nível nominal de probabilidade da Tabela 8 é de 0,00064 e NA significa que o marcador SSR foi monomórfico para a respectiva população.

A estruturação da variabilidade foi visualizada através de dendrogramas no programa NTSYS (ROHLF *et al.*, 1989), construídos pela matriz de distâncias genéticas modificada por Rogers (WRIGHT 1978) e pelo critério de agrupamento UPGMA (método de média aritmético não ponderada), e para isto foi utilizado utilizando o programa TFPGA (MILLER *et al.*, 1997). A estabilidade dos agrupamentos foi testada através de 10.000 reamostragens bootstrap. O valor cofenético foi calculado através de um teste de Mantel com 1000 permutações no NTSYS com o objetivo de estimar quanto da matriz original foi explicada pelo dendrograma obtido. Foi feito também uma análise de componentes principais utilizando o mesmo programa.

Foi utilizado o programa STRUCTURE (PRITCHARD *et al.*, 2000) a fim de definir o número de grupos (K) mais provável nas amostras, através de métodos bayesianos, não necessitando de informações hierárquicas *a priori* para definir os grupos de indivíduos geneticamente relacionados com o objetivo de visualizar a estrutura genética do banco germoplasma.

A determinação do número K mais provável em relação aos propostos foi realizada utilizando valores de ΔK , segundo Evanno *et al.* (2005).

2.3 Resultados e Discussão

2.3.1 Marcadores SSR desenhados

A partir da biblioteca genômica construída foram obtidas e sequenciadas 84 colônias, sendo que, em apenas seis delas não foram encontrados microssatélites, o que representa um ótimo índice de enriquecimento das colônias (92,86%).

No total foram identificados 97 microssatélites, sendo 84 dinucleotídeos (86,6%), quatro trinucleotídeos (4,12%), nove tetranucleotídeos (9,28%), como pode ser observado na Figura 3.

Também foram encontrados na biblioteca genômica 79 SSR pentanucleotídeos e 25 hexanucleotídeos. No entanto, todos esses os motivos apresentaram apenas duas repetições e foram descartados por representar pouca probabilidade de polimorfismo. Esse resultado também observado por SIGRIST *et al.* (2009) e ocorre devido aos parâmetros de busca empregados no programa Webtroll para localizar os motivos SSR, no qual repetições de no mínimo 10 pares de bases já são consideradas como SSR.

Figura 3 - Proporção dos motivos dos SSR encontrados na biblioteca genômica construída para *Ricinus communis*

A partir da biblioteca genômica gerada, sintetizou-se 41 marcadores SSR apresentados na tabela 4, considerados superiores em relação aos marcadores SSR descartados. Os marcadores SSR sintetizados apresentam porcentagem média de GC igual a 47,56%, amplificando produtos com tamanho médio de 207,49 pb, com temperatura de anelamento média de 59,85°C na PCR, desenhados a partir de sequências com tamanho médio de 844,69 nucleotídeos. Dos 41 SSR sintetizados, 29 (70,73%) são classificados como perfeitos, oito (19,51%) são compostos e apenas quatro (9,75%) são interrompidos.

Tabela 4: Marcadores SSR sintetizados, com os respectivos motivos, temperaturas de anelamento, tamanhos do produto de PCR formado e as porcentagem de CG

(continuação)

Locos	Sequência	T °C	Motivo	Produto	%CG	Classificação
Rco01	F: GCGAAGAAACCAAAATGGAG R: CATGAAAGAAACCCAACACG	58	(CT) ₁₈	284	45	Perfeito
Rco02	F: CTAGCTTTGGGGCACAGTC R: GGAAAATAGGTGCGTATGAAAC	59	(AC) ₁₂	230	49.4	Perfeito
Rco03	F: GATGTGAGCCATTATGCTG R: TCAGAAATACCTCTAGGCGACA	62	(GA) ₁₁ (AG) ₁₁	260	47.7	Interrompido
Rco04	F: TAGTGATTCTCTTGCTTACG R: CCTTACAATGCCAGTATCAA	56	(AC) ₁₆	190	40	Perfeito
Rco05	F: AGCCCAGAAATTGGAAAAGA R: CAAACCCAAGCAAACCTCA	56	(TG) ₆ (GA) ₂₂ (GAA) ₄	270	43.7	Composto
Rco06	F: GGGTGAAAATGAAGAGATTGG R: ATAACCCGTGAAGCATGGAC	60	(TG) ₁₁	310	45	Perfeito
Rco07	F: CATGAGGCTACGGGCATAA R: ATCGTTGCAGTCATTTAGTGG	59	(TG) ₁₀	150	47.7	Perfeito
Rco08	F: CGTGTGTCTGTGTGCATGTC R: CCTCAACCCCTTGTGTTTC	61	(TG) ₁₀	320	52.5	Perfeito
Rco09	F: CCAACTCCCTTGTCTGCAA R: GTGAATGGCAAGCAGCAAT	57	(AC) ₁₁	190	50	Perfeito
Rco10	F: AGTGATTCTCTCGCTTACG R: GCCCCTTCTCCTTTGTTCT	61	(GA) ₂₂	160	48.8	Perfeito
Rco11	F: GCGTGGACTAACTTCAAGCA R: CCCCATTAGCATCGAGAAAG	60	(TC) ₁₀ (GT) ₆	250	50	Composto
Rco12	F: AAGACTGCACCTCCTCCTA R: TGCTGGAACAAACCCTGATA	58	(TG) ₈ (GA) ₉	220	48.8	Composto
Rco13	F: GGTGCTTCCAGAAATTCAGTT R: GGAGGGGAAAGACAGGATTC	61	(GA) ₂₃	250	48.9	Perfeito
Rco14	F: TTGATCCACTGATAGCCAAGA R: GAGAAGGGGTAGTGGTCGTG	63	(ATC) ₈	150	50	Perfeito
Rco15	F: CACGCACGTTAAAGCAAAC R: GCGAAGAAACCAAAATGGAG	58	(AG) ₁₈	260	45	Perfeito
Rco16	F: GGGTGATCGTCTGAGGGATA R: CACGACCATGTGGATGACTC	62	(TC) ₅ (TC) ₁₉ (CA) ₁₆	180	55	Composto/ Interrompido
Rco17	F: CTGAGTCATTGGCATCTACCAG R: GTGTGTGTGTGTGTGTGGA	65	(AG) ₈	210	51.2	Perfeito
Rco18	F: AGGGGGATAAGCGTGATATG R: CCGTTATGAAAAGGAAAGCA	58	(CA) ₁₇	250	45	Perfeito
Rco19	F: TGCTTGTAATTGCTTCTCC R: TGCAACTCAAAAAGTGAAGAG	61	(CTTT) ₃	326	41.9	Perfeito

Tabela 4: Marcadores SSR sintetizados, com os respectivos motivos, temperaturas de anelamento, tamanhos do produto de PCR formado e as porcentagem de CG

(conclusão)						
Locos	Sequência	T °C	Motivo	Produto	%CG	Classificação
Rco20	F: CCAAAGGAATGTGGGACTC R: TGTGGAGAGGATGAAGAGGAA	61	(TC) ₂₃	320	48.8	Perfeito
Rco21	F: CCTGTCCAGCAGCAAATG R: CCGTGTGAGTGGTTCCAT	57	(TG) ₈	180	54.1	Perfeito
Rco22	F: ATCCGCCGACAATAGCAG R: GCAACACTCTCTTCCCTGAA	58	(AAAC) ₃ (AC) ₉ (TC) ₅	250	52.8	Composto
Rco23	F: CATGGATGTAGAGGGTCGAT R: CAGCCAAGCCAAAGATTTC	59	(GA) ₁₅ (AG) ₈	320	47.5	Interrompido
Rco24	F: GAGAGAGAGGGCATTGAATTG R: AGCAAGCATACGCAAGACAG	61	(GCTA) ₃	170	48.8	Perfeito
Rco25	F: CCTTGCATGTTCTACAATCC R: CCCATGAAACAACACAAATCA	60	(TG) ₆ (TGTC) ₄ (CT) ₆	180	42.8	Composto
Rco26	F: TTGCTTGTCAAAGGGGAGTT R: TCATTTTGAGGGAGAAACCA	57	(CT) ₁₉	310	42.5	Perfeito
Rco27	F: TGAAGTATTAGACAACGACCA R: GCAACAACCCCACTTCTCTC	62	(AAAG) ₅	230	47.95	Perfeito
Rco28	F: TGGATTTCGAGCACTCTCT R: GCGTGGACTAACAAGATTGG	60	(TCT) ₉	150	50	Perfeito
Rco29	F: GGAGAAAAGAAAGGGAGAAGG R: GCCAAAAGCACACTTAATTTGA	61	(GA) ₇	250	42	Perfeito
Rco30	F: TGAAGTATTGGAGCTTGGAGA R: GGTCCCACACATTCATACACA	61	(AG) ₁₉	232	45.25	Perfeito
Rco31	F: ACAATGCGTGTGTCTGTGTG R: CCTCAACCCCTTGTGTTTC	60	(TG) ₁₁	240	50	Perfeito
Rco32	F: GCGTGGACTAACATTCTTTG R: TCCACCTACTATTGATGCTTGG	62	(TCTTTC) ₂ (TC) ₁₂ (TCTA) ₄ (AC) ₁₀	200	44.2	Composto
Rco33	F: ACATACATGCAGGGAGACCA R: TCTGCTTTAATGGCTGATCG	59	(TCT) ₁₁	240	47.5	Perfeito
Rco34	F: TCGGTAAAGGGTATGGGTTG R: CACACTTCAATTCGAGACC	60	(GT) ₁₁	300	50	Perfeito
Rco35	F: GGAAGAATTGGGTTGGAAGT R: AACAAACACAGGTGCATCAT	57	(AG) ₁₆	290	42.5	Perfeito
Rco36	F: CTTTAGGCACCATCATCATCC R: CATGTTGTTTTTGGCAGCTC	58	(AGATA) ₂ (TG) ₁₃ (GA) ₁₇	220	48.8	Composto
Rco37	F: GGAGGAGGAAGTAGACAAACCA R: GAGAAGGAAGGAGCCAAGTG	64	(GT) ₆	170	52.5	Perfeito
Rco38	F: TGAAGTATTAGACAACGACCA R: GCAACAACCCCACTTCTCTC	62	(AAAG) ₅	230	47.95	Perfeito
Rco39	F: TCCATTAGACAAGCATCACACA R: CAGCCAGGAGAACCAGTAGG	63	(AC) ₇	220	50.45	Perfeito
Rco40	F: AACTGGATAAAGGGGTATTTGG R: GCTTTTGGTAGCAGGTTTGA	61	(TC) ₅ (CT) ₇	200	41.9	Interrompido
Rco41	F: CATGTTGTTTTTGGCAGCTC R: CGTTCACACTCATCAATCCA	58	(CT) ₁₇ (CA) ₁₁	224	45	Composto

Todos os marcadores SSR sintetizados apresentaram amplicons, porém 15 foram monomórficos nos genótipos utilizados no estudo.

2.3.2 Condições de amplificação dos locos microssatélites

Os marcadores SSR sintetizados para o presente estudo foram inicialmente testados em diferentes temperaturas de anelamento e diferentes concentrações de DNA, utilizando apenas dois acessos do banco de germoplasma da Embrapa (CNPAM 20001-89 e BRA 10111) em géis de agarose 2% corados com Syber Safe, apresentados nas figuras 4A, 4B e 4C, para obter a melhor condição de amplificação dos fragmentos desejados. Os ciclos de amplificação foram feitos no termociclador BIO-RAD (MyCycler) e seguem o seguinte padrão: iniciando com dois minutos a 94°C, seguidos por 32 ciclos de um minuto de desnaturação a 94°C, um minuto de pareamento a 60°C (na maioria dos SSR obtidos) e um minuto de extensão a 72°C. A extensão final foi feita por 10 minutos a 72°C. O volume final das reações de PCR foi de 25 µl sendo composto por:

- 2 µl de DNA (10 ng/ µl) de cada acesso;
- 2 µl de uma solução de *primers* contendo: [forward (10 µM) + reverse (10 µM)];
- 1 µl de dNTP's (2,5mM);
- 1,5 µl de solução tampão (50mM de KCl, 10mM de TRIS-HCl – pH 8,9);
- 0,8 µl de MgCl₂ (50mM)
- 1 U de Taq DNA polimerase;
- 6,6 µl de água ultrapura autoclavada.

Foram obtidos 26 marcadores SSR polimórficos, dos quais, 12 estão disponibilizados em anexo (BAJAY *et al.*, 2009).

2.3.3 Análises estatístico-genéticas

Figura 4A: Perfil de gel de agarose 2% corado com Syber Safe, mostrando a amplificação de 15 locos SSR (Rco1 a Rco15), com apenas dois acessos (CNPAM 20001-89 e BRA 10111) por marcador SSR

Figura 4B: Perfil de gel de agarose 2% corado com Syber Safe, mostrando a amplificação de 15 locos SSR (Rco15 a Rco30), com apenas dois acessos (CNPAM 20001-89 e BRA 10111) por marcador SSR

Figura 4C: Perfil de gel de agarose 2% corado com Syber Safe, mostrando a amplificação de 11 locos SSR (Rco30 a Rco41), com apenas dois acessos (CNPAM 20001-89 e BRA 10111) por marcador SSR

Dentre os 26 SSR que foram polimórficos, destaca-se o Rco33 que foi o único com motivo trinucleotídeo. Rco1, Rco19 e Rco36 foram polimórficos apenas para os acessos do banco de germoplasma da EMBRAPA. Apenas cinco locos (19,23%) foram amplificados com temperaturas de anelamento diferente de 60°C. Rco26 foi o loco que apresentou o maior número de alelos (nove). Quatro locos tiveram apenas dois alelos e o loco que apresentou maior amplitude alélica foi Rco40 (contendo seis alelos). A Figura 5 representa um exemplo dos géis de acrilamida 7% obtidos para a genotipagem dos locos.

Tabela 5 - Descrição dos 26 *primers* utilizados no estudo

Locos	Sequência	T °C	Motivo	Número de alelos	Amplitude alélica
Rco01	F: GCGAAGAAACCAAAATGGAG R: CATGAAAGAAACCCAACACG	60	(CT) ₁₈	2	280-284
Rco02	F: CTAGCTTTGGGGCACAGTC R: GGAAATAGGTGCGTATGAAAC	60	(AC) ₁₂	5	210-232
Rco03	F: GATGTGAGCCCATATTGCTG R: TCAGAAATACCTCTAGGCGACA	60	(GA) ₁₁ (AG) ₁₁	2	260-270
Rco05	F: AGCCAGAAATTGGAAAAGA R: CAAACCCAAGCAAACCTCA	60	(TG) ₆ (GA) ₂₂ (GAA) ₄	6	264-310
Rco06	F: GGGTGAAAATGAAGAGATTGG R: ATAACCCGTGAAGCATGGAC	60	(TG) ₁₁	7	266-310
Rco08	F: CGTGTGTCTGTGTGCATGTC R: CCTCAACCCTTTGCTGTTTC	60	(TG) ₁₀	6	320-336
Rco09	F: CCAACTCCCTTGCTGCAA R: GTGAATGGCAAGCAGCAAT	60	(AC) ₁₁	5	170-194
Rco11	F: GCGTGGACTAAGCTCAAGCA R: CCCCATTAGCATCGAGAAAG	60	(TC) ₁₀ (GT) ₆	2	240-250
Rco12	F: AAGACTGCACCTCCTCCTA R: TGCTGGAACAAACCTGATA	62	(TG) ₈ (GA) ₉	3	220-230
Rco13	F: GGTGCTTCCAGAAATTCAGTT R: GGAGGGGAAAGACAGGATTG	62	(GA) ₂₃	6	226-254
Rco15	F: CACGCACGTTAAAGCAAACT R: GCGAAGAAACCAAAATGGAG	60	(AG) ₁₈	4	250-260
Rco18	F: AGGGGGATAAGCGTGATATG R: CCGTTATGAAAAGGAAAGCA	60	(CA) ₁₇	4	244-256
Rco19	F: TGCTTGTGTAATTGCTTCTCC R: TGCAACTCAAAACTGGAAGAG	60	(CTTT) ₃	3	290-326
Rco20	F: CCAAAAGGAATGTGGGACTC R: TGTGGAGAGGATGAAGAGGAA	60	(TC) ₂₃	6	300-360
Rco22	F: ATCCGCCGACAATAGCAG R: GCAACACTCTCTCCCTGAA	62	(AAAC) ₃ (AC) ₉ (TC) ₅	3	230-250
Rco23	F: CATGGATGTAGAGGGTCGAT R: CAGCCAAGCCAAAGATTTTC	62	(GA) ₁₅ (AG) ₈	4	300-320
Rco26	F: TTGCTTGTCAAAGGGGAGTT R: TCATTTTGAGGGAGAAACCA	60	(CT) ₁₉	9	280-324
Rco29	F: GGAGAAAAGAAAGGGAGAAGG R: GCCAAAAGCACACTTAATTGA	60	(GA) ₇	3	250-260
Rco30	F: TGAAACTTTGGAGCTTGGAGA R: GGTCCACACATTTCATACACA	60	(AG) ₁₉	5	220-240
Rco31	F: ACAATGCGTGTGTCTGTGTG R: CCTCAACCCTTTGCTGTTTC	60	(TG) ₁₁	3	240-260
Rco33	F: ACATACATGCAGGGAGACCA R: TCTGCTTTAATGGCTGATCG	50	(TCT) ₁₁	2	237-240
Rco34	F: TCGGTTAAGGGTATGGGTTG R: CACACTTCATTTGCAGACC	60	(GT) ₁₁	3	296-300
Rco35	F: GGAAGAATTGGGTTGGAAGT R: AACAAACACAGGTGCATCAT	60	(AG) ₁₆	4	288-298
Rco36	F: CTTTAGGCACCATCATATCC R: CATGTTGTTTTGGCAGCTC	60	(AGATA) ₂ (TG) ₁₃ (GA) ₁₇	3	220-260
Rco40	F: AACTGGATAAAGGGTATTTGG R: GCTTTTTGGTAGCAGTTTGA	60	(TC) ₅ (CT) ₇	6	200-280
Rco41	F: CATGTTGTTTTGGCAGCTC R: CGTTCACACTCATCAATCCA	60	(CT) ₁₇ (CA) ₁₁	5	224-254

Figura 5 – Perfil de dois géis de acrilamida 7%, correspondentes aos loco Rco06 (sete alelos) e Rco29 (três alelos) respectivamente

A distribuição das frequências alélicas para os 26 locos SSR comparando os genótipos das 3 populações de mamona analisadas nesta pesquisa encontra-se na Figura 6.

As oscilações nas frequências alélicas, perdas e fixação de alelos, podem sugerir devido à ação da deriva genética ou da seleção. No caso de marcadores moleculares neutros como SSR, essas oscilações se desenvolvem através de processos aleatórios. Dessa maneira, a análise das frequências alélicas é de grande importância, pois pode refletir melhor os efeitos aleatórios do que a maioria dos parâmetros utilizados no estudo de genética de populações, uma vez que tais parâmetros não refletem diretamente frequências alélicas eventualmente muito baixas (OLIVEIRA *et al.*, 2002)

As análises realizadas com os 26 locos SSR para o banco de germoplasma da EMBRAPA e do IAC e para a população de híbridos de BOTUCATU indicaram um baixo polimorfismo, foi observado um total de 111 alelos (4,27 alelos/ loco). Dos quais 37 alelos (33,33%), apresentaram uma frequência inferior a 5 %, sendo considerados alelos raros. 41 alelos (36,93%) são privados, sendo observados em somente uma das três populações analisadas. A riqueza alélica média foi de 2,554.

Figura 6 - Histograma das frequências alélicas dos 26 locos de SSR, estimados para cada uma das três populações de *Ricinus communis* utilizadas no estudo. O eixo Y indica a frequência alélica e o eixo X indica o tamanho do alelo (continuação)

Figura 6 - Histograma das frequências alélicas dos 26 locos de SSR, estimados para cada uma das três populações de *Ricinus communis* utilizadas no estudo. O eixo Y indica a frequência alélica e o eixo X indica o tamanho do alelo (continuação)

Figura 6 - Histograma das frequências alélicas dos 26 locos de SSR, estimados para cada uma das três populações de *Ricinus communis* utilizadas no estudo. O eixo Y indica a frequência alélica e o eixo X indica o tamanho do alelo (conclusão)

Figura 7 - Número de alelos por loco polimórfico das três populações presentes no estudo

Como pode ser observado na Figura 7, o número máximo de alelos por população (seis) foi encontrado no germoplasma da Embrapa. O marcador Rco41 apresenta apenas três alelos nas três populações, assim como os marcadores Rco11, Rco12, Rco22 e Rco33 apresentam apenas dois alelos nas três populações. Rco01, Rco19 e Rco31 foram polimórficos apenas nos acessos do banco de germoplasma da EMBRAPA.

Figura 8 - Número total de alelos nas três populações presentes no estudo

De acordo com a Figura 8, os acessos da EMBRAPA e do IAC apresentam um número de alelos total de alelos semelhante, porém, depois da correção levando em conta o tamanho amostral (Tabela 6), o banco de germoplasma da EMBRAPA obteve uma riqueza alélica significativamente maior do que o banco do IAC. Os sete genótipos de Botucatu apresentaram menor quantidade de alelos e menor riqueza alélica, como já era esperado visto que a população é composta por apenas sete genótipos.

A população de 38 acessos que representa o banco de germoplasma da Embrapa apresentou maior número de alelos (82), a maior quantidade de alelos raros (23) e privados (29), como observado nas Figuras 9 e 10, assim como a maior riqueza alélica (2,256) e a maior diversidade genética entre as três populações (Figura 11).

Para os 76 acessos banco de germoplasma do IAC, apesar da maior quantidade de genótipos analisados, observou-se um número pouco menor de alelos (78), apresentando valores intermediários de alelos raros (18) e privados (9), riqueza alélica de 2,256.

Os genótipos de Botucatu apresentaram o menor número de alelos (52), a menor quantidade de alelos raros (7) e privados (3), também a menor riqueza alélica, esses resultados são esperados devido ao pequeno tamanho amostral ($n=7$). Esses genótipos também mostraram maior variação na diversidade genética através dos 26 locos analisados.

Tabela 6 - Riqueza alélica por loco e por população

	EMBRAPA	IAC	BOTUCATU	MÉDIA
Rco01	1,499	1,000	1,000	1,184
Rco02	1,767	1,423	1,000	2,450
Rco03	1,608	1,604	1,000	1,970
Rco05	3,307	2,406	2,000	2,810
Rco06	2,770	2,767	4,476	4,087
Rco08	2,265	2,595	1,835	3,601
Rco09	1,821	3,534	1,000	3,476
Rco11	1,844	1,895	2,000	1,990
Rco12	1,945	1,978	1,835	2,841
Rco13	3,448	2,042	2,956	3,137
Rco15	2,491	2,746	2,832	2,720
Rco18	2,928	2,228	2,670	2,737
Rco19	1,896	1,000	1,000	1,983
Rco20	2,970	1,843	2,000	2,565
Rco22	1,795	1,959	1,835	2,766
Rco23	2,778	2,051	2,000	2,851
Rco26	2,853	2,592	2,970	3,822
Rco29	1,765	1,694	2,000	1,883
Rco30	3,457	2,737	1,835	3,113
Rco31	1,695	1,986	1,000	2,869
Rco33	1,695	1,547	2,000	1,684
Rco34	1,945	1,882	1,000	2,091
Rco35	2,962	2,193	1,909	2,636
Rco36	1,801	1,000	1,000	1,323
Rco40	1,499	1,405	1,835	1,512
Rco41	1,855	2,099	2,820	2,307
MÉDIA	2,256	2,008	1,916	2,554

Figura 9 - Números de alelos privados das três populações de *Ricinus communis* utilizadas no estudo a partir dos 26 locos de SSR, estimados para as três populações de *Ricinus communis* utilizadas no estudo

Figura 10 - Números de alelos raros das três populações de *Ricinus communis* utilizadas no estudo a partir dos 26 locos de SSR, estimados para as três populações de *Ricinus communis* utilizadas no estudo

Figura 11 - Diversidade genética por loco polimórfico das três populações presentes no estudo

Segundo a classificação de PIC de Botstein *et al.* (1980), os marcadores SSR desenvolvidos para o presente estudo indicam uma média de 0.4335 para valor de PIC, sendo considerado um valor moderadamente informativo ($0,25 < \text{PIC} < 0,5$). O *primer* Rco36 apresentou o menor valor de PIC (0,0832) enquanto que o *primer* Rco06 foi o mais polimórfico, com um valor de PIC igual a 0.7226 (Tabela 7).

Os valores de H_T' (heterozigosidade total) e de D_{ST}' (diversidade entre as amostras) são estimativas independentes do número de amostras e no presente estudo foram encontrados valores considerados relativamente altos: 0,581 e 0,211 respectivamente (Tabela 7). Desta maneira, os resultados indicam que os marcadores SSR desenvolvidos nesse estudo são informativos e existe diversidade genética significativa nas amostras analisadas.

Com base nos resultados, foram observados 10 alelos exclusivos (presentes em apenas um genótipo). Os acessos da Embrapa apresentaram sete desses alelos. O marcador Rco29 foi responsável por dois desses alelos. O marcador Rco40 teve dois alelos exclusivos nos acessos do IAC (PB56II e PB44II) e apenas um nos genótipos de Botucatu (MRZ. 03).

Entre os 26 locos polimórficos, podemos destacar dez deles como os mais informativos em análises de diversidade genética em mamona. Na Figura 5 observa-se que os locos Rco02, Rco06 e Rco09 são muito úteis em estudos de diversidade em mamona, pois apresentam seus alelos de maior frequência localizados em genótipos que pertencem a populações distintas e geneticamente distantes. Os locos Rco08, Rco12 e Rco26 também são informativos, pois apresentam valores relativamente altos de PIC (Tabela 7). Rco05, Rco13 e Rco30 apresentam valores relativamente elevados de riqueza alélica. Rco15 é um dos poucos marcadores SSR desenvolvidos nesse estudo que apresenta heterozigosidade observada destacada.

É importante salientar que os índices de Nei são mais adequadas aos estudos com bancos de germoplasma pois não tem os pressupostos de ancestralidade comum.

Tabelas 7 - Estimativas por loco de índices relacionados à variabilidade genética dos 121 genótipos analisados distribuídos em três populações

Locos	PIC	H_e	H_o	F_{IS}	F_{IT}	F_{ST}	H_S	D_{ST}'	H_T'	G_{ST}'	G_{IS}
Rco01	0,048	0,049	0,000	0,851	0,909	0,389	0,052	0,001	0,053	0,019	1,000
Rco02	0,431	0,500	0,025	0,852	0,906	0,362	0,109	0,577	0,686	0,842	0,798
Rco03	0,352	0,457	0,000	0,849	0,905	0,371	0,134	0,469	0,604	0,777	1,000
Rco05	0,463	0,475	0,109	0,856	0,913	0,400	0,452	0,001	0,453	0,001	0,668
Rco06	0,723	0,758	0,079	0,851	0,909	0,390	0,682	0,227	0,909	0,249	0,815
Rco08	0,671	0,717	0,033	0,847	0,906	0,382	0,417	0,396	0,813	0,487	0,937
Rco09	0,667	0,723	0,232	0,869	0,920	0,388	0,353	0,483	0,836	0,578	0,591
Rco11	0,371	0,494	0,060	0,853	0,909	0,387	0,418	0,135	0,553	0,243	0,448
Rco12	0,575	0,653	0,085	0,854	0,910	0,383	0,396	0,350	0,746	0,469	0,842
Rco13	0,532	0,564	0,052	0,850	0,909	0,391	0,573	0,272	0,845	0,322	0,928
Rco15	0,519	0,602	0,240	0,869	0,921	0,402	0,597	0,006	0,603	0,009	0,570
Rco18	0,524	0,609	0,000	0,843	0,905	0,396	0,532	0,121	0,654	0,186	1,000
Rco19	0,091	0,095	0,017	0,852	0,906	0,365	0,096	0,483	0,579	0,834	0,818
Rco20	0,393	0,412	0,054	0,852	0,910	0,396	0,468	0,084	0,553	0,152	0,928
Rco22	0,539	0,611	0,066	0,853	0,909	0,380	0,341	0,362	0,704	0,515	0,859
Rco23	0,506	0,556	0,017	0,845	0,907	0,398	0,547	0,068	0,615	0,110	0,968
Rco26	0,668	0,705	0,034	0,848	0,906	0,383	0,560	0,395	0,956	0,414	0,953
Rco29	0,237	0,253	0,000	0,848	0,908	0,393	0,337	0,054	0,391	0,137	1,000
Rco30	0,527	0,557	0,034	0,847	0,908	0,401	0,496	0,030	0,526	0,057	0,927
Rco31	0,582	0,660	0,009	0,847	0,904	0,376	0,252	0,531	0,783	0,678	0,978
Rco33	0,203	0,230	0,017	0,850	0,909	0,392	0,311	0,126	0,437	0,287	0,694
Rco34	0,298	0,326	0,157	0,864	0,917	0,394	0,237	0,039	0,276	0,142	0,620
Rco35	0,485	0,561	0,000	0,843	0,906	0,401	0,504	0,024	0,528	0,046	1,000
Rco36	0,083	0,087	0,008	0,851	0,909	0,390	0,089	0,008	0,096	0,080	0,901
Rco40	0,124	0,129	0,050	0,855	0,912	0,391	0,175	0,002	0,176	0,009	0,845
Rco41	0,426	0,524	0,034	0,848	0,908	0,393	0,474	0,250	0,724	0,345	0,953
Média	0,434	0,473	0,054	0,851	0,909	0,389	0,369	0,211	0,581	0,364	0,845

As heterozigosidades médias esperadas H_E (0,473) e H_S (0,369) foram muito maiores do que a heterozigosidade observada H_O (0,054), indicando taxa relativamente alta de endogamia nas três populações estudadas, confirmadas através do alto valor médio de F_{IS} (0,851) e G_{IS} (0,845). A população dos acessos da EMBRAPA apresentou o maior F_{IS} (0,906), seguida da população dos acessos do IAC (0,822) e finalmente Botucatu (0,796). O valor médio obtido de F_{ST} (0,389) e de G_{ST}' (0,364) foi superior a 0,25, indicando que existem fortes indícios de estruturação na amostra total.

O Teste de aderência ao Equilíbrio de Hardy-Weinberg apresentado na Tabela 8 detectou que ocorreu, em média, desvios na proporção do Equilíbrio de Hardy-Weinberg em todos os locos e nas três populações definidas. Esse resultado era esperado uma vez que o Equilíbrio de Hardy-Weinberg só é atingido em populações panmíticas. Além disso, foi observado em estudos anteriores (GURGEL 1945) que a mamona apresenta sistema reprodutivo misto, com predominância de autogamia, o que explica o excesso de homozigotos.

Os genótipos de Botucatu apresentaram maior quantidade de locos heterozigóticos observados do que as outras amostras, portanto, a maioria dos marcadores SSR acusou, individualmente, que esses genótipos aderem às proporções do Equilíbrio de Hardy-Weinberg. Inclusive o loco Rco5 estima um F_{IS} igual à zero nessa população. No entanto, o pequeno tamanho amostral dessa população não permite inferir informações precisas sobre esse tipo de análise. O loco Rco09 na população da EMBRAPA e os locos Rco34 e Rco40 na população do IAC também acusaram individualmente significância a 5% às proporções do Equilíbrio de Hardy-Weinberg para suas respectivas populações.

Segundo Mota (2003), o teste aplicado é bastante conservador, em decorrência da correção de Bonferroni para testes múltiplos, que consiste em dividir o nível nominal de probabilidade do teste (5%) pelo número de testes (n° amostras x n° locos). Assim, o nível nominal ajustado probabilidade assume valores muito baixos para a rejeição da hipótese de nulidade. De qualquer modo, os resultados do teste exato mostram-se mais confiáveis que testes baseados em X^2 ou razão de verossimilhança (G2-valor), quando a análise envolve amostras pequenas e locos multialélicos, onde é muito comum a ocorrência de classes genotípicas de baixa frequência.

Tabela 8 - Teste exato de aderência ao Equilíbrio de Hardy-Weinberg para cada loco dentro de cada população

EHW	EMBRAPA	IAC	BOTUCATU
Rco01	0.0006*	NA	NA
Rco02	0.0006*	0.0006*	NA
Rco03	0.0006*	0.0006*	NA
Rco05	0.0006*	0.0006*	1.0000 ^{ns}
Rco06	0.0006*	0.0006*	0.0019 ^{ns}
Rco08	0.0006*	0.0006*	0.0756 ^{ns}
Rco09	0.0013 ^{ns}	0.0006*	NA
Rco11	0.0006*	0.0006*	0.7583 ^{ns}
Rco12	0.0006*	0.0006*	0.0712 ^{ns}
Rco13	0.0006*	0.0006*	0.0147 ^{ns}
Rco15	0.0006*	0.0006*	0.1885 ^{ns}
Rco18	0.0006*	0.0006*	0.0064 ^{ns}
Rco19	0.0006*	NA	NA
Rco20	0.0006*	0.0006*	0.0205 ^{ns}
Rco22	0.0006*	0.0006*	0.0744 ^{ns}
Rco23	0.0006*	0.0006*	0.0096 ^{ns}
Rco26	0.0006*	0.0006*	0.0019 ^{ns}
Rco29	0.0006*	0.0006*	0.0115 ^{ns}
Rco30	0.0006*	0.0006*	0.0744 ^{ns}
Rco31	0.0006*	0.0006*	NA
Rco33	0.0006*	0.0006*	0.2872 ^{ns}
Rco34	0.0006*	0.6186	NA
Rco35	0.0006*	0.0006*	0.0891 ^{ns}
Rco36	0.0006*	NA	NA
Rco40	0.0006*	0.0827 ^{ns}	0.0827 ^{ns}
Rco41	0.0006*	0.0006*	0.0038 ^{ns}
Geral	0.0006*	0.0006*	0.0006*

Nota: NA significa que o marcador SSR é monomórfico para a respectiva população

Alelos nulos podem causar os desvios de heterozigosidade esperada de acordo com EHW. Pois a ocorrência de alelos nulos leva ao aumento da homozigose e redução da heterozigose, visto que indivíduos heterozigotos contando o alelo nulo serão interpretados como homozigotos.

Entretanto, presença de alelos nulos não deve ser a explicação mais provável para o alto valor de endogamia encontrado para a espécie. Valores altos de f observados em todos os locos favorecem a interpretação de que os genótipos analisados estão submetidos a forte endogamia, provavelmente devido ao próprio processo de replicação dos acessos ao longo das gerações.

Cinco diferentes dendrogramas foram construídos a partir dos dados moleculares com o objetivo de visualizar a distância genética entre as amostras. A identificação dos genótipos no dendrograma foi definida com a cor determinada pelo programa STRUCTURE para o grupo em que se enquadra o indivíduo.

De acordo com o dendrograma da Figura 12, pode-se destacar a formação de dois grandes grupos. O primeiro grupo é composto pela maioria dos acessos descritos como “CNPAM”, nesse grupo estão as cultivares Pernambucana e BRS 149 Nordestina. A cultivar Pernambucana se destaca por sua produtividade no semiárido brasileiro, sendo reunida próxima ao acesso CNPAM 2000-87. O acesso CNPAM 2000-09 foi agrupado junto a BRS 149 Nordestina, que é uma cultivar que apresenta porte médio e frutos semideiscentes e foi obtida por meio de seleção individual com testes de progênie da variedade local Baianita (não amostrada). CNPAM 99-13 e CNPAM 14 foram reunidos e apresentam teor de óleo superior a 50%. Nesse primeiro grupo também estão juntos a maioria dos acessos descritos como “BRA”, os acessos BRA 10100 e BRA 5916 são tão semelhantes (distância genética igual à zero) que podem ser considerados clones para os locos amostrados. O segundo grupo é muito mais diversificado, contém o acesso Pajé que foi o mais isolado entre os acessos analisados do banco de germoplasma da Embrapa, contendo dois alelos exclusivos. Todos os quatro acessos denominados “CSRN” estão distribuídos no dendrograma de maneira dispersa, isso significa que esses acessos possuem base genética diversificada, a mesma situação ocorre com os quatro acessos denominados “CSRD”. Os acessos CNPAM 20001-57 e CSRN 3379 foram agrupados no dendrograma e, assim como o agrupamento relativo aos acessos CNPAM 2001-50, CNPAM 2000-79 e CSRD 2, apresentam porte baixo. As cultivares BRS Nordestina, CNPAM 2000-9 e Pernambucana 02 também foram agrupadas no dendrograma e apresentam porte alto.

Com relação ao dendrograma da Figura 13, é possível verificar que os acessos PB56 II e PB68 II são muito semelhantes geneticamente. É possível observar a formação de um grupo contendo IAC 80, que é uma cultivar antiga, de porte alto e frutos semideiscentes, que está agrupada muito próxima ao acesso PB75 II, IAC 80 também está agrupada com os acessos PB 10 II e PB 38 (ambos são descendentes do cruzamento Anã x Viradouro), os acessos PB 29 e PB32 (ambos de porte baixo) também estão nesse grupo porque são descendentes do genitor IAC 80 (Cmm88/26). Outra cultivar comercial a IAC2028, de porte baixo e frutos indeiscentes forma um agrupamento bastante destacado junto com os acessos PB49 II, TS03 e PB34 II. IAC Guarani é uma das cultivares mais antigas, de porte médio e frutos indeiscentes, que é utilizada até hoje pelos agricultores, no dendrograma os acesso mais relacionados com essa cultivar são PB 26 II e TS 17. IAC 226 é uma cultivar também de porte médio e frutos indeiscentes que é geneticamente relacionada aos acessos PB 28 II e TS 27. Os acessos PB 45 II, PB 46 II e PB 47 (ambos de porte baixo) foram reunidos no dendrograma e são descendentes de um mesmo genitor (H78 – não amostrado). Assim como os acessos reunidos PB 7 II e PB 8 II são descendentes do mesmo cruzamento {IAC 38 x H34(Jau)}. Os acessos reunidos PB 42 II e PB 43 II são descendentes do cruzamento Anã x Viradouro e apresentam porte médio. Os acessos reunidos no dendrograma PB 72II e PB 73II são oriundos do genitor {S80226 (Tatuí)} e apresentam porte médio. Os acessos P 33 II, PB 35 e PB 04 II estão agrupados no dendrograma e apresentam porte anão.

Os sete genótipos de Botucatu foram agrupados pelo dendrograma da Figura 14. Como esperado, Parcela 13XParcela 13 (cv. Guarani) é muito próximo geneticamente da Parcela 17.1 (Cv. Guarani). Esses genótipos estão relacionados com Israel e Sara (híbrido) desenvolvido pela EMBRAPA que se destaca pela menor incidência de mofo-cinzeno. O acesso MRZ.02 x Parcela 49 está agrupado com China. MRZ. 03 está mais isolado no dendrograma.

Não foi identificada nenhuma duplicata entre os dois bancos de germoplasma através da análise do dendrograma da Figura 16. Isso reforça a importância de se manter e conservar todos os acessos presentes em cada um dos bancos de germoplasma.

Figura 12 - Dendrograma dos acessos que correspondem ao banco de germoplasma da Embrapa, construído pelas distâncias de Rogers-W e pelo agrupamento UPGMA. Incluindo os valores de bootstrap e o valor cofenético, representado como r . As cores marcadas na identificação dos acessos correspondem exatamente às cores geradas pelo gráfico do STRUCTURE (Figura 18)

Figura 14 - Dendrograma dos genótipos de Botucatu, construído pelas distâncias de Rogers-W e pelo agrupamento UPGMA. Incluindo os valores de bootstrap e o valor cofenético, representado como r

Figura 15 - Dendrograma de todas as amostras agrupada nas três populações, construído pelas distâncias de Rogers-W e pelo agrupamento UPGMA. Incluindo os valores de bootstrap e o valor cofenético, representado como r

$r = 0.88436$

Figura 16 - Dendrograma de todas as amostras, construído pelas distâncias de Rogers-W e pelo agrupamento UPGMA. Incluindo os valores de bootstrap e o valor cofenético, representado como r . As cores marcadas na identificação dos acessos correspondem exatamente às cores geradas pelo gráfico do STRUCTURE (Figura 24)

Figura 17 – Análise dos componentes principais envolvendo todos os genótipos. Identificação dos acessos: Tabela 1 (1 a 38 EMBRAPA), Tabela 2 (39 a 114 IAC), Tabela 3 (115 a 121 BOTUCATU)

A análise de componentes principais do NTSYS (Figura 17) foi feita com todos os acessos para facilitar a visualização em três dimensões da estruturação observada nas amostras. Essa análise também formou dois grupos principais, separando os acessos da EMBRAPA dos acessos do IAC junto com os genótipos de BOTUCATU. Os acessos da EMBRAPA: CRD 20 (n° 2), BRA 8877 (n° 35) e CSRD 10P (n° 38) ficaram situados entre os dois grupos formados, sugerindo que são os acessos do banco de germoplasma da EMBRAPA mais próximos geneticamente dos acessos do banco de germoplasma do IAC.

Utilizando o programa STRUCTURE 2.3, foram testados 11 arquivos de parâmetros contendo valores de $k=1$ à $k=11$, com cinco corridas para cada valor de K , admitindo um modelo sem mistura (pois as amostras de uma população são isoladas das outras), frequências alélicas correlacionadas, 100.000 burnins e 200000 simulações de Monte Carlo de cadeias de Markov (MCMC). A seleção do número K mais provável foi realizada através dos valores de ΔK , de acordo com o método proposto por EVANNO *et al.* (2005). A partir do valor de K selecionado, foi gerados quatro gráficos utilizando o programa STRUCTURE 2.2, um para cada uma das três populações pré-definidas e outro gráfico englobando todo o conjunto de genótipos.

Em todos os gráficos gerados pelo programa STRUCTURE, o valor de K selecionado com o maior valor de ΔK foi de dois (com exceção do gráfico para Botucatu, que foi descartado). Isso significa que esse programa formou dois grupos entre as amostras em todos os quatro gráficos gerados.

FIGURA 18 - Valores de ΔK para cada valor de K, calculado nos acessos do bando de germoplasma da EMBRAPA, de acordo com o proposto por Evanno *et al.* (2005). O maior valor de ΔK corresponde ao K ótimo

FIGURA 19 - Teste de atribuição para os acessos do banco de germoplasma da EMBRAPA de *Ricinus communis* avaliados ($K=3$). Acessos representados pelas barras verticais coloridas. A mesma cor em acessos diferentes indica que eles pertencem ao mesmo grupo. Cores diferentes no mesmo indivíduo indica a porcentagem do genoma compartilhado com cada grupo. Para a identificação dos acessos (Tabela 1). O eixo das ordenadas mostra o coeficiente de participação do indivíduo em cada grupo (Q)

Dentro do banco de germoplasma da EMBRAPA, ocorreu exatamente o mesmo agrupamento na análise de dendrograma (Figura 12) e no gráfico do STRUCTURE (Figura 19), formando dois principais grupos. No dendrograma, essa divisão em dois grupos tem valor satisfatório de Bootstrap (64,78%) suportando esse nó. No gráfico gerado pelo STRUCTURE, todas as barras estão cheias, significando que a probabilidade de todos os indivíduos estarem agrupados corretamente nos dois grupos é de 100%.

FIGURA 20 - Valores de ΔK para cada valor de K, calculado nos acessos do bando de germoplasma do IAC de acordo com o proposto por Evanno *et al.* (2005). O maior valor de ΔK corresponde ao K ótimo

FIGURA 21 - Teste de atribuição para os acessos do banco de germoplasma do IAC de *Ricinus communis* avaliados (K=2). Acessos representados pelas barras verticais coloridas. A mesma cor em acessos diferentes indica que eles pertencem ao mesmo grupo. Cores diferentes no mesmo indivíduo indica a porcentagem do genoma compartilhado com cada grupo. Para a identificação dos acessos (Tabela 2). O eixo das ordenadas mostra o coeficiente de participação do indivíduo em cada grupo (Q)

Já no banco de germoplasma da IAC, não se pode observar uma correlação muito clara entre a análise do dendrograma (Figura 13) e o gráfico do STRUCTURE. Alguns acessos apresentam duas cores, o que significa que eles não podem se encaixar completamente em somente um dos dois grupos sugeridos pelo programa. O acesso PB04II apresentou 50% de probabilidade de seu genoma estar associado a um dos grupos e 50% de estar associado ao outro. O primeiro grupo formado (verde) contém a maioria das cultivares: IAC 2028, IAC 226 e IAC Guarani. Já o segundo grupo contém apenas a cultivar IAC 80.

FIGURA 22- Valores de $\ln P(D)$ para cada valor de K , calculado nos acessos genótipos de BOTUCATU, de acordo com o proposto por EVANNO *et al.* (2005). O maior valor de ΔK corresponde ao K ótimo

A população de Botucatu tem um tamanho amostral muito pequeno, portanto a aplicação do método de Evanno *et al.* (2005) não permitiu a detecção do valor de K que melhor acomoda os genótipos (dados não apresentados). Isso pode ser explicado devido ao fato de que este método não poder ser aplicado para $K = 1$, sendo o menor valor possível $K=2$.

De fato, utilizando o método sugerido por Pritchard *et al.* (2000), observa-se através da Figura 22 que o valor de K mais adequado para essa população é igual a 1, para o qual foi obtido o maior valor de $\ln P(D)$ e o menor desvio padrão. Representando a falta de estruturação dentro dessa pequena população.

FIGURA 23 - Valores de ΔK para cada valor de K, calculado em todos os genótipos, de acordo com o proposto por Evanno *et al.* (2005). O maior valor de ΔK corresponde ao K ótimo

FIGURA 24 - Teste de atribuição para todos os genótipos de *Ricinus communis* avaliados (K=2). Acessos representados pelas barras verticais coloridas. A mesma cor em acessos diferentes indica que eles pertencem ao mesmo grupo. Cores diferentes no mesmo indivíduo indica a porcentagem do genoma compartilhado com cada grupo. Identificação dos acessos (ver Tabela 1, 2 e 3. 1 a 38 - EMBRAPA, 39 a 114 - IAC, 115 a 121 - BOTUCATU). O eixo das ordenadas mostra o coeficiente de participação do indivíduo em cada grupo (Q)

Os resultados do dendrograma que representa os grupos de acessos em três populações (Figura 15), do dendrograma contendo todas as amostras individualmente (Figura 16), da análise feita com componentes principais pelo NTSYS (Figura 17) e do gráfico do STRUCTURE feito com todos os genótipos (Figura 24) permitem inferir que ocorre uma estruturação no conjunto total das amostras analisadas, formando dois grupos geneticamente distantes. O primeiro grupo (vermelho) é composto principalmente pelos acessos que representam o banco de germoplasma da EMBRAPA; o segundo grupo (verde) é formado pelos acessos representantes do banco de germoplasma do IAC e também pelos genótipos da UNESP – BOTUCATU. Essa conclusão é corroborada pelo alto valor de F_{ST} encontrado. Os acessos Parcela 13XParcela 13 (cv. Guarani), Parcela 17.1 (Cv. Guarani) e MRZ.02 x Parcela 49 da população de Botucatu foram colocados pelo gráfico do STUCTURE (Figura 24) como pertencentes ao primeiro grupo junto com os acessos da EMBRAPA, o que pode significar alguma semelhança em termos de frequências alélicas entre esses acessos.

A H_e média encontrada (0,473) e o número médio de alelos por loco (4,27) em genótipos brasileiros, através do presente estudo, foi maior do que a H_e média (0,188) e o número médio de alelos por loco (3,1) descritos por Allan *et al.* (2008), que analisou a diversidade de 41 acessos de bancos de germoplasma de 35 outros países, utilizando nove locos SSR. No entanto o F_{IT} obtido com os genótipos brasileiros (0,909) foi muito maior do que os acessos dos outros países (0,594), analisados por Allan *et al.* (2008).

A manutenção de bancos ativos de germoplasma de mamona possibilita a preservação do patrimônio genético da mamoneira e o fornecimento de matéria prima para as atividades de melhoramento e fomento da cadeia produtiva da cultura. Para que ocorra um manejo mais eficiente desses bancos é necessário um profundo conhecimento sobre a variabilidade genética contida nos mesmos.

3 CONCLUSÕES

Os genótipos avaliados representam grande parte da variabilidade genética de mamona disponível no Brasil e os marcadores SSR sintetizados nesse trabalho poderão ser utilizados para a análise de genótipos de mamoneira, em outros estudos de diversidade ou mapeamento genético.

A partir dos resultados obtidos no presente estudo, é possível afirmar que os programas de melhoramento de mamona devem utilizar genitores provenientes dos dois bancos de germoplasma para obter o máximo de variação genética disponível no país para a cultura da mamona. Visto que os acessos analisados possuem uma quantidade significativa de alelos raros e cada banco tem um *pool* genético bem diferenciado.

Os genótipos analisados apresentaram um excesso de homozigose em todos os locos, sugerindo uma forte endogamia, que provavelmente é resultado da autofecundação da mamona, mas pode também ser explicado pela presença de alelos nulos.

A construção de biblioteca genômica para a síntese de marcadores SSR em *Ricinus communis* L. é inédita no país e as informações geradas são de suma importância para a identificação, exploração e a conservação da variabilidade genética dessa espécie, além de ser de extrema importância para os programas de melhoramento da mamoneira.

REFERÊNCIA

- AKTAR, M.; MAHMOOD, I. Control of plant-parasitic nematodes with organic and inorganic amendments in agricultural soil. **Applied Soil Ecology**, Aligarh, v. 4, n. 3, p. 243-247, 1996.
- ALLAN G.; WILLIAMS A.; RABINOWICZ P.D.; CHAN A. P.; RAVEL J.; KEIM P. Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. **Genetical Research Crop Evolution**, Gatersleben, v. 55, p. 365–378, 2007.
- ALMEIDA, F. de A. C.; MORAIS, A.M.; CARVALHO, J.M.F.C.; GOUVEIA, J.P.G. Crioconservação de sementes de mamona das variedades nordestina e pernambucana. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v.6, n. 2, p.295-302, 2002.
- ALVES, R. M. **Caracterização genética de populações de cupuaçuzeiro, *Theobroma grandiflorum* (Willd. ex. Spreng.) Shum. por marcadores microssatélites e descritores botânico-agronômicos**. 2002. 159 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – ESALQ (Escola Superior de Agricultura Luiz de Queiroz), Piracicaba, 2002.
- ALZATE-MARIN, A.L.; CERVIGNI, G. D. L.; MOREIRA, M. A.; BARROS, E. G. Marker assisted selection in the development of disease resistant plants, with emphasis on common bean and soybean. **Fitopatologia brasileira**, Brasília, v.30 n. 4, p. 333-342, 2005.
- ANDERSON, J.A., CHURCHILL, G.A.; AUTRIQUE, J.E.; TANKSLEY S.D.; SORRELLS M.E. Optimizing parental selection for genetic linkage maps. **Genome**, Toronto, v. 36, p. 181-186, 1993.
- ANTHONISEN, D.; SCHIRMER M.; SILVA S. D. A. Caracterização de cultivares de mamona utilizando marcadores moleculares RAPD. In: ENCONTRO DE INICIAÇÃO CIENTÍFICA E PÓS-GRADUAÇÃO DA EMBRAPA CLIMA TEMPERADO, 2006, Pelotas. **Anais...** Pelotas: Embrapa Clima Temperado, 2006. p. 79-82.
- ATSMON, D. M. P. Castor. In: Robbelen G, Downey RK, Ashri A (Eds) **Oil crops of the world: their breeding and utilization**, New York: McGraw-Hill Publishing Company, 1989. p. 438–447.
- AZEVEDO, D.M.P. de.; LIMA, E.F.; BATISTA, F.A.S. **Recomendações técnicas para o cultivo da mamoneira (*Ricinus communis* L.) no Brasil**. Campina Grande: EMBRAPA-CNPA, 1997. 52p. (EMBRAPA-CNPA. Circular Técnica, 25).

AZEVEDO, D.M.P.; BELTRÃO N.E. de M.; **O Agronegócio da mamona no Brasil**. Campina Grande: EMBRAPA-Algodão, 2007. 350 p.

AZZINI, A.; SAVY FILHO, A.; SALGADO, A. L. de B.; ARNALDI, F. Z. Deslignificação dos resíduos agrícolas da cultura da mamona para produção de celulose e papel. **Bragantia**, Campinas, v. 43, n. 2, , p. 519-530, 1984.

BAJAY, M. M.; PINHEIRO J. P.; BATISTA, C. E. A.; NÓBREGA, M. B. M.; ZUCCHI M. I. Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. **Conservation Genetic Resources**, Durham, 2009. Disponível em: <http://www.springerlink.com/content/t6677g608151t1l3/>. Acesso em: 25 nov. 2009.

BALDANZI, M.; FAMBRINI, M.; PUGLIESI, C. Redesign of the castor bean plant body plan for optimal combining harvesting. **Annals of Applied Biology**, Warwickshire, v.142, n. 3, p. 299-306, 2003.

BELTRÃO, N.E.M. **Mamona**: árvore do conhecimento e sistemas de produção para o Semi-Árido Brasileiro. Campina Grande: EMBRAPA-Algodão, 2003. 19 p. (EMBRAPA-CNPA. Circular técnica, 70).

BERTOZZO F. **Avaliação da seleção para aumento da porcentagem de flores pistiladas em mamona (*Ricinus communis* L.)**. 2009. 43 p. Dissertação (Mestrado em Ciências Agrônômicas)- Faculdade de Ciências Agrônômicas da Unesp - Campus de Botucatu, Botucatu, 2009.

BORÉM, A.; MIRANDA G. V. **Melhoramento de plantas**. Viçosa: Editora Universidade Federal de Viçosa, 2005. 524 p.

BORÉM, A.; CAIXETA, E.T. **Marcadores moleculares**. Viçosa: Editora Universidade Federal de Viçosa, 2006. 374 p.

BOTSTEIN, D.; WHITE, R.L.; SKOLNIEK, M.; DAVIS, R.V. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal Human Genetic**, Boston, n. 32, p. 314-331, 1980.

BILLOTTE, N.; LAGODA, P.J.L.; RISTERUCCI, A.M.; BAURENS, C. Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. **Fruits**, Les Ulis, v.54, p.277-288, 1999.

CARLINI, C. R.; SÁ, M. F. G. Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. **Toxicon**: official journal of the international society on toxicology, San Diego, v. 40, p. 1515-1539, 2002.

CARVALHO, B.C.L. **Manual do cultivo da mamona**. Salvador: EBDA, 2005. 65 p.

CASTOR BEAN GENOME DATABASE. Disponível em: <<http://castorbean.tigr.org/>>. Acesso em: 25 nov. 2009.

CONAB - Companhia Nacional de Abastecimento. Disponível em: <www.conab.gov.br/conabweb/>. Acesso em: 25 nov. 2009.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2004. p.223-375.

DEVEY, M.E.; BELL, J.C.; SMITH, D.N.; NEALE, D.B.; MORAN, G.F. A genetic linkage map for *Pinus radiata* based on RFLP, RAPD and microsatellite markers. **Theoretical and Applied Genetics**, Berlin, v.92, p.673-679, 1996.

DOYLE, J.J.; DOYLE, J.L. Isolation of DNA from fresh tissue. **Focus**, Rockville, v. 12, n. 27, p. 13-15, 1990.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. Disponível em <www.embrapa.br>. Acesso em: 25 nov. 2009.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology**, Vancouver, n. 14, p. 2611-2620, 2005.

FANAN, S.; MEDINA P. F.; DE CAMARGO M. B. P.; GALBIERI R. Descrição de características agrônômicas e avaliação de épocas de colheita na produtividade da mamoneira cultivar IAC 2028. **Bragantia**, Campinas, v.68, n. 2, 2009, p.415-422.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares na análise genética**. Brasília: EMBRAPA-CENARGEN, 1998. 220 p.

FREIRE, E.C.; LIMA, E.F.; ANDRADE, F.P.; MILANI, M.; NOBREGA, M.B.M. Melhoramento genético. In: AZEVEDO, D.M.P.; BELTRAO, N.E.M. (Ed.). **O agronegócio da mamona no Brasil**. Brasília: Embrapa Informação Tecnológica, 2007. p. 169-194.

FREITAS S. M.; FREDO C. E. **Biodiesel à base de óleo de mamona**: algumas considerações. **Informações econômicas**, São Paulo, v.35, n. 1, 2005. p. 37-42.

FONTQUER, P. **Plantas medicinales**: el dioscorides. 5ª edição Barcelona: Labor, 1979. p. 187-188.

GOUDET, J. **Fstat version 2.9.3.2**. 2002. Institute of Ecology, Lausann,. Disponível em: <<http://www2.unil.ch/popgen/softwares/fstat.htm>>. Acesso em: 25 nov. 2009.

GOLDSTEIN D. B.; SCHLÖTTERER C. **Microsatellites**: evolution and applications. New York: Oxford University Press, 1999. p. 34-41.

GURGEL, J. T. do A. **Estudos sobre a mamoneira (*Ricinus communis* L.)**. 1945. 92 p. Tese (Livre-Docência) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 1945.

HAYDEN, M. J.; SHARP, P. J. Targeted development of informative microsatellite (SSR) markers. **Nucleic Acids Research**, London, V. 29, n. 8, p. 1-6, 2001.

HEMERLY, F. X. **Mamona: comportamento e tendências no Brasil**. Brasília: EMBRAPA -DID, 1981. 69 p. (EMBRAPA-DTC. Documentos, 2).

HOLANDA, A. **Biodiesel e inclusão social**. Brasília: Coordenação de Publicações. 2004. p.13-60.

KOUTROUBAS, S.D.; PAPAKOSTA, D.K.; DOITSINIS, A. Water requirements for castor oil crop (*Ricinus communis* L.) in a Mediterranean climate. **Journal of Agronomy & Crop Science**, Berlin, p. 33-41, 2000.

KRUG, C.A.; MENDES, P.T. ; SOUZA, G.F. DE. Melhoramento da mamoneira (*Ricinus communis* L.) III. Primeira série de ensaios de variedades (1937/38 – 1938/39). **Bragantia**, Campinas, v.3, n. 5, p. 85-122, 1943.

LEWIS, P.; ZAYKIN, D. **Genetic data analysis: computer program for the analysis of allelic data (software)** 2000. version 1.1.
Disponível em: <<http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php>>.
Acesso em: 25 nov. 2009.

LIMA, P.C.R. **Biodiesel: um novo combustível para o Brasil**. Consultoria Legislativa.
Disponível em:
<http://www2.camara.gov.br/publicacoes/estnottec/tema3/2005_177.pdf>. Acesso em: 25 nov. 2009.

LITT, M.; HAUGE, X. e SHARMA, V. Shadow bands seen when type polymorphic dinucleotide repeats: some causes and cures. **BioTechnique**, Portland, v. 15, p. 280-284, 1993.

MCCOUCH S.R.; CHEN X.; PANAUD O.; TEMNYKH S.; XU Y.; CHO Y. G.; HUANG N.; ISHII T.; BLAIR M. Microsatellite marker development, mapping and applications in rice genetics and breeding. **Plant Molecular Biology**, Dordrecht, v. 35, p. 89–99, 1997.

MILLER, M. **Tools for population genetic analyses (TFPGA)** (software), version 1.3: a windows program for analyses of allozyme and molecular population genetic data, 1997. Disponível em : <<http://www.marksgeneticsoftware.net/tfpga.htm>>.
Acesso em: 25 nov. 2009.

MOREIRA, J. de A. N.; LIMA, E. F.; FARIAS, F. J. C.; AZEVEDO, D. M. P. **Melhoramento da mamoneira (*Ricinus communis* L.)**. Campina Grande: Embrapa Algodão, 1996. 30p. (EMBRAPA-CNPA. Documento, 44).

MOSHKIN, V.A. **Castor**. New Delhi: Amerind Publishing Co. Pvt. Ltd., 1986. 315 p.

MYCZKOWSKI, M. L. **Seleção para aumento da porcentagem de flores femininas na População FCA-Unesp-PB de mamona (*Ricinus communis* L.)**. 2006. 42 p. Tese (Doutorado em Agronomia/Agricultura) – Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2006.

NASS, L.L. Utilização de recursos genéticos vegetais no melhoramento. In NASS, L.L.; MELO, I.S.; VALADARES-INGLIS, M.C. (Eds.). **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, 2001, p.29-56.

NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals, **Genetics**, Pittsburgh, v. 89, n. 3, p. 583-590, 1978.

NÓBREGA, M. B. de M. **Avaliação de genótipos de mamona (*Ricinus communis* L.) em cruzamentos dialélicos parciais Piracicaba**. 2008. 77 p. Tese (Doutorado em Genética e Melhoramento), ESALQ (Escola Superior de Agricultura Luiz de Queiroz), Piracicaba, 2008.

OLIVEIRA, N. S. **Variabilidade genética em *Ricinus communis* (Euphorbiaceae) através de DAF**. 2004. 73 p. Dissertação (Mestrado em Genética) da Universidade Federal de Pernambuco, Pernambuco, 2004.

PEREIRA, H. **Pequena contribuição para um dicionário de plantas úteis do Estado de São Paulo: indígenas e aclimadas**. São Paulo: Typographia Brasil de Rothschild & Co., 1929. 489 p.

PERES, A. Historia del medicamento el ricino. 5. ed. **Fitomedicina**: Barcelona, 1997, p. 64-69.

POWELL, W.; MORGANTE, M.; MCDEVITT, R.; VENDRAMIN, G.G.; RAFASLKI, J. A. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to population genetics of pines. **Proceedings of the National Academy of Science of the USA**, Washington, v.92, 1995, p.7759-7763.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, Washington v. 155, n.º2, p. 945-959, 2000.

PURSEGLOVE, J.W. **Tropical crops: Dicotyledons**. London: Longman, Green and Co. Ltd., 1968. 346 p.

RAFALSKY, J.A.; MORGANTE, M.; POWELL, W.; VOGEL, J. M.; TINGEY, S.V. Generating and using DNA markers in plants. In: BIRREN, B.; LAI, E. (Ed.). **Analysis of non mammalian genomes - a practical guide**, New York: Academic Press, 1996. p. 75-134.

RAJORA, O.P.; RAHMAN, M.H.; BIUCHERT, G.P.; DANCİK, B.P. Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario. **Molecular Ecology**, Vancouver, v.9, p.339-348, 2000.

ROHLF, F.J. **NTSYS-Pc**: numerical taxonomy and multivariate analysis system. New York: Exeter Publisher, 1989. 210p.

SANTOS, R. F. Análise econômica. In: AZEVEDO, D. M. P. de; LIMA, E. F. **O agronegócio da mamona no Brasil**. Brasília: Embrapa Informação Tecnológica, 2001. p. 17-35.

SAVY FILHO, A. Melhoramento da mamona. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: Editora da Universidade Federal de Viçosa, 1999a. p. 385-407.

SAVY FILHO, A. Hibridação em Mamona. In: BORÉM, A. (Ed.). **Hibridação artificial de plantas**. Viçosa: Editora da Universidade Federal de Viçosa, 1999b. p. 331–342.

SAVY FILHO, A. **Mamona (*Ricinus communis*)**, desenvolvimento de tecnologia de produção. (Prêmio Péter Murányi), Sumaré, Disponível em: < <http://www.fundacaopetermuranyi.org.br/main.asp?pag=2007>>. 2007. Acesso em: 25 nov. 2009.

SAVY FILHO, A. **Mamona**: tecnologia agrícola. Campinas: EMOPI, 2005. 105 p.

SCHLÖTTERER C, TAUTZ D. Slippage synthesis of simple sequence DNA. **Nucleic Acids Research**, London, v. 20, p. 211–215, 1992.

SIGRIST, M. S. **Divergência genética em *Curcuma longa* L. utilizando marcadores microssatélites e agromorfológicos**. 2009. 93 p. Tese (Mestrado em Agricultura Tropical e Subtropical) - Instituto Agronômico Campinas, 2009.

MOTA, J. W. S. da. **Análise da diversidade genética de germoplasma de *Theobroma cacao* L. da Amazônia brasileira por microssatélites**. 2003. 107 p. Tese (Doutorado em Fisiologia Vegetal) – Universidade Federal de Viçosa. Viçosa. 2003.

SOUSA, R.F.; MOTTA, J.D.; GONZAGA, E. da N.; FERNANDES, M.F.; SANTOS, M.J. Aptidão agrícola do assentamento Venâncio Tomé de Araújo para a cultura da mamona (*Ricinus communis* L.). **Revista de Biologia e Ciências da Terra**, Campina Grande, v.4, n. 1, p. 1-17, 2004.

SUGANUMA, E., CIAMPI, A. Y. **Análise genética populacional de jatobá (*Hymenaea ssp Leguminosaea*) utilizando microssatélites. 2001.** Disponível em: <www.redbio.org/porta1/encuentros/enc_2001/posters/03/03pdf/03-007.pdf> Acesso em: 25 nov. 2009.

SUJATHA M, REDDY TP Differential cytokinin effects on the stimulation of in vitro shoot proliferation from meristematic explants of castor (*Ricinus communis* L.). **Plant Cell Reports**, Berlin, n. 17, p. 561–566, 1998.

TÁVORA, F.J.A. **A cultura da mamona**. Fortaleza: EPACE, 1982. 111p.

VALLS, J.F.M. Caracterização de recursos genéticos vegetais . In NASS, L.L. (Ed.) **Recursos Vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia. 2007. p. 281-305.

VIEIRA R. M.; LIMA E. F. Importância sócio-econômica e melhoramento genético da mamoneira no Brasil. In QUEIROZ, M. A. DE; GOEDERT, C. O.; RAMOS S. R. R. (Ed). **Recursos Genéticos e Melhoramento de Plantas para o Nordeste Brasileiro**. Disponível em: <www.cpatas.embrapa.br>. Acesso em: 25 nov. 2009.

WEISS, E. A. **Oilseed crops**. London: Longman, 1983. 660 p.

WRIGHT, S. **Evolution and the genetics of populations: variability within and among natural populations**. Chicago: University of Chicago Press, 1978. 573 p.

YORINORI, J.T; KIIHL, R.A.S. Melhoramento de plantas visando resistência a doenças. In NASS, L.L.; MELO, I.S.; VALADARES, INGLIS, M.C. (Eds.). **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, p. 715-735, 2001.

ANEXOS

Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production

Miklos Maximiliano Bajay José Baldin Pinheiro
Carlos Eduardo Araujo Batista Márcia Barreto Medeiros Nóbrega
Maria Imaculada Zucchi

Received: 24 June 2009 / Accepted: 30 June 2009
© Springer Science+Business Media B.V. 2009

Abstract Castor (*Ricinus communis* L.) is an important oleaginous plant from both economic and social points of view. The seeds contain an oil with excellent properties for industrial uses. This paper presents the main results of a study aiming to develop microsatellite markers for castor. Twelve new polymorphic microsatellite markers were isolated and characterized in 38 genotypes accessions from the castor germplasm of the Brazilian Agricultural Research Company (EMBRAPA). Knowledge on the genetic diversity of castor can be used to gain a better understanding on genetic diversity conservation, and germplasm management, guiding breeding programs and conservation strategies.

Keywords Germplasm • Genetic diversity • SSRs and conservation

Castor (*Ricinus communis* L., Euphorbiaceae, $2n = 20$), whose origin center is probably Ethiopia, is cultivated in many tropical and subtropical regions of the world (Govaerts et al. 2000). The seeds, leaves and stem of castor are poisonous when consumed by humans and livestock, because of the presence of ricin, the major toxic alkaloid present in the plant (Zimmerman et al. 1958). In the other hand, the seeds have a very high concentration (more 45%) of excellent oil, ideal for the production of biodiesel (Jeong and Park 2009). Is the only vegetable oil soluble in alcohol, presenting high viscosity and requiring less heating than others oils during the production of biodiesel (Beltrão 2003). Castor oil has many other uses ranging from cosmetics and, medicines to lubricants utilized in aircraft and space rockets.

Loss of genetic diversity is a problem in the case of some important species for agriculture, since ancient cultivars (or landraces) and wild relatives of domesticated species are being lost as modern varieties become adopted by farmers. This has led to calls for genetic conservation of crop germplasm (Frankel and Bennet 1970). Castor diversity is well represented in Brazilian germplasms collections, but there is little information on the molecular characterization of the accessions, and diversity loss is a serious risk. Managers of collections strive to accumulate and maintain the biological diversity of crops and other economically important plant species (Allan et al. 2007).

The development of microsatellite markers for this species would greatly facilitate the work of several researches involved in its germplasm diversity conservation. Aiming to provide researchers with these tools, in this work we provide 12 microsatellite markers for castor.

Genomic DNAs were extracted from fresh leaf samples using the CTAB method (Doyle and Doyle 1987). A microsatellite-enriched library was obtained using

M. M. Bajay • J. B. Pinheiro • C. E. A. Batista
Departamento de Genética, Universidade de São Paulo-USP,
Av. Pádua Dias, 11, Vila Independência, C. P. 83, Piracicaba,
SP 13400-970, Brazil

M. B. M. Nóbrega
Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional
de Pesquisa de Algodão, OSWALDO CRUZ, 1143.
CENTENARIO, C. P. 174, Campina Grande CEP 58428-095,
Brazil

M. I. Zucchi (✉)
Centro de Pesquisa e Desenvolvimento em Recursos Genéticos
Vegetais, Instituto Agrônomo IAC, Av. Barão de Itapura,
1481, Centro 13020902, C. P. 28, Campinas, SP 13075-630,
Brazil
e-mail: mizucchi@iac.sp.gov.br

Table 1 Characteristics of 12 microsatellite loci from *Ricinus communis* L. forward (F) and reverse (R) primer sequence, repeat motif, temperature of annealing number of alleles (N), product size range in base pairs, observed (H_o) and expected (H_e) heterozygosities, FIS, PIC and P value HWE

Locus	GenBank accession	Primer nucleotide sequence (5'–3')	Repeat motif	T (LC)	Size range (bp)	No. of aleles	H_o	H_e	FIS	PIC	P value HWE
Rco2	GF100470	F: CTAGCTTTGGGGCACAGTC R: GGAAAATAGGTGCGTATGAAAC	(AC) ₁₂	60	130–152	4	0.0526	0.1953	0.737	0.1881	0.0011*
Rco3	GF100460	F: GATGTGAGCCCATTATGCTG R: TCAGAAAATACCTCTAGGCGACA	(GA) ₂₂	60	230–240	2	0.0000	0.1884	1.000	0.1703	0.0000*
Rco8	GF100461	F: CGTGTGTCTGTGTGCATGTC R: CCTCAACCCTTTGCTGTTC	(TG) ₁₀	60	280–288	4	0.1081	0.3579	0.705	0.3235	0.0002*
Rco11	GF100471	F: GCGTGGAATACTTCAAGCA R: CCCATTAGCATCGAGAAAG	(TC) ₁₀ (GT) ₆	60	240–250	2	0.0789	0.3168	0.757	0.2663	0.0000*
Rco12	GF100462	F: AAGACTGCACCTCCTCCTA R: TGCTGGAACAAACCCTGATA	(TG) ₈ (GA) ₉	62	220–224	2	0.1053	0.4114	0.750	0.3265	0.0000*
Rco13	GF100463	F: GGTGCTTCCAGAAATTCAGTT R: GGAGGGGAAAGACAGGATTC	(GA) ₂₃	62	226–254	5	0.0833	0.7126	0.886	0.6597	0.0000*
Rco15	GF100464	F: CACGCACGTTAAAGCAAAC R: GCGAAGAAACCAAAATGGAG	(AG) ₁₈	60	220–230	4	0.0000	0.4945	1.000	0.431	0.0000*
Rco22	GF100465	F: ATCCGCCGACAATAGCAG R: GCAACACTCTCTTCCCTGAA	(AAAC) ₃ (AC) ₉ (TC) ₅	62	220–230	2	0.0789	0.2836	0.728	0.2433	0.0000*
Rco23	GF100466	F: CATGGATGTAGAGGGTCGAT R: CAGCCAAGCCAAAGATTTTC	(GA) ₁₅ (AG) ₈	62	240–260	3	0.0526	0.6163	0.917	0.5427	0.0001*
Rco26	GF100467	F: TTGCTTGTCAAAGGGGAGTT R: TCATTTTGAGGGAGAAACCA	(CT) ₁₉	62	260–274	5	0.0526	0.4865	0.895	0.4583	0.0000*
Rco29	GF100468	F: GGAGAAAAGAAAGGGAGAAGG R: GCCAAAAGCACACTTAATTTGA	(GA) ₇	60	210–220	2	0.0000	0.2659	1.000	0.2307	0.0000*
Rco30	GF100469	F: TGAAACTTTGGAGCTTGGAGA R: GGTCCCACACATTCATACACA	(AG) ₁₉	60	220–240	5	0.1081	0.6709	0.843	0.6275	0.0000*

Forward (F) and reverse (R) primer sequence, temperature of annealing (T), number of alleles (N), observed (H_o) and expected (H_e) heterozygoties, FIS, PIC and P value HWE *
Departs significantly from HWE at $P \leq 0.05$ after Bonferroni correction

adapted protocols from Billotte et al. (1999). Genomic DNA of one genotype of *R. communis* was digested with *Rsa* I (Invitrogen), enriched in microsatellite fragments using (CT)₈ and (GT)₈ motifs. The enriched fragments were cloned into pGEM-T (Promega) and ligation products were used to transform Epicurian Coli XL1-Blue *Escherichia coli* competent cells. The positive clones were selected using the β -galactosidase gene and then grown overnight with ampicillin. A total of 96 clones were sequenced in an ABI 377 automated sequencer (PE Applied Biosystems) using BigDye terminator cycle sequencing kit (Applied Biosystems). About 12 pairs of primers were designed for SSR flanking regions using Primer 3 and tested in 38 morphologically divergent genotypes of *R. communis* from the germplasm collection of EMBRAPA. PCR reactions were performed in a 25 μ l reaction volume of buffer [20 mM] Tris-HCl (pH 8.4), 50 mM KCl and 1.5 mM MgCl₂] containing 50 ng of genomic DNA, 0.8 μ M of each primer, 150 μ M dNTPs and 1 U Taq DNA polymerase (Invitrogen) using a BIO-RAD My Cycler thermocycler. The PCR program consisted of an initial denaturing step at 94°C for 1 min followed by 35 cycles of amplification [94°C (1 min), 1 min at the specific annealing temperature of each primer pair (Table 1), 72°C (1 min)] and a final elongation step at 72°C for 10 min.

Amplification products were resolved by electrophoresis in 7% denaturing polyacrylamide gels and visualized by silver-staining (Creste et al. 2001). The allele scoring was done using the 10 pb DNA Ladder (Invitrogen) as size standards.

The number of alleles observed for each loci ranged from two to five, with an average of 3.3 alleles per locus.

Descriptive statistics and the test for Hardy–Weinberg Equilibrium were performed using Tools for Genetic Population Analysis (TFPGA) (Miller 1997). The observed and expected heterozygosities ranged from 0.000 to 0.108 (0.060 on average) and 0.188 to 0.712 (0.416 on average), respectively. The twelve loci depart significantly from Hardy–Weinberg Equilibrium ($P < 5\%$) after Bonferroni correction. However, results were obtained from a germplasm collection, which may not be in HWE. Indeed, there is absolutely no reason to expect HWE in germplasm collections, once genotypes are collected in different populations and, since the moment of sampling, they did not breed anymore.

The low observed heterozygosity values suggest the predominance of autogamy in this species, which is known to have a mixed mating system, being both self- and cross-pollinated by wind (Allan et al. 2007). Hardy–Weinberg equilibrium is expected in panmictic populations. However, if the species is selfing (autogamous) or mixed mating system, HWE equilibrium will not be attained, population will reach Wright's equilibrium (Ritland and Jain 1981; Weir 1996).

The linkage disequilibrium was tested adjusted P -value for 5% nominal level using the Fstat (Goudet 1995) and no disequilibrium was detected among all loci.

The development of molecular markers SSR for *Ricinus communis* L. is essential for the ongoing research on this culture in Brazil and worldwide. The generated information is of great importance for the identification, rational exploitation and conservation of the genetic variability of this species.

References

- Allan G, Willians A, Rabinowicz PD, Chan AP, Ravel J, Keim P (2007) Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. Genet Resour Crop Evol. doi: [10.1007/s10722-007-9244-3](https://doi.org/10.1007/s10722-007-9244-3)
- Beltrão NEM, Melo FB, Cardoso GD, Severino LS (2003) Mamona: Arvore do Conhecimento e Sistemas de Produção para o Semi-Arido Brasileiro. Campina Grande: Embrapa – CNPA, 2003. 19 p (Embrapa CNPA. Circular Técnica, 70)
- Billotte N, Lagoda PJR, Risterucci AM, Baurens FC (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits 54:277–288
- Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299–306
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15
- Frankel OH, Bennet E (1970) Genetic resources in plants—their exploration and conservation. F.A. Davis, Philadelphia
- Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486
- Govaerts R, Frodin DG, Radcliffe-Smith A (2000) World checklist and bibliography of Euphorbiaceae (with Pandaceae). Redwood Books Limited, Trowbridge, Wiltshire
- Jeong GT, Park DH (2009) Optimization of biodiesel production from castor oil using response surface methodology. Appl Biochem Biotechnol 156:431–441
- Miller MP (1997) Tools for Population Genetic Analysis (TFPGA), Version 1.3. Department of Biological Sciences, Northern Arizona University, AZ, USA
- Ritland K, Jain S (1981) A model for the estimation of outcrossing rate and gene frequency using independent loci. Heredity 47:35–52
- Weir BS (1996) Genetic Data Analysis II: Methods for Discrete Population Genetic Data (Sinauer, Sunderland, MA)
- Zimmerman LH, Miller MD, Knowles PF (1958). Castorbeans in California. Univ. Calif. Agr. Expt. Sta., Davis, Cir.