

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



PREDIÇÃO DA REAÇÃO DE INCOMPATIBILIDADE SEXUAL
DO CACAUEIRO (*Theobroma cacao* L.) USANDO FERRAMENTAS
DE MAPEAMENTO POR ASSOCIAÇÃO

MARCOS RAMOS DA SILVA

ILHÉUS – BAHIA – BRASIL
Fevereiro de 2015

MARCOS RAMOS DA SILVA

**PREDIÇÃO DA REAÇÃO DE INCOMPATIBILIDADE SEXUAL
DO CACAUEIRO (*Theobroma cacao* L.) USANDO FERRAMENTAS
DE MAPEAMENTO POR ASSOCIAÇÃO**

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

Área de Concentração: Genética e
Biologia Molecular

ILHÉUS – BAHIA – BRASIL
Fevereiro de 2015

S586 Silva, Marcos Ramos da
Predição da reação de incompatibilidade sexual do
cacaueiro (*Theobroma cacao* L.) usando ferramentas de
mapeamento por associação / Marcos Ramos da Silva. –
Ilhéus, BA: Uesc, 2015.
xi, 89 f.: il.

Orientador: Didier Pierre Clément.
Dissertação (Mestrado) – Universidade Estadual de
Santa Cruz. Programa de Pós-graduação em Genética e
Biologia Molecular.
Inclui referências.

1. Cacaueiro. 2. Cacaueiro – Melhoramento genético. 3.
Interação genótipo-meio ambiente. 4. Cacao – Cultivo. 5.
Vegetação – Mapeamento. I. Título.

CDD 633.74

MARCOS RAMOS DA SILVA

**“PREDIÇÃO DA REAÇÃO DE INCOMPATIBILIDADE SEXUAL EM CACAUEIRO
(THEOBROMA CACAO L) UTILIZANDO FERRAMENTAS DE MAPEAMENTO POR
ASSOCIAÇÃO”**

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

Área de Concentração: Genética e
Biologia Molecular

APROVADA: 26 de fevereiro de 2015.



Prof. Dr. Didier Pierre Clément
(Ceplac/Cirad - orientador)



Profa. Dra. Mariana Araujo Barreto
(UESC)



Prof. Dr. Carlos Priminho Pirovani
(UESC)



Prof. Dr. Messias Gonzaga Pereira
(Uenf)

Aos meus pais e avós, Francisco Sabino da Silva (in memoriam) (pai), Raquel Ramos da Silva (mãe), Leliro Lelis da Silva (avô) e Terezinha Ramos da Silva (avó), pelos momentos de carinho, confiança, incentivo aos estudos e todos os valores que aprendi e levo como referência em todos os momentos da minha vida.

Às minhas irmãs, tias e primos que sempre se fizeram presente nessa caminhada, não deixando que eu me sentisse sozinho nunca.

À minha namorada, Jamile Silva, que lutou junto comigo durante esse tempo. Tendo paciência em me esperar e nunca me deixando sozinho.

Aos meus orientadores, Uilson Vanderlei Lopes e Didier Pierre Clément, pela amizade, por contribuírem imensamente com seus ensinamentos durante todos esses anos de convivência e por acreditarem no meu potencial.

DEDICO

AGRADECIMENTOS

À Deus, que me permitiu estar aqui hoje e que me deu sabedoria para persistir e lutar diante dos desafios.

Ao Dr. Uilson Lopes pela orientação, amizade, pela contribuição em todos os momentos, compreensão e confiança em minha competência.

Ao Dr. Didier Clément, pela orientação, constante disponibilidade, presteza, boa vontade e confiança.

À Universidade Estadual de Santa Cruz (UESC), onde comecei a construir minha carreira acadêmica, passando da graduação ao mestrado. Em especial ao Programa de Genética e Biologia Molecular da UESC e aos professores, pelas contribuições valiosas para a minha formação.

À coordenação e aos funcionários do Colegiado do Programa de Pós-graduação em Genética e Biologia Molecular, em especial a Fabricia, Mara, Kátia e o Dr. Márcio Gilbert pela atenção, paciência, imprescindível apoio e boa vontade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Dr^a Fabienne Micheli pela concessão da bolsa de mestrado, apoio importante para a realização desse trabalho.

À CEPLAC e todos os funcionários do CEPEC, que tive a oportunidade de conviver, em especial a Reinaldo, Nádia, Marlene, Glória, Francisco, Mará, Nivaldo, Augusto, Zé Reis, Antônio e Silvio.

À Dr^a Karina Gramacho, pela atenção e o espaço oferecido no Fitomol, o que possibilitou minha convivência em um ambiente de constante aprendizagem.

Aos meus colegas da CEPLAC, Analine, Mariana Carvalho, Geovane, Nara, Israel, Rafael, Paulo, Thiago, Uine, Belmiro, Leandro, Iara, Lívia, Mariana Barreto, Taty, Louíse, Francisca, Rafaela do Mar, Rodrigo, Rangeline, Kaleandra, Yaska, Marlon, Ramon, Júlia, Mayara, Keidmille, Tâmilis e Keyla, por todos os momentos bons compartilhados na companhia de vocês durante esse tempo.

Aos meus colegas do CBG, especialmente Lívia Santana, Alessandro, Ryan, Juliano, Lara, Àkyla e Raner. Levarei comigo boas lembranças da amizade de vocês.

Às minhas duas amigas, Acácia Brasil e Maria Rafaela pelos seis anos de amizade e companheirismo (quatro anos de graduação e dois anos de mestrado juntos). Espero ter vocês duas sempre na minha vida.

Aos meus amigos, Carina Nobre, Camilla Nobre, Kleber Gonçalves, Ruth Ferreira e Cristiane Pereira, que direta o indiretamente me ajudaram na realização deste trabalho.

Muito obrigado!

Índice

EXTRATO	viii
ABSTRACT	x
1.INTRODUÇÃO	1
2. REVISÃO BIBLIOGRÁFICA	4
2.1 Botânica da espécie <i>Theobroma cacao</i>	4
2.2 Incompatibilidade sexual em angiospermas.....	7
2.3 Marcadores moleculares e suas aplicações no melhoramento genético de plantas	11
2.4 Métodos de seleção baseados em marcadores moleculares	15
3. CAPÍTULO 1- Caracterização da incompatibilidade sexual em uma população de melhoramento do cacaueteiro (<i>Theobroma cacao</i> L.)	22
Resumo.....	22
Introdução	23
Materiais e Método	25
Resultados	28
Discussão.....	38
Referências Bibliográficas	41
4. CAPÍTULO 2- Mapeamento por Associação de genes relacionados à incompatibilidade sexual do cacaueteiro (<i>Theobroma cacao</i> L.)	46
Resumo.....	46
Introdução	47
Materiais e Método	50
Resultados	55
Discussão.....	64
Agradecimentos.....	69
Referências Bibliográficas	69
5. CONCLUSÕES GERAIS	75
6. REFERÊNCIAS BIBLIOGRÁFICAS COMPLEMENTARES	76

EXTRATO

SILVA, Marcos Ramos, MSc, Universidade Estadual de Santa Cruz, Ilhéus, Fevereiro de 2015.
Predição da reação de incompatibilidade sexual do cacauero (*Theobroma cacao* L.) usando ferramentas de Mapeamento por Associação. Orientador: Didier Pierre Louis Clément. Co-orientador: Uilson Vanderlei Lopes

A incompatibilidade sexual do cacauero causa grandes perdas em produtividade, podendo chegar a 100% em certos casos. Infelizmente, a identificação de genótipos autocompatíveis em populações de melhoramento, usando os métodos convencionais (polinizações no campo), é um processo que demanda tempo (4-5 anos) e muitos recursos. Métodos de seleção baseados no uso de marcadores moleculares é uma alternativa para seleção precoce e eficiente de plantas para diferentes características. Entre esses métodos é possível destacar o Mapeamento por Associação. Assim, o presente estudo buscou caracterizar uma população de melhoramento em relação à compatibilidade/incompatibilidade e estabelecer um método de predição e seleção precoce de genótipos autocompatíveis usando ferramentas de Mapeamento por Associação. A dissertação foi dividida em dois capítulos. O **primeiro capítulo** buscou caracterizar fenotipicamente uma população de melhoramento quanto à compatibilidade/incompatibilidade sexual. Para alcançar esse objetivo foram feitos dois tipos de polinização controlada: autopolinização de 333 plantas e polinização cruzada com pólen de genótipos autocompatíveis (como controle). No total, 9.147 flores foram polinizadas e avaliadas em relação ao pegamento aos 7, 15, 21 e 30 dias após as polinizações. A partir dessa metodologia foi possível detectar que a compatibilidade sexual em cacauero exibe efeito gradual, com gradiente de retenção floral variando de 0 a 100%. Porém, houve uma predominância de genótipos com baixo pegamento na população, podendo ser resultado de uma alta frequência de alelos de incompatibilidade na população. Por outro lado, genótipos com alta retenção de flores (90-100%) foram identificados e estes possuem potencial para serem selecionados com o objetivo de aumentar a frequência de alelos de compatibilidade no programa de melhoramento. Além disso, foi detectado quatro padrões de retenção floral. Dois padrões (grupos 1 e 2) apresentaram alta retenção floral, um deles possuindo um tipo semelhante ao do controle. Os outros dois padrões (grupo 3 e 4) apresentaram uma baixa retenção floral. Assim, levando-se em

consideração a variação contínua do percentual de flores retidas pelos genótipos da população estudada e os diferentes padrões detectados, foi possível concluir que a compatibilidade sexual deve ser tratada como uma variável quantitativa contínua, enquanto a incompatibilidade como uma variável qualitativa discreta. O **segundo capítulo** buscou identificar regiões genômicas associadas à incompatibilidade usando a técnica de Mapeamento por Associação, visando à seleção precoce de plantas autocompatíveis. Para atingir esse objetivo, 295 genótipos provenientes de 109 cruzamentos foram selecionados. Estes genótipos foram caracterizados fenotipicamente, sendo que 8.199 flores foram polinizadas e avaliadas quanto à retenção floral aos 15 dias após as polinizações. Além disso, folhas destes 295 genótipos foram coletadas e o DNA extraído para genotipagem por sequenciamento (GBS). A GBS gerou 29.030 marcadores do tipo SNP, que após tratamento dos mesmos, resultaram em 11.792 SNPs utilizáveis. As análises de associação entre marcador e fenótipo foram realizadas no programa TASSEL, com os 11.792 SNPs e a percentagem de flores retidas 15 dias após as autopolinizações. Com base nessa metodologia foi possível identificar 12 marcadores associados à incompatibilidade, todos localizados na extremidade proximal do cromossomo 4. Esses marcadores apresentaram valores dos coeficientes de determinação (r^2) que explicam 6,68 a 20,79% da variação fenotípica e os efeitos dos 12 marcadores sobre a retenção floral variaram de 2,79 a 17,12. Posteriormente, foi feita uma validação desses 12 marcadores. Para isto, a população de estudo foi subdividida em dois grupos tomados ao acaso dentre os 295 genótipos, o primeiro com 236 (População de Modelagem) e o segundo com 59 genótipos (População de Validação), repetindo este processo 30 vezes. Na população de modelagem foram estimadas as equações que foram aplicadas na população de validação. As estimativas do nível de retenção floral obtidas na população de modelagem foram então comparadas com valores fenotípicos observados na população de validação. Ao todo, 75.300 equações foram obtidas, com 1 a 6 e 12 SNPs. A melhor equação de predição foi gerada pela combinação de três marcadores (SNP11326, SNP11343 e SNP11396), com média dos coeficientes de correlação igual 0,58. Assim, no presente estudo foi possível identificar a região envolvida no controle da incompatibilidade em cacaueteiro e 12 SNPs foram identificados associados a genes controlando esta característica. Equações de predição do nível de compatibilidade combinando alguns destes marcadores foram sugeridas e apresentam um potencial para serem validadas em outras populações.

Palavras-chave: Equações de predição; Marcadores moleculares; SNPs; Retenção floral; Seleção precoce.

ABSTRACT

SILVA, Marcos Ramos, MSc, Universidade Estadual de Santa Cruz, Ilhéus, February 2015.
Prediction of the sexual self-incompatibility reaction in cacao using Association Mapping tools. Advisor: Didier Pierre Louis Clément. Co-Advisor: Uilson Vanderlei Lopes.

Sexual incompatibility of cacao causes major losses in productivity, reaching 100% in some cases. Unfortunately, the identification of self-compatible genotypes in breeding populations using conventional methods (by field pollinations) is a process that takes time (4-5 years) and many resources. Selection methods based on the use of molecular markers is an alternative for early and efficient selection of plants for different characteristics. Among these methods it can highlighted the Association Mapping. Thus, this study aimed to characterize a breeding population in respect to the compatibility/incompatibility and to establish an early prediction and selection method of self-compatible genotypes using Association Mapping tools. The dissertation was divided into two chapters. The **first chapter** has sought to characterize phenotypically a breeding population in relation to the sexual compatibility/incompatibility. To achieve this goal two types of controlled pollinations were made: self-pollination of 333 plants and cross-pollination with pollen self-compatible genotypes (as a control). In total, 9,147 pollinated flowers were evaluated at 7, 15, 21 and 30 days after pollination in relation to the flower setting. With this approach it was possible to detect that the sexual compatibility in cacao presents a gradual effect, with floral setting ranging from 0 to 100%. However, there was a predominance of genotypes with low setting in the population, probably as a result of the high frequency of incompatibility alleles in the population. Moreover, genotypes with high flower retention (90-100%) have been identified and these have the potential to be selected aiming to increase the frequency of compatibility alleles in the breeding program. Furthermore, it was detected four patterns of floral settings. Two patterns (groups 1 and 2) showed high floral setting, one of them a pattern similar to the control. The other two patterns (groups 3 and 4) showed a low floral setting. Thus, taking into account the continuous variation in the percentage of flowers set by the genotypes of the study population and the different patterns detected, it was concluded that sexual compatibility should be treated as a continuous quantitative variable, while the incompatibility as a discrete qualitative variable. The **second chapter** aimed to identify genomic regions associated with the incompatibility using the

Association Mapping technique aiming the early selection of self-compatible plants. To achieve this goal, 295 genotypes from 109 crosses were selected. These genotypes were characterized phenotypically, and 8,199 flowers were pollinated and evaluated for floral retention at 15 days after pollination. Furthermore, leaves of these 295 genotypes were collected and DNA extracted for genotyping by sequencing (GBS). GBS generated 29,030 SNP markers, which, after treatment, resulted in 11,792 usable SNPs. The association analysis between marker and phenotype was performed in TASSEL program with the 11,792 SNPs and the percent of flower set 15 days after the self-pollinations. Based on this methodology 12 markers were identified to be associated with incompatibility, all located at the proximal end of chromosome 4. These markers showed values of coefficients of determination (r^2) explaining 6.68 to 20.79% of the phenotypic variation and the effects associated to these 12 markers on floral setting ranged from 2.79 to 17.12. Subsequently, a validation of these 12 markers was done. For this, the study population was divided into two groups randomly taken from among 295 genotypes, with the first 236 (Modeling Population) and the second with 59 genotypes (Validation of Population), repeating this process 30 times. In the modeling population equations were estimated that were then applied in the validation population. Estimates of floral setting levels obtained in the modeling population were then compared to phenotypic values observed in the validation population. In total, 75.300 equations were obtained, with 1 to 6 and 12 SNPs. The best prediction equation was generated by the combination of three markers (SNP11326, SNP11343 and SNP11396), with an average of the correlation coefficients equal to 0.58. Thus, in the present study it was possible to identify the region involved in the control of incompatibility in cacao and 12 SNPs were identified associated with genes controlling this trait. Prediction equations of the level of compatibility combining some of these markers have been suggested and have the potential to be validated in other populations.

Key words: Prediction equations; Molecular markers; SNPs; Floral retention; Early selection.

1. INTRODUÇÃO

A incompatibilidade sexual é um mecanismo fisiológico, com base genética que tem o potencial de ampliar o fluxo gênico, aumentando a variabilidade genética e o nível de heterozigose dentro de populações de uma determinada espécie (BREWBAKER, 1957; NETTANCOURT, 1977, 2000). Porém, em programas de melhoramento de cacaueteiro, a incompatibilidade sexual é uma importante barreira à produtividade de frutos e sementes, bem como um fator responsável pelo insucesso de cruzamentos ou autopolinização de germoplasmas (POSNETTE, 1940, 1945; JACOB; ATANDA, 1975; BARTLEY; YAMADA, 1982). Em casos extremos, por exemplo, em plantios monoclonais de genótipos autoincompatíveis, a incompatibilidade pode afetar até 100% da produção.

Embora muitos programas de melhoramento de cacaueteiro tenham colocado um esforço considerável para eliminar esta característica de suas populações de estudo, há uma carência de estudos sobre métodos para identificar genótipos autocompatíveis, em particular em estádios precoces. O método mais utilizado para identificar genótipos autocompatíveis em populações de melhoramento é o da percentagem de retenção floral após polinização artificial (ROSENQUIST, 1950; AREVALO; CARLETTO; OCAMPO, 1972; TERREROS; CHAVARRO; ROJAS, 1982). De modo geral, 15 a 40 flores são polinizadas e as avaliações da retenção floral são realizadas em 3, 7 e 15 dias após as polinizações. Plantas que apresentam um percentual de retenção entre 5 a 30% das flores polinizadas são consideradas autocompatíveis (LOPES; CARLETTO, 1995).

Entretanto, esse método de caracterização do fenótipo da incompatibilidade por meio de polinização artificial apresenta várias limitações. Uma destas limitações é o fato da incompatibilidade em cacaueteiro não ser perfeita, chamada por alguns autores de pseudocompatibilidade (CORAL; SORIA, 1972), ou seja, em algumas situações cacaueteiros autoincompatíveis produzem sementes quando autopolinizados. Outros autores têm sugerido a existência de diferentes graus de incompatibilidade em cacaueteiro (CARLETTO; SORIA, 1973). Isto tem dificultado o estabelecimento de um percentual mínimo de retenção floral para

diferenciar os genótipos em relação ao fenótipo da compatibilidade. Fatores adicionais como erros de polinizações, perturbação de insetos, chuva e temperatura podem interferir na caracterização da retenção floral (COPE, 1939; CARLETTO; SORIA, 1973; CARLETTO, 1975).

Além disso, a identificação de genótipos autocompatíveis pelo método de polinização artificial só ocorre após a planta atingir a maturidade e precisa de um número razoável de flores (15-40 flores). Considerando que o período juvenil em cacauero é bastante longo (3 ou mais anos), tempo e recursos são perdidos; genótipos autocompatíveis e autoincompatíveis são levados para o campo e permanecem durante anos, antes de permitir a eliminação dos autoincompatíveis. Desse modo, métodos precoces de identificação de indivíduos autocompatíveis em populações de melhoramento, ainda no estágio de plântula, são desejáveis. Isto possibilita que genótipos autoincompatíveis sejam identificados e eliminados ainda no viveiro, reduzindo-se o número de indivíduos levados para o campo e assim o tamanho e os custos dos ensaios.

Partindo disso, é necessário o estabelecimento de um método eficiente e rápido para a identificação de plantas autocompatíveis. Neste contexto, a utilização de marcadores moleculares pode ser uma ferramenta útil para permitir a seleção precoce de genótipos autocompatíveis, ainda no primeiro mês de vida. Em cacauero, alguns estudos com marcadores associados à incompatibilidade sexual foram realizados usando populações de cruzamentos biparentais. Entre esses, é possível destacar os primeiros trabalhos, usando isoenzimas, que permitiram identificar marcadores ligados a um dos locos (loco *S*) da incompatibilidade sexual em cacauero (WARREN; MISIR; KALAI, 1994). Posteriormente, Yamada et al. (2010), usando marcadores moleculares RAPDs (*Randomly Amplified Polymorphic DNA*), AFLP (*Amplified Fragment Length Polymorphism*) e microssatélites ou SSRs (*Simple Sequence Repeat*) identificaram 19 potenciais marcadores relacionados a incompatibilidade. Outros estudos realizados usando RFLPs (*Restriction Fragment Length Polymorphism*), RAPDs e marcadores microssatélites permitiram a detecção de marcadores associados a regiões genômicas associadas à incompatibilidade em cacauero, no grupo de ligação 4 (PHILLIPS-MORA et al., 1995; ROYAERT et al., 2011).

Outros métodos de mapeamento de genes usando populações de melhoramento diversas (germoplasmas, progênies, dentre outras) têm sido propostos, dentre eles o Mapeamento por Associação (THORNSBERRY et al., 2001; GUPTA et al., 2005; SOTO-CERDA; CLOUTIER, 2012). Este método tem sido usado em cacauero para identificar regiões genômicas associadas a outros caracteres de interesse agrônomico (PIRES, 2003; MARCANO et al., 2007;

MARCANO et al., 2009), mas não associadas à incompatibilidade. Assim, o presente estudo busca estabelecer um método eficiente de predição da incompatibilidade sexual de cacauero usando ferramentas de Mapeamento por Associação, bem como caracterizar a incompatibilidade em diversos genótipos de populações segregantes.

2. REVISÃO BIBLIOGRÁFICA

2.1 Botânica da espécie *Theobroma cacao*

Theobroma cacao L. (cacaueiro) é uma espécie diploide ($2n=20$), e genoma haploide de 0,43 pg (GLICENSTEIN; FLITZ, 1989; FIGUEIRA et al., 1992). Esta espécie assim como outras do gênero até recentemente pertenciam à família Sterculiaceae, porém foi reclassificada como pertencente à família Malvaceae (ALVERSON et al., 1998). As espécies do gênero *Theobroma* são originárias das florestas nativas da América Sul (CUATRECASAS, 1964), com distribuição geográfica relativamente ampla em regiões tropicais. Em condições naturais o cacaueiro pode alcançar entre 20 e 25 metros de altura, enquanto sob cultivo pode variar entre 2 e 5 metros (LACHENAUD; SOUNIGO; SALLÉE, 2005). No Brasil, a Bahia e o Pará se destacam como os principais produtores de cacau. O cacaueiro possui uma grande importância econômica em diversos países, principalmente para a produção de amêndoas a serem usadas na fabricação do chocolate. O cacaueiro é uma planta perene e preferencialmente alógama, com flores hermafroditas (TOXOPEUS, 1972).

2.1.1. Sistema reprodutivo em cacaueiro

As flores têm um pedicelo largo e fino, de 1,0 a 1,5 cm de comprimento, e são compostas por cálice e corola (LEÓN, 2000). O cálice apresenta cinco sépalas brancas ou rosadas unidas pela base. A corola é constituída de cinco pétalas brancas que se encontram na flor alternadas com as sépalas, apresentando um formato estreito na base, depois se alarga no ápice, tornando-se côncava formando um pequeno capuz chamado de cógula que recobre as anteras. As pétalas apresentam ainda um prolongamento de coloração amarela denominada de lígula. Tanto as sépalas quanto as pétalas variam de 6 a 8 mm de comprimento (LEÓN, 2000).

Flores de cacauieiro possuem tanto androceu quanto gineceu. O androceu é constituído por cinco estaminóides inférteis que rodeiam o pistilo e por cinco estames férteis alternados, que se encontram protegidos pelas pétalas e cada um possui duas anteras onde ficam depositados os grãos de pólen (LEÓN, 2000). Os grãos de pólen são pegajosos (o que dificulta o transporte pelo vento), binucleados, 17-23 µm em diâmetro, formato esférico e, com três poros cada um (BOUHARMONT, 1960). Cada flor pode produzir até 14 mil grãos de pólen (MASSAUX et al., 1976), sendo necessário apenas 200 grãos de pólen para alta eficiência no processo de frutificação (FALQUE et al., 1995; FALQUE; LESDALONS; ESKES, 1996). A deiscência das anteras começa quase ao mesmo tempo em que a flor se abre, tornando o pólen viável imediatamente, permanecendo assim por período de 2 a 3 dias (PARVAIS et al., 1977; REFFYE et al., 1978).

O gineceu é constituído essencialmente pelo pistilo. No ápice do pistilo existe uma área receptiva chamada de estigma, por onde os grãos de pólen iniciam o processo de germinação, formando os tubos polínicos. O estigma é conectado com o ovário por meio de uma porção alongada em forma de tubo fino, o estilete, estrutura que possibilita o tubo polínico germinar e crescer até chegar ao ovário, parte basal e achatada do pistilo (LEÓN, 2000). O ovário é súpero, formado por cinco cavidades denominadas de lóculos que se encontram unidas entre si. Em cada lóculo do ovário ficam armazenados os óvulos (estrutura reprodutiva que depois de fertilizada por um gameta masculino, dá origem ao zigoto, posteriormente formando as sementes, durante o processo da reprodução sexuada). Dependendo do genótipo cada ovário pode ter até 74 óvulos (BARTLEY, 2005).

Em cacauieiro, o sistema de polinização natural é entomófilo, que geralmente é realizada por um inseto do gênero *Forcipomyia*, uma mosca da família Cerapotogonidae (Diptera) (BILLES, 1942; GLENDINNING, 1972; CHAPMAN; SORIA, 1983). Outros insetos, tais como trips, formigas e afídeos podem também eventualmente promover a polinização (ENTWISTLE, 1972; CHAPMAN; SORIA, 1983; YOUNG, 1994). Porém, em cacauieiro o processo de polinização apresenta algumas limitações considerando-se a relação entre número de flores produzidas e o número de flores fertilizadas por planta. Um único cacauieiro pode produzir até 125.000 a 150.000 flores por ano (LACHENAUD; MOSSU, 1985; LEÓN, 2000), entretanto 25 a 50% não são polinizadas (PARVAIS et al., 1977; REFFYE et al., 1978; TOVAR; ORTIZ, 1991; ADJALOO; ODURO, 2013) e apenas 2 a 5 % das flores polinizadas resultam em frutos maduros (ALVIM, 1984; LEÓN, 2000), em grande parte devido à existência de flores autoincompatíveis ou limitações do polinizador natural da espécie (LEÓN, 2000).

Os principais estudos sobre a biologia reprodutiva do cacauero foram realizados por Cheesman (1927; 1932), Bouharmont (1960) e Ford (2004). Após um evento de polinização (transferência de pólen da antera para o estigma), os grãos de pólen germinam, formando os tubos polínicos (crescimento celular do grão de pólen em direção ao óvulo, o que possibilita a fecundação nas plantas espermatófitas), que crescem através da camada epidérmica da parede interna do canal estilar (CHEESMAN, 1927; BOUHARMONT, 1960; FORD, 2004). Esse canal estilar se conecta com a cavidade do interior do ovário. Cada tubo polínico passa através do tecido placentário e individualmente cada tubo penetra em um óvulo. Porém, para ocorrer o processo de penetração no óvulo, o tubo polínico passa por uma abertura chamada de micrópila e continua a crescer através de uma das células sinérgidas (CHEESMAN, 1927). Neste ponto, os gametas masculinos passam através da extremidade do tubo polínico, liberando os núcleos espermáticos (também chamados de anterozóides). Nesta etapa, as células sinérgidas são impulsionadas para o interior do saco embrionário, dando início ao processo de dupla fertilização (CHEESMAN, 1927). Durante esse processo as antípodas degeneram-se rapidamente. O resultado da fusão da primeira célula espermática com a oosfera constitui o zigoto diploide ($2n$), que posteriormente originará o embrião. Enquanto isso, a fusão dos dois núcleos polares e a segunda célula espermática formará o endosperma triploide ($3n$), que tem a função de servir como uma fonte nutritiva para o desenvolvimento do embrião (FORD, 2004). Antes da fertilização forma-se uma camada de amido em torno dos núcleos polares com a função de fornecer energia e nutrientes para a fertilização e desenvolvimento celular (CHEESMAN, 1927, 1932; FORD, 2004).

Estudos sugerem que a dupla fertilização em cacauero não ocorre antes de 24 horas após a polinização (BOUCHARMONT, 1960). A fusão entre a oosfera e o gameta masculino (mais especificamente a fusão dos núcleos gaméticos) ocorre no saco embrionário entre 24 e 48 horas após a polinização (BOUCHARMONT, 1960). Nesse intervalo, ocorre também o processo de fusão dos núcleos polares com o segundo gameta masculino. Após a dupla fecundação, o óvulo aumenta em tamanho nos dias subsequentes, e o desenvolvimento do endosperma inicia cerca de 3-4 dias após a polinização. Cinquenta dias após a polinização, o zigoto divide-se pela primeira vez e começa a formar o embrião. Após 90 dias, o embrião começa seu crescimento e é completado em até 110 dias após a polinização. O embrião atinge a maturidade entre 4 e 5 meses depois, com a maturação dos frutos, aproximadamente, 5 a 6 meses após a polinização. Porém em determinados genótipos de cacauero todo o processo de formação do zigoto ou produção de sementes é interrompido pela ocorrência do fenômeno da incompatibilidade

sexual, levando a flor a entrar em um processo de abscisão precoce (BOUHARMONT, 1960; FALQUE, 1994a).

2.2 Incompatibilidade sexual em angiospermas

A incompatibilidade sexual é a incapacidade de uma planta fértil e hermafrodita produzir frutos ou sementes após o processo de polinização (NETTANCOURT, 1977). Em geral, a incompatibilidade sexual é um mecanismo que ocorre em angiospermas (NETTANCOURT, 1977, 2000). Estudo realizado por pesquisadores verificou-se que 76 a 86% das espécies de cinco florestas neotropicais e 42% de uma floresta de altitude apresentam incompatibilidade sexual (GIBBS, 1990). De modo geral, este sistema envolve interações entre o gametófito e o esporófito, e se baseia essencialmente na capacidade da planta em rejeitar o próprio pólen, no caso da autoincompatibilidade sexual, favorecendo desta forma a alogamia, promovendo a variabilidade genética (BREWBAKER, 1957; NETTANCOURT, 1977, 2000). Além da autoincompatibilidade, também existe a incompatibilidade cruzada, que ocorre quando o pistilo de uma planta rejeita o pólen de plantas que são geneticamente semelhantes em relação aos alelos em um loco em particular (NETTANCOURT, 1977, 2000).

Em geral, a incompatibilidade é controlada geneticamente pelo loco S , com diferentes formas alélicas alternativas, que geralmente estão associados a outros genes que contribuem para sua expressão (NETTANCOURT, 1997). Em alguns casos, o controle genético pode ser exercido por mais de um gene (controle polifatorial), ou seja, mais de um loco pode estar envolvido (NETTANCOURT, 1997). Com base no controle genético, a incompatibilidade pode ser classificada em dois tipos de sistemas: esporofítico e gametofítico (NETTANCOURT, 1997). No sistema gametofítico, geralmente, ocorre um processo de reconhecimento do grão de pólen ao longo do canal estilar. Nesse sistema, a incompatibilidade ocorre quando o alelo presente no gametófito masculino (grão de pólen ou tubo polínico) é o mesmo de um daqueles presentes no pistilo (estigma, estilete e ovário). Em síntese, a incompatibilidade é determinada pelo genoma haplóide do gametófito masculino. Assim, grãos de pólen de constituição alélica S_1 e S_2 quando depositados por meio de autopolinizações ou polinizações cruzadas no pistilo de uma planta receptora com o genótipo S_1S_2 , será rejeitado, ocorrendo, neste caso, 100% de incompatibilidade (NETTANCOURT, 1977, 2000). Em geral, cruzamentos envolvendo progenitores com formas alélicas alternativas para o loco S podem apresentar diferentes graus de incompatibilidade, por exemplo, os cruzamentos $S_1S_2 \times S_1S_2$, $S_1S_3 \times S_1S_2$ e $S_3S_4 \times S_1S_2$ resultarão em 100%, 50% e 0% de incompatibilidade, respectivamente. Nos dois primeiros

cruzamentos, $S_1S_2 \times S_1S_2$ e $S_1S_3 \times S_1S_2$, isto significa que 100% e 50% dos óvulos não chegarão a formar o zigoto ou haverá uma falha no seu desenvolvimento, respectivamente, levando a abscisão floral de 100% das flores polinizadas em ambos os casos. No terceiro cruzamento, $S_3S_4 \times S_1S_2$, isto quer dizer que em 100% dos óvulos haverá formação do zigoto ou não haverá falha no seu desenvolvimento, conseqüentemente 100% das flores polinizadas serão retidas na planta. O sistema gametofítico ocorre em espécies das famílias Solanaceae, Rosaceae, Papaveraceae, Liliaceae, Onagraceae, Leguminosae, Commelinaceae e Amaryllidaceae (GIBBS, 1986).

Por outro lado, no sistema esporofítico o processo de reconhecimento e rejeição, geralmente ocorre quando os grãos pólen começam a germinar ao longo do canal estilar. Nesse sistema, a incompatibilidade é determinada pelo genótipo da planta doadora de pólen. Plantas com grãos de pólen com constituição alélica S_1S_2 são incompatíveis com o pistilo de plantas que têm o alelo S_1 ou o S_2 (NETTANCOURT, 1977, 2000). Em outras palavras, a incompatibilidade não é determinada pelo alelo que o grão de pólen carrega, mas sim pelos alelos presentes no tecido diploide da planta doadora de pólen. Desse modo, cruzamentos entre os genótipos $S_1S_2 \times S_1S_2$, $S_1S_2 \times S_1S_3$ e $S_1S_2 \times S_2S_3$ serão 100% incompatíveis, ou seja, em 100% dos óvulos não ocorrerá à formação do zigoto ou haverá uma falha no seu desenvolvimento, o que leva ao processo de abscisão de 100% das flores polinizadas nos três tipos de cruzamentos, enquanto que o cruzamento $S_1S_2 \times S_3S_4$ será 100% compatível, logo haverá a formação do zigoto ou não haverá falha no seu desenvolvimento em 100% dos óvulos. Este sistema de incompatibilidade está presente em espécies das famílias Brassicaceae e Asteraceae. A incompatibilidade esporofítica foi também proposta para *Corylus avellana* (família Betulaceae) (THOMPSON, 1979), *Cerastium arvense* (Caryophyllaceae) (LUNDQVIST, 1979) e *Ipomoea leucantha* (família Convolvulaceae) (KOWYAMA; SHIMANO; KAWASE, 1980).

2.2.1 Incompatibilidade sexual em cacauero

Assim como ocorre em várias espécies de angiospermas, em cacauero ocorre com frequência o fenômeno da incompatibilidade sexual. Após a descoberta da incompatibilidade em cacauero, onde constataram-se que algumas plantas não produziam frutos por meio do processo de polinização (HARLAND, 1925; POUND, 1932), diversos outros estudos foram realizados e mostraram que a incompatibilidade ocorre em diversos acessos das coleções de germoplasmas desta espécie (COPE, 1939; POSNETTE, 1945; BARTLEY, 1973; COPE, 1976; BARTLEY; YAMADA, 1982; YAMADA et al., 1982).

A base genética da incompatibilidade em cacaueteiro é ainda pouco conhecida. Os estudos realizados até o momento propõem que a incompatibilidade nesta espécie é geneticamente controlada pelo loco *S* (KNIGHT; ROGERS, 1955), com interferência de dois outros locos, A e B (COPE, 1962), apresentando no mínimo cinco alelos no loco *S* que exercem uma relação de dominância e codominância entre eles ($S_1 > S_2 = S_3 > S_4 > S_5$). Estudos mais recentes com marcadores moleculares realizados por Yamada et al. (2010) propõem, assim como Cope (1962), que mais de um loco está envolvido na incompatibilidade sexual. Porém, há ainda uma extensa discussão sobre o tipo de controle genético envolvido na incompatibilidade. Alguns autores propõem que a incompatibilidade em cacaueteiro é do tipo esporofítico, ou seja, é determinada por ambos os alelos que constituem o genótipo diplóide da planta doadora de pólen (KNIGHT; ROGERS, 1953, 1955). Enquanto isso, em outro estudo foi proposto que a incompatibilidade é controlada pelo sistema gametofítico, onde é determinada pelo genoma haplóide do gametófito masculino (BOUHARMONT, 1960). Entretanto, Cope (1962), com base em um estudo amplo com cortes histológicos de ovários oriundo de diversos genótipos de cacaueteiro, baseado na fusão ou não fusão dos núcleos gaméticos, propôs que a incompatibilidade nesta espécie apresenta ambos os tipos de controles - gametofítico e esporofítico. Esse mesmo autor mostrou também que a incompatibilidade sexual em cacaueteiro apresenta gradações na sua expressão, pois o processo de reação que propicia abscisão floral pode ocorrer em genótipos incompatíveis que exibem 25%, 50% ou 100% de não fusão dos núcleos gaméticos presente em um ovário (COPE, 1962).

Outro ponto de debate sobre a incompatibilidade em cacaueteiro está relacionado ao local de reconhecimento e rejeição do grão de pólen/tubo polínico. Nesta espécie, especificamente, há uma extensa discussão se o local de reconhecimento e rejeição ocorre no estigma, no estilete ou no ovário. Os primeiros estudos de incompatibilidade nessa espécie apontavam que o local de rejeição ocorria no estigma (POSNETTE, 1940). Porém, estudos posteriores propuseram dois sítios de ocorrência da reação de incompatibilidade: o primeiro deles seria no ato da germinação dos grãos de pólen (interação pólen-estigma) e o segundo durante o processo de fusão dos gametas (contato tubo polínico-óvulo) (ANEJA et al., 1992; ANEJA et al., 1994; BAKER; HASENSTEIN; ZAVADA, 1997).

Todavia, outros estudos demonstram que em cacaueteiro ocorre um tipo diferente de incompatibilidade, estando localizado especificamente no saco embrionário, sendo resultado da interação tubo polínico e óvulo (BENNETT; COPE, 1959; COPE, 1939, 1940, 1962; KNIGHT; ROGERS, 1953, 1955; FORD, 2004; FORD; WILKINSON, 2012). É importante ressaltar que na maioria das espécies de angiospermas, o processo de incompatibilidade ocorre pela inibição

do crescimento do tubo polínico no estigma (NETTANCOURT, 1977; SEAVEY; BAW, 1986). Porém, em cacauero o tipo de incompatibilidade sexual (BENNETT; COPE, 1959; COPE, 1958, 1962; KNIGHT; ROGERS, 1953, 1955; FORD; WILKINSON, 2012) é considerado anômalo (NETTANCOURT, 1977). Este tipo de incompatibilidade é conhecido como *late-acting* (de ação tardia) (SEAVEY; BAW, 1986) e foi descoberto por meio de estudos com cortes histológicos, nos quais foi evidenciado entre genótipos autocompatíveis e autoincompatíveis um comportamento semelhante no processo de germinação dos grãos de pólen (COPE, 1940; FORD, 2004). Neste tipo de incompatibilidade sexual, os grãos de pólen germinam no estigma e crescem através do canal estilar, chegando até o ovário, onde penetra nos óvulos, passando pela micrópila e liberando as duas células espermáticas. Neste momento, uma das células espermáticas entra em contato com a oosfera (célula-ovo), e a outra célula espermática se move em direção aos núcleos polares (célula central), até este ponto não ocorrem diferenças visíveis na reação entre materiais autocompatíveis e autoincompatíveis (COPE, 1939, 1940, 1962 BOUHARMONT, 1960; FORD, 2004). A diferença torna-se evidente, quando do contato entre as membranas das células gaméticas, onde ocorre uma inibição do processo de formação do zigoto e do endosperma, devido à não fusão dos núcleos das células gaméticas, ocorrendo uma falha no processo de singamia (BENNETT; COPE, 1959). Estudos com técnicas de microscopia confocal realizados por Ford e Wilkinson (2012) mostraram um comportamento divergente dos núcleos espermáticos no ovário, onde em alguns casos ocorre realmente uma falha na fusão dos gametas como relatado por Bennett e Cope (1959), porém, em outros casos, o tubo polínico consegue chegar até o saco embrionário presente no ovário, mas as membranas das células gaméticas masculina não chegam a entrar em contato com as membranas das células gaméticas femininas. Esse tipo anômalo de incompatibilidade sexual se faz presente também em outras espécies da família Malvaceae, como em *Cola nitida* (JACOB, 1973), *Sterculia chicha* (TARODA; GIBBS, 1982), *Dombeya acutangula* (GIGORD; LAVIGNE; SHYKOFF, 1998), *Theobroma grandiflorum* (RAMOS et al., 2005) e *Theobroma speciosum* (SOUZA; VENTURIERI, 2010).

A inibição ovariana posterior à penetração nos óvulos, mas anterior à fusão dos gametas, é apenas um dos tipos de incompatibilidade de ação tardia. Estudos realizados por Seavey e Bawa (1986) mostraram a existência de dois outros tipos de incompatibilidade de ação tardia: a) inibição do crescimento do tubo polínico no ovário, porém anterior à penetração nos óvulos, presente em espécies como *Lilium candidum* e *Lilium szovitsianum* (BROCK, 1954), *Hemerocallis thunbergii* (STOUT; CHANDLER, 1933; BREWBAKER; GORREZ, 1967) e *Hemorocallis citrina* (STOUT; CHANDLER, 1933) e em *Medicago sativa* (BRINK;

COOPER, 1938; COOPER; BRINK, 1940; SAYERS; MURPHY, 1966); b) incompatibilidade após a fusão dos núcleos gaméticos, ocorrendo uma falha no desenvolvimento do zigoto, presentes em espécies como *Liquidambar styraciflua*, *Epilobium obcordatum*, *Dalbergia retusa*, *Dipteryx panamensis* e *Myrospermum frutescens* (SEAVEY; BAWA, 1986).

Em cacauero, o efeito da incompatibilidade sexual pode ser alterado ou superado. A incompatibilidade pode ser contornada ou quebrada pela mistura entre grãos de pólen de flores de genótipos autocompatíveis e autoincompatíveis ou pela mistura de pólen de *Herrania* com pólen autoincompatível do próprio cacauero (GLENDINNING, 1960; LANAUD et al., 1987). O tratamento de flores de cacauero com dióxido de carbono (CO₂), também pode quebrar a incompatibilidade sexual. Aneja et al. (1992, 1994) ao avaliarem o efeito do tratamento com CO₂ na germinação do pólen e frutificação de um genótipo autoincompatível constataram a quebra parcial da incompatibilidade sexual em cacauero. Hormônios também afetam a reação de incompatibilidade em cacauero. Baker, Hasenstein e Zavada (1997) detectaram diferenças significativas na concentração de hormônios entre genótipos autocompatíveis e autoincompatíveis após a polinização de flores. Hanenstein e Zavada (2001) demonstraram que a concentração de auxinas e citocininas podem ser determinantes na quebra da incompatibilidade sexual. Estas maneiras de contornar a incompatibilidade têm grande relevância nos programas de melhoramento do cacauero, mas têm aplicações limitadas em larga escala para plantações estabelecidas em campo. Por esse motivo, é importante desenvolver métodos detecção precoce e eficiente de genótipos autocompatíveis.

2.3 Marcadores moleculares e suas aplicações no melhoramento genético de plantas

Nos programas de melhoramento de plantas, a eficiência no processo de seleção e maximização dos ganhos genéticos são atributos de fundamental importância para a geração de variedades com características de interesse agrônomicas adequadas. Neste contexto, a utilização de marcadores moleculares pode ser uma ferramenta com potencial de reduzir os ciclos de melhoramento para obtenção de novas variedades, diminuindo-se assim o número de cruzamentos, gastos com obtenção, avaliação e recombinação de progênes, pois permitem a seleção de plantas em qualquer estágio de desenvolvimento (FERREIRA; GRATTAPAGLIA, 1995; BORÉM; CAIXETA, 2009).

Os marcadores genéticos podem ser classificados como: marcadores morfológicos e marcadores moleculares. Os marcadores morfológicos por meio da variação fenotípica das

características têm sido utilizados na identificação de plantas desde a época de Mendel. Alguns estudos foram desenvolvidos em milho (*Zea mays* L.) (COE; NEUFFER, 1993) e trigo (*Triticum aestivum* L.) utilizando essa metodologia (HART; GALE; McINTOSH, 1993). Entretanto, apesar de existir um número variável de marcadores morfológicos em plantas (ex. nanismo, cor de pétala ou morfologia foliar), diversas limitações têm sido apresentadas com sua utilização nos programas de melhoramento, como baixo nível de polimorfismo, dificuldade para a construção de mapas genéticos e a forte influência dos efeitos de ambiente (BOREN; CAIXETA, 2009).

Com o advento de novas tecnologias na área de Genética e Biologia Molecular, principalmente o surgimento da tecnologia da reação em cadeia da polimerase (PCR) e da automatização do processo de sequenciamento da molécula de DNA, diversos tipos de marcadores foram desenvolvidos. Entre as vantagens dos marcadores moleculares estão o alto potencial em detectar polimorfismo, a identificação direta do genótipo sem influência do ambiente e a detecção de variações no DNA em qualquer estágio de desenvolvimento da planta, possibilitando a seleção precoce de genótipos. A estratégia de utilizar marcadores possui um grande potencial para reduzir o tempo e energias necessárias para desenvolver grandes populações segregantes por várias gerações. Marcadores moleculares têm sido utilizados nos programas de melhoramento como importantes ferramentas na detecção de alterações na molécula de DNA e se diferenciam entre si quanto à sua habilidade em detectar polimorfismo, ao custo de aplicação, facilidade de uso e à consistência de resultados (FERREIRA; GRATTAPAGLIA, 1995; BORÉM; CAIXETA, 2009).

Os marcadores moleculares podem ser de dois tipos: enzimáticos e os baseados em DNA. Um deles são os marcadores baseados em enzimas (Aloenzimas ou isoenzimas), denominados de marcadores isoenzimáticos. Esses marcadores são co-dominantes, e baseiam-se no fato de que os organismos produzem enzimas com diferentes formas moleculares, mas com a mesma função enzimática. No melhoramento de plantas existem diversos estudos com marcadores isoenzimáticos, alguns são relacionados à detecção de ligação gênica entre marcadores isoenzimáticos e caracteres de interesse econômico (GARVIN et al., 1989; TANKSLEY et al., 1982; GAUER; SLINKARD, 1990) e variabilidade genética (OLIVEIRA et al., 2002). Entretanto, atualmente, estes têm sido de uso limitado devido ao baixo polimorfismo.

Outros tipos de marcadores moleculares são os baseados em ácidos nucleicos (DNA). Esse tipo de marcador pode ser dividido em duas classes: hibridização e amplificação do DNA via reação em cadeia de polimerase (PCR) (MULLIS; FALOONA, 1987). A primeira classe é

formada por marcadores que apresentam polimorfismo no tamanho dos fragmentos do DNA, resultado do processo diferenciado de clivagem do DNA por enzimas de restrição. Essa variação no comprimento do fragmento, gerados em diferentes indivíduos, deu origem à classe de marcadores denominada RFLP (*Restriction Fragment Length Polymorphism*) (GRODZICKER et al., 1974; BOTSTEIN et al., 1980).

Uma segunda classe de marcadores baseados em ácidos nucleicos surgiu após a descoberta da amplificação de fragmentos de DNA via PCR, possibilitando a visualização de fragmentos específicos da molécula de DNA. A técnica de PCR permite a síntese *in vitro* de milhões de cópias de um segmento específico de DNA por meio da enzima DNA polimerase. Neste contexto, surgiram marcadores como o RAPD (*Random Amplified Polymorphism DNA*) (WILLIAMS et al., 1990), produzidos pela amplificação aleatória de fragmentos de DNA. Posteriormente, surgiram os marcadores AFLP (*Amplified Fragment Length Polymorphism*) (VOS et al., 1995), que são baseados na amplificação de regiões da molécula de DNA via PCR visando à identificação de variações em um conjunto de fragmentos selecionados e digeridos com enzima de restrição. Estes marcadores, os RAPDs e os AFLPs, são marcadores dominantes, ou seja, os alelos de um mesmo loco são revelados pela presença ou ausência de uma banda, resultado da amplificação de um fragmento de determinado tamanho no gel, ou seja, não é possível distinguir indivíduos homocigotos dos heterocigotos. Outro tipo de marcador baseado em PCR são os microssatélites ou SSR (*Simple Sequence Repeats*) (LITT; LUTY, 1989), que consistem de pequenas sequências repetidas em tandem. São marcadores codominantes, multialélicos e distribuídos por todo o genoma.

O sequenciamento de genomas e de genes expressos permitiu evidenciar uma nova classe de marcadores baseados na detecção de polimorfismos resultantes de variações de um único nucleotídeo ou SNP (*Single Nucleotide Polymorphism*) (BROOKES, 1999). Essas variações são resultado de substituições, deleções e inserções presentes em uma frequência alélica correspondente em pelo menos 1% da população, presentes tanto em regiões expressas quanto em regiões não expressas. Os SNPs são marcadores bi-alélicos e em grande abundância no genoma (CHING et al., 2002; WEINER; HUDSON, 2002). Associado ao sequenciamento de genomas, o advento das tecnologias de genotipagem em larga escala permitiu a ampliação e utilização dos SNPs em diversas pesquisas em plantas, principalmente no mapeamento de genes de interesse.

Com a redução dos custos de sequenciamento e o surgimento de novas tecnologias, além dos SNPs, outros marcadores moleculares baseados na técnica de hibridização por microarranjo foram desenvolvidos: marcador do tipo DArT (*Diversity Arrays Technology*)

(JACCOUD et al., 2001). Essa técnica permite detectar variações na molécula de DNA localizadas nos sítios das enzimas de restrição. O processo de genotipagem por meio dos marcadores DArT requer uma série de etapas: primeiro, é necessária a construção de um painel de diversidade (JACCOUD et al., 2001; SANSALONI et al., 2010); segundo, o DNA é clivado com enzimas de restrição, produzindo vários fragmentos que serão utilizados na construção de bibliotecas e posteriormente será formado o microarranjo (JACCOUD et al., 2001; SANSALONI et al., 2010). Cada ponto do microarranjo contendo um fragmento DNA pode ser considerado um loco. Neste tipo de marcador o polimorfismo é detectado pela presença ou ausência de hibridização de um fragmento específico (JACCOUD et al., 2001; SANSALONI et al., 2010), portanto, é um marcador de caráter dominante.

As diversas aplicações dos marcadores moleculares no melhoramento vegetal dependem dos objetivos de cada pesquisa. Os marcadores moleculares podem ser utilizados na identificação e caracterização de acessos em bancos de germoplasma, teste de paternidade, melhoramento assistido, predição de fenótipos, análise de pedigree etc. (BORÉM; CAIXETA, 2009; MILLACH, 1998). Em longo prazo, os marcadores moleculares podem ser ferramentas úteis ao melhorista nos estudos de mapeamento genético de diversas espécies (FERREIRA; GRATTAPAGLIA, 1995; FALEIRO et al., 2003).

Em cacauero, diversos estudos foram realizados utilizando marcadores. No caso de marcadores morfológicos, dois foram descritos a mancha axilar e a cor de cotilédone. Esses marcadores genéticos foram utilizados em estudos de incompatibilidade sexual e no desenvolvimento de cacaueros haplóides (DUBLIN, 1972; LANAUD et al., 1987; FALQUE, 1994b). Por outro lado, a utilização de marcadores moleculares em cacauero, tanto os baseados em isoenzimas quanto em ácidos nucleicos, tiveram uma ampla aplicação. Os marcadores baseados em isoenzimas, por exemplo, foram utilizados em análises de paternidade (YAMADA; LOPES, 1999), diversidade genética (YAMADA et al., 2001), origem, relações genéticas, estrutura de populações (LANAUD, 1986) e na elucidação da genética da incompatibilidade sexual (LANAUD et al., 1987; WARREN; MISIR; KALAI, 1994). Os marcadores baseados em ácidos nucleicos como RFLP foi utilizado em estudos de diversidade por meio de sondas de genes ribossomais (FIGUEIRA; JANICK; GOLDSBROUGH, 1992; LAURENT; RISTERUCCI; LANAUD, 1993; FIGUEIRA; JANICK; BEMILLER, 1994), construção de mapas genômicos (LANAUD, et a., 1995; CROUZILLAT et al., 1996; RISTERUCCI et al., 2000). Marcadores moleculares como o RAPD foram usados em vários estudos de diversidade (FIGUEIRA; JANICK; GOLDSBROUGH, 1992; RUSSEL et al., 1993; LERCETEAU et al., 1997; MARITA et al., 2001) e incompatibilidade sexual (YAMADA et

al., 2010); os AFLPs foram usados na construção e saturação de mapas genômicos (QUEIROZ et al., 1998; RISTERUCCI et al., 2000); os SSRs para a construção de mapas genéticos visando a identificação de QTLs associados à resistência a doenças (ex. podridão parda, vassoura-de-bruxa), incompatibilidade sexual, número de óvulos e caracteres que afetam a qualidade do fruto (CLÉMENT et al., 2003a; CLÉMENT et al., 2003b QUEIROZ et al., 2003; FALEIRO et al., 2006; BROWN et al., 2005; BROWN et al., 2007; YAMADA et al., 2010; FOUET et al., 2011; ROYAERT et al., 2011). Além disso, os marcadores SNPs e os microssatélites foram utilizados no desenvolvimento de um mapa consenso visando identificar regiões envolvidas no controle de diferentes caracteres que afetam a produção (ALLEGRE et al., 2011). Desse modo, fica evidente a ampla utilização dos marcadores moleculares nos diversos tipos de estudos em cacauero, podendo auxiliar nas estratégias de seleção de características de importância econômica.

2.4 Métodos de seleção baseados em marcadores moleculares

O melhoramento genético por meio da identificação de indivíduos geneticamente superiores e pela criação de novas combinações genotípicas tem o potencial de aumentar a produtividade e proporcionar o melhoramento de características de interesse para o homem. Neste processo, a seleção tem um papel importante ao fazer inferência sobre os valores genéticos dos indivíduos e selecioná-los com base nesses valores. Todavia, no melhoramento animal e de plantas, a seleção tradicional se baseia essencialmente na capacidade de prever o valor genético de um indivíduo com base apenas em informações fenotípicas. Um programa de melhoramento baseado exclusivamente na análise de fenótipos apresenta limitações na sua eficiência devido a fatores como a natureza poligênica das características de importância agrônoma, influência do ambiente e o tempo necessário para avaliação fenotípica em culturas perenes como o cacauero, por exemplo. Partindo disso, os programas de melhoramento têm buscado acelerar esse processo de seleção promovendo ganhos genéticos com alta eficiência e maior precisão.

A partir de pesquisas na área de Biologia e Genética molecular, foram identificados marcadores de DNA com potencial de promover seleção com alta eficiência, rapidez na obtenção de ganhos genéticos e baixo custo. Partindo desse pressuposto, foram surgindo diversos métodos de seleção baseados em marcadores moleculares, podendo destacar três

principais: (1) Mapeamento por ligação visando a Seleção Assistida por Marcadores (SAM); (2) Mapeamento por Associação (MA); e (3) Seleção Genômica (SG).

2.4.1 Mapeamento por ligação visando a Seleção assistida por marcadores (SAM)

A Seleção Assistida por Marcadores (SAM) foi proposta por Lande e Thompson (1990), a qual se baseia em utilizar simultaneamente dados fenotípicos e genotípicos de marcadores moleculares em ligação gênica com alguns locos controladores de características quantitativas (*Quantitative trait loci* - QTL), permitindo obter e selecionar os indivíduos com maior valor genético. Os marcadores utilizados na SAM, geralmente, são identificados por meio dos estudos de mapeamento baseados em ligação, utilizando populações previamente estabelecidas e derivadas de poucos cruzamentos (geralmente biparentais). Para esta técnica, as populações de mapeamento utilizadas são as populações F₂, obtidas por meio de autofecundação do híbrido F₁ oriundo do cruzamento de dois indivíduos contrastantes para a característica de interesse, e as populações de retrocruzamento, oriundas do cruzamento do híbrido F₁ com um dos pais (COLLARD et al., 2005).

Com o objetivo de aumentar os ganhos de seleção e reduzir os ciclos de melhoramento por meio da localização de genes e regiões genômicas que controlam características de interesse econômico (tanto qualitativas quanto quantitativas), os estudos clássicos de QTL (*Quantitative trait loci*) baseados em mapas de ligação visando a SAM, permitiu detectar e mapear genes em algumas culturas como milho (DOEBLEY; STEC; HUBBARD, 1997), tomate (FRIDMAN et al., 2000; FRARY, et al., 2000; LIU et al., 2002), *Arabidopsis* (JOHANSON et al., 2000; EL-DIN EL-ASSAL et al., 2001), arroz (YANO et al., 2000; TAKAHASHI et al., 2001). Entretanto, para outras culturas esta abordagem apresentou algumas limitações nos programas de melhoramento (BERNARDO, 2008). Primeiro, a associação entre marcadores-QTLs está restrita a populações de mapeamento biparentais, ou seja, em princípio apresenta utilidade apenas para aquela família avaliada, além disto, o número de recombinações entre os dois locos que estão localizados em um mesmo grupo ligação é limitado, o que resulta em baixa resolução para o mapeamento de QTLs (BERNARDO, 2008; HEFFNER; SORRELLS; JANNINK, 2009). Segundo, existe ainda o fato de que apenas poucos alelos por loco do QTL podem ser avaliados simultaneamente. Por último, devido à natureza poligênica e à alta influência ambiental nos caracteres quantitativos, apenas um pequeno número de QTLs de grande efeito é detectado e não se consegue explicar grande parte da variação genética da característica em estudo (GRATTAPAGLIA et al., 2009).

Em cacau, a estratégia de mapeamento por ligação visando a SAM, foi utilizada em diversos estudos. Lanaud et al. (1995) utilizou essa abordagem utilizando uma população de 100 indivíduos originados a partir de um cruzamento entre os genótipos UPA-402 e UF-676 visando mapear genes de resistência a doenças e genes relacionados a caracteres que afetam a qualidade do cacau. Diversos outros trabalhos foram realizados utilizando cruzamentos biparentais, estes estudos permitiram a identificação de QTLs ligados à resistência a doenças (podridão parda, vassoura-de-bruxa, murcha-de-ceratocystis e monilíase), incompatibilidade sexual e número de óvulos (LANAUD et al., 2000; CLEMENT et al., 2003a; CLEMENT et al., 2003b; BROWN et al., 2005; FALEIRO et al., 2006; ALBUQUERQUE, 2006; BROWN et al., 2007; YAMADA et al., 2010; ROYAERT et al., 2011).

2.4.2 Mapeamento por Associação

Buscando contornar os problemas encontrados com os estudos clássicos de QTLs visando a SAM, foi proposto o método de Mapeamento por Associação. Esta técnica também tem como objetivo mapear QTLs, baseando-se em análises de associação entre variações na sequência de DNA e variações fenotípicas da característica estudada, utilizando nesses estudos populações naturais, coleções de germoplasma ou conjunto de genótipos oriundos de diferentes cruzamentos, com potencial de explorar dados fenotípicos historicamente acumulados nos programas de melhoramento (GUPTA et al., 2005; ZHU et al., 2008; MYLES et al., 2009).

Nos estudos de Mapeamento por Associação é comum a utilização de duas estratégias. A primeira é chamada de *genome-wide* e tem como objetivo explorar todo o genoma de uma espécie por meio de um grande número de marcadores amplamente distribuídos, buscando identificar locos ou regiões genômicas associadas a um determinado gene (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003; MORGANTE; SALAMINI, 2003; ABDURAKHMONOV; ABDUKARIMOV, 2008). A segunda, chamada de *candidate-gene* (gene candidato), busca especificamente genotipar e explorar uma determinada região do genoma, onde possivelmente está localizado o gene de interesse (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003; MORGANTE; SALAMINI, 2003; ABDURAKHMONOV; ABDUKARIMOV, 2008).

No geral, o Mapeamento por Associação é um tipo abordagem utilizada para explorar a variação genética existente nas populações de determinada espécie nos programas de melhoramento. Esta técnica tem potencial para avaliar grande parte da diversidade alélica por

loco simultaneamente e pode aumentar a eficiência do mapeamento de alta resolução por causa da utilização de genótipos ou germoplasmas com vários eventos de recombinação gerados a partir de um grande número de meioses ocorridas ao longo da história de desenvolvimento da população de estudo (KHAN; KORBAN, 2012). Além disso, as associações detectadas são em populações geradas a partir de um grande número de cruzamentos e não somente para um ou poucos cruzamentos específicos (YU et al. 2006), ampliando o potencial do uso de marcadores moleculares em programas de melhoramento. Desse modo, o Mapeamento por Associação é uma técnica promissora para identificar QTLs que têm efeitos em um amplo espectro de germoplasmas. A técnica de Mapeamento por Associação baseia-se essencialmente em desequilíbrios de ligação existentes nas populações que ocorreram previamente e historicamente ao longo do tempo nos programas de melhoramento (RISCH; MERIKANGAS, 1996; NORDBORG; TAVARÉ, 2002; GEBHARDT et al., 2004; WRAY; VISSCHER, 2008). O desequilíbrio de ligação é a associação não ao acaso de alelos em diferentes locos, ocorrendo numa frequência maior que a esperada (SOTO-CERDA; CLOUTIER, 2012).

A recombinação é a principal força que reduz o desequilíbrio de ligação. Em uma determinada progênie, os indivíduos que a compõe recebem os alelos que estavam fisicamente ligados em seus pais (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003). Esta transferência de alelos fisicamente ligados é reduzida no decorrer de cada geração na proporção em que frequência de recombinação é aumentada até que entrem em equilíbrio de ligação (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003; GUPTA et al., 2005). Quanto menor a distância física e genética entre dois locos, maior é a probabilidade deles permanecerem ligados por muitas gerações na população, pois frequência de recombinação é menor (GUPTA et al., 2005). Diferentemente do genoma humano e de outros mamíferos, o genoma de plantas apresenta regiões com altas taxas de recombinação e outras regiões de longas sequências com baixa taxa de recombinação (FERREIRA; GRATTAPAGLIA, 2006).

Outros fatores interferem na identificação dos locos envolvidos no controle da expressão fenotípica de uma dada característica. Isto porque associações significativas podem ocorrer entre alelos de locos distantes, ou mesmo, situados em cromossomos diferentes (ABDURAKHMONOV; ABDUKARIMOV, 2008; HUTTLEY et al., 1999). Estas associações podem ser devidas a causas como: a presença de estrutura na população, seleção, mutação, deriva genética, migração e o grau de parentesco entre os indivíduos da população estudada, (WRAY; VISSCHER, 2008; KHAN; KORBAN, 2012). Esses fatores podem aumentar ou diminuir o desequilíbrio de ligação nas populações de plantas. O desequilíbrio de ligação é aumentado devido a uma nova mutação, ao sistema de acasalamento (autopolinização), ao

isolamento genético (por exemplo, quando ocorre diminuição ou ausência de fluxo gênico), a estrutura de população, a relação de parentesco entre os indivíduos, ao efeito fundador ou deriva genética, a seleção (natural ou artificial), a epistasia, dentre outras (GUPTA et al., 2005; ORAGUZIE et al., 2007). O desequilíbrio de ligação é reduzido quando se tem uma alta taxa de mutação e recombinação, polinização cruzada, etc. (GUPTA et al., 2005; ORAGUZIE et al., 2007).

Dentre os fatores que afetam o desequilíbrio de ligação, destacados acima, a estrutura da população e o grau da relação de parentesco são os principais responsáveis pelas associações falso-positivas (PRITCHARD, 2001; PARISSEAU; BERNARDO, 2004; YU et al. 2006). A estrutura populacional é considerada a principal causa de associação falso-positivas entre marcadores e caracteres fenotípicos (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003). Alguns pesquisadores verificaram que a maioria das populações de plantas apresenta algum grau de estruturação, o que resulta em correlações entre locos que não estão ligados fisicamente e geneticamente (MACKAY; POWELL, 2007). Isto ocorre, principalmente, devido à presença de frequências alélicas distintas com ascendência diferente em uma população misturada ou estruturada. Estudos realizados em germoplasmas de milho mostraram evidências de que a relação de parentesco causa desequilíbrio de ligação em locos que não estão associados com o fenótipo estudado (STICH et al., 2005). A relação de parentesco entre os indivíduos da população pode gerar tanto desequilíbrio de ligação entre locos ligados, quanto pode gerar também desequilíbrio de ligação entre os pares de locos não ligados, principalmente quando se têm vários grupos de plantas oriundos de diferentes cruzamentos (MACKAY; POWELL, 2007).

O sucesso do Mapeamento por Associação depende da capacidade de separar o desequilíbrio de ligação oriundo do efeito da recombinação, do desequilíbrio de ligação decorrente de outros fatores (REMYNGTON et al., 2001). Neste contexto, métodos estatísticos foram desenvolvidos visando corrigir ou eliminar a influência das associações falso-positivas nas análises. Uma das estratégias mais utilizadas no mapeamento por associação é verificar a existência de grupos geneticamente distintos dentro da população (estruturação) e, em seguida fazer a correção para o efeito de grupo intrapopulacional, para que então sejam testadas as associações marcador-fenótipo (PRITCHARD et al., 2000). O objetivo principal é eliminar os efeitos de estruturação, de modo que apenas as associações causadas pela ligação física e genética entre os alelos irão permanecer (MALOSETTI et al., 2007).

A utilização de modelos mistos em estudos de Mapeamento por Associação em plantas tem se mostrado eficiente e têm sido considerados um dos métodos mais eficazes para diminuir

as chances de associações espúrias entre marcadores e fenótipos (YU et al., 2006). No método de modelos mistos para análise de associação podem ser incorporadas as informações do *pedigree*, da estrutura populacional ou qualquer outra informação que tem o potencial para controlar as possíveis falsas associações (MALOSETTI, 2007). Atualmente, existem diversos programas que são utilizados nas análises de Mapeamento por Associação usando modelos mistos (ABDURAKHMONOV e ABDUKARIMOV, 2008; ZHU et al., 2008; SOTO-CERDA; CLOUTIER, 2012). Esses programas têm como foco principal avaliar o desequilíbrio de ligação, a presença ou não de estrutura na população, a relação de parentesco e a análise associativa propriamente dita.

Desde a sua introdução em estudos com plantas (THORNSBERRY et al., 2001), o Mapeamento de Associação tem ganhado destaque devido aos avanços nas tecnologias de genotipagem e à melhoria das análises estatísticas. Os primeiros estudos de Mapeamento por Associação foram realizados em milho (*Zea mays*) (BAR-HEN et al., 1995), arroz (*Oryza sativa*) (VIRK et al., 1996) e aveia (*Avena sativa*) (BEER et al., 1997). Com os avanços nas técnicas de biologia e genética molecular, associadas aos programas de análises estatísticas tem se ampliado os estudos em várias outras espécies como *Arabidopsis* (ARANZANA et al., 2005; ERSOZ et al., 2007; BRACHI et al., 2010; LI et al., 2010), soja (*Glycine max*) (WANG et al., 2008), trigo (*Triticum turgidum*) (LAIDO et al., 2014), dentre outras. Em cacaueteiro, existem três estudos utilizando Mapeamento por Associação, que objetivaram identificar marcadores associados a locos envolvidos no controle de caracteres de produção (PIRES, 2003; MARCANO et al., 2007; MARCANO et al., 2009).

2.4.3 Seleção genômica (SG)

Buscando aumentar a eficiência no processo de seleção de características de interesse, associado com os avanços de tecnologias de genotipagem e a descoberta de novos tipos de marcadores (SNPs e DArTs), uma nova estratégia de seleção denominada Seleção Genômica (GW) ou Seleção Genômica Ampla (GWS) foi implementada. Essa técnica foi proposta por Meuwissen et al. (2001) com o objetivo de estabelecer um método de predição precoce do fenótipo de uma população, baseando-se exclusivamente em informações de marcadores de DNA que tiveram seus efeitos genéticos previamente estimados e validados. Este método de seleção é utilizado com sucesso nos programas de melhoramento animal (SCHAEFFER, 2006; GODDARD; HAYES, 2007; LONG et al., 2007; CALUS et al., 2008, SOLBERG et al., 2008).

Em plantas, apesar de existirem poucos estudos usando a técnica de Seleção Genômica, sabe-se do potencial desta técnica para os programas de melhoramento, devido à capacidade de capturar genes de efeito maior e menor, aumentando a eficiência das análises de correlação entre o valor genético real e aquele estimado a partir de informações genotípicas e/ou fenotípicas para uma seleção precoce eficiente (HELFNER; SORRELS; JANNINK, 2009; GRATTAPAGLIA; RESENDE, 2011).

A Seleção Genômica distingue em determinados aspectos das técnicas de Seleção Assistida e Mapeamento por Associação (GRATTAPAGLIA; RESENDE, 2011). Primeiramente, é que na SG não precisa localizar os QTLs antes do processo de seleção e a mesma é viável mesmo na ausência de genes de efeito maior. Segundo, a SG possui potencial para explicar 100% da variação genética aditiva. Terceiro, a alta densidade de marcadores obtidos aumenta a eficiência na detecção de desequilíbrio de ligação (DL) na população, pois dado o grande número de marcadores moleculares gerados, a probabilidade de alguns desses estarem em DL com o(s) loco(s) que controla(m) um determinado caráter é alta.

A Seleção Genômica (SG) tem como proposta estimar simultaneamente (ou seja, não se faz teste de significância individual prévia do efeito dos marcadores) os efeitos genéticos de dezenas ou milhares de marcadores moleculares, com o objetivo de permitir capturar os efeitos de todos os locos, tanto de grande, quanto de pequeno efeito, envolvidos na variação genética de um dado caráter (MEUWISSEN et al., 2001). Para tanto, em nível populacional, é necessário que os alelos dos marcadores e os alelos dos genes que controlam uma determinada variação fenotípica estejam em desequilíbrio de ligação (MEUWISSEN; HAYES; GODDARD, 2001).

Partindo disso, a técnica de Seleção Genômica se destaca como uma importante ferramenta com potencial de permitir uma seleção de caracteres de interesse agrônomo confiável e de alta acurácia em programas de melhoramento genético, uma vez que pode ser aplicada em todas as linhagens ou famílias de uma espécie num programa de melhoramento (MEUWISSEN; HAYES; GODDARD, 2001; RESENDE et al., 2008; HEFFNER; SORRELS; JANNINK, 2009; JANNINK; LORENZ; IWATA, 2010; GRATTAPAGLIA; RESENDE, 2011). Esta técnica geralmente tem utilizado marcadores moleculares codominantes, como os SNPs e os SSR, e marcadores dominantes reprodutíveis como os DArTs.

3. CAPÍTULO 1

Caracterização da compatibilidade sexual em uma população de melhoramento de cacaueteiro (*Theobroma cacao* L.)¹

Marcos Ramos da Silva, Wilson Monteiro, Didier Clément, Uilson Lopes.

Resumo

A incompatibilidade sexual é um fator limitante da produção potencial de cacaueteiro, sendo um importante critério de seleção em programas de melhoramento. Em estudos anteriores, alguns autores apontam para uma possível existência de efeito quantitativo na incompatibilidade e que esta pode influenciar na determinação e seleção de genótipos autocompatíveis. Partindo disso, o presente estudo buscou avaliar como determinar a compatibilidade/incompatibilidade em cacaueteiro por meio da retenção floral e compreender este fenômeno tratando-o tanto como variável quantitativa contínua quanto como uma variável qualitativa discreta. Para isto, foi utilizado o método de polinização manual controlada de 333 genótipos (66 germoplasmas e 267 descendentes resultantes de diferentes cruzamentos), caracterizados quanto à compatibilidade/incompatibilidade sexual, em uma população do programa de melhoramento do CEPEC. Esta população também vem sendo usada para estudos de mapeamento de genes de incompatibilidade. Levando-se em consideração a influência dos fatores ambientais, foram feitas polinizações cruzadas com um genótipo autocompatível, usado como testemunha. No total, 9.147 flores foram polinizadas e avaliadas em relação ao pegamento aos 7, 15, 21 e 30 dias após as polinizações. A partir dos resultados obtidos foi possível constatar um gradiente de retenção floral de 0 a 100%. Porém, houve uma predominância de genótipos com baixo pegamento, afetando drasticamente o percentual de flores retidas na população, podendo ser resultado de uma alta frequência de alelos de incompatibilidade na população. Por outro lado, genótipos com alta retenção de flores (90-100%) foram identificados e estes possuem potencial para serem selecionados com o objetivo de aumentar a frequência de alelos de compatibilidade

¹ Autorização CGEN No. 10/2012 (Processo 02001.007611/2012-46)

no programa de melhoramento. Utilizando um teste de Qui-Quadrado foi possível determinar que 45 clones (incluindo acessos do grupo Scavina, TSH, Parinari, Pound e IMC) e 206 descendentes são autoincompatíveis e 21 clones (incluindo acessos do grupo CCN, BE, ICS, SIC e SIAL) e 61 descendentes são autocompatíveis. Além disso, quatro padrões de comportamento em relação à retenção floral foram detectados. Dois padrões (grupos 1 e 2) apresentaram alto pegamento, um deles possuindo um tipo semelhante ao do controle. Os outros dois padrões (grupo 3 e 4) apresentaram um baixo pegamento. Levando-se em consideração a variação contínua do percentual de flores retidas pelos genótipos da população estudada e os diferentes padrões detectados, foi possível concluir que a compatibilidade nesta espécie deve ser tratada como uma variável quantitativa contínua. Por outro lado, a incompatibilidade deve ser tratada como uma variável qualitativa discreta.

Palavras-chave: Cacau; Retenção floral; Padrões de comportamento; Variável quantitativa; Variável qualitativa.

Introdução

O cacauieiro (*Theobroma cacao* L.) é uma espécie perene, preferencialmente alógama (POUND, 1938; TOXOPEUS, 1972). O mecanismo de polinização natural de cacauieiro possui aspectos peculiares, a estrutura da flor dificulta a polinização. Primeiro, os grãos de pólen são muito pegajosos para serem transportados facilmente pelo vento. Segundo, as anteras são protegidas pelas pétalas que formam a cógula ao redor da mesma, dificultando a dispersão do pólen. Terceiro, existe um círculo de estaminóides que rodeiam o pistilo, protegendo o contato com o estigma, o que limita ainda mais as chances de autofecundação. Por isso, nesta espécie, o processo de transferência de pólen das anteras para o estigma fica quase limitado à atividade de agentes polinizadores específicos, geralmente, espécies de moscas do gênero *Forcipomyia* sp. (BILLES, 1942; SORIA et al., 1975). Quarto, na espécie a incompatibilidade sexual é frequentemente encontrada entre os germoplasmas. Como resultado destes e de outros fatores, muitas das flores produzidas não são eficientemente polinizadas ou fecundadas e somente 2-5% das flores chegam a produzir frutos (ALVIM, 1984; LEÓN, 2000).

A incompatibilidade sexual em cacauieiro é a capacidade de uma planta com flores férteis rejeitar o próprio pólen (autoincompatibilidade) ou o pólen de plantas que são geneticamente semelhantes aos alelos presentes no loco *S* (incompatibilidade cruzada), levando a uma falha na formação do zigoto e do endosperma (COPE, 1962; SEAVEY; BAWA, 1986). No cacauieiro, este fenômeno foi observado pela primeira vez por Harland (1925), em Trinidad, onde verificou-se diferenças entre plantas de cacauieiro quanto ao número de frutos, e que em alguns casos não chegavam a produzi-los. Pound (1932) realizou o primeiro estudo de

incompatibilidade em cacau e descobriu que flores de determinados genótipos apresentavam tanto autoincompatibilidade, quanto incompatibilidade cruzada.

A presença da incompatibilidade sexual em populações de cacau apresenta aspectos negativos tanto para os melhoristas, como para os produtores de cacau. Para o melhorista, a incompatibilidade limita a realização de autofecundações e de determinados cruzamentos, necessários à produção de progênes de interesse. Para o produtor, a incompatibilidade causa a redução na produtividade, uma vez que variedades incompatíveis produzem um baixo número de fruto por planta e baixo número de sementes por fruto (POSNETTE, 1940; MORERA et al., 1994; VIEIRA; MULLER, 2001; PIRES et al., 2012). Em estudos recentes em populações de cacau na Bahia, foi estimada uma perda de 38% na produção de frutos associada á incompatibilidade (PIRES et al., 2012). Estudos sobre a caracterização da compatibilidade sexual podem contribuir para a geração de variedades de cacau capazes de produzir mais frutos por planta e maior número de sementes por fruto. Entretanto, há poucos estudos em cacau abordando a caracterização da incompatibilidade em populações de melhoramento.

Alguns estudos propõem a existência de gradações na expressão do fenótipo da compatibilidade/incompatibilidade sexual em cacau. Segundo Cheesman (1938), a existência de diferentes níveis de compatibilidade sexual apresentados pelo cacau pode ser responsável pelas grandes alterações dos níveis de produção. Cope (1939) propôs que o sucesso das polinizações artificiais é variável de material para outro, devido ao fato de alguns destes apresentarem diferentes níveis de compatibilidade. Coral (1970) sugeriu que existem diferentes graus de incompatibilidade, e que estes são controlados por genes. Carletto (1972) e Carletto e Soria (1973) realizaram um estudo visando comprovar a existência de graus de incompatibilidade utilizando os métodos de abscisões e pegamento de Knight e Rogers (1955) e de cortes histológicos de Cope (1962). Porém, nos estudos realizados por esses autores não foi encontrada nenhuma comprovação genética para a proposta de Coral (1970) sobre a existência de graus de incompatibilidade em cacau. Por outro lado, Lanaud et al (1987) verificaram, por meio da mistura de pólen incompatível e compatível, que existem variações na percentagem de sementes geradas a partir de autofecundação, propondo a existência de um efeito quantitativo na incompatibilidade em cacau.

Porém, até o momento, os métodos de caracterização da compatibilidade/incompatibilidade em cacau, tratam esse fenômeno como uma variável meramente qualitativa (BARTLEY; COPE, 1973; LÓPEZ, 1982; PINTO et al., 1998; LEAL, 2005; CASTILHO, 2005), o que tem levado a divergências nos resultados encontrados pelos

diferentes métodos usados para determinar se um dado genótipo é autocompatível ou autoincompatível, levando portanto a erros no processo de selecionar genótipos para um alto pegamento. De modo geral, a determinação da compatibilidade/incompatibilidade de um dado genótipo é realizada através da percentagem de retenção/abscisão floral após polinização manual controlada (CARLETTO, 1946; HARDY, 1961). Para cada genótipo, são polinizadas entre 15 e 40 flores (LOPES; CARLETTO, 1995). Posteriormente, avaliações em relação à retenção floral são feitas, geralmente, no terceiro, sétimo e 15º dia após polinização (TERREROS; CHAVARRO; OCAMPO, 1982; YAMADA et al., 1982; YAMADA; BARTLEY, 1984). Porém, existem divergências na determinação de um limiar de retenção floral para fazer inferência sobre a compatibilidade de um determinado genótipo. López (1982) considera que um material é autocompatível se este apresentar um índice de retenção floral de no mínimo 2%. Por outro lado, autores como Bartley e Cope (1973), Terreros, Chavarro e Ocampo (1982), Lopes e Carletto (1995) e León (2000), consideram que esse índice de retenção deva ser de 5%. Entretanto, para Pinto et al. (1998) é necessário um índice de retenção de pelo menos 10%. Outros autores ainda recomendam outros limiares de retenção como 14% (LEAL, 2005) e até 40% (CASTILHO, 2005). Essa divergência entre os autores sobre os diferentes limiares aponta para indícios que a compatibilidade/incompatibilidade deve ser tratada como uma variável quantitativa contínua, e não como uma variável meramente qualitativa discreta.

Assim, o presente estudo buscou identificar o melhor critério para determinar a incompatibilidade por meio da retenção floral e verificou se este fenômeno deve ser tratado como uma variável qualitativa discreta ou quantitativa contínua.

Materiais e Método

População de estudo

Trezentos e trinta e três genótipos de cacauzeiro foram utilizados no presente estudo. Esses genótipos incluem 66 acessos da coleção de germoplasmas (clones fundadores) e 267 indivíduos descendentes daqueles acessos, selecionados ao acaso dentro de 109 progênes da terceira geração de seleção recorrente (ensaios TP-0221, TP-0222, TP-0323 e TP-0324), do programa de melhoramento do Centro de Pesquisas do Cacau (CEPEC), da Comissão

Executiva do Plano de Lavoura Cacaueira (CEPLAC), em Ilhéus, Bahia, Brasil. Os ensaios TP-0221, TP-0222, TP-0323 e TP-0324 foram estabelecidos em 2002 e 2003, com 41, 15, 30 e 23 progênies, respectivamente. Os dois primeiros ensaios foram estabelecidos sem repetições com parcelas de 75 e 80 plantas em média respectivamente. Já os dois últimos ensaios foram estabelecidos segundo um delineamento em blocos casualizados, com 3 repetições de 20 plantas/parcela e 4 repetições de 15 plantas/parcela, respectivamente. As progênies dos quatro ensaios foram geradas segundo um delineamento fatorial bastante desbalanceado.

Caracterização fenotípica da incompatibilidade sexual

Polinizações artificiais foram realizadas nos 333 genótipos entre os meses de abril de 2013 e junho de 2014. Dois tipos de polinizações foram feitas: autopolinizações e polinizações cruzadas (controle). Para caracterizar a incompatibilidade na população de estudo, 7.256 flores foram autopolinizadas (10-57 flores por genótipo). Como controle, foram feitas 1.891 polinizações cruzadas (10-25 flores por dia), utilizando-se pólen de clones reconhecidamente autocompatíveis (CCN-51, SIC-17 e SIC-864), visando isolar o efeito de outros fatores causadores de queda de flores, não associados à incompatibilidade (chuva, insetos).

Para evitar polinizações indesejadas, os botões florais entumecidos, prestes a abrir no dia seguinte, foram protegidos no dia anterior à polinização artificial com um pequeno tubo de mangueira transparente, de 1 polegada de diâmetro e 7 cm de comprimento, cuja extremidade final era coberta por uma tela fina (tecido fino denominado de voal, preso por uma liga de borracha), e a outra extremidade aberta foi fixada na planta por meio de massa de modelar, a fim de proteger os botões. Processo este similar ao usado por outros autores (KNIGHT; ROGERS, 1955; CARLETTO, 1946; HARDY, 1961; TERREROS; CHAVARRO; ROJAS, 1982; LOPES; CARLETTO, 1995; YAMADA et al., 2010; ROYAERT et al., 2011).

Na manhã do dia seguinte, entre 8 e 11 horas da manhã, depois da retirada do tubo, os botões florais abertos foram polinizados artificialmente. Para isto, na flor receptora que se encontrava completamente aberta e o estigma receptivo, foi realizado a retirada dos estaminóides, utilizando pinça, para facilitar a polinização. Algumas flores da mesma planta, também protegidas, foram selecionadas como doadoras de pólen, desde que suas anteras estivessem liberando o pólen. Com a pinça, era retirada a cógula da flor doadora de pólen para liberar os estames e, em seguida, as anteras com pólen eram esfregadas suavemente no estigma da flor receptora. Cada flor polinizada foi novamente protegida e identificada por um alfinete

com uma fita colorida. Os protetores foram retirados um dia após a polinização, uma vez que a fertilização ocorre entre 10 e 24 horas após a polinização. Para verificar o índice de pegamento, avaliações foram realizadas 7, 15, 21 e 30 dias após as polinizações. A caracterização da incompatibilidade sexual foi realizada com base na percentagem de retenção das flores autopolinizadas, comparada com a retenção das polinizações cruzadas.

Análises estatísticas

Dois critérios foram utilizados para a caracterização dos genótipos em relação à incompatibilidade sexual. No primeiro critério (Método 1), as percentagens de flores retidas aos 15 dias após as autopolinizações e polinizações cruzadas (controle), foram comparadas estatisticamente entre si por meio do teste de Qui-quadrado, buscando testar a proporção de 1:1, a 5% de probabilidade com 1 grau de liberdade (STEEL; TORRIE, 1980):

$$\chi^2 = [(n_{11}n_{22} - n_{12}n_{21})^2 n_{..}] / [n_{1.}n_{.2} \cdot n_{.1}n_{.2}]$$

onde,

n_{11} é o número a flores de autopolinizadas;

n_{22} é o número de flores com polinização cruzada que sofreram abscisões;

n_{12} é o número flores autopolinizadas que sofreram abscisões;

n_{21} é o número de flores com polinização cruzada.

No segundo critério (Método 2), a incompatibilidade foi caracterizada considerando-se somente o pegamento, aos 15 dias, nas autopolinizações (ou seja a retenção floral no controle foi ignorada). Foram considerados como autoincompatíveis genótipos com um pegamento inferior a um dado percentual, variando de 1 a 60% . Este é o método comumente usado para declarar a compatibilidade de um dado genótipo em cacauero (BARTLEY; COPE, 1973; TERREROS; CHAVARRO; OCAMPO, 1982; LOPES; CARLETTO, 1995; LEÓN, 2000). Os resultados oriundos com as análises dos Métodos 1 e 2, foram estatisticamente comparados pelo teste de McNemar (CAMPOS, 1983), a 0,05 de probabilidade, assumindo diferentes percentuais de corte, ou limiar, no Método 2 variando de 1 a 60%.

Outra análise foi realizada buscando explorar o comportamento de abscisão floral ao longo tempo. Para isto, uma análise multivariada com variáveis canônicas foi feita utilizando-se procedimento CANDISC implementado no programa SAS (SAS, 1989). Esta análise permitiu agrupar os genótipos da população estudada em relação aos diferentes comportamentos de retenção floral aos 7, 15, 21 e 30 dias após as polinizações.

Resultados

Retenção floral

Nas polinizações cruzadas com um genótipo autocompatível, utilizadas como controle, ao longo dos 15 meses, observou-se uma grande variação na retenção floral, com uma amplitude de 59,72% (Figura 1.1). Isto certamente influencia a caracterização de um genótipo como autocompatível/incompatível quando a retenção floral de um controle (ou seja, fatores não relacionados à incompatibilidade) é ignorada (Método 2). Existem dias com alto pegamento, como aqueles dos meses de outubro a dezembro de 2013; outros dias com baixo pegamento como aqueles dos meses de abril de 2013 e 2014.

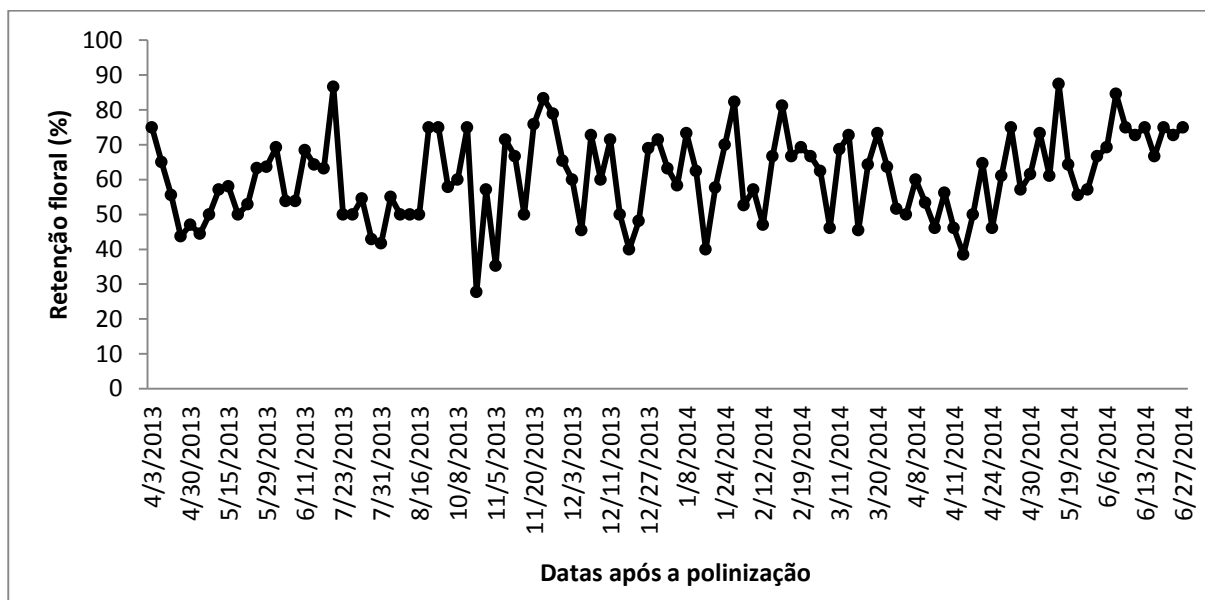


Figura 1.1- Percentagem de retenção floral aos 15 dias resultantes das polinizações cruzadas com pólen de clones autocompatíveis de cacauero no período de execução do estudo (Abril/2013 a Junho/2014).

Dentre as 7.256 flores autopolinizadas na população de estudo, a maioria sofreu uma queda brusca até os 7 dias após as polinizações (Figura 1.2). Nas polinizações cruzadas, esse percentual de queda foi menor. Por outro lado, em ambos os casos a taxa de queda reduziu drasticamente após os 7 dias (Figura 1.2). A retenção floral caiu de 100% para aproximadamente 60-70% nos cruzamentos com o genótipo autocompatível, e para aproximadamente 20% nas autofecundações. Nas comparações entre as percentagens de flores retidas até os 15 dias das autopolinizações e das polinizações cruzadas foi possível constatar que há uma variação grande na retenção floral das autofecundações, maior que a variação do controle (Figura 1.3).

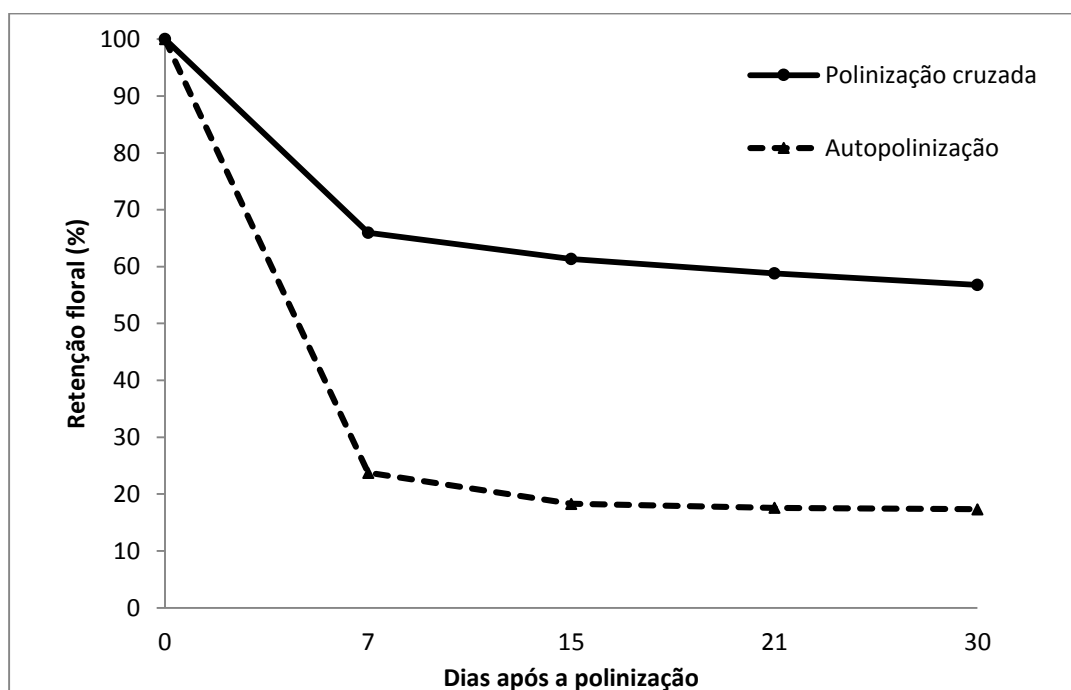


Figura 1.2 - Percentagem de flores retidas na população de cacaueteiro em estudo aos 7, 15, 21 e 30 dias após as polinizações.

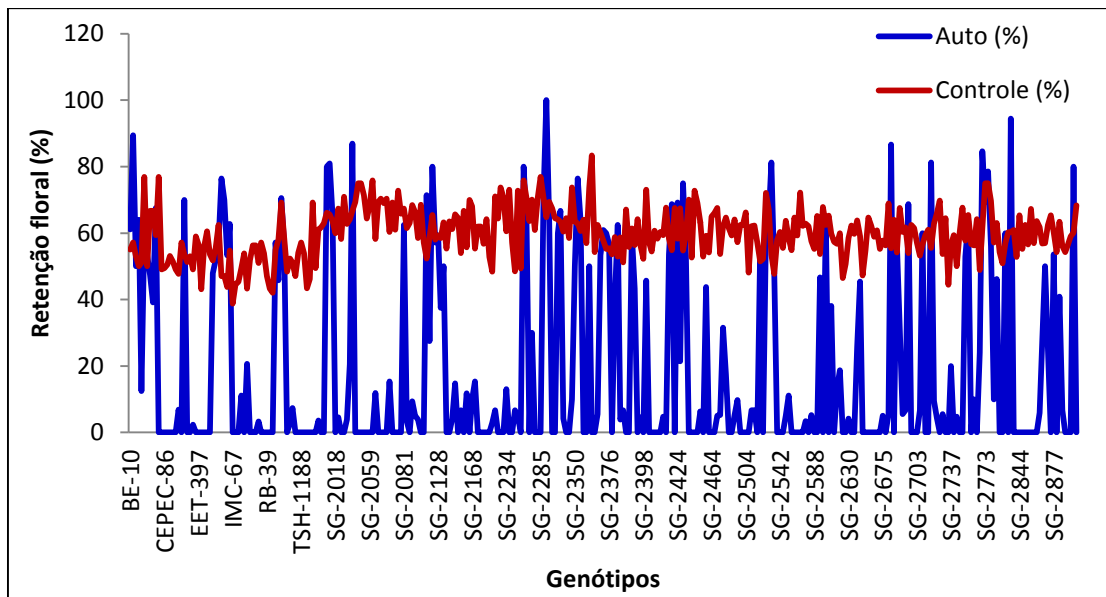


Figura 1.3 - Percentagem de flores retidas na população de cacaueteiro em estudo 15 dias após as polinizações.

Com o objetivo de identificar os genótipos autocompatíveis na população de estudo, foi feito um teste de Qui-quadrado. Com base neste teste foi possível verificar que dentre os 66 germoplasmas, 45 foram considerados autoincompatíveis e 21 autocompatíveis (Tabela 1.1). Germoplasmas dos grupos Be, ICS, CCN, CC, SIAL, SIC e SGu foram identificados como autocompatíveis; por outro lado, germoplasmas dos grupos TSH, TSA, Scavina, C.Sul, IMC, Na, Pound e Parinari foram todos considerados autoincompatíveis. Nos descendentes ocorre também a predominância da incompatibilidade, isto é, dentre os 267 genótipos, 206 foram autoincompatíveis e apenas 61 autocompatíveis (resultados não mostrados).

Posteriormente foi aplicado um teste (Teste de Mc Nemar) para identificar o limiar de retenção floral, levando em consideração os métodos 1 (usando teste estatístico) e 2 (usando um limiar de retenção floral) de classificação de genótipos em relação a incompatibilidade. Esse teste permitiu verificar que o menor número de trocas (3-8 trocas) entre os dois métodos ocorreu quando o limiar (*threshold*) utilizado no método 2 fica entre 32-45% (Figura 1.4).

Tabela 1.1-Retenção floral de 66 acessos de cacauero do banco de germoplasma do CEPEC/CEPLAC e das polinizações cruzadas (controle) aos 7, 15, 21 e 30 dias após a polinização

Germoplasma	N° Flores polinizadas		N° Flores retidas		χ^2	p-valor	Retenção floral (%)		Classificação Método 1
	Auto	Cruz	Auto	Cruz			Auto	Cruz	
Be-10	36	89	22	49	0,38	0,54	61,11	55,06	C
Be-4	19	14	17	8	4,59	0,03	89,47	57,14	C
Be-8	16	17	8	9	0,03	0,87	50,00	52,94	C
Be-9	25	20	16	10	0,89	0,34	64,00	50,00	C
CA-4	48	90	6	46	19,87	0,00	12,50	51,11	I
CA-5	25	13	15	10	1,09	0,30	60,00	76,92	C
CAB-28	30	40	16	20	0,08	0,78	53,33	50,00	C
CC-10	19	12	9	8	1,11	0,29	47,37	66,67	C
CCN-10	23	12	9	8	2,39	0,12	39,13	66,67	C
CCN-51	37	32	25	19	0,50	0,48	67,57	59,38	C
CEPEC-42	22	13	0	10	23,69	0,00	0,00	76,92	I
CEPEC-52	32	53	0	26	22,62	0,00	0,00	49,06	I
CEPEC-86	35	95	0	47	27,12	0,00	0,00	49,47	I
CEPEC-89	40	97	0	49	31,46	0,00	0,00	50,52	I
CEPEC-90	35	81	0	43	29,52	0,00	0,00	53,09	I
CEPEC-94	29	68	0	35	23,35	0,00	0,00	51,47	I
C.Sul-3	31	55	0	27	22,18	0,00	0,00	49,09	I
C.Sul-7	29	44	2	21	13,50	0,00	6,90	47,73	I
C.Sul-8	19	14	0	8	14,33	0,00	0,00	57,14	I
EEG-29	30	28	21	15	1,66	0,20	70,00	53,57	C
EET-156	44	113	0	58	35,82	0,00	0,00	51,33	I

EET-377	41	115	0	61	35,71	0,00	0,00	53,04	I
EET-390	42	53	1	26	25,09	0,00	2,38	49,06	I
EET-392	40	78	0	46	38,66	0,00	0,00	58,97	I
EET-397	51	39	0	22	38,08	0,00	0,00	56,41	I
EET-399	38	51	0	22	21,77	0,00	0,00	43,14	I
EET-400	23	25	0	14	18,18	0,00	0,00	56,00	I
EET-45	33	76	0	46	34,56	0,00	0,00	60,53	I
EET-59	45	84	0	45	37,02	0,00	0,00	53,57	I
EET-94	48	58	23	30	0,15	0,70	47,92	51,72	C
ICS-1	25	18	13	10	0,05	0,82	52,00	55,56	C
ICS-32	28	40	16	25	0,20	0,66	57,14	62,50	C
ICS-6	17	17	13	8	3,11	0,08	76,47	47,06	C
ICS-9	20	17	14	8	2,01	0,16	70,00	47,06	C
ICS-95	15	16	8	7	0,28	0,59	53,33	43,75	C
ICS-98	43	53	27	29	0,64	0,42	62,79	54,72	C
IMC-67	36	49	0	19	17,98	0,00	0,00	38,78	I
IMC-76	42	56	0	25	25,17	0,00	0,00	44,64	I
Ma-15	30	64	0	29	19,66	0,00	0,00	45,31	I
Na-33	27	77	3	38	12,24	0,00	11,11	49,35	I
P-12	25	39	0	21	20,04	0,00	0,00	53,85	I
P-7	29	67	6	29	4,46	0,03	20,69	43,28	I
Pa-169	18	27	0	14	13,55	0,00	0,00	51,85	I
Pa-285	22	32	0	18	18,56	0,00	0,00	56,25	I
Pa-30	18	32	0	18	15,82	0,00	0,00	56,25	I
Pa-300	30	51	1	26	19,30	0,00	3,33	50,98	I
RB-30	24	14	0	8	17,37	0,00	0,00	57,14	I
RB-36	14	39	0	21	12,49	0,00	0,00	53,85	I
RB-39	11	17	0	8	7,25	0,01	0,00	47,06	I
Sca-12	33	67	0	29	20,12	0,00	0,00	43,28	I
Sca-6	35	50	0	21	19,52	0,00	0,00	42,00	I

SGu-54	14	51	8	28	0,02	0,88	57,14	54,90	C
SIAL-70	24	31	11	18	0,81	0,37	45,83	58,06	C
SIC-19	17	13	12	9	0,01	0,94	70,59	69,23	C
SIC-864	32	29	18	17	0,03	0,85	56,25	58,62	C
TSA-641	35	64	0	31	24,68	0,00	0,00	48,44	I
TSA-644	37	63	1	33	25,64	0,00	2,70	52,38	I
TSA-654	41	90	3	46	23,07	0,00	7,32	51,11	I
TSA-656	12	17	0	8	7,80	0,01	0,00	47,06	I
TSA-792	57	98	0	53	46,84	0,00	0,00	54,08	I
TSH-1188	31	56	0	32	28,02	0,00	0,00	57,14	I
TSH-516	16	13	0	7	11,36	0,00	0,00	53,85	I
TSH-565	33	53	0	23	19,55	0,00	0,00	43,40	I
TSH-774	38	78	0	36	25,43	0,00	0,00	46,15	I
U-10	17	13	0	9	16,81	0,00	0,00	69,23	I
UF-296	38	87	0	43	28,63	0,00	0,00	49,43	I

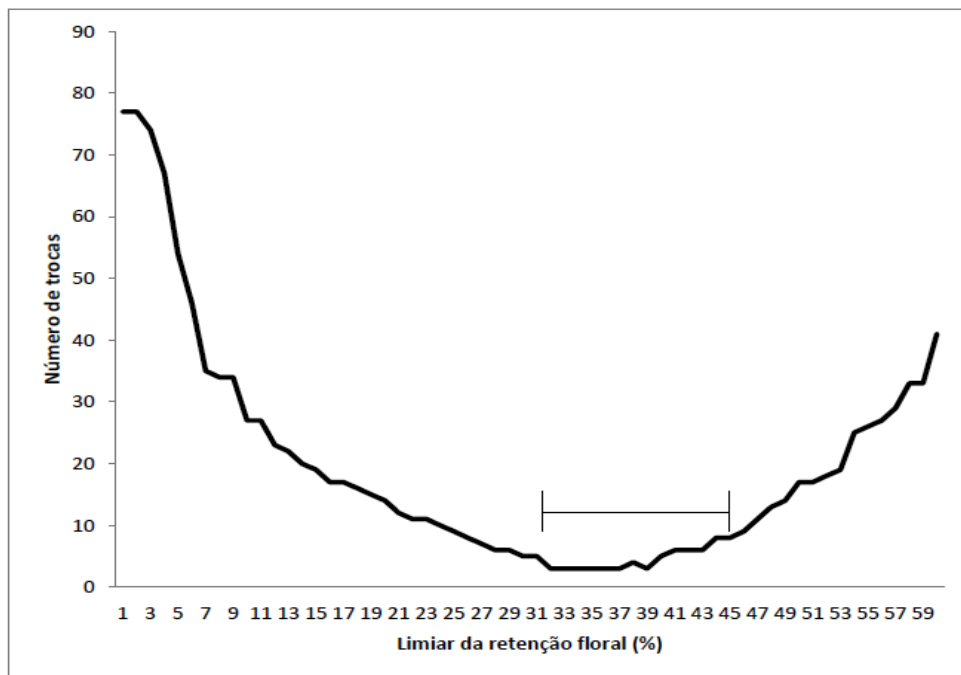


Figura 1.4 - Número de trocas na classificação de genótipos de cacaueteiro quanto à incompatibilidade, usando-se os métodos 1 e 2. A barra horizontal mostra a faixa de retenção floral que os resultados produzidos pelos métodos 1 e 2 não são estatisticamente diferentes pelo teste de McNemar ($p=0,05$).

Outro aspecto considerado neste estudo foi a exploração da variação gradual existente nas percentagens de pegamento na população estudada. Quinze dias após as autopolinizações foi observado um padrão similar na curva de distribuição entre germoplasmas e os descendentes referente à retenção floral (Figura 1.5a e b). Tanto os acessos do banco de germoplasma do CEPEC, quanto seus descendentes tiveram um grande número de genótipos (37 germoplasmas e 137 descendentes) que não conseguiram reter nenhuma flor (0%) e com grupos menores de genótipos distribuídos em diferentes classes, ou seja, com diferentes percentagens de pegamento que variaram de 1-100%. A presença de muitos genótipos com 0% de pagamento explica também o baixo percentual de retenção floral das autopolinizações observada na Figura 1.2. Alguns genótipos tiveram retenção floral próxima ou igual a 100%, logo, com potencial para produzir variedades com alta retenção floral.

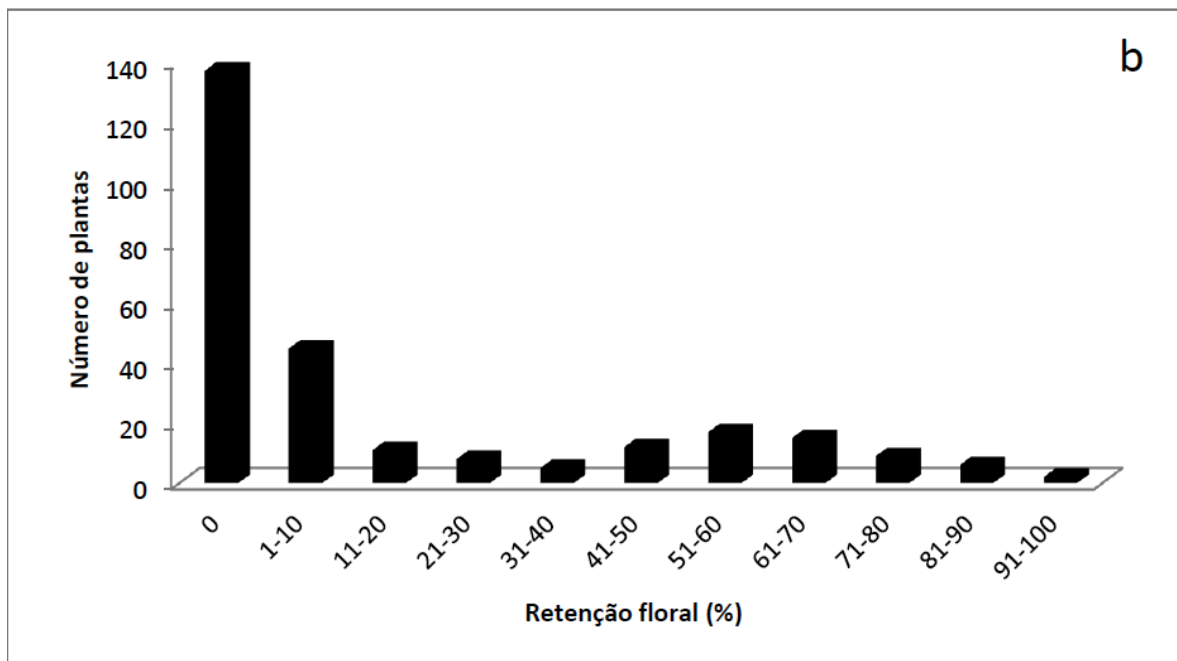
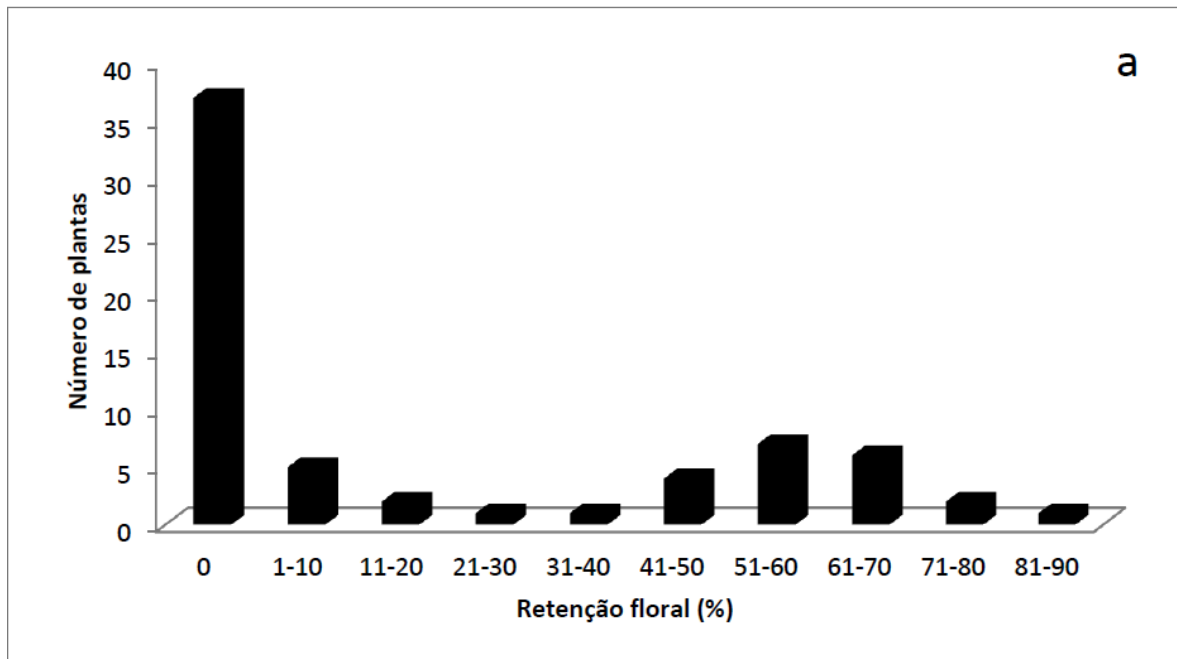


Figura 1.5 - Histograma de frequência do número de germoplasmas (a) e de plantas das progênes (b) de cacaueteiro alocadas em diferentes classes de retenção floral aos 15 dias (66 germoplasmas e 267 plantas das progênes).

Retenção floral ao longo do tempo

Na população estudada (66 germoplasmas e 267 descendentes) foram observados diferentes níveis de comportamento referente à retenção floral. Sete dias após as polinizações houve uma acentuada queda na retenção floral nos vários genótipos estudados, entretanto, a

retenção floral final nestes 333 genótipos varia de 0 a 100%, de modo quase contínuo, entretanto, com uma clara descontinuidade próximo de 30% de retenção (Figura 1.6). Aos 15, 21 e 30 dias após as autopolinizações, independente da percentagem de flores retidas, a queda foi estabilizada, ou seja, não houve alteração acentuada nos valores do pegamento.

Observa-se também na Figura 1.6, que há diferentes padrões de retenção floral ao longo do tempo de avaliação após a polinização. Em alguns genótipos há uma queda suave na retenção floral entre o tempo zero e os 7 dias após a polinização, resultando em genótipos com alta retenção floral. Em outros genótipos há uma drástica queda da retenção floral, no mesmo período (0-7 dias), resultando em uma baixa retenção floral final. Entretanto, há um grupo de 29 genótipos com comportando anômalo em relação aos dois outros tipos, nos quais foi observada uma queda na retenção floral até os 7 dias, mas o percentual de retenção continuou caindo até os 15 dias, apresentando uma reação de incompatibilidade mais tardia.

Por meio da análise de agrupamento foi possível detectar quatro grupos com padrões de comportamento estatisticamente diferentes (Figura 1.7). Os grupos 1, 2 e 3 se distinguem pela existência genótipos com diferentes níveis de compatibilidade em relação à retenção floral. O grupo 4 foi formado essencialmente por genótipos autoincompatíveis, ou seja, com 0% compatibilidade em relação a retenção floral. O grupo 1 foi formado por 5 acessos (Be-4, EEG-29, ICS-9, ICS-6 e SIC-19) e 21 descendentes e teve uma retenção floral de aproximadamente 80%, sendo assim candidatos à seleção em relação a alta retenção floral. O grupo 2 foi formado por 16 acessos (ex. Be-8, Be-9, Be-10, ICS-1, ICS-32, ICS-95, ICS-98, CCN-51e CCN-10) e 43 descendentes, com retenção floral de aproximadamente 60%, e não é estatisticamente diferente do controle autocompatível. O grupo 3 foi formado por 6 acessos (Ca-4; C.Sul-8; EET-45; Na-33; P-7) e 31 descendentes, com uma retenção floral de aproximadamente 30% até o 7º dia, que reduziu para 15% no 15º dia e estabilizou-se (Figura 1.7), sendo observado um tipo de reação tardia quanto à incompatibilidade, que se prolonga após o 7º dia após a polinização. O grupo 4 foi formado por 39 acessos (ex. IMC-67, RB-30, Sca-6, Sca-12, TSH-1188 e TSH-516) e 172 descendentes, com uma queda drástica na retenção floral até o 7º dia, chegando a 0% aos 15 dias, ou seja, são genótipos que não apresentaram nenhum de nível de compatibilidade.

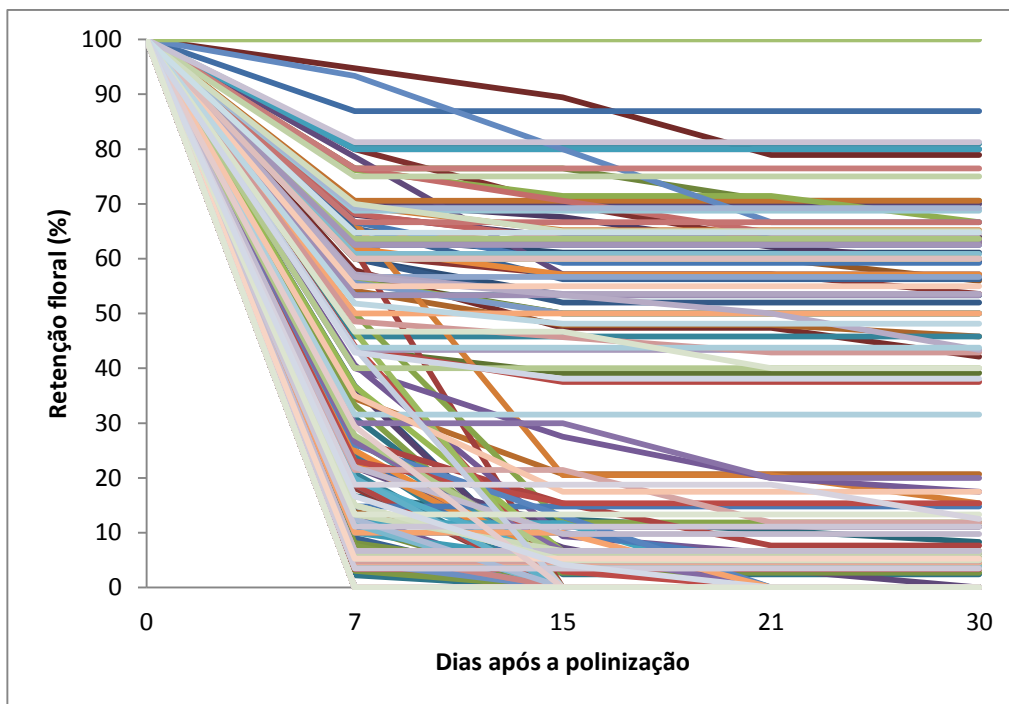


Figura 1.6 - Padrões de retenção floral ao longo do tempo após a polinização com genótipos (germoplasmas e progênies) de cacauero.

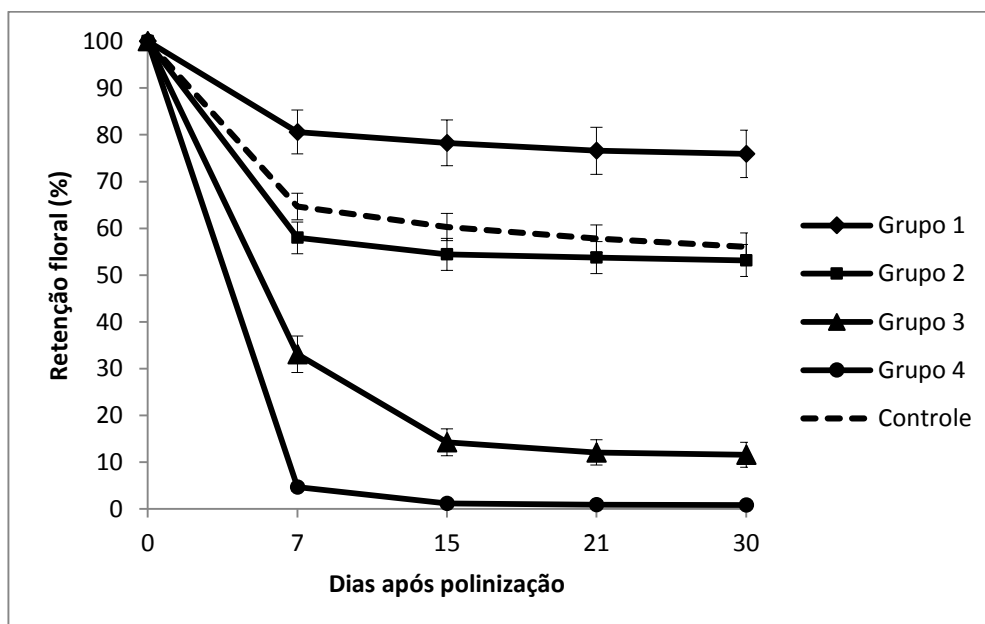


Figura 1.7- Retenção floral média de genótipos de cacauero (germoplasmas e progênies) alocados em quatro grupos e do controle (polinizações cruzadas com pólen autocompatível), após uma análise multivariada de agrupamento.

Discussão

Neste estudo, os resultados encontrados permitem uma melhor compreensão sobre a incompatibilidade no cacaueteiro e oferece possíveis alternativas para caracterizar esse fenômeno na espécie. Para isto, é necessário levar em consideração alguns aspectos importantes. O primeiro deles é que o ambiente afeta a retenção floral (ex. fatores sazonais ou alterações fisiológicas no órgão reprodutor), o que pode severamente afetar os resultados da caracterização da incompatibilidade ou a avaliação da retenção floral, se controles não são incluídos. Isto é válido também para os estudos que buscam associação de marcadores com a incompatibilidade, neste caso é recomendável fazer um ajuste das retenções das autopolinizações em relação às polinizações cruzadas (controle). Segundo, dado a ampla variação da retenção floral de diversos genótipos, não é fácil caracterizar um indivíduo como autocompatível/autoincompatível, com base em um percentual de corte ou limiar. Por outro lado, métodos estatísticos (ex. qui-quadrado) que comparam a retenção floral de um dado genótipo com aquela de genótipo reconhecidamente autocompatível, como foi feito neste estudo (Método 1), pode ser uma solução para minimizar os erros de determinação de genótipos quanto à incompatibilidade. Entretanto, na limitação ou impossibilidade de incluir polinizações controle (pelo custo ou pela pouca disponibilidade de flores), aparentemente um limiar/ponto de corte pode ser adotado para análise do caráter, mas esse limiar precisa ser maior (32-45%), em vez do que vem sendo adotado de 2% por vários autores (LÓPEZ, 1982), 5% (BARTLEY; COPE, 1973; TERREROS; CHAVARRO; OCAMPO, 1982; LOPES; CARLETTO, 1995; LÉON, 2000), 10% (PINTO et al., 1995) e 14% (LEAL, 2005).

Neste trabalho, a compatibilidade/incompatibilidade foi caracterizada utilizando-se um critério que não usa um limiar para determinar a compatibilidade como os métodos anteriores, mas sim testes estatísticos de probabilidade. Por esse método, a compatibilidade de um genótipo é declarada por meio da comparação estatística do percentual de flores retidas nas autopolinizações em comparação com uma testemunha. Com este método foi possível identificar 251 genótipos autoincompatíveis e 82 autocompatíveis, na população estudada. Dentre estes estão 45 germoplasmas autoincompatíveis e 21 autocompatíveis. Estes resultados entram em consenso com os relatos de outros pesquisadores (POSNETTE, 1945; BARTLEY, 1970; CORAL; SORIA, 1974; YAMADA et al., 1982; ARCINIEGAS; PHILLIPS-MORA, 2006). Porém, divergem de resultados encontrados por autores (BARTLEY, 1958; ERIQUEZ; SORIA, 1967; KKO, 1993). Esta divergência é resultado da ampla variação dos

métodos de determinação de genótipos autocompatíveis, devido ao uso de diferentes limiares da retenção floral.

Embora muitos dos acessos do banco de germoplasmas possam ser autoincompatíveis com base no método 1, alguns vem sendo usados mais intensamente no programa de melhoramento do CEPEC em virtude de sua resistência a vassoura-de-bruxa (TSH, Sca, etc.). Além disso, esses acessos são de diversas origens (Brasil, Trinidad, Peru) e sabendo que a incompatibilidade em cacaueteiro é controlada geneticamente por um loco com 5 ($S_1 > S_2 = S_3 > S_4 > S_5$) (KNIGHT; ROGERS, 1955) ou 6 ($S_1 > S_2 = S_3 > S_4 > S_5 > S_6$) formas alélicas alternativas (COPE, 1962), espera-se assim que na população haja vários alelos segregando. Desse modo, a população utilizada neste estudo apresenta um grande potencial para capturar grande parte da diversidade alélica existente da incompatibilidade em cacaueteiro.

Os resultados obtidos neste estudo fornecem subsídios para inferir que a compatibilidade e incompatibilidade sexual podem ser diferenciadas em duas classes distintas de alta e a baixa retenção floral, porém com aspectos peculiares. Na compatibilidade foi observado um gradiente de retenção floral variando de 1 a 100 % com genótipos distribuídos em diversas classes. Entretanto, a incompatibilidade foi diferenciada em uma única classe com grande número de genótipos com pegamento zero. Isto mostra que a compatibilidade deve ser tratada como uma variável quantitativa nesta espécie, com efeito contínuo na retenção floral que pode estar sofrendo influência de vários genes de efeito menor, levando em consideração a curva de distribuição das diferentes classes apresentadas na figura 1.5. Isso pode ser resultado a existência do fenômeno da pseudocompatibilidade em cacaueteiro, fazendo com genótipos que carregam alelos de incompatibilidade consigam reter flores em determinadas situações. Enquanto isso, a incompatibilidade sexual, ausência total de retenção floral, deve ser tratada como uma variável qualitativa discreta, com genes de efeito maior influenciado fortemente na expressão fenotípica. Portanto, nesta espécie ocorre a presença de alta compatibilidade (32-100% de retenção flora), baixa compatibilidade (1-31% de retenção floral) e a incompatibilidade (0% de retenção floral).

Neste estudo, foi detectado um forte efeito da incompatibilidade tanto nos progenitores (acessos da coleção do banco de germoplasma) quanto nas progênies. Isto está ocorrendo uma vez que os alelos de incompatibilidade têm sido reportados como dominantes sobre os alelos de compatibilidade, o que leva a uma curiose no sentido de baixo pegamento, e ao frequente uso de clones incompatíveis como fundadores das populações estudadas. Por isso, são necessários métodos eficientes na caracterização da incompatibilidade de modo que permitam

os melhoristas aumentarem a frequência de alelos de compatibilidade por seleção de genótipos autocompatíveis.

Por outro lado, devido ao efeito contínuo da compatibilidade sexual (1-100%), para selecionar alelos de compatibilidade, é necessário à utilização de métodos que permitam o melhorista selecionar plantas para um alto pegamento com maior eficiência e precisão. Por exemplo, plantas com retenção abaixo de 32% apresentam maiores chances de carregarem alelos de incompatibilidade, logo, com potencial afetar o pegamento em uma população, enquanto que acima desse percentual a frequência alelos de compatibilidade é maior. Além disso, foi possível também identificar genótipos com percentual de retenção próximo de 100%, sendo assim, bons candidatos à seleção para alta retenção floral.

Em análise multivariada utilizando variáveis canônicas foi possível identificar diferentes padrões de comportamento em relação à abscisão floral, a existência quatro padrões distintos na população estudada foram observados. Alguns têm retenção média acima do controle, outros similares ao controle e, ainda outros, bem abaixo da média. Dois padrões apresentam um alto pegamento (grupos 1 e 2), com genótipos com grande potencial para seleção de alelos de compatibilidade. Por outro lado, foi identificado também dois padrões (grupo 3 e 4) com baixo pegamento, incluindo diferentes acessos do banco de germoplasma do CEPEC com alelos diferentes de incompatibilidade reportados por outros autores (YAMADA et al., 2013a; YAMADA et al., 2013b; YAMADA et al., 2014). Um desses padrões, grupo 3, apresentou um padrão diferente dos demais grupos (1 e 2), o efeito de abscisão floral se expressa antes dos 7 dias após a polinização mas continua expressando após este período, levando a uma reação mais tardia, o que pode ser resultado do efeito da pseudocompatibilidade, levando a um padrão diferente dos demais. O outro padrão, grupo 4 é formado essencialmente por genótipos autoincompatíveis, em que a limitação é maior em superar a incompatibilidade. Desse modo, é possível perceber que para os genótipos do grupo 1 e 2 o efeito de queda das flores se expressa até sete dias. Porém, para os determinados genótipos do grupo 3 essa expressão ocorre mais tardiamente ou mais lentamente (até 15 dias). Estes resultados corroboram a deficiência em tratar a compatibilidade sexual como uma variável meramente qualitativa (binária), pois diferentes padrões podem ser identificados e a retenção floral varia de um modo contínuo, com genótipos de cacauete posicionados em todo o espectro de variação possível (1 a 100% de retenção floral), com genótipos apresentando baixa e alta compatibilidade. Por outro lado, o grupo que apresenta incompatibilidade sexual (grupo 4), ausência de retenção floral até os 15 dias, pode ser tratado como uma variável

qualitativa. Assim, a compatibilidade sexual pode ser caracterizada considerando-a como variável quantitativa contínua, ao passo que a incompatibilidade pode ser tratada como variável qualitativa discreta.

Os resultados encontrados nesse estudo corroboram com achados por outros pesquisadores (COPE, 1939; GLENDINNING, 1960; 1972; CORAL, 1970; LANAUD et al., 1987). Cope (1939) propõe com base em estudos de polinizações artificiais que existem diferenças na percentagem de pegamento entre genótipos autocompatíveis. Coral (1970) realizou um estudo comparativo das teorias sobre o controle genético da compatibilidade/incompatibilidade, aplicando o método de polinização artificial em 5 acessos, (ICS-1, Sca-6, Sca-12, IMC-67, UF-613, UF-667) e observou diferenças significativas em relação à percentagem de retenção floral. Neste estudo, o genótipo ICS-1 apresentou uma taxa de 75% de pegamento, enquanto o UF-667 teve 50% pegamento; o Sca-6, Sca-12 e o IMC-67 apresentaram zero de pegamento. Partindo disso, o autor propôs a existência de graus de incompatibilidade e que este seria controlado por genes. Entretanto, Carletto e Soria (1973) não confirmaram os resultados observados por Coral (1970). Por outro lado, alguns autores têm relatado que alguns genótipos têm a capacidade de superar mais fácil a incompatibilidade que outros (LANAUD et al., 1987), o que pode explicar a ocorrência de certos níveis de retenção floral em alguns genótipos de cacauero.

Desse modo, o presente estudo permitiu fazer inferências sobre como determinar a compatibilidade sexual em cacauero de modo que minimize os erros na identificação de genótipos autocompatíveis por meio da retenção floral. Além disso, compreendeu que a compatibilidade apresenta diferentes padrões de comportamento que a leva esse fenômeno ser melhor tratado como uma variável quantitativa contínua e a incompatibilidade como uma variável qualitativa discreta.

Referências Bibliográficas

ALVIM, P.T. **Flowering of cocoa**. Cocoa Growers' Bulletin, v. 35, p. 22-3, 1984.

ARCINIEGAS, A.; PHILLIPS-MORA, W. Caracterización de Genotipos Superiores de Cacao Seleccionados por el Programa de Mejoramiento Genético del CATIE por Su Rendimiento Y/O Resistencia a Moniliasis In: Proceedings of the 15th International Cocoa Research Conference, San Jose, Costa Rica. **Proceedings...** Costa Rica: San Jose, 2006.

BARTLEY, B. G. **A preliminary catalogue of the cacao germplasm held at the University of the West Indies, Trinidad**. University of the West Indies, St. Augustine, Trinidad, 1970.

BARTLEY, B. G. D.; COPE, F. W. Practical aspects of self-incompatibility in *Theobroma cacao* L. In: Moav, R. (org.). **Agricultural genetics - selected topics**. John Wiley & Sons, New York, p. 109-134, 1973.

BARTLEY, B. G. Trinitario-Scavina hybrids - new prospects for cocoa improvement. In: Proceedings of the 1957 Cocoa Conference, London. **Proceedings...** London: U. K, 1958. p. 36-40.

BILLES, D. J. Pollination of *Theobroma cacao* L. in Trinidad, B.W.I. **Tropical Agriculture**, v. 18, p. 151-156, 1942.

CAMPOS, H. D. **Estatística experimental não-paramétrica**. ESALQ, Piracicaba, Brasil, 1983. 261p.

CARLETTO, G. A. A polinização manual controlada na flor do cacau. **Boletim Técnico do Instituto do cacau**, v. 6, p. 5-30, 1946.

CARLETTO, G. A. **Estudo dos mecanismos determinantes de graus de autocompatibilidade em *Theobroma cacao* L.** Dissertação. 1972. 49 f. – Instituto Internacional de Ciências Agrícolas do OEA, Turrialba, 1972.

CARLETTO, G. A.; SORIA, J. Testes de graus de incompatibilidade em vários cacauzeiros (*Theobroma cacao* L.). **Revista Theobroma**, v. 3, p. 26-35, 1973.

CASTILLO, J. **Informe de Viaje a Brasil. 22 de Octubre al 5 de Noviembre 2004**. Centro de Estudios de cacao Almirante. Comissao Executiva do Plano da la Voura Cacaueira (CEPLAC). CATIE (Centro Agronomico Tropical de Investigacion y Ensenanza) (entrevista) Turrialba, 2005.

CHEESMAN, E. E. The botanical progame of 1937. Imperial College of Tropical Agriculture. Trinidad. **Ann. Rep. on Cacao Rep.**, v. 7, p. 1-2, 1938.

COPE, F. W. Compatibility and fruit setting in cacao. **Annual Report on Cacao Research**, v. 8, p. 17-20, 1939.

COPE, F. W. The mechanisms of pollen incompatibility in *Theobroma cacao* L. **Heredity**, v.17, p. 157-182, 1962.

CORAL, F. J. **Estudo comparativo das teorias sobre o controle genético das incompatibilidades do cacauzeiro (*Theobroma cacao*)**. 1970. 73 f. Tese – Instituto Internacional de Ciências Agrícolas do OEA, Turrialba, 1970.

CORAL, F. J.; SORIA, V. J. Diferentes graus de autocompatibilidade em *Theobroma cacao* L. In: Proceedings of the 4th International Cocoa Research Conference, 1972, St. Augustine, Trinidad. **Proceedings...** Trinidad: St. Augustine, 1974. p. 77-81.

ENRIQUEZ, G. A.; SORIA, V. J. **Cacao Cultivars Register**. IICA Teaching and Research Centre, Costa Rica, 1967.

GLENDINNING, D. R. Natural pollination of cocoa. **New Phytol**, v. 71, p. 719-729, 1972.

GLENDINNING, D. R. Selfing of self-incompatible cocoa. **Nature**, v. 187, p. 170-170, 1960.

HARDY, F. **Manual de cacao**. IICA, Turrialba, 1961.

HARLAND, S. C. Some botanical problems of cacao. **Tropical Agriculturist**, v. 64 p. 289-291, 1925.

KKO. **Information from the KKO cocoa germplasm database**. Information supplied on computer diskette by C. Cilas, CIRAD-CP, France, 1993.

KNIGHT, R.; ROGERS, H. H. Incompatibility in *Theobroma cacao*. **Heredity**, v. 9, p. 69-77, 1955.

LANAUD, C.; SOUNIGO, O.; AMEFIA, Y. K.; PAULIN D.; LACHENAUD, P.; CLEMENT, D. New data on the mechanisms of incompatibility in cocoa and its consequences on breeding. **Café Cacao Thé**, v. 3, p. 278-282, 1987.

LEAL, A. M. A. **Caracterización de árboles superiores de cacao (*Theobroma cacao* L.) seleccionados por el programa de mejoramiento genético del CATIE**. 2005. 144 f. Dissertacao (Mestrado em Producao Vegetal). - Centro Agronomico Tropical de Investigacion y Ensenanza. Costa Rica, 2005.

LEÓN, J. **Botánica de los Cultivos Tropicales**. Costa Rica: San Jose, IICA. Ed. 3, 2000. 678p.

LOPES, U. V.; CARLETTO, G. A. Tamanho de amostra para estudos de caracterização de germoplasma de cacauero quanto à incompatibilidade. **Informe de Pesquisas 1987/1990**, v. 2, p. 56-57, 1995.

LOPEZ, B. O. **Revisión de Literatura sobre la incompatibilidad en Cacao**. Turrialba, CR. CATIE. p. 14, 1982.

MORERA, J.; MORA, A.; LAINEZ, J. Evaluation del cruce interclonal de cacao 'Catongo x Pound 12' en CATIE, Costa Rica. I. Compatibilidad. In: International Cocoa Research Conference, 11. **Proceedings...** London, 1994. p. 467-471.

PINTO, L. R. M.; YAMADA, M. M.; AHNERT, D. Recomendações para determinação da incompatibilidade sexual do cacauero. **Boletim Técnico, CEPLAC/CEPEC**, 24p. 1998.

PIRES, J. L.; ROSA, E. S.; MACEDO, M. M. Avaliação de clones de cacauero na Bahia. **Agrotrópica**, v. 24, p. 79-84, 2012.

POSNETTE, A. F. Incompatibility in Amazon cacao. **Tropical Agriculture**, v. 22, p. 184-187, 1945.

POSNETTE, A. F. Self-incompatibility in cocoa (*Theobroma spp.*). **Tropical Agriculture**, v. 17, p. 67-71, 1940.

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

POUND, F. J. Studies of fruitfulness in cacao. II. Evidence for partial sterility. **Annual Report on Cacao Research**, v. 1, p. 21-28, 1932.

ROYAERT, S.; PHILLIPS-MORA, W.; LEAL, A. M. A.; CARIAGA, K.; BROWN, J. S.; KUHN, D. N.; SCHNELL, R. J.; MOTAMAYOR, J. C. Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L.. **Tree Genetics & Genomes**, v. 7, p. 1159-1168, 2011.

SAS INSTITUTE. **Advanced general linear models with an emphasis on mixed models**. Cary: Statistical Analysis System Institute, 1996. 614 p.

SEAVEY, S.R.; BAWA, K.S. Late-acting self-incompatibility in angiosperms. **The Botanical Review**, v. 52, p. 195-219, 1986.

SORIA, J.; TONOSAKI, S.; MORENO, J. A polinização do cacauzeiro pela *Forcipomyia* ao vivo. **Cacau Atualidades**, v. 12, p. 14-18, 1975.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. McGraw-Hill Book Company, Wisconsin, 1980. 625 p.

TERREROS, J. R.; CHAVARRO, G.; OCAMPO, F. R. Determinación de los genotipos de incompatibilidad o compatibilidad en varios cultivares de cacao (*Theobroma cacao* L.). **Revista ICA**, v. 17 p. 93-99, 1982.

TERREROS, J. R.; CHAVARRO, G.; ROJAS, F. O. Determinación de los genotipos de incompatibilidad o compatibilidad en varios cultivares de cacao. **Cacaotero Colombiano**, v. 24, p. 27-37, 1982.

TOXOPEUS, H. Cocoa breeding; a consequence of mating system, heterosis and population structure. In: Conference on Cocoa and Coconuts in Malaysia, **Proceedings...** Kuala Lumpur, 1972. p. 3-12.

VIEIRA, D. R.; MÜLLER, M. W. Estudo comparativo da produção de cacauzeiros autocompatíveis e autoincompatíveis do híbrido SIAL 169 x P7 no Recôncavo da Bahia. In: **Proc. 13th Int. Cocoa Res. Conf.**, Kota Kinabalu, Malaysia, p. 175-181, 2001.

YAMADA, M. M.; BARTLEY, B. G. D. Incompatibilidade em cacauzeiro (*Theobroma cacao* L.): clones EEG 13 e EEG 65. **Revista Theobroma**, v. 14, p. 167-169, 1984.

YAMADA, M. M.; BARTLEY, B. G. D.; CASTRO, G. C. T.; MELO, G. R. P. Herança do fator incompatibilidade em *Theobroma cacao* L.. I. Relações fenotípicas na família PA (Parinari). **Revista Theobroma**, v. 12, p. 163-167, 1982.

YAMADA, M. M.; FALEIRO, F. G.; CLEMENT, D.; LOPES, U. V.; PIRES, J. L.; MELO, G. R. P. Relationship between molecular markers and incompatibility in *Theobroma cacao* L.. **Agrotrópica**, v. 22, p. 71-74, 2010.

YAMADA, M. M.; FALEIRO, F. G.; LOPES, U. V.; PIRES, J. L.; SANTOS, R. F. Reações de compatibilidade de clones de cacauzeiros da série CEPEC. **Agrotropica**, v. 26, p. 61-64, 2014.

YAMADA, M. M.; LOPES, U. V.; PIRES, J. L.; SANTOS, R. F. Identificação dos alelos de compatibilidade em clones da família Parinari de cacau. **Agrotropica**, v. 25, p. 121-124, 2013a.

YAMADA, M. M.; MONTEIRO, W. R.; LOPES, U. V.; PIRES, J. L.; BARTLEY, B. G.; SANTOS, R. F.; Determination of incompatibility alleles of Scavina 6 (*Theobroma cacao* L.) clone in farm selections. **Agrotropica**, v. 25, p. 93-96, 2013b.

4. CAPÍTULO 2

Mapeamento por Associação de genes relacionados à incompatibilidade sexual do cacauero (*Theobroma cacao* L.)²

Marcos Ramos da Silva, Didier Clément, Karina Peres Gramacho, Wilson Monteiro, Xavier Argout, Claire Lanaud, Uilson Lopes.

Resumo

A incompatibilidade sexual é um mecanismo biológico com base genética que limita a produção nas plantações de cacauero, sendo, portanto, um importante critério para seleção negativa em programas de melhoramento. Entretanto, o método atualmente disponível baseado na retenção floral após polinização, além de subjetivo, é demorado, bastante influenciado por fatores não genéticos e requer um grande número de flores. Por outro lado, a identificação de marcadores moleculares em regiões genômicas que controlam a incompatibilidade pode ser uma alternativa para permitir a seleção precoce e eficiente de plantas autocompatíveis. Partindo disso, o presente estudo buscou identificar marcadores associados com a incompatibilidade por meio de ferramentas de Mapeamento por Associação visando obter um método de predição e seleção precoce de plantas autocompatíveis. Para isto, foi utilizado no estudo uma população de 295 genótipos. Esta população foi caracterizada fenotipicamente, sendo que 8.199 flores foram polinizadas manualmente e avaliadas quanto à retenção floral aos 15 dias após as polinizações. Além disso, amostras foliares de cada genótipo da população foram coletadas e DNA extraído para genotipagem por sequenciamento, gerando ao final 29.030 marcadores do tipo SNP. Após limpeza e processamento dos dados, uma análise de Mapeamento por Associação foi feita entre os

² Autorização CGEN No. 10/2012 (Processo 02001.007611/2012-46)

11.792 SNPs e as percentagens de retenção floral após autopolinização dos 295 genótipos, utilizando-se o programa TASSEL. Dos 11.792 marcadores utilizados, 12 foram responsáveis por associações significativas com o fenótipo da incompatibilidade. Os 12 marcadores detectados estão todos localizados na extremidade proximal do grupo de ligação 4. Esses marcadores explicam de 6,68 a 20,79% da variação fenotípica e os efeitos dos marcadores variaram de 2,79 a 17,12. A equação de predição do nível de retenção floral desenvolvida usando-se 236 indivíduos tomados ao acaso e obtidas com 12 marcadores foi validada contra valores observados em 30 amostras ao acaso, de 59 outros indivíduos. Constatou-se que a melhor equação é a obtida com a combinação de três marcadores (SNP11326, SNP11343 e SNP11396), com uma correlação igual a 0,58^{**}. Esta equação tem grande potencial para ser usada na seleção precoce de plantas autocompatíveis.

Palavras-chave: Cacau; Incompatibilidade sexual; Seleção precoce; SNPs; Equações de predição.

Introdução

O cacauero (*Theobroma cacao* L.) é um dos membros do grupo das 22 espécies do gênero *Theobroma*, sendo junto com *Theobroma grandiflorum* (conhecido como cupuaçu) as únicas desse gênero exploradas economicamente. As amêndoas dos frutos de cacauero possui grande importância econômica em diversos países, sendo utilizadas principalmente para a produção de chocolate. *Theobroma cacao* é uma espécie diplóide ($2n=20$) (GLICENSTEIN; FRITZ, 1989), com genoma de 430 a 445 Mbp (ARGOUT et al., 2010; MOTAMAYOR et al., 2013), pertencente à família Malvaceae (ALVERSON et al., 1998), apresentando uma distribuição geográfica natural que abrange as florestas tropicais úmidas das Américas Central e do Sul, localizadas no hemisfério ocidental, desde o sul do México até a floresta Amazônica (CUATRECASAS, 1964). O cacauero é uma planta perene, arbórea, monóica, cauliflora (flores formadas no tronco), alógama, com taxa de cruzamento natural entre 50 e 100% (TOXOPEUS, 1972). As flores são hermafroditas e pentâmeras, com os órgãos reprodutivos (estames e pistilo) isolados na mesma flor por um círculo de estaminóides e pelas pétalas que envolvem as anteras, necessitando assim de agentes polinizadores para realizar a transferência de pólen para o estigma, geralmente moscas do gênero *Forcipomyia* (BILLES, 1942; SORIA et al., 1975).

A capacidade produtiva do cacauero pode ser bastante afetada por fatores genéticos como a incompatibilidade sexual presente em populações de cacauero (LOCKWOOD, 1977; MORERA et al., 1994). A incompatibilidade sexual em cacauero foi descoberta por Harland (1925) e confirmada por Pound (1932) e, tem sido sugerido, tratar-se de um mecanismo

fisiológico resultante da expressão do gene presente no loco *S* (KNIGHT; ROGERS, 1955), com possíveis interferências de dois outros locos, A e B (COPE, 1962), com diferentes formas alélicas alternativas, no mínimo cinco, exercendo uma relação de dominância e codominância ($S_1 > S_2 = S_3 > S_4 > S_5$) (KNIGHT; ROGERS, 1955). Em cacauero, a incompatibilidade ocorre no ovário, quando o tubo polínico e o óvulo entram em contato, e a reação de rejeição resulta da não fusão dos núcleos gaméticos, provocando uma falha no processo de singamia, ocasionando a abscisão da flor (BENNET; COPE 1959; COPE, 1962). Embora alguns autores tenham também citado que a mesma inicia no estigma (ANEJA et al., 1994; BAKER; HASENSTEIN; ZAVADA, 1997).

Em cacauero, a caracterização da compatibilidade é feita através da percentagem de retenção/abscisão floral após polinização manual controlada no campo. O método, geralmente, é baseado em autopolinizações de 15-40 flores, protegidas 24 horas antes e desprotegidas 24 horas depois de polinizadas (TERREROS; CHAVARRO; OCAMPO, 1982; YAMADA et al., 1982; YAMADA; BARTLEY, 1984). Porém, esse método apresenta algumas limitações que podem interferir no processo de caracterização da compatibilidade. Primeiro, são as influências de fatores ambientais, tais como chuva, temperatura, vento e ataque de insetos, que afetam a expressão do fenótipo (CARLETTO, 1975). Segundo, o uso deste método requer plantas adultas florindo, o que retarda sobremaneira a identificação de plantas autocompatíveis em populações segregantes de melhoramento. Assim a identificação de genótipos autocompatíveis só ocorre após a planta ser levada para o campo e atingir 3 a 5 anos de idade, desperdiçando-se recursos e tempo.

Em cacauero, o uso de marcadores moleculares tem sido sugerido na identificação precoce de plantas em relação a características de interesse agrônomo (FALEIRO et al., 2006; BROWN et al., 2005), incluindo incompatibilidade (PHILLIPS-MORA et al., 1995; YAMADA et al., 2010; ROYAERT et al., 2011). Estes autores, como primeira etapa da Seleção Assistida por Marcadores (SAM), buscaram identificar marcas moleculares associadas à incompatibilidade, usando populações segregantes envolvendo poucos pais (cruzamento biparentais) (populações F_1 ou F_2). Phillips-Mora et al. (1995) identificaram um marcador fortemente ligado à incompatibilidade sexual do cacauero, no grupo de ligação 4, usando uma F_1 (Catongo x Pound-12). Yamada et al. (2010), usando uma população F_2 (autofecundação do TSH-516 = Scavina-6 x ICS-1), conseguiram detectar 19 marcadores potenciais, 5 deles no cromossomo 4. Royært et al. (2011) identificaram 4 marcadores em uma população F_1 (Pound-7 x UF-273), todos localizados no grupo de ligação 4.

Outra estratégia de identificação de marcas associadas a caracteres de interesse, não mais usando populações com poucos parentais (F_1 , F_2 , retrocruzamento, etc.), é o Mapeamento por Associação. Esse método foi inicialmente utilizado por Pritchard e Przeworski (2001) para identificar regiões genômicas associadas com doenças em humanos e, posteriormente, Gupta et al. (2005) propôs o uso da técnica para o mapeamento de genes em espécies de plantas de interesse agrônomico. Mapeamento por Associação tem sido proposto como uma alternativa para superar as limitações do mapeamento de QTLs baseado essencialmente em indivíduos aparentados e geralmente, com populações resultantes de cruzamentos biparentais (GUPTA et al., 2005; ZHU et al., 2008; MYLES et al., 2009). Desse modo, no Mapeamento por Associação, correlações entre genótipo e fenótipo são realizadas em uma população de referência mais ampla, formada por indivíduos que não têm uma relação de parentesco muito próxima (BRESEGHELLO; SORRELS, 2006; ERSOZ et al., 2007; MYLES et al., 2009). Os estudos de Mapeamento por Associação são baseados na capacidade de identificar marcadores que estão em desequilíbrio de ligação (KHAN; KORBAN, 2012), isto é, ocorrendo mais frequentemente juntos do que de forma esperada para locos em equilíbrio de ligação. Desse modo, marcadores moleculares intimamente ligados a características de interesse irão apresentar associação significativa entre marcador e fenótipo, sendo mais dificilmente separados por recombinação durante a meiose ao longo da história evolutiva da cultura.

Mapeamento por Associação em plantas foi realizado com sucesso em espécies modelo, tais como milho e *Arabidopsis*. Um dos primeiros estudos nesta área, em plantas, teve como objetivo a identificação de genes envolvidos no tempo de floração em linhagens de milho (THORNSBERRY et al., 2001) e *Arabidopsis* (OLSEN et al., 2004). Posteriormente, outros pesquisadores, como Aranzana et al. (2005), utilizaram o Mapeamento por Associação para estudar tanto o tempo de floração, quanto outras características como a resistência a patógenos em *Arabidopsis*.

Contudo, a maioria dos estudos de Mapeamento por Associação foi realizada em plantas de ciclo curto, embora alguns estudos em plantas perenes tenham sido realizados como em pinheiros (*Pinus taeda* e *Pseudotsuga menziesii*), uva (*Vitis vinífera*), dentre outras (GONZÁLEZ-MARTÍNEZ et al., 2008; ECKERT et al., 2009; EMANUELLI et al., 2010). Em cacaueteiro, existem três estudos que usaram esta técnica (PIRES, 2003; MARCANO et al., 2007; MARCANO et al., 2009). Pires (2003), usando 291 acessos de uma coleção de germoplasmas, fez um estudo de associação entre marcadores moleculares RAPD e

características fenotípicas, tais como resistência à vassoura-de-bruxa e podridão parda, produtividade, características físicas de fruto e semente, teor e dureza da gordura e concentração de ácidos graxos e triglicerídeos. Marcano et al. (2007) utilizando a técnica de Mapeamento por Associação em 150 acessos de uma coleção de germoplasmas e 291 genótipos de uma população de mapeamento, conseguiram identificar quinze regiões genômicas envolvidas no controle do peso do fruto e da semente. Posteriormente, Marcano et al. (2009), usando uma população de 257 indivíduos descendentes do criollo, identificaram regiões do genoma envolvidas no número de frutos por planta, número de sementes por fruto, peso da semente úmida, dentre outras.

Assim, visando obter um método eficiente de predição e seleção precoce para detectar plantas autocompatíveis de cacau ainda em viveiro, o presente estudo teve como objetivo utilizar a metodologia de Mapeamento por Associação para identificar marcadores moleculares SNPs relacionados com a incompatibilidade sexual nesta espécie e propor equações usando tais marcadores que permitam a seleção precoce.

Materiais e Método

População de estudo

No presente estudo 295 genótipos foram utilizados. Dentre estes foram incluídos 39 acessos da coleção de germoplasmas do Centro de Pesquisas do Cacau (CEPEC), em Ilhéus, Bahia, Brasil, bem como 256 indivíduos escolhidos ao acaso em 109 progênies (1-8 indivíduos por progênie), em diferentes gerações do programa de melhoramento do cacau do CEPEC, predominantemente da 3ª geração e estabelecidas nos ensaios de progênies TP-0221, TP-0222, TP-0323 e TP-0324. O ensaio TP-0221 envolve 41 progênies, sem repetição, com uma média de 75 plantas/progênie; o ensaio TP-0222 envolve 15 progênies, também sem repetições, com uma média de 80 plantas/progênie; o ensaio TP-0323 envolve 31 progênies, em um delineamento em blocos casualizados, com três repetições e 20 plantas/parcela e o ensaio TP-0324 envolve 23 progênies, em um delineamento em blocos casualizados, com 15 plantas/parcela. Estas progênies segregam para diferentes caracteres avaliados (componentes de produção, resistência a doenças), incluindo incompatibilidade.

Caracterização fenotípica

Duzentos e noventa e cinco genótipos foram caracterizados quanto à incompatibilidade usando o método de polinização manual controlada. Para evitar polinizações indesejadas (por meio de insetos), flores das plantas foram protegidas 24 h antes da polinização e desprotegidas 24 h após a mesma. Aproximadamente 8.000 flores foram polinizadas manualmente no intervalo entre 8h30min e 11h30min, no período entre 03/04/2013 e 27/06/2014.

Neste estudo, foram feitos dois tipos de polinizações: autopolinizações (para caracterizar o fenótipo da incompatibilidade) e polinizações cruzadas (como controle). Nas autopolinizações, 13 a 57 flores foram polinizadas (6.308 flores polinizadas; média de 21 flores/genótipo). Nas polinizações cruzadas, 10 a 25 flores (1.891 flores polinizadas; média de 15 flores/dia) foram polinizadas com pólen de um material autocompatível (CCN-51, SIC-17 e SIC-864). Nas polinizações (auto e cruzada), os botões florais foram isolados com um tubo de mangueira transparente com uma de suas extremidades cobertas por um tecido, para evitar a passagem de insetos, e a outra extremidade aberta foi fixada na haste ou ramo da planta usando massa de modelar ou borracha e alfinetes. No momento da polinização, quando a flor estava aberta, o tubo foi retirado e os estaminóides que rodeiam o pistilo da flor foram removidos para facilitar o acesso ao estigma. Algumas flores das plantas foram utilizadas como doadoras de pólen. Nestas as pétalas foram removidas para liberar estames e anteras. As flores utilizadas como doadoras de pólen foram aquelas liberando pólen no momento da polinização. Com o auxílio de pinça, o pólen foi depositado sobre o estigma da flor, passando suavemente a antera. Cada flor polinizada foi novamente protegida, datada e identificada por fitas. Aos 15 dias após as polinizações, as flores foram avaliadas quanto à frequência de retenção floral.

Extração de DNA

Amostras foliares de 295 plantas foram coletadas em estágio intermediário de maturação e DNA extraído no Laboratório Fitopatologia Molecular do CEPEC. O protocolo de extração utilizado foi o método Matab proposto por Risterucci (2000), com adaptações de Santos et al. (2014) para a preservação de tecidos de amostras para extração de DNA e para a extração manual do grande número de plantas utilizando esferas. Amostras de tecido foliar

(aproximadamente 300 mg) foram maceradas com o auxílio de esferas metálicas em tubos eppendorfs, utilizando nitrogênio líquido. Foi adicionado ao macerado 850 µL de tampão de extração de DNA (1.4 M NaCl, 100 mM Tris HCl pH 8.0, 20 mM EDTA, 10 mM Na₂SO₃, 1% PEG 6000, 2% MATAB). O extrato foi então homogeneizado durante 10 segundos com vórtex e em seguida foi incubado durante 30 min a 74°C. Posteriormente, foram acrescentados 700µL de clorofórmio-álcool isoamílico (24:1 v/v). Os tubos contendo as amostras foram então submetidos à centrifugação por 30 minutos e o sobrenadante retirados para novos tubos. Adicionou-se 700 µL de isopropanol gelado. Os DNAs de cada amostra foram precipitados e retirados do tubo e transferidos para novos tubos contendo 200µL de tampão TE (50 mM Tris-HCl, 10 mM de EDTA, pH 7,0) com RNase. A análise da integridade e pureza das amostras de DNA foi realizada por eletroforese em gel de agarose a 1%, corado com GELRED 1ng/µL. Posteriormente, as amostras de DNA foram purificadas com o método de colunas utilizando o protocolo do kit "Macherey-Nagel Plasmid DNA Purificação".

Genotipagem por sequenciamento (GBS) e processamento dos dados

As amostras de DNA foram então genotipadas por sequenciamento utilizando a tecnologia DarTseq (*Diversity Arrays Technology Sequencing*), descrita por JACCOUD et al., (2001)(<http://www.diversityarrays.com/dart-application-dartseq>). A metodologia DArTseq se baseia no corte do DNA com duas enzimas de restrição, PstI e MseI, que cortam o genoma em regiões codificadoras, gerando vários fragmentos de sequências curtas, de forma que cada loco é representado por uma dezena ou mais de sequências. O sequenciamento desses fragmentos foi feito, então, em plataforma Illumina de sequenciamento de nova geração ou *Next Generation Sequencing* (NGS). O polimorfismo detectado entre os indivíduos resultou de variações de um único nucleotídeo (*single nucleotide polymorphism* - SNP) entre sequências em comum para todos os indivíduos, presentes na população numa frequência mínima de 1%.

Após a corrida de sequenciamento e todas as etapas de limpeza e processamento das sequências obtidas, dados brutos contendo 29.030 marcadores SNPs foram exportados para planilhas Excel para uma melhor manipulação e controle de qualidade. Partindo disso, foram eliminados os marcadores do cromossomo zero, os monomórficos, *Call Rate* ≤ 90% e os sem similaridade com o genoma de cacau (<http://cocoagendb.cirad.fr/gbrowse/cgi->

[bin/gbrowse/theobroma/](#)) (ARGOUT et al., 2010). Após essa limpeza ficaram 18.526 SNPs distribuídos nos 10 grupos de ligação do cacauero.

Análises estatísticas

Como fenótipo foram usadas as frequências de retenção floral aos 15 dias após as autopolinizações, ajustadas pela retenção floral dos cruzamentos compatíveis (controle) realizados nos mesmos dias em que as autopolinizações foram feitas. O ajuste foi feito através da obtenção de médias de mínimos quadrados (LSMEANS), em uma análise de covariância, tendo a percentagem de retenção floral das autofecundações como variável e as dos cruzamentos controle como covariável. As análises de associação entre marcador e fenótipo foram realizadas usando o programa TASSEL (*Trait Analysis by aSSociation, Evolution and Linkage*) versão 5.1.0 (BRADBURY et al., 2007) (<http://www.maizegenetics.net/#!tassel/c17q9>), conforme a metodologia proposta por Yu et al. (2006). Para aumentar a confiabilidade da análise de associação (eliminando possíveis erros de genotipagem), foi realizada uma “filtragem” no número de marcadores no programa TASSEL. Com essa filtragem foram selecionados apenas os marcadores com uma frequência alélica mínima $\geq 0,05$, resultando em 11.792 SNPs. No programa TASSEL, a análise de associação foi baseada na abordagem de modelos mistos (*Mixed Linear Model – MLM*). Neste método, preconizado por Yu et al. (2006), marcadores aleatórios são utilizados para estimar as matrizes que representam a estrutura populacional (Q) e os coeficientes de parentesco (K) da população. Essas matrizes são incorporadas nas análises usando modelos mistos para estimar as associações entre os marcadores e o fenótipo. Q é uma matriz de coancestralidade do tipo $n \times p$ (onde n é o número de indivíduos e p é o número de subpopulações definidas) e K é uma matriz de parentesco ($n \times n$), na qual, os elementos fora da diagonal são uma estimativa da probabilidade de identidade por descendência, baseada em marcadores moleculares.

Para estimativa da estrutura genética e obtenção da matriz Q foi empregado o método bayesiano utilizando o software STRUCTURE (PRITCHARD et al., 2000). Estudos de inferência de estrutura de populações usando o STRUCTURE versão 2.3.4 foram realizados para subconjuntos de dados de 300 SNPs, escolhidos aleatoriamente, ou seja, 200 marcadores a mais que os sugeridos por Turakulov e Easteal (2003) para avaliar estrutura populacional. Para o cálculo do número de subpopulações mais provável (k) foram utilizados os seguintes parâmetros: 10 corridas independentes com k variando entre 2 e 10, com 100.000 iterações (*burn-ins*) e 100.000 simulações de Monte Carlo Cadeias de Markov (MCMC). Nas análises,

foi escolhido o modelo de mistura com os valores das frequências alélicas dos 295 genótipos correspondente a cada grupo. A seleção do número k mais provável em relação aos propostos foi realizada por meio dos valores de ΔK (<http://taylor0.biology.ucla.edu/structureHarvester/>) (EVANNO et al., 2005). Por sua vez, a matriz de coancestralidade (matriz K ou *Kinship*) entre os genótipos foi obtida pelo software TASSEL versão 5.1.0. As matrizes Q (estrutura populacional) e a matriz K (parentesco) obtidas foram então inseridas como covariáveis no modelo misto para determinar as associações entre marcadores e o fenótipo da incompatibilidade sexual.

A identificação de marcadores fortemente associados à incompatibilidade foi determinada por alguns parâmetros. Primeiro, a significância da associação entre um marcador e o fenótipo foi determinada pelos níveis de significância ou valores de p (*p-values*) obtidos no TASSEL. Segundo, para eliminar associações falso-positivas, os valores de p obtidos foram submetidos a múltiplos testes usando o software QVALUE (STOREY, 2002) por meio de um pacote estatístico implementado no programa R. Este pacote utiliza os valores de p e calcula os valores de q . O valor q de um teste (também chamado de *False Discovery Rate* – FDR) mede a proporção de falsos positivos ocorridos quando um valor obtido por outro teste em particular é chamado de significativo. Semelhante a um valor de p , o valor q pode ser considerado como uma medida de significância estatística. O limite do valor p correspondente foi dado pelo QVALUE de 0,05 para garantir um FDR de 5%.

Com o objetivo de testar se existe uma relação de dominância ou codominância entre os alelos de loco SNP significativo na análise de associação, foi feita uma análise de variância. Para isto, foram utilizadas 30 amostras de 59 genótipos tomadas ao acaso dentre os 295 genótipos da população. Essas 30 amostras foram utilizadas para estimar os efeitos dos marcadores no programa TASSEL. Posteriormente, foram testadas as médias destes efeitos para cada genótipo SNP por meio de uma análise de variância (ANOVA) no programa SAS utilizando o procedimento PROC GLM (SAS, 1989).

Outra etapa desse estudo foi validar uma equação de predição do nível de compatibilidade usando os marcadores identificados pela análise de mapeamento por associação. Para validação desses marcadores, a população de 295 indivíduos foi subdividida em dois grupos: População de Modelagem ou Treinamento e População de Validação. O primeiro grupo refere-se a uma população de 236 indivíduos tomados ao acaso dentre os 295 disponíveis (População de Treinamento ou Modelagem, equivalente a 80% da população original) que foi utilizada para estimar os efeitos dos marcadores detectados no teste de

associação como significativamente associados ao fenótipo da incompatibilidade. O efeito de cada marcador na retenção floral (b_i), ou efeito médio do alelo, foi estimado através de uma regressão linear simples dos valores estimados para cada genótipo (ex. AA, AT, TT) em função do número de alelos em menor frequência (MAF) no loco considerado. Usando os efeitos estimados na população de modelagem, a equação de predição do valor genético de um dado genótipo foi estimada como: $VG_i = \sum_{j=1}^n (m_{ij}b_j)$, em que VG_i é o valor genético predito para o i -ésimo indivíduo da população quanto ao nível de retenção floral, m_{ij} é o número de alelos em menor frequência (MAF) presente no j -ésimo loco SNP no i -ésimo indivíduo, b_j é o efeito do j -ésimo loco SNP na retenção floral estimado na população de modelagem e n é o número de locos SNPs significativos e com grande efeito na retenção floral (neste estudo $n \leq 12$). Estas equações foram então aplicadas a 59 outros indivíduos tomados ao acaso (População de Validação, equivalente a 20% da população original), estimando-se os valores genéticos para aqueles indivíduos. No total 30 re-amostragens aleatórias foram realizadas (30 populações de modelagem e 30 de validação) e, para cada uma delas, equações de predição foram estimadas na população de modelagem e aplicadas na população de validação. Os valores genéticos estimados para cada um dos 59 indivíduos, em cada uma das 30 populações de validação, foram então correlacionados (correlação de Pearson) com os valores de retenção floral observados naqueles 59 indivíduos, obtendo-se assim, 30 estimativas de coeficientes de correlação.

Visando avaliar se um número menor de marcadores (menor que 12) poderia eventualmente ser usado na seleção de indivíduos autocompatíveis, equações de predição considerando-se combinações de 1, 2, 3, 4, 5, 6 e 12 marcadores foram geradas usando-se a metodologia descrita anteriormente. Assim, para cada uma das combinações, 30 coeficientes de correlação foram estimados. No final, 75.300 equações foram obtidas a partir de 2.510 combinações possíveis com 1, 2, 3, 4, 5, 6 e 12 marcadores.

Resultados

Caracterização fenotípica

Ao avaliar a percentagem de flores retidas por genótipo, verificou-se que 153 genótipos apresentaram queda de 100% das flores aos 15 dias após as autopolinizações. Enquanto isso, 142 genótipos apresentaram uma percentagem de retenção floral variável,

entre 1 e 100% (Figura 2.1). Os valores das percentagens de retenção floral dos 295 genótipos ajustados com os valores de percentagens do controle foram utilizados para a análise de associação.

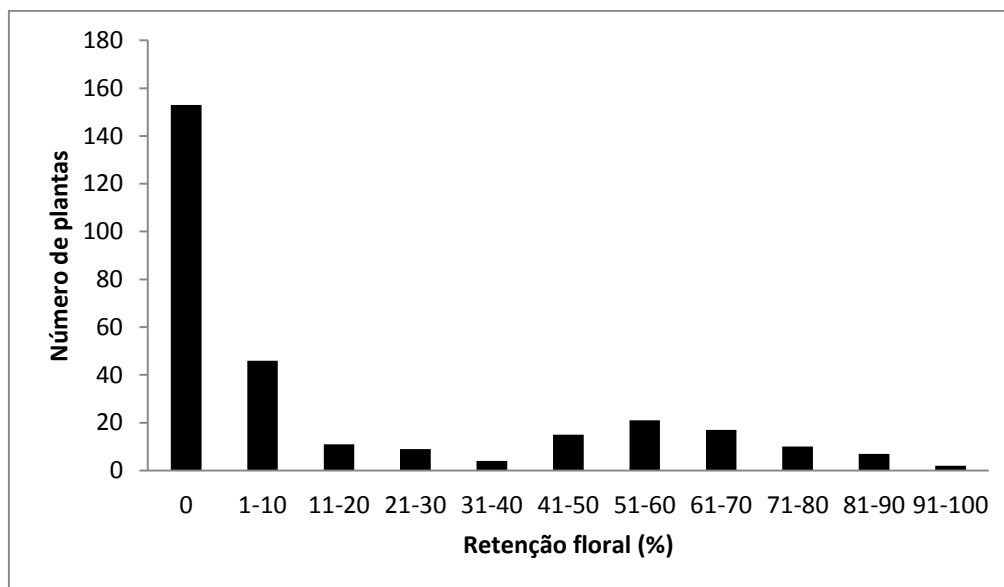


Figura 2.1- Histograma de frequência do número de genótipos de cacauero nas diferentes classes de retenção floral 15 dias após a polinização (295 genótipos no total).

Estrutura populacional

Para que estudos de associação sejam efetivos, são necessárias informações prévias referentes à população de estudo. O conhecimento sobre a estrutura populacional é um dos mais relevantes, de modo a evitar associações espúrias entre marcadores e fenótipo. Utilizando-se o programa STRUCTURE e o método sugerido por Evanno et al. (2005), observou-se que o número de subpopulações mais provável na população estudada (k) é 2 (Figura 2.2). Assim, a percentagem de mistura de cada indivíduo obtido para $k = 2$ foi usado em análises de associação subsequentes (Figura 2.3).

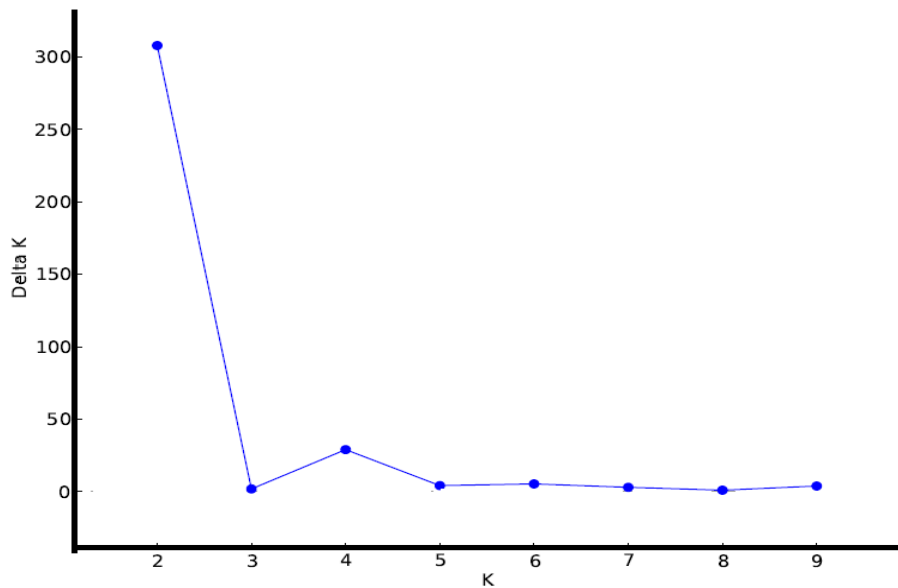


Figura 2.2 - Valores de Δk para cada valor de k , calculado de acordo com o método proposto por Evanno et al. (2005). O maior valor de Δk corresponde ao número ótimo de subgrupos.

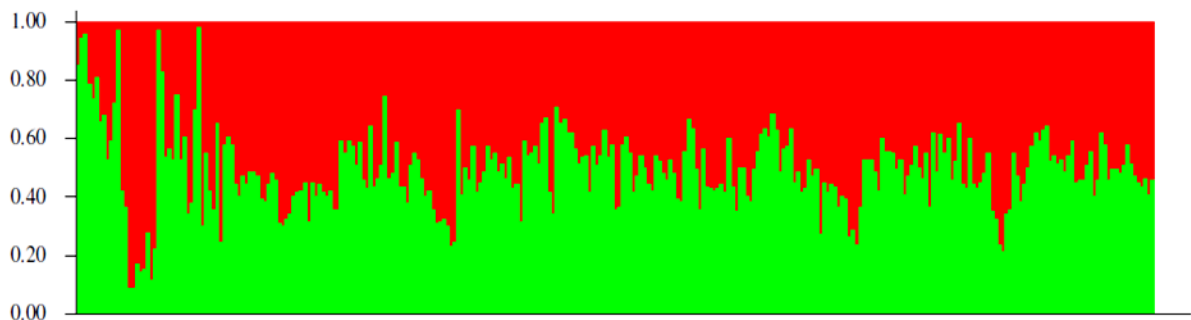


Figura 2.3 - Teste de atribuição para os genótipos de cacauero avaliados ($k=2$). Diferentes genótipos representados por barras verticais, sendo os de mesma cor pertencentes ao mesmo grupo. Diferentes cores em um mesmo indivíduo indicam a mistura de pool gênico.

Mapeamento por associação

A análise de associação feita entre 11.792 SNPs distribuídos nos 10 grupos de ligação (GL) do cacauero (644 a 1.721 marcadores por cromossomo) (Figura 2.4) e a percentagem de flores retidas nas plantas 15 dias após as autopolinizações, permitiu identificar 12 marcadores associados de forma significativa à incompatibilidade em cacauero, considerando-se um valor de $p < 4,24 \times 10^{-6}$ e um valor de $q < 0,05$ (Tabela 2.1), sendo que quatro marcadores

apresentaram um valor de $q < 0,0001$. Pelos valores de r^2 percebe-se que os 12 SNPs individualmente explicam de 6,68 a 20,79% da variação fenotípica estudada. As frequências alélicas mínimas (MAF) desses marcadores na população variaram de 23 a 46%. Estes 12 marcadores estão todos localizados no cromossomo 4 (Figura 2.5), mais precisamente na extremidade proximal desse grupo entre a posição 236.512 e 846.776 do genoma, ou seja, a uma distância máxima entre eles de 610.264 pares de base (Figura 2.6).

Com base na distância física no genoma, esses marcadores podem ser agrupados em dois grupos ou *clusters*. O primeiro *cluster* é formado por sete marcadores, entre eles o marcador mais significativo, SNP11326, localizado na posição 515.593 pares de base no GL4 e os outros seis marcadores circundantes, SNP13818 (o segundo mais significativo), SNP27554 (terceiro mais significativo), SNP11343 (quinto mais significativo), SNP13820, SNP11328 e SNP11340, estão distantes do primeiro 53.840, 43.244, 72.965, 50.541, 102 e 65.605 pares de bases, respectivamente (Tabela 2.1). É importante destacar que os sete marcadores deste *cluster* estão dentro de uma distância física de 126.805 pares de bases, sendo os três marcadores mais significativos dentro de uma distância física de 97.084 pares de bases. O segundo *cluster* é formado por quatro marcadores, SNP27990 (quarto mais significativo), SNP11287, SNP11281 e SNP11288, localizados entre a posição 236.512 e 296.881 pares de base, ou seja, a uma distância de 60.369 pares de bases. Os dois *clusters* formados pelos 11 marcadores cobrem uma região de aproximadamente 352.046 pares de base. Além disso, outro marcador fora dos dois *clusters*, o SNP11396, está localizado na posição 846.776 pares de base no grupo de ligação 4, com uma distância de 331.183 pares de bases em relação ao marcador mais significativo.

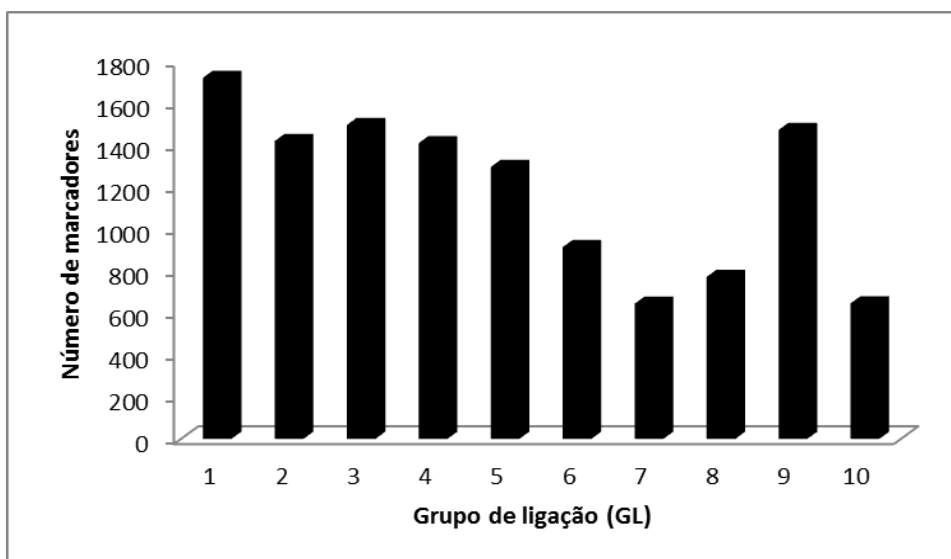


Figura 2.4- Distribuição dos 11.792 marcadores nos 10 grupos de ligação de cacauero.

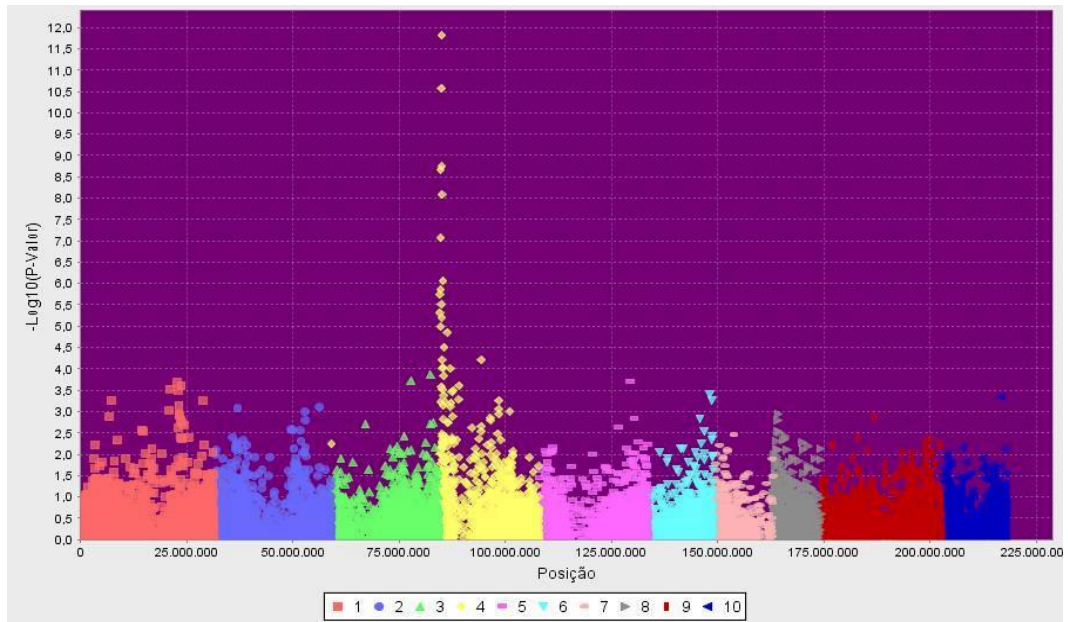


Figura 2.5- Manhattan Plot dos valores de significância das associações entre os marcadores e o fenótipo da incompatibilidade sexual, distribuídos nos 10 grupos de ligação (representado por cores diferentes) do genoma de cacaueteiro.

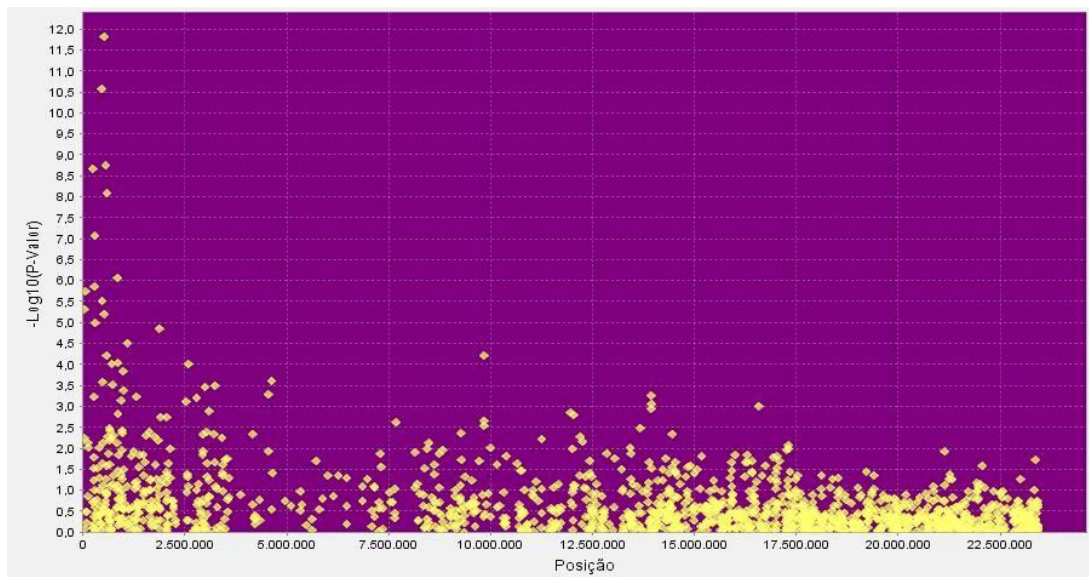


Figura 2.6- Manhattan Plot dos valores de significância das associações entre os marcadores e o fenótipo da incompatibilidade, distribuídos no grupo de ligação 4 do genoma de cacaueteiro.

Tabela 2.1 - Marcadores associados significativamente com incompatibilidade em genótipos avaliados em relação à retenção de flores 15 dias após as autopolinizações

Marcador	GL	Posição	p-valor	q-valor	r ² (%)	MAF (%)	Genótipo	Obs	Efeito (%)
SNP11326	4	515.593	1,52x10 ⁻¹²	1,70x10 ⁻⁰⁸ *****	20,79	46	GG	76	29,74
							GT	101	1,83
							TT	98	0,00
SNP13818	4	461.753	2,67x10 ⁻¹¹	1,49x10 ⁻⁰⁷ *****	17,31	45,07	AA	60	32,88
							AG	145	7,27
							GG	89	0,00
SNP27554	4	558.837	1,83x10 ⁻⁰⁹	6,01x10 ⁻⁰⁶ *****	14,11	31,06	TT	48	30,99
							TC	142	7,54
							CC	104	0,00
SNP27990	4	236.512	2,15x10 ⁻⁰⁹	6,01x10 ⁻⁰⁶ *****	14,48	44,98	TT	109	22,18
							TA	89	3,09
							AA	81	0,00
SNP11343	4	588.558	8,29x10 ⁻⁰⁹	1,85x10 ⁻⁰⁵ *****	12,99	40,48	GG	38	27,64
							GT	106	17,12
							TT	149	0,00
SNP11287	4	291.210	8,68x10 ⁻⁰⁸	1,62x10 ⁻⁰⁴ ****	11,25	43,54	AA	68	24,77
							AG	120	9,41
							GG	106	0,00
SNP11396	4	846.776	8,62x10 ⁻⁷	1,38x10 ⁻⁰³ ***	9,59	25,4	TT	20	33,43
							TA	109	8,19
							AA	165	0,00
SNP11281	4	283.463	1,41x10 ⁻⁰⁶	1,97x10 ⁻⁰³ ***	9,45	48,07	CC	92	15,99
							CT	113	0,00
							TT	81	-3,22
SNP13820	4	465.052	3,10x10 ⁻⁰⁶	3,15x10 ⁻⁰³ ***	8,92	23,6	TT	184	19,35
							AA	34	13,81
							TA	66	0,00
SNP11328	4	515.695	4,01x10 ⁻⁰⁶	3,74x10 ⁻⁰³ ***	8,61	34,38	AA	116	16,40
							AG	120	0,00
							GG	56	-2,29
SNP11288	4	296.881	1,04x10 ⁻⁰⁵	8,95x10 ⁻⁰³ **	8,02	43,81	GG	97	15,53
							GT	133	0,00
							TT	61	-2,31
SNP11340	4	581.198	6,23x10 ⁻⁰⁵	4,10x10 ⁻⁰² *	6,68	35,22	CC	129	12,65
							TT	43	0,00
							CT	119	-2,45

O limiar de significância razoável do valor de p é 4,24x10⁻⁶, que é de 5% após a correção múltipla do teste de Bonferroni (0,05 / 11.792). Os valores de q significativamente diferente de zero a *0,05; **0,01; ***0,005; ****0,0005; *****0,0001 de probabilidade. GL=grupo de ligação; Posição=localização do marcador no genoma; r²=coeficiente de determinação; MAF=Frequência alélica mínima; Obs=Número de observações; Efeito= efeito dos genótipos de cada SNP na retenção floral.

Para cada um dos 12 locos SNPs foi avaliado se existe relação de dominância ou codominância entre os alelos. Com base nas médias dos efeitos estimados para cada genótipo de cada SNP foi constatado que de modo geral a maioria apresentou dominância parcial ou incompleta no sentido de baixa retenção floral, exceto para o loco SNP11287 que a relação entre os alelos é de codominância (Tabela 2.2).

Tabela 2.2 - Efeito médio dos genótipos SNPs na retenção floral em 30 amostras de 236 plantas de cacauero.

SNP11326		SNP13818		SNP27554		SNP27990		SNP11343		SNP11287	
Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)
GG	28,07 ^a	AA	31,24 ^a	TT	29,11 ^a	TT	21,48 ^a	GG	25,67 ^a	AA	21,83 ^a
PM	13,91 ^b	PM	15,62 ^b	PM	14,44 ^b	PM	10,74 ^b	GT	15,49 ^b	PM	10,60 ^b
GT	2,08 ^c	AG	7,02 ^c	TC	7,42 ^c	TA	2,75 ^c	PM	12,29 ^c	AG	8,61 ^b
TT	-0,16 ^d	GG	0,00 ^d	CC	-0,23 ^d	AA	0,00 ^d	TT	-1,09 ^d	GG	-0,62 ^c
SNP11396		SNP11281		SNP13820		SNP11328		SNP11288		SNP11340	
Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)	Genótipo	Efeito (%)
TT	31,84 ^a	CC	15,31 ^a	TT	18,58 ^a	AA	15,43 ^a	GG	14,87 ^a	CC	12,35 ^a
PM	15,83 ^b	PM	6,52 ^b	AA	13,16 ^b	PM	6,60 ^b	PM	6,71 ^b	PM	6,15 ^b
TA	7,56 ^c	CT	0,94 ^c	PM	9,29 ^c	AG	0,05 ^c	GT	0,57 ^c	TT	-0,06 ^c
AA	-0,19 ^d	TT	-2,26 ^d	TA	0,00 ^d	GG	-2,17 ^d	TT	-1,38 ^d	CT	-1,55 ^d

Médias dos efeitos seguidas pela letra não diferem estatisticamente entre si pelo teste de Tukey (p=0,05). PM= Ponto Médio

Validação dos marcadores

A eficiência da predição da retenção floral baseada nos efeitos dos locos SNPs significativos foi avaliada com base na correlação destes valores preditos e os valores observados, considerando-se as combinações de 1, 2, 3, 4, 5, 6 e 12 SNPs no modelo de predição. As estimativas dos coeficientes de correlação entre os valores preditos da percentagem de retenção floral, com base equações de predição considerando 1 a 12 SNPs, com os valores observados foram calculadas (Tabela 2.3). As médias dos coeficientes de correlação dentre 2.510 combinações possíveis com 1, 2, 3, 4, 5, 6 e 12 variaram de 0,29 a 0,58 e todas foram significativamente diferentes de zero, a 0,001 de probabilidade. Avaliando o potencial de predição dos 12 SNPs individualmente os coeficientes variaram de 0,29 a 0,50, sendo o SNP11326 o de maior eficiência (média das correlações=0,50) e o SNP13820 o de menor (média das correlações=0,29). Ao combinar os 12 SNPs na equação

de predição, a média dos coeficientes de correlação foi menor (média=0,49). A equação de predição do nível de retenção floral usando esses 12 marcadores é:

$$VG = 14,87(N_{\text{snp11326}}) + 16,44(N_{\text{snp13818}}) + 15,49(N_{\text{snp27554}}) + 11,09(N_{\text{snp27990}}) + 13,82(N_{\text{snp11343}}) + 12,38(N_{\text{snp11287}}) + 17,12(N_{\text{snp11396}}) + 9,61(N_{\text{snp11281}}) + 2,79(N_{\text{snp13820}}) + 9,34(N_{\text{snp11328}}) + 8,92(N_{\text{snp11288}}) + 6,33(N_{\text{snp11340}})$$

Onde VG é o valor genético predito do nível de retenção floral; e N_{snp11326}, N_{snp13818}, N_{snp27554}, N_{snp27990}, N_{snp11343}, N_{snp11287}, N_{snp11396}, N_{snp11281}, N_{snp13820}, N_{snp11328}, N_{snp11288} e N_{snp11340} são os números de alelos com menor frequência (MAF) naqueles locos SNP, ou seja, G, A, T, T, G, A, T, C, T, A, G, C, respectivamente.

Por outro lado, os modelos de predição com 3, 4, 5 e 6 SNPs foram os que apresentaram os maiores coeficientes de correlação nas 30 amostras usadas ($r=0,58^{**}$), com um aumento de 9% na média da correlação em relação com a combinação de 12 marcadores (Tabela 2.3). Entretanto, associando a melhor correlação entre valor genético predito e o observado e a redução nos custos com a genotipagem desses SNPs, a equação de predição recomendada seria a obtida a partir da combinação com SNP11326, SNP11343 e SNP11396, pois com essa combinação é possível obter o mesmo sucesso no processo de seleção que com 4, 5 e 6 SNPs. A equação de predição usando três SNPs é:

$$VG = 14,87 (N_{\text{snp11326}}) + 13,82 (N_{\text{snp11343}}) + 17,12 (N_{\text{snp11396}})$$

Onde os termos da equação são os mesmos previamente definidos.

Em situações em que o melhorista pretende utilizar apenas um marcador, é recomendável utilizar o SNP11236, pois na predição individual, foi o que teve a maior média dos coeficientes de correlação dentre os 12 marcadores. Além disso, esse foi o marcador mais significativo (considerando-se os valores de p e q) e com o maior valor de r^2 , explicando sozinho aproximadamente 21% da variação fenotípica. Para este marcador a equação de predição seria $VG = 14,87(N_{\text{snp11326}})$.

Tabela 2. 3- Médias dos coeficientes de correlação de Pearson entre os valores de retenção floral e os valores genéticos observados em 30 amostras de 59 genótipos (população de validação), preditos com base em equações de predição desenvolvidas usando outras 30 amostras de 236 genótipos (população de modelagem), considerando-se o modelo 1, 2, 3, 4, 5, 6 e 12 SNPs.

Modelo	N° de combinações	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	Correlação	DP
1 SNP	12	SNP11326	-	-	-	-	-	0,50	0,09
		SNP11818	-	-	-	-	-	0,49	0,10
		SNP27554	-	-	-	-	-	0,46	0,11
		SNP11343	-	-	-	-	-	0,45	0,09
		SNP11396	-	-	-	-	-	0,44	0,10
		SNP11287	-	-	-	-	-	0,41	0,09
		SNP27990	-	-	-	-	-	0,40	0,10
		SNP11328	-	-	-	-	-	0,38	0,08
		SNP11288	-	-	-	-	-	0,37	0,09
		SNP11340	-	-	-	-	-	0,35	0,07
		SNP11281	-	-	-	-	-	0,33	0,10
SNP13820	-	-	-	-	-	0,29 ^a	0,09		
2 SNPs	66	SNP11326	SNP11343	-	-	-	-	0,56	0,08
		SNP11326	SNP11396	-	-	-	-	0,54	0,09
		SNP11326	SNP11818	-	-	-	-	0,54	0,09
		SNP11343	SNP11818	-	-	-	-	0,54	0,08
		SNP11343	SNP27554	-	-	-	-	0,53	0,09
		SNP11281	SNP13820	-	-	-	-	0,31 ^a	0,10
3 SNPs	220	SNP11326	SNP11343	SNP11396	-	-	-	0,58	0,08
		SNP11326	SNP11343	SNP11818	-	-	-	0,57	0,08
		SNP11326	SNP11396	SNP11818	-	-	-	0,56	0,09
		SNP11326	SNP11343	SNP27554	-	-	-	0,56	0,09
		SNP11343	SNP11396	SNP11818	-	-	-	0,56	0,08
		SNP11281	SNP11340	SNP13820	-	-	-	0,33 ^a	0,09
4 SNPs	495	SNP11326	SNP11343	SNP11396	SNP11818	-	-	0,58	0,08
		SNP11326	SNP11343	SNP11396	SNP13820	-	-	0,58	0,08
		SNP11326	SNP11343	SNP11818	SNP13820	-	-	0,58	0,08
		SNP11326	SNP11340	SNP11343	SNP11818	-	-	0,57	0,08
		SNP11326	SNP11340	SNP11343	SNP11396	-	-	0,57	0,09
		SNP11281	SNP11288	SNP11340	SNP13820	-	-	0,34 ^a	0,09
5 SNPs	792	SNP11326	SNP11343	SNP11396	SNP11818	SNP13820	-	0,58	0,08
		SNP11326	SNP11340	SNP11343	SNP11396	SNP11818	-	0,58	0,08
		SNP11288	SNP11326	SNP11343	SNP11396	SNP11818	-	0,58	0,08
		SNP11326	SNP11343	SNP11396	SNP11818	SNP27554	-	0,57	0,08
		SNP11281	SNP11326	SNP11343	SNP11396	SNP11818	-	0,57	0,08
		SNP11281	SNP11288	SNP11328	SNP11340	SNP13820	-	0,35 ^a	0,09
6 SNPs	924	SNP11326	SNP11340	SNP11343	SNP11396	SNP11818	SNP13820	0,58	0,08
		SNP11326	SNP11343	SNP11396	SNP11818	SNP13820	SNP27554	0,58	0,08
		SNP11288	SNP11326	SNP11343	SNP11396	SNP11818	SNP13820	0,58	0,08
		SNP11326	SNP11340	SNP11343	SNP11396	SNP11818	SNP27554	0,57	0,09
		SNP11287	SNP11326	SNP11340	SNP11343	SNP11396	SNP11818	0,57	0,09
		SNP11281	SNP11288	SNP11328	SNP11340	SNP13820	SNP27990	0,35 ^a	0,10
12 SNPs	1	-	-	-	-	-	-	0,49	0,09

Todas as correlações foram significativamente diferentes de zero a 0,001 de probabilidade. DP=desvio padrão. Para cada tipo de combinação (2 a 6 SNPs) é apresentado os 5 maiores coeficientes. Os valores seguidos pela letra “a” representam os menores coeficientes dentre todas as combinações para cada tipo de modelo.

Discussão

A incompatibilidade sexual é um dos grandes limitantes da produtividade do cacau. Estudos realizados por diferentes pesquisadores demonstraram que a incompatibilidade causa grandes perdas para a produção de cacau (POSNETTE, 1940; MORERA et al., 1994; VIEIRA; MULLER, 2001; PIRES et al., 2012). Posnette (1940), ao avaliar os efeitos da incompatibilidade na produção por 23 anos (1914-1937), verificou que a média de frutos por ano em plantas autocompatíveis foi de 91,8, diferente das plantas autoincompatíveis que tiveram uma média inferior, ou seja, de apenas de 42,7. Assim como Posnette (1940), outros pesquisadores como Morera et al. (1994); Vieira e Muller (2001); Pires et al. (2012) constataram também que a produção de genótipos autocompatíveis é superior em comparação aos autoincompatíveis. Desse modo, a incompatibilidade sexual torna-se um importante critério de seleção.

Entretanto, os métodos disponíveis para a identificação de genótipos autoincompatíveis apresentam pouca eficácia, frequentemente gerando dúvidas quanto aos resultados produzidos, além de só poderem ser aplicados depois que as plantas atingem a idade adulta e produzem uma quantidade razoável de flores (3-5 anos). Assim, é importante buscar novos métodos que possam substituir o atualmente usado. Por outro lado, o uso de marcadores moleculares tem sido amplamente usado como método precoce de seleção de plantas, de várias espécies, para diferentes características (EDWARDS; PAGE, 1994; SANCHEZ et al., 2000; DREHER et al., 2003). Para os melhoristas de cacau os marcadores moleculares têm o potencial de permitir a identificação precoce de indivíduos em populações de melhoramento, ainda no estágio de plântula. Isto possibilita que genótipos autoincompatíveis sejam identificados e eliminados ainda no viveiro, reduzindo-se o número de indivíduos levados para o campo e assim o tamanho e custos dos ensaios.

No entanto, para uma seleção confiável, a utilização de marcadores próximos aos genes, aumenta a probabilidade de se estabelecer correlações entre fenótipo e genótipo. Partindo disso, a busca e identificação de regiões do genoma associadas à expressão do fenótipo da incompatibilidade sexual são desejáveis, uma vez que permite a identificação de marcadores próximos aos genes. Para isto, é necessário localizar o(s) loco(s) que participa(m) do controle da característica de interesse entre milhares de genes distribuídos nos cromossomos que compõem o genoma da espécie *Theobroma cacao*.

Porém, para potencializar o processo de busca e identificação das tais regiões é desejável usar um grande número de marcadores moleculares (para aumentar as chances de

encontrar marcadores fortemente ligados com o loco que controla a característica de interesse), o que demanda tempo e elevados custos para obtenção de tais marcadores. Contudo, nos últimos anos houve avanços nas tecnologias de genotipagem tanto em termos de redução no tempo quanto em custo de avaliação de polimorfismos genéticos, favorecendo a obtenção de um grande número de marcadores (milhares), permitindo uma maior cobertura do genoma de uma espécie, o que por consequência tem subsidiado a obtenção de estimativas mais acuradas dos efeitos genéticos associados a regiões específicas do genoma (METZKER 2010; ROBERTS et al., 2013).

Neste estudo, a obtenção e utilização de um grande número de marcadores do tipo SNP permitiu identificar marcas fortemente associadas com a expressão do fenótipo da incompatibilidade sexual. As análises de associação entre 11.792 marcadores moleculares e o fenótipo da incompatibilidade sexual permitiram identificar 12 marcadores significativos. Todos esses marcadores estão localizados na extremidade proximal no grupo de ligação 4 (GL 4). Mais especificamente 7 SNPs (entre eles estão os três mais significativos) estão localizados entre as posições 461.753 e 588.558 pares de base do genoma; 4 SNPs entre a região 236.512 e 296.881 pares de base e um outro SNP na posição 846.776 pares de base, indicando que o controle genético é exercido por genes de efeito maior presentes nessa extremidade do cromossomo 4. Estes resultados corroboram com os achados em estudos anteriores que utilizaram os estudos clássicos de QTL visando a Seleção Assistida, em que os autores encontraram marcadores associados à incompatibilidade no GL 4, indicando fortes indícios de que essa característica é controlada por genes presentes neste cromossomo (PHILIPS-MORA; RODRIGUEZ; FRITZ, 1995; YAMADA et al., 2010; ROYAERT et al., 2011).

Philips-Mora, Rodriguez e Fritz (1995) mapearam os genes de incompatibilidade no final do GL 4. Estes autores identificaram um marcador ligado à incompatibilidade, usando duas populações: uma F₁ resultante do cruzamento entre Catongo (autocompatível) e Pound-12 (autoincompatível) (55 plantas) e outra obtida por retrocruzamento (66 plantas). Com base neste estudo, os autores propõem que a incompatibilidade sexual é controlada por um gene presente na extremidade proximal do cromossomo 4.

Yamada et al. (2010) identificaram 19 marcadores associados à incompatibilidade em uma população de mapeamento. Estes autores verificaram a relação entre marcadores moleculares e o fenótipo da incompatibilidade em uma população F₂ com 68 plantas obtidas da autofecundação do TSH-516 (resultante do cruzamento entre Scavina-6 e ICS-1). Neste

estudo, os pesquisadores utilizaram 342 marcadores moleculares (232 RAPD; 77 AFLP; 33 microssatélites), destes, 19 foram significativos (7 RAPD; 7 AFLP; 5 microssatélites), com 5 deles localizados no GL 4: três RAPD (Sae18.1095; sAU17.960; sT12.1320) e dois AFLP (iagccag.177; sacacac.311).

Royaert et al. (2011) identificaram quatro marcadores associados com a incompatibilidade. Estes pesquisadores realizaram um estudo em uma população de mapeamento com 256 plantas (F₁) provenientes de um cruzamento entre um genótipo autoincompatível (Pound-7) e um autocompatível (UF-273). Neste estudo foi feito um mapa genético contendo 170 marcadores, distribuídos nos dez grupos de ligação de cacaueteiro. A análise de associação entre marcador e fenótipo permitiu identificar um marcador (mTcCIR222) fortemente associado com a incompatibilidade, bem como três marcadores circundantes (mTcCIR168, mTcCIR115 e mTcCIR158), todos localizados na extremidade da parte proximal do GL 4. O marcador mTcCIR222 está localizado 1.061.552 no genoma, enquanto os três marcadores circundantes, mTcCIR168, mTcCIR115 e mTcCIR158, estão posicionados no genoma na posição 139.573; 1.255.075; 1.611.146 pares de base respectivamente. A distância física entre o marcador mais significativo, mTcCIR222, e os três demais, mTcCIR168, mTcCIR115 e mTcCIR158, corresponde a 921.979; 193.523 e, 549.594 pares de bases, respectivamente, cobrindo uma região de 1.471.573 pares de bases. Os resultados encontrados por esses pesquisadores entram em consenso com os resultados obtidos no presente estudo. Porém, em comparação aos marcadores encontrados por Royaert et al. (2011), nesse estudo foi encontrado um grupo de 7 marcadores com uma distância física menor entre o marcador principal, SNP11326, e os demais marcadores, SNP13818 ;SNP27554; SNP11343; SNP13820; SNP11328 e SNP11340, o que corresponde respectivamente a 53.840; 43.244; 72.965; 50.541; 102 e 65.605 pares de bases, cobrindo uma região de 126.000 pares de bases no genoma, menor em comparação com o outro estudo. Mesmo considerando a distância física entre os 12 marcadores significativos detectados neste estudo, a região é menor, correspondendo a 610.264 pares de bases. Estes resultados podem contribuir para aumentar a resolução da região onde estão os genes de efeito maior envolvidos no controle da incompatibilidade sexual.

Esse aumento na resolução da genômica se deve ao tipo de metodologia utilizada neste estudo, pois diferente dos demais trabalhos sobre incompatibilidade, este utilizou o método de Mapeamento por Associação, e não os estudos clássicos de QTL baseados em mapas de ligação visando a Seleção Assistida. Isto porque o Mapeamento por Associação

tem o potencial de considerar todas as gerações de recombinação ao longo da história da população ou grupo estudado de uma espécie. Desse modo, as regiões onde estão localizados genes-alvos no genoma tendem a ser menores em relação ao mapeamento convencional (GUPTA et al., 2005), permitindo identificar marcadores com forte desequilíbrio de ligação com o loco que controla a incompatibilidade. Mesmo a ocorrência de vários eventos de recombinação não foi capaz de separar a ligação física entre os locos dos marcadores e os locos envolvidos no controle do fenótipo. O presente estudo, por exemplo, utilizou uma população com pelo menos três ciclos de recombinação acumulados por meio de diversos cruzamentos realizados ao longo do tempo. Desse modo, os marcadores identificados têm um grande potencial de estarem fortemente associados com o loco que controla o fenótipo da incompatibilidade. Além disso, por meio da técnica de Mapeamento por Associação, este estudo conseguiu explorar grande parte da diversidade existente no programa de melhoramento do cacaueteiro do CEPEC. Ou seja, neste estudo incluíram-se as progênies de 109 cruzamentos entre 39 clones fundadores (acessos da coleção de germoplasmas), vários deles com alelos diferentes de incompatibilidade e que de alguma forma contempla a maioria dos fundadores correntemente usados no programa de melhoramento do CEPEC.

Ao contrário dos estudos anteriores (PHILIPS-MORA; RODRIGUEZ; FRITZ, 1995; YAMADA et al., 2010; ROYAERT et al., 2011), neste estudo a incompatibilidade sexual foi tratada como uma variável quantitativa, pois não houve classificação dos genótipos em relação a ser autocompatível ou autoincompatível (Capítulo 1). Em vez disso, levou-se em consideração a variação contínua nas percentagens de retenção floral aos 15 dias após as autopolinizações, com seus valores ajustados pelo controle.

Outro aspecto importante que deve ser destacado neste trabalho foi em relação às estimativas dos efeitos de substituição alélica sobre as percentagens de retenção floral. Foi constatado na população que a frequência de determinados alelos em um dos 12 locos identificados pode causar um aumento no percentual de flores retidas. Entretanto, na população estudada, a frequências de alelos que causam um baixo pegamento é maior, o que explica o grande número de genótipos autoincompatíveis na população. Por outro lado, se o melhorista seleciona plantas levando em consideração os alelos que causam um alto pegamento para cada loco dos 12 SNPs, ou pelo menos os mais importantes destes, por consequência estará aumentando a frequência de genótipos autocompatíveis em uma população.

Entretanto, visando obter uma seleção confiável e eficiente de plantas autocompatíveis, os 12 marcadores foram testados e validados. Ao testar os efeitos de substituição alélica desses marcadores, foi possível detectar por meio de 7 modelos diferentes a capacidade desses SNPs em prever o valor genético dos genótipos da população. Os sete modelos testados, dentre todas as combinações possíveis, tiveram uma correlação positiva com valores observados, com uma média variando de 0,29 a 0,58, indicando que os efeitos de substituição alélica dos marcadores testados provocam variações nas porcentagens de flores retidas nas plantas da população, sendo um indício que esses marcadores estão fortemente associados ao loco do gene de incompatibilidade. Em relação aos sete modelos testados, é importante destacar que dos modelos com 3, 4, 5 e 6 SNPs, apresentaram 10 tipos de combinação com as maiores médias dos valores das correlações obtidas. Porém, para o melhorista é mais vantajoso utilizar a equação gerada a partir da combinação dos com três SNPs (SNP11326; SNP11343; SNP11396) ($r= 0,58$), pois o sucesso no processo de predição é o mesmo que das outras combinações, pois os marcadores que compõem essas combinações estão correlacionados de tal maneira que não há ganhos em termos de predição do valor genético. Isto pode ser uma evidência que esses marcadores estejam sendo herdados em blocos de SNPs (haplótipos) ao longo das gerações. Por outro lado, na combinação com 12 marcadores houve perdas em termos de sucesso na predição, pois existem marcadores, apesar de serem significativos, são de menores efeitos e erram em muitos casos na predição do valor genético, o que interfere na acurácia da equação, afetando consequentemente os valores dos coeficientes de correlação. Isto mostra que alguns marcadores detectados como os SNP11340, SNP11281 e SNP13820 não são tão eficientes para predição do valor genético em relação a retenção floral. Por outro lado, no modelo desenvolvido somente com um marcador, o SNP11326 (o mais significativo), teve uma correlação positiva com média de 0,50.

Assim, as marcas moleculares identificadas neste estudo podem ser utilizadas como uma importante ferramenta para o melhorista, no desenvolvimento de variedades autocompatíveis, bem como, na seleção precoce de plantas ainda em estágios iniciais de vida. Equações obtidas pela combinação entre determinados SNPs têm grande potencial para prever com sucesso o valor genético de um dado genótipo em relação à incompatibilidade. Além disso, as informações sobre a posição dos 12 marcadores no genoma podem ser utilizadas futuramente para a localização exata e o isolamento de genes que controlam a expressão do fenótipo da incompatibilidade sexual em cacauero.

Agradecimentos

Agradecimento ao CIRAD pelo o financiamento para a genotipagem por sequenciamento (GBS) da população de plantas utilizada neste estudo (Projeto ATP-SEPANG).

Referências Bibliográficas

ALVERSON, W.S.; KAROL, K.G.; BAUM, D.A.; CHASE, M.W.; SWENSEN, S.M.; MCCOURT, R.; SYTSMA, K.J. Circumscription of the Malvales and relationships to the other Rosidae: evidence from rbcL sequence data. **American Journal of Botany**, v. 85, p. 876-887, 1998.

ANEJA, M.; GIANFAGNA, T.; NG, E.; BADILLA, I. Carbon dioxide treatment partially overcomes self-incompatibility in a cocoa genotype. **HortScience**, v. 29, p. 15-17, 1994.

ARANZANA, M.; KIM, S.; ZHAO, K.; BAKKER, E.; HORTON, M.; JACOB, K.; LISTER, C.; MOLITOR, J.; SHINDO, C.; TANG, C.; TOOMAJIAN, C.; TRAW, B.; ZHENG, H.; BERGELSON, J.; DEAN, C.; MARJORAM, P.; NORDBORG, M. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. **PLoS Genetics**, v. 1, p. 0531-0539, 2005.

ARGOUT, X.; SALSE, J.; AURY, JEAN-MARC.; GUILTINAN, M. J.; DROC G.; GOUZY, J.; ALLEGRE M.; CHAPARRO, C.; LEGAVRE, T.; MAXIMOVA, S. N.; ABR OUK, M.; MURAT, F.; FOUET, O.; POULAIN, J.; RUIZ, M.; ROGUET Y.; RODIER-GOUD, M.; BARBOSA-NETO, J. F.; SABOT, F.; KUDRNA, D.; AMMIRAJU, J. S. S.; SCHUSTER, S. C.; CARLSON, J. E.; SALLET, E.; SCHIEX, T.; DIEVART, A.; KRAMER, M.; GELLEY, L.; SHI, Z.; BÉRARD, A.; VIOT, C.; BOCCARA, M.; RISTERUCCI, A. M.; GUIGNON, V.; SABAU, X.; AXTELL, M. J.; MA, Z.; ZHANG, Y.; BROWN S.; BOURGE, M.; GOLSER W.; SONG X.; CLEMENT, D.; RIVALLAN, R.; TAHI, M.; AKAZA, J. M.; PITOLLAT, B.; GRAMACHO, K.; D'HONT, A.; BRUNEL, D.; INFANTE, D.; KEBE, S.; COSTET, P.; WING, R.; MCCOMBIE, W R.; GUIDERDONI, E.; QUETIER, F.; PANAUD, O.; WINCKER, P.; BOCS, S.; LANAUD C. The genome of *Theobroma cacao*. **Nature Genetics**, v. 4, p. 101-108, 2010.

BAKER, R. P.; HASENSTEIN, K. H.; ZAVADA, M. S. Hormonal changes after compatible and incompatible pollination in *Theobroma cacao* L. **HortScience**, v. 32 p.1231-1234, 1997.

BENNETT, M. C.; COPE, F. W. Nuclear fusion and non-fusion in *Theobroma cacao* L. **Nature**, v. 183, p. 1540, 1959.

BILLES, D. J. Pollination of *Theobroma cacao* L.. **Tropical Agriculture**, v. 18, p. 151-156, 1942.

BRADBURY, P. J.; ZHANG, Z.; KROON, D. E.; CASSTEVEN, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, p. 2633-2635, 2007.

BRESEGHELLO, F.; SORRELLS, M. E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. **Genetics**, v. 172, p. 1165-1177, 2006.

BROWN, J. S.; SCHNELL, R. J.; MOTAMAYOR, J. C.; LOPES, U.; KUHN, D. N.; BORRONE, J. W. Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an F2 population using SSR markers and candidate genes. **Journal of the American Society for Horticultural Science**, v. 130, p.366-373, 2005.

CARLETTO, G. A. Variação sazonal da incompatibilidade em vários clones de cacau. **Informe Técnico do CEPEC 1975**. Ilhéus, BA, Brasil. P. 22-38, 1975.

COPE, F. W. The mechanism of pollen incompatibility in *Theobroma cacao* L. **Heredity**, v. 17, p. 157-182, 1962.

CUATRECASAS, J. Cacao and its allies; a taxonomic revision of the genus *Theobroma*. **Contrib. U.S. Nat. Herb**, v. 35, p. 379-614, 1964.

DREHER, K.; KHAIRALLAH, M.; RIBAUT, J. M.; MORRIS, M. Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. **Molecular Breeding**, v. 11, p. 221-234, 2003.

ECKERT, A. J.; BOWER, A. D.; WEGRZYN, J. L.; PANDE, B.; JERMSTAD, K. D. Association genetics of coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. **Genetics**, v. 182, p. 1289-1302, 2009.

EDWARDS, M. D.; PAGE, N. J. Evaluation of marker-assisted selection through computer simulation. **Theoretical and Applied Genetics**, v. 88, p. 376-382, 1994.

EMANUELLI, F.; BATTILANA, J.; COSTANTINI, L.; LEE CUNFF, L.; BOURSIQUOT, J. M.; THIS, P.; GRANDO, M. S. A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). **BMC Plant Biol**, v. 10, p. 241, 2010.

ERSOZ, E.; YU, J.; BUCKLER, E. Application of linkage disequilibrium and association mapping in crop plants. **Genomics-Assisted Crop Improvement**, v. 1 p. 97-119, 2007.

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

FALEIRO, F. G.; QUEIROZ, V. T.; LOPES, U. V.; GUIMARÃES, C. T.; PIRES, J. L.; YAMADA, M. M.; ARAÚJO, I. S.; PEREIRA, M. G.; SOUZA FILHO, G. A.; BROWN, J. S.; SCHNELL, R.; FERREIRA, C. F.; BARROS, E. G.; MORERA, M. A. Mapping QTLs

for witches'broom (*Crinipellis pernicioso*) resistance in cacao (*Theobroma cacao* L.). **Euphytica**, v. 149, p. 227-235, 2006.

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

GLICENSTEIN, L. J.; FRITZ, D. J. Ploidy level in *Theobroma cacao* L.. **Journal of Heredity**, v. 80, p. 464-467, 1989.

GONZÁLEZ-MARTÍNEZ, S. C.; HUBER, D.; ERSOZ, E.; DAVIS, J. M.; NEALE, D. B. Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. **Heredity**, v. 101, p. 19- 26, 2008.

GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology**, v. 57, p. 461-485, 2005.

HARLAND, S. C. Some botanical problems of cacao. **Tropical Agriculturist**, v. 64 p. 289-291, 1925.

JACCOUD, D.; PENG, K.; FEINSTEIN, D.; KILIAN, A. Diversity Arrays: a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, v. 29, p. 4-25, 2001.

KHAN, M. A.; HAN, Y.; KORBAN, S. S. A high-throughput apple SNP genotyping assay using the GoldenGate platform. **Gene**, v. 494, p. 196-201, 2012.

KNIGHT, R.; ROGERS, H. H. Incompatibility in *Theobroma cacao* L. **Heredity**, v. 9, p. 69-77, 1955.

LOCKWOOD, G. Studies on the effects of cross-incompatibility on the cropping of cocoa in Ghana. **Journal of Horticultural Science**, v. 52, p. 113-126, 1977.

MARCANO, M.; MORALES, S.; HOYER, M. T.; COURTOIS, B.; RISTERUCCI, A. M.; FOUET, O.; PUGH, T.; CROS, E.; GONZALEZ, V.; DAGERT, M.; LANAUD, C. A genomewide admixture mapping study for yield factors and morphological traits in a cultivated cocoa (*Theobroma cacao* L.) population. **Tree Genetics & Genomes**, v. 5, p. 329-337, 2009.

MARCANO, M.; PUGH, T.; CROS, E.; MORALES, S.; PORTILLO PÁEZ, E. A.; COURTOIS, B.; GLASZMANN, J. C.; ENGELS, J. M. M.; PHILLIPS, W.; ASTORGA, C.; RISTERUCCI, A. M.; FOUET, O.; GONZÁLEZ, V.; ROSENBERG, K.; VALLAT, I.; DAGERT, M.; LANAUD, C. Adding value to cocoa (*Theobroma cacao* L.) germplasm information with domestication history and admixture mapping. **Theoretical and Applied Genetics**, p. 114, p. 877- 884, 2007.

METZKER, M. L. Sequencing Technologies-the next generation. **Nature Reviews Genetics**, v. 11, p. 31-46, 2010.

MORERA, J.; MORA, A.; LAINEZ, J. Evaluation del cruce interclonal de cacao 'Catongo x Pound 12' en CATIE, Costa Rica. I. Compatibilidad. In: International Cocoa Research Conference, 11, **Proceedings. J. de Lafforest Transla-Inter**, London, p. 467-471, 1994.

MOTAMAYOR, J.; MOCKAITIS, K.; SCHMUTZ, J.; HAIMINEN, N.; LIVINGSTONE, D.; CORNEJO, O.; FINDLEY, S. D.; ZHENG, P.; UTRO, F.; ROYAERT, S.; SASKI, C.; JENKINS, J.; PODICHETI, R.; ZHAO, M.; SCHEFFLER, B. E.; STACK, J. C.; FELTUS, F. A.; MUSTIGA, G. M.; AMORES, F.; PHILLIPS, W.; MARELLI, J. P.; MAY, G. D.; SHAPIRO, H.; MA, J.; BUSTAMANTE, C. D.; SCHNELL, R. J.; MAIN, D.; GILBERT, D.; PARIDA, L.; KHHN, D. N. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. **Genome Biology**, v. 14, p. 1-24, 2013.

MYLES, S.; PEIFFER, J.; BROWN, P.; ERSOZ, E.; ZHANG, Z.; COSTICH, D.; BUCKLER, E. Association mapping: critical considerations shift from genotyping to experimental design. **The Plant Cell**, v. 21, p. 2194-2202, 2009.

PHILIPS-MORA, R.; RODRIGUEZ, H.; FRITZ, P. J. **Marcadores de ADN: Teoría, aplicaciones y protocolos de trabajo, con ejemplos de investigaciones en cacao**(*Theobroma cacao*). Costa Rica, Turrialba: CATIE, 1995.184p.

PIRES, J. L. **Avaliação quantitativa e molecular de germoplasma para o melhoramento do cacauero com ênfase na produtividade, qualidade de frutos e resistência a doenças**. 2003. 342 f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa, 2003.

PIRES, J. L.; ROSA, E. S.; MACEDO, M. M. Avaliação de clones de cacauero na Bahia. **Agrotrópica**, v. 24, p. 79-84, 2012.

POSNETTE, A. F. Self-incompatibility in cocoa (*Theobroma spp.*). *Tropical Agriculture*, v. 17, p. 67-71, 1940.

POUND, F. J. Studies of fruitfulness in cacao. II. Evidence for partial sterility. **Annual Report on Cacao Research**, v. 1, p. 21-28, 1932.

PRITCHARD, J. K.; STEPHENS, M.; ROSENBERG, N. A.; DONNELLY, P. Association mapping in structured populations. **American Journal of Human Genetics**, v. 67, p. 170-181, 2000.

PRITCHARD, J.K.; PRZEWORSKI, M. Linkage disequilibrium in humans: models and data. **American Journal of Human Genetics**, v. 69, p. 1-14, 2001.

RISTERUCCI, A. M.; GRIVET, L.; N'GORAN, J. A. K.; PIERETTI, I.; FLAMENT, M. H.; LANAUD, C. A high density linkage map of *Theobroma cacao* L. **Theoretical e Applied Genetics**, v. 101, p. 948 -955, 2000.

ROBERTS, R. J.; CARNEIRO, M. O.; SCHATZ, M. C. The advantages of SMRT sequencing. **Genome Biology**, v. 14, p. 1-4, 2013.

ROYAERT, S.; PHILLIPS-MORA, W.; LEAL, A. M. A.; CARIAGA, K.; BROWN, J. S.; KUHN, D. N.; SCHNELL, R. J.; MOTAMAYOR, J.C. Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L.. **Tree Genetics & Genomes**, v. 7, p. 1159-1168, 2011.

SANCHEZ, A. C.; BRAR, D. S.; HUANG, N.; LI, Z.; KHUSH, G. S. Sequence tagged site marker assisted selection for three bacterial blight resistance gene in rice. **Crop Science**, v. 40, p. 792-797, 2000.

SANTOS, R. M. F.; LOPES, U. V.; CLEMENT, D.; PIRES, J. L.; LIMA, E. M.; MESSIAS, T. B.; GRAMACHO, K. P. A protocol for large scale genomic DNA isolation for cacao genetics analysis. **African Journal of Biotechnology**, v. 13, p. 814-820, 2014.

SAS INSTITUTE. **Advanced general linear models with na emphasis on mixed models**. Cary: Statistical Analysis System Institute, 1996. 614p.

SORIA, J.; TONSAKI, S.; MORENO, J. A polinização do cacauzeiro pela *Forcipomyia* ao vivo. **Cacau Atualidades**, v. 12, p. 14-18, 1975.

STOREY, J. D. A direct approach to false discovery rates. **Journal of the Royal Statistical Society**, v. 64, p. 479-498, 2002.

TERREROS, J. R.; CHAVARRO, G.; OCAMPO, F. R. Determinacion de los genotipos de incompatibilidad o compatibilidad en varios cultivares de cacao (*Theobroma cacao* L.). **Revista ICA**, v. 17, p. 93-99, 1982.

THORNSBERRY, J. M.; GOODMAN, M. M.; DOEBLEY, J.; KRESOVICH, S.; NIELSEN, D.; BUCKLER, E. S. Dwarf8 polymorphisms associate with variation in flowering time. **Nature Genetics**, v. 28, p. 286-289, 2001.

TOXOPEUS, H. Cocoa breeding; a consequence of mating system, heterose and population struture. In: **Proceedings of the Conference on Cocoaand Coconuts in Malaysia; 1971**, Kuala Lumpur, p. 3-12, 1972.

TURAKULOV, R.; EASTEAL, S. Number of SNP loci needed to detect population structure. **Human Heredity**, v. 55, p. 37-45, 2003.

VIEIRA, D. R.; MÜLLER, M. W. Estudo comparativo da produção de cacauzeiros autocompatíveis e autoincompatíveis do híbrido SIAL 169 x P7 no Recôncavo da Bahia. In: **Proc. 13th Int. Cocoa Res. Conf.**, Kota Kinabalu, Malaysia, p. 175-181, 2001.

YAMADA, M. M.; BARTLEY, B. G. D. Incompatibilidade em cacauzeiro (*Theobroma cacao* L.): clones EEG 13 e EEG 65. **Revista Theobroma**, v.14, p. 167-169, 1984.

YAMADA, M. M.; BARTLEY, B. G. D.; CASTRO, G. C. T.; MELO, G. R. P. Heranca do fator incompatibilidade em *Theobroma cacao* L.. I. Relacoes fenotipicas na familia PA (Parinari). **Revista Theobroma**, v. 12, p. 163-167, 1982.

YAMADA, M. M.; FALEIRO, F. G.; CLEMENT, D.; LOPES, U.V.; PIRES, J. L.; MELO, G. R. P. Relationship between molecular markers and incompatibility in *Theobroma cacao* L.. **Agrotrópica**, v. 22, p. 71-74, 2010.

YU, J.; PRESSOIR, G.; BRIGGS, W.H.; BI, I.V.; YAMASAKI, M.; DOEBLEY, J.F.; MCMULLEN, M.D.; GAUT, B.S.; NIELSEN, D.M.; HOLLAND, J.B.; KRESOVICH, S.; BUCKLER, E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, v. 38, p. 203-208, 2006.

ZHU, C.; GORE, M.; BUCKLER, E. S.; YU, J. Status and prospects of association mapping in plants. **Plant Genome**, v. 1, p. 5-20, 2008.

5. CONCLUSÕES GERAIS

Com este trabalho, foi possível concluir que em cacaueteiro a compatibilidade sexual exibe um efeito contínuo em relação à retenção floral, variando de 0 a 100%, com quatro padrões de comportamento estatisticamente diferentes. Tratar a compatibilidade/incompatibilidade sem considerar os efeitos do ambiente sobre o fenótipo e determinar a compatibilidade de um dado genótipo com base apenas em um limiar ou ponto de corte, pode levar a possíveis erros, o que compromete, por exemplo, estudos de identificação de regiões genômicas associadas a essa característica. Entretanto, a utilização de um método que compara com um controle os percentuais de flores retidas para cada dia de polinização ou considera um limiar superior (32 a 45%) a aquele que vem sendo adotado (2%, 5%, 10%, 14%), pode ser uma solução para minimizar os possíveis erros na caracterização do fenótipo da incompatibilidade.

Além disso, neste estudo, foi possível compreender que a incompatibilidade sexual em cacaueteiro é controlada por genes de efeito maior localizados na extremidade proximal do cromossomo 4, mais especificamente entre as posições 236.512 e 846.776 pares de base do genoma dessa espécie. Doze marcadores moleculares associados a esses genes foram identificados. Esses marcadores explicam de 6,68 a 20,79% da variação fenotípica e seus efeitos variam de 2,79 a 17,12. Equações de predição baseadas nos efeitos dos 12 marcadores foram desenvolvidas e testadas. Dentre as 2.510 combinações com os 12 marcadores, verificou-se que a melhor combinação foi com três marcadores, SNP11326, SNP11343 e SNP11396, com uma correlação igual a 0,58^{**}. Esta equação tem grande potencial para ser utilizada na predição e seleção precoce de plantas, em estágio de plântula, ainda em viveiro.

Por fim, como perspectiva, a região genômica associada à incompatibilidade sexual, detectada neste trabalho com sucesso pela técnica de Mapeamento por Associação, oferece uma melhor resolução do local onde estão os genes que controlam essa característica. Para as próximas pesquisas visando mapear genes, a localização dessas regiões detectadas pelos doze marcadores na extremidade proximal do grupo de ligação 4, pode facilitar na identificação e no isolamento dos genes que controlam a incompatibilidade nesta espécie.

6. REFERÊNCIAS BIBLIOGRÁFICAS COMPLEMENTARES

ABDURAKHMONOV, I. Y.; ABDUKARIMOV, A. Application of association mapping to understanding genetic diversity of plant germplasm resources. **International Journal of Plant Genomics**, v. 2008, p. 1-17, 2008.

ADJALOO, M. K.; ODURO, W. Insect assemblage and the pollination system of cocoa (*Theobroma cacao* L). **Journal Applied Biosciences**, v. 62, p. 4582-4594, 2013.

ALBUQUERQUE, P. S. B. **Mapas de ligação e identificação de locos controladores de características quantitativas (QTLs) associados à resistência a *Crinipellis perniciosa* em acessos de cacauero (*Theobroma cacao*) originários da Amazônia Brasileira**. 2006. 134f. Tese (Doutorado em Fitopatologia) – Universidade de São Paulo, USP, São Paulo, 2006.

ALLEGRE, M.; ARGOUT, X.; BOCCARA, M.; FOUET, O.; ROGUET, Y.; BERARD, A.; THÉVENIN, J. M.; CHAUVEAU, A.; RIVALLAN, R.; CLEMENT, D.; COURTOIS, B.; GRAMACHO, K. P.; BAULAND-AUGE, A.; TAHI, M.; PATHMANATHAN, U.; BRUNEL, D.; LANAUD, C. Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. **Dna Research**, p. 23-35, 2011.

ALVERSON, W. S.; KAROL, K. G.; BAUM, D. A.; CHASE, M.W.; SWENSEN, S. M.; MCCOURT, R.; SYTSMA, K. J. Circumscription of the Malvales and relationships to the other Rosidae: evidence from rbcL sequence data. **American Journal of Botany**, v. 85, p. 876-887, 1998.

ALVIM, P. T. Flowering of cocoa. **Cocoa Growers' Bulletin**, v. 25, p. 23-31, 1984.

ANEJA, M.; GIANFAGNA, T.; NG, E.; BADILLA I. Carbon dioxide and temperature influence pollen germination and fruit set in cocoa. **HortScience**, v. 27 p. 1038-1040, 1992.

ANEJA, M.; GIANFAGNA, T.; NG, E.; BADILLA, I. Carbon dioxide treatment partially overcomes self-incompatibility in a cocoa genotype. **HortScience**, v. 29, p. 15-17, 1994.

ARANZANA, M.; KIM, S.; ZHAO, K., BAKKER, E.; HORTON, M.; JACOB, K.; LISTER, C.; MOLITOR, J.; SHINDO, C.; TANG, C.; TOOMAJIAN, C.; TRAW, B.; ZHENG, H.; BERGELSON, J.; DEAN, C.; MARJORAM, P.; NORDBORG, M. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. **PLoS Genetics**, v. 1, p. 0531-0539, 2005.

AREVALO, R. A.; CARLETTO, G. A.; OCAMPO, R. F. Determinacion de los genotipos de incompatibilidad e compatibilidad de varios clones de cacao. **Revista Theobroma**, v. 2, p. 33-38, 1972.

- BAKER, R. P.; HASENSTEIN, K. H.; ZAVADA, M. S. Hormonal changes after compatible and incompatible pollination in *Theobroma cacao* L. **HortScience**, v. 32 p.1231-1234, 1997.
- BAR-HEN, A.; CHARCOSSET, A.; BOURGOIN, M.; GUIARD, J. Relationship between genetic markers and morphological traits in a maize inbred lines collection. **Euphytica**, v. 84, p. 145-154, 1995.
- BARTLEY, B. G. D. **The genetic diversity of cacao and its utilization**. CABI Publishing. Wallingford, UK. 2005. 341p.
- BARTLEY, B. G. D.; COPE, F. W. Practical aspects of self-incompatibility in *Theobroma cacao* L. **Agricultural genetics**, p. 109-134, 1973.
- BARTLEY, B. G. D.; YAMADA, M. M. Fertilidade sexual e incompatibilidade. **Informe Técnico do CEPEC 1982**, Ilhéus, p. 14-19, 1982.
- BEER, S.; SIRIPOONWIWAT, W.; O'DONOUGHUE, L.; SOUZA, E.; MATHEWS, D.; SORRELS, M. Associations between molecular markers and quantitative traits in oat germplasm pool: can we infer linkages? **Journal of Agricultural Genomics**, v. 3, p. 1, 1997.
- BENNETT, M.C.; COPE, F. W. Nuclear fusion and non-fusion in *Theobroma cacao* L. **Nature**, v. 183, p. 1540, 1959.
- BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, v.48, p.1649-1664, 2008.
- BILLES, D. J. Pollination of *Theobroma cacao* L. in Trinidad, B.W.I. **Tropical Agriculture**, v. 18, p. 151-156, 1942.
- BORÉM, A.; CAIXETA, E.T. **Marcadores Moleculares**. Minas Gerais: Folha de Viçosa, 2009. 532 p.
- BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M.; DAVIS, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **The American Journal of Human Genetics**, v. 32, p. 314-331, 1980.
- BOUHARMONT, J. **Recherches cytologiques sur la fructification et l'incompatibilité chez *Theobroma cacao* L.** Publication de l'Institut National pour l'Etude Agronomique du Congo, 1960. 117p.
- BRACHI, B.; FAURE, N.; HORTON, M.; FLAHAUW, E.; VAZQUEZ, A.; NORDBORG, M.; BERGELSON, J.; CUGUEN, J.; ROUX, F. Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. **PLoS Genetics**, v. 6, p. 1-17, 2010
- BREWBAKER, J. L. Pollen cytology and self-incompatibility systems in plants. **The Journal of Heredity**, v. 48, p. 271-277, 1957.
- BREWBAKER, J. L.; GORREZ, D. D. Genetics of self-incompatibility in the monocot genera *Ananas* (pineapple) and *Gasteria*. **American Journal of Botany**, v. 54. p. 611-616, 1967.

BRINK, R. A.; COOPER, D. C. Partial self-incompatibility in *Medicago sativa*. **Proceedings of the National Academy of Science U. S. A.**, v. 24, p. 497-499, 1938.

BROCK, R. D. Fertility in *Lilium* hybrids. **Heredity**, v. 8, p. 409-420, 1954.

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

BROWN, J. S.; PHILLIPS-MORA, W.; POWER, E. J.; KROL, C.; CERVANTES-MARTINEZ, C.; MOTAMAYOR, J. C.; SCHNELL, R. J. Mapping QTLs for Resistance to Frosty Pod and Black Pod Diseases and Horticultural Traits in *Theobroma cacao* L. **Crop Science**, v. 47, p. 1851-1858, 2007.

BROWN, S. J.; SCHENLL, R. J.; MOTAMAYOR, J. C.; LOPES, U. V.; KUHN, D. N.; BORRONE, J. W. Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an F₂ population using SSR markers and candidate genes. **J. Americ. Soc. Hort. Sci.**, v. 130, p. 366-373, 2005.

CALUS, M. P. L.; MEUWISSEN, T. H. E.; ROOS, A. P.W.; VEERKAMP, R. F. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, v. 178, p. 553-561, 2008.

CARLETTO, G. A. **Estudo dos mecanismos determinantes de graus de autocompatibilidade em *Theobroma cacao* L.** Dissertação. 1972. 49 f. – Instituto Internacional de Ciências Agrícolas do OEA, Turrialba, 1972.

CARLETTO, G. A. Variação sazonal da incompatibilidade em vários clones de cacau. **Informe Técnico do CEPEC** 1975, Ilhéus, p. 22-38, 1975.

CARLETTO, G. A.; SORIA, J. Testes de graus de incompatibilidade em vários cacaueiros (*Theobroma cacao* L.). **Revista Theobroma**, v. 3, p. 26-35, 1973.

CHAPMAN, R. K.; SORIA, S. J. Comparative *Forcipomyia* (Diptera, Ceratopogonidae) pollination of cacao in Central America and Southern Mexico. **Revista Theobroma**,v. 13, p. 129-139, 1983.

CHEESMAN, E. E. Fertilisation and embryogeny in *Theobroma cacao* L. **Annals of Botany**, v. 41, p. 107-126, 1927.

CHEESMAN, E. E. The economic botany of cacao: a critical survey of the literature to the end of 1930. **Tropical Agriculture Supplement**, v. 9, p. 1-16, 1932.

CHING, A.; CALDWELL, K. S.; JUNG, M.; DOLAN, M.; SMITH, O. S.; TINGEY, S. MORGANTE, M.; RAFALSKI, A. J. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetics**, v.3, p. 19, 2002.

CLÉMENT D.; RISTERUCCI A. M.; MOTAMAYOR J. C.; N'GORAN J.; LANAUD C. Mapping quantitative trait loci for bean traits and ovule number in *Theobroma cacao* L. **Genome**, v. 46, p. 103-111, 2003b.

CLÉMENT, D.; RISTERUCCI, A. M.; MOTAMAYOR, J. C.; N GORAN, J.; LANAUD, C. Mapping QTL for yield components, vigor, and resistance to *Phytophthora palmivora* in *Theobroma cacao* L. **Genome**, v. 46, p. 204-212, 2003a.

COE, E. H.; NEUFFER, M. G. Gene loci and linkage map of corn (maize) (*Zea mays* L.) (2n = 20). O'BRIEN, S. J. **Genetic Maps**. New York: Cold Spring Harbor, 1993. p. 157-156.

COLLARD, B. C. Y.; JAHUFER, M. Z. Z.; BROUWER, J. B.; PANG, E. C. K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement. **Euphytica**, v. 142, p. 169-196, 2005.

COOPER, D. C.; BRINK, R. A. Partial self-incompatibility and the collapse of fertile ovules as factors affecting seed formation in alfalfa. **Journal of Agricultural Research**, v. 60, p. 453-472, 1940.

COPE, F. W. Cacao - *Theobroma cacao* (Sterculiaceae) In: Simmonds, N.W., ed. **Evolution of crop plants**, p. 285-289, 1976.

COPE, F. W. Compatibility and fruit setting in cacao. **Annual Report on Cacao Research**, v 8, p. 17-20, 1939.

COPE, F. W. The mechanism of pollen incompatibility in *Theobroma cacao* L. **Heredity**, v. 17, p. 157-182, 1962.

COPE, F.W. Studies in the mechanism of self-incompatibility in cacao II. **9th Annual Report on Cacao Research (1939)**,Trinidad, p. 19-23, 1940.

CORAL, F. J.; SORIA, J. Diferentes graus de autocompatibilidade em *Theobroma cacao* L. Proceedings of the 4th International Cocoa Research Conference. **Proceedings...** Government of Trinidad and Tobago, St. Augustine, Trinidad, 1972. p. 77-81.

CROUZILLAT, D.; LERCETEAU, E.; PETIARD, V.; MORERA, J.; RODRIGUEZ, H.; WALKER, D.; PHILLIPS, W.; RONNING, C.; SCHNELL, R.; OSEI, J.; FRITZ, P. *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. **Theoretical and Applied Genetic**, v. 93, p. 205-214, 1996.

CUATRECASAS, J. Cacao and its allies; a taxonomic revision of the genus *Theobroma*. **Contributions From The United States National Herbarium**, v. 35, p. 379-614, 1964.

DOEBLEY, J.; STEC, A.; HUBBARD, L. The evolution of apical dominance in maize. **Nature**, v. 386, p. 485-488, 1997.

DUBLIN, P. Polyembryonic et haploidie chez *Theobroma cacao*. **Café Cacao Thé**, v. 16, p. 295-311, 1972.

EL-DIN EL-ASSAL, S.; ALONSO-BLANCO, C.; PEETERS, A. J.; RAZ, V.; KOORNNEEF, M. A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. **Nature Genetics**, v. 29, p. 435-440, 2001.

ENTWISTLE, P. F. Pests of Cocoa. First Edition. Longman, London, 1972.

ERSOZ, E. S.; YU, J.; BUCKLER, E. S. **Applications of linkage disequilibrium and association mapping in crop plants**. In: VARSHNEY, R. K.; R. TUBEROSA (Org.). Genomics-assisted crop improvement. Dordrecht: Springer, 2007. p. 97-119.

FALEIRO, F. G.; FALEIRO, A. S. G.; CORDEIRO, M. C. R., KARIA, C. T. **Metodologia para operacionalizar a extração de DNA de espécies nativas do cerrado**. Planaltina: Embrapa Cerrados, 2003. 6 p.

FALEIRO, F. G.; QUEIROZ, V. T.; LOPES, U. V.; GUIMARAES, C. T.; PIRES, J. L.; YAMADA, M. M.; ARAUJO, I. S.; PEREIRA, M. G.; SOUZA FILHO, G. A.; BROWN, J. S.; SCHNELL, R.; FERREIRA, C. F.; BARROS, E. G.; MORERA, M. A. Mapping QTLs for witches' broom (*Crinipellis pernicioso*) resistance in cacao (*Theobroma cacao* L.). **Euphytica**, v. 149, p. 227-235, 2006.

FALQUE, M. Fécondation et développement des fruits et des graines chez le cacaoyer (*Theobroma cacao* L.). **Influence de l'irradiation du pollen et de l'intensité de la pollinisation**. 1994. 110 f. Tese (Doutorado) - Institut National Polytechnique de Toulouse, Toulouse, 1994a.

FALQUE, M. Pod ans seed development and phenotype of the M1 plants after pollination and fertilization with irradiated pollen in cacao (*Theobroma cacao* L.). **Euphytica**, v. 75, p. 19-25, 1994b.

FALQUE, M.; LESDALONS, C.; ESKES, A. B. Comparison of two cacao (*Theobroma cacao* L.) clones for the effect of pollination intensity on fruit set and seed content. **Sexual Plant Reproduction**, v. 9, p. 221-227, 1996.

FALQUE, M.; VINCENT, A.; VAISSIÈRE, B. E.; ESKES, A. B. Effect of pollination intensity on fruit and seed set in cacao (*Theobroma cacao* L.). **Sexual Plant Reproduction**, v. 8, p. 354-360, 1995.

FERREIRA, M. E.; GRATTAPAGLIA, D. Genética de associação em plantas. In: BORÉM, A.; CAIXETA, E. T. (Org.). **Marcadores Moleculares**. Viçosa: UFV, 2006. p. 273-306.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores RAPD e RFLP em análise genética**. Brasília: EMBRAPA-CENARGEN, 1995. 220 p.

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

FIGUEIRA, A.; JANICK, J.; BEMILLER, J. N. Partial characterization of cacao pod and stem gums. **Carbohydrate Polymers**, v. 24, p. 133-138, 1994.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, v. 54, p. 357-374, 2003.

- FORD, C. S. **Investigations into the genetic and physiological basis of self-incompatibility in cocoa (*Theobroma cacao* L.)**. 2004. 288f. Tese (Doutorado) - The University of Reading, Inglaterra, 2004.
- FORD, C. S.; WILKINSON, M. J. Confocal observations of late-acting self-incompatibility in *Theobroma cacao* L. **Sexual Plant Reproduction**, v. 25, p. 169-183, 2012.
- FOUET, O.; ALLEGRE, M.; ARGOUT, X.; JEANNEAU, M.; LEMAINQUE, A.; PAVEK, S.; BOLAND, A.; RISTERUCCI, A. M.; LOOR, G.;TAHI, M.; SABAU, X.; COURTOIS, B.; LANAUD, C. Structural characterization and mapping of functional EST-SSR markers in *Theobroma cacao*. **Tree Genetics & Genomes**, v. 74, p. 799-817, 2011.
- FRARY, A.; NESBITT, T. C.; GRANDILLO, S.; KNAAP, E.; CONG, B.; LIU, J. P.; MELLER, J.; ELBER, R.; ALPERT, K. B.; TANKSLEY, S. D. *fw2.2*: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. **Science**, v. 289, p. 85-88, 2000.
- FRIDMAN, E.; PLEBAN, T.; ZAMIR, D. Recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, p. 4718-4723, 2000.
- GARVIN, D. F.; ROOSE, M. L.; WAINES, J. G. Isozyme genetics and linkage in tepary bean, *Phaseolus acutifolius* A. Gray. **Journal of Heredity**, v. 81, p. 455-461, 1989.
- GAUER, P. M.; SLLNKARD, A. Inheritance and linkage of Isozyme coding genes In chickpea. **Journal of Heredity**, v. 81, p. 455- 461, 1990.
- GEBHARDT, C.; BALLVORA, A.; WALKEMEIER, B.; OBERHAGEMANN, P.; SCHULER, K. Assessing genetic potential in germplasm collection of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. **Molecular Breeding**, v. 13, p. 93-102, 2004.
- GIBBS, P. E. Do homomorphic and heteromorphic self-incompatibility systems have the same sporophytic mechanism? **Plant Systematics and Evolution**, v. 154, p. 285-323, 1986.
- GIBBS, P. E. Self-incompatibility in flowering plants: a neotropical perspective. **Revista Brasileira de Botânica**, v. 13, p. 125-136, 1990.
- GIGORD, L.; LAVIGNE, C.; SHYKOFF, J. A. Partial self-incompatibility and inbreeding depression in a native tree species of La Reunion (Indian Ocean). **Oecologia**, v. 117, p. 342-352, 1998.
- GLENDINNING, D. R. Natural pollination of cocoa. **New Phytol**, v. 71, p. 719-729, 1972.
- GLENDINNING, D. R. Selfing of self-incompatible cocoa. **Nature**, v. 187, p. 170-170, 1960.
- GLICENSTEIN, L. J.; FRITZ, D. J. Ploidy level in *Theobroma cacao* L.. **Journal of Heredity**, v. 80, p. 464-467, 1989.

- GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal Animal of Breeding and Genetics**, v. 124, p. 323-330, 2007.
- GRATTAPAGLIA, D.; PLOMION, C.; KIRST, M.; SEDEROFF, R. Genomics of growth traits in forest trees. **Current Opinion in Plant Biology**, v. 12, p. 148-156, 2009.
- GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, p. 241-255, 2011.
- GRODZICKER, T.; WILLIAMS, J.; SHARP, P.; SAMBOOK, J. Physical mapping of temperature-sensitive mutations adenoviruses. **Quant. Biol.**, v. 39, p. 439-446, 1974.
- GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: present status and future prospects. **Plant Molecular Biology**, v. 57, p. 461-485, 2005.
- HARLAND, S. C. Some botanical problems of cacao. **Tropical Agriculturist**, v. 64 p. 289-291, 1925.
- HART, G. E.; GALE, M. D.; McINTOSH, R. A. Linkage maps of *Triticum aestivum* (hexaploid wheat, $2n = 42$, genomes A, B and D) and *T. tauschii* ($2n = 14$, genome D). **Genetic Maps**, p. 6204-6219, 1993.
- HASENSTEIN, K. H.; ZAVADA, M. S. Auxin modification of the incompatibility response in *Theobroma cacao*. **Physiologia Plantarum**, v. 112, p. 113-118, 2001.
- HEFFNER, E.L.; SORRELLS, M.E.; JANNINK, J.L. Genomic Selection for Crop Improvement. **Crop Sci**, v. 49, p. 1-12, 2009.
- HUTTLEY, G.A.; SMITH, M. W.; CARRINGTON, M.; O'BRIEN, S. J. A scan for linkage disequilibrium across the human. **Genetics**, v. 152, p. 1711-1722, 1999.
- JACCOUD, D.; PENG, K.; FEINSTEIN, D.; KILIAN, A. Diversity Arrays: a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, v. 29, p. 1-7, 2001.
- JACOB, V. J. Self-incompatibility mechanism in *Cola nitida*. **Incompatibility Newsletter**, v. 3, p. 60-61, 1973.
- JACOB, V. J.; ATANDA, O. A. Compatibility and fruit setting in *Theobroma cacao* L. **Revista Theobroma**, v. 5, p. 12-18, 1975.
- JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, v. 9, p. 166-177, 2010.
- JOHANSON, U.; WEST, J.; LISTER, C.; MICHAELS, S.; AMASINO, R.; DEAN, C. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. **Science**, v. 290, p. 290-344, 2000.

- KHAN, M. A.; KORBAN, S. S. Association mapping in forest trees and fruit. **Journal of Experimental Botany**, v. 63, p. 1-16, 2012.
- KNIGHT, R.; ROGERS, H. H. Sterility in *Theobroma cacao* L. **Nature**, v. 172 p.164, 1953.
- KNIGHT, R.; ROGERS, H.H. Incompatibility in *Theobroma cacao* L. **Heredity**, v. 9, p. 69-77, 1955.
- KOWYAMA, Y.; SHIMANO, N.; KAWASE, T. Genetic analysis of incompatibility in the diploid *Ipomoea* species closely related to the sweet potato. **Theoretical and Applied Genetics**, v. 58, p. 149-155, 1980.
- LACHENAUD, P.; MOSSU, G. Etude comparative de l'influence de deux modes de conduite sur lés facteurs du redement d'une cacaoyère. **Café Cacao Thé**, v. 29, p. 21-30, 1985.
- LACHENAUD, P.; SOUNIGO, O.; SALLÉE, B. Les cacaoyers spontanés de Guyane. Nouvelles prospections. **Acta Botanica Gallica**, v. 3, p. 325-346, 2005.
- LAIDO, G.; MARONE, D.; RUSSO, M. A.; COLECCHIA, S. A.; MASTRANGELO, A. M.; DE VITA, P.; PAPA, R. Linkage disequilibrium and genome-wide association mapping in tetraploid wheat (*Triticum turgidum* L.). **PLoS ONE**, v. 9, p. 1-18 2014.
- LANAUD, C. Genetic studies of *Theobroma cacao* L. with the help of enzymatic markers. I. Genetic control and linkage of nine enzymatic markers. **Café Cacao Thé**, v. 30, p. 259-270, 1986.
- LANAUD, C.; MOTAMAYOR, J. C.; RISTERUCCI, A. M. Implications of new insight into the genetic structure of *Theobroma cacao* L. for breeding strategies. In: Proceedings of the International Workshop on New Technologies and Cocoa Breeding, **Proceedings...** Ingenic: Malaysia, 2000. p. 93-111.
- LANAUD, C.; RISTERUCCI, A. M.; N'GORAN, J. A. K.; CLEMENT, D.; FLAMENT, M. H.; LAURENT, V.; FALQUE, M. A genetic linkage map of *Theobroma cacao* L. **Theoretical and Applied Genetics**, v.91, p. 987-993, 1995.
- LANAUD, C.; SOUNIGO, O.; AMEFIA, Y. K.; PAULIN D.; LACHENAUD, P.; CLEMENT, D. New data on the mechanisms of incompatibility in cocoa and its consequences on breeding. **Café Cacao Thé**, v. 3, p. 278-282, 1987.
- LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in improvement of quantitative traits. **Genetics**, v.124, p.743-756, 1990.
- LAURENT, V.; RISTERUCCI, A. M.; LANAUD, C. Variability for nuclear ribosomal genes within *Theobroma cacao*. **Heredity**, v. 71, p. 96-103, 1993.
- LEÓN, J. **Botánica de los Cultivos Tropicales**. Costa Rica, San José: IICA. 2000 678p.
- LERCETEAU, E.; ROBERT, T.; PÉRTIARD, V.; CROUZILLAT, D. Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers. **Theoretical and Applied Genetics**, v.95, p.10-19, 1997.

LI, Y.; HUANG, Y.; BERGELSON, J.; NORDBORG, M.; BOREVITZ, J. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, p. 21119-21204, 2010.

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

LIU, J.; VAN ECK, J.; CONG, B.; TANKSLEY, S. D. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. **Proceedings of the National Academy of Sciences of the United States of America**, 99, 13302-13306, 2002.

LONG, N.; GIANOLA, D.; ROSA, G. J. M.; WEIGEL, K. A.; AVENDANO, S. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Journal of Animal Breeding and Genetics**, v. 124, p. 377- 389, 2007.

LOPES, U. V.; CARLETTO, G. A. Tamanho de amostra para estudos de caracterização de germoplasma de cacauero quanto à incompatibilidade. **Informe de Pesquisas 1987/1990**, Ilhéus, p. 56-57, 1995.

LUNDQVIST, A. One-locus sporophytic self-incompatibility in the Carnation family, Caryophyllaceae. **Hereditas**, v. 91, p. 307, 1979.

MACKAY, I.; POWEL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, v. 12, p. 57-63, 2007.

MALOSETTI, M.; VAN DER LINDEN, C. G.; VOSMAN, B.; VAN EEUWIJK, F. A. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. **Genetics**, v. 175, p. 879-889, 2007.

MARCANO, M.; MORALES, S.; HOYER, M. T.; COURTOIS, B.; RISTERUCCI, A. M.; FOUET, O.; PUGH, T.; CROS, E.; GONZALEZ, V.; DAGERT, M.; LANAUD, C. A genomewide admixture mapping study for yield factors and morphological traits in a cultivated cocoa (*Theobroma cacao* L.) population. **Tree Genetics & Genomes**, v. 5, p. 329-337, 2009.

MARCANO, M.; PUGH, T.; CROS, E.; MORALES, S.; PORTILLO PÁEZ, E. A.; COURTOIS, B.; GLASZMANN, J. C.; ENGELS, J. M. M.; PHILLIPS, W.; ASTORGA, C.; RISTERUCCI, A. M.; FOUET, O.; GONZÁLEZ, V.; ROSENBERG, K.; VALLAT, I.; DAGERT, M.; LANAUD, C. Adding value to cocoa (*Theobroma cacao* L.) germplasm information with domestication history and admixture mapping. **Theoretical and Applied Genetics**, p. 114, p. 877- 884, 2007.

MARITA, J. M.; NIENHUIS, J.; PIRES, J. L.; AITKEN, W. M. Analysis of genetic diversity in *Theobroma cacao* with emphasis on witches' broom disease resistance. **Crop Science**, v.41, p. 1305-1316, 2001.

MASSAUX, F.; TCHIENDJI, C.; MISSE, C.; DECAZY, B. Etude du transport du pollen de cacaoyer par marquage au 32P. **Café CacaoThé**, v. 20, p. 163-170, 1976.

- MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.
- MILLACH, S. C. K. Marcadores moleculares em plantas. Porto Alegre: S.C.K. 1998. 141 p.
- MORERA, M. A. Identification of a major QTL in cocoa (*Theobroma cacao* L.) associated with resistance to witches' broom disease. **Plant Breeding**, v.122, p. 268-272, 2003.
- MORGANTE, M.; SALAMINI, F. From plant genomics to breeding practice. **Current Opinion in Biotechnology**, v. 14, p. 214-219, 2003.
- MULLIS, K. B.; FALOONA, F. A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. **Methods Enzymol**, v. 155, p. 335-350, 1987.
- MYLES, S.; PEIFFER, J.; BROWN, P.; ERSOZ, E.; ZHANG, Z.; COSTICH, D. E.; BUCKLER, E. S. Association mapping: critical considerations shift from genotyping to experimental design. **Plant Cell**, v. 21, p. 2194-2202, 2009.
- NETTANCOURT, D. **Incompatibility and incongruity in wild and cultivated plants**. Berlin: Springer, 2000. 320 p.
- NETTANCOURT, D. Incompatibility in angiosperms. **Sexual Plant Reproduction**, v.10, p.185-199, 1997.
- NETTANCOURT, D. **Incompatibility Systems in Angiosperms**. Berlin: Springer, 1977. 230 p.
- NORDBORG, M.; TAVARÉ, S. Linkage disequilibrium: What history has to tell us. **Trends Genetics**, v. 18, p. 83-90, 2002.
- OLIVEIRA, A. C. B.; SEDIAYAMA, M. A. N.; SEDIYAMA, T.; T.; FINGER, F. L.; CRUZ, C. D. Variabilidade genética em batata-doce com base em marcadores isoenzimáticos. **Horticultura Brasileira**, v. 132, p. 1811-1824, 2002.
- OLSEN, K. M.; HALLDORSDDOTTIR, S. S.; STINCHCOMBE, J. R.; WEINIG, C.; SCMITT, J.; PURUGGAN, M. D. Linkage disequilibrium mapping of *Arabidopsis* CRY2 flowering time alleles. **Genetics**, v. 167, p. 1361-1369, 2004.
- ORAGUZIE, N. C.; WILCOX, P. L.; RIKKERINK, E. H. A.; SILVA, H. N. Linkage disequilibrium. In: ORAGUZIE, N. C.; WILCOX, P. L.; RIKKERINK, E. H. A.; GARDINER, S. E.; SILVA, H. N. **Association Mapping in Plants**. New York: Springer, 2007. p. 11-39.
- PARISSEAUX, B.; BERNARDO, R. In silico mapping of quantitative trait loci in maize. **Theoretical and Applied Genetics**, v. 109, p. 508-514, 2004.
- PARVAIS, J. P.; REFFYE, P.; LUCAS, P. Observations sur la pollinisation libre chez *Theobroma cacao* analyse mathématique des données et modélisation. **Café Cacao Thé**, v. 21, p. 253-262, 1977.
- PHILLIPS-MORA, W.; RODRIGUEZ, H.; FRITZ, P. J. **Marcadores de ADN: Teoria, aplicaciones y protocolos de trabajo, com exemplos de investigaciones em cacaco** (*Theobroma cacao*). Costa Rica, Turrialba: CATIE, 1995.184p.

PIRES, J. L. **Avaliação quantitativa e molecular de germoplasma para o melhoramento do cacauero com ênfase na produtividade, qualidade de frutos e resistência a doenças**. 2003. 342 f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa, 2003.

POSNETTE, A. F. Incompatibility in Amazon cacao. **Tropical Agriculture**, v. 22, p. 184-187, 1945.

POSNETTE, A. F. Self-incompatibility in cocoa (*Theobroma spp.*). **Tropical Agriculture**, v. 17, p. 67-71, 1940.

POUND, F. J. Studies of fruitfulness in cacao. II. Evidence for partial sterility. **Annual Report on Cacao Research**, v. 1, p. 21-28, 1932.

PRITCHARD, J. K. Deconstructing maize population structure. **Nature Genetics**, v. 28, p. 203-204, 2001.

PRITCHARD, J. K.; STEPHENS, M.; ROSENBERG, N. A.; DONNELLY, P. Association mapping in structured populations. **American Journal of Human Genetics, Pittsburgh**, v. 67, p. 170-181, 2000.

QUEIROZ, V. T.; GUIMARÃES, C. T.; AHNERT, D.; SCHUSTER, I.; DAHER, R. T.; PEREIRA, M. G.; MIRANDA, V. R. M.; LOGUÉRCIO, L. L.; BARROS, E. G.; QUEIROZ, V. T.; MOTA, J. W. S.; GUIMARÃES, C. T.; BARROS, E. G.; MORERA, M. A. Mapeamento genético e estudo da resistência a vassoura-de-bruxa em cacau (*Theobroma cacao* L.). **Genetics and Molecular Biology**, v. 21, p. 195, 1998.

RAMOS, A. R.; VENTURIERI, G. A.; CUCO, S.M.; CASTRO, N. M. The site of self-incompatibility in cupuassu (*Theobroma grandiflorum*). **Revista Brasileira de Botânica**, v. 28, p. 569-578, 2005.

REFFYE, P.; PARVAIS, J. P.; MOSSU, G.; LUCAS, P. Influence des aléas de la polinisation sur les rendements du cacaoyer modele mathématique et simulation **Café Cacao Thé**, v. 22, p. 254-274, 1978.

REMYNTOON, D. L.; THORNSBERRY, J. M.; MATSUOKA, Y.; WILSON, L. M.; WHITT, S. R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M. M.; BUCKLER, E. S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. **Genetics**, v. 98, p. 11479-11484, 2001.

RESENDE, M.D.V.; LOPES, P.S.; SILVA, R.L.; PIRES, I.L. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v. 56, p. 63-77, 2008.

RISCH, N.; MERIKANGAS, K. The future of genetic studies of complex human diseases. **Science**, v. 273, p. 1516-1517, 1996.

RISTERUCCI, A. M.; GRIVET, L.; N'GORAN, J. A. K.; PIERETTI, I.; FLAMENT, M. H.; LANAUD, C. A high-density linkage map of *Theobroma cacao* L.. **Theoretical and Applied Genetics**, v.100, p. 948-955, 2000.

ROSENQUIST, E. A. Cocoa selection and breeding in Malaya. **Malayan Agricultural Journal**, v. 33, p. 181-193, 1950.

ROYAERT, S.; PHILLIPS-MORA, W.; LEAL, A. M. A.; CARIAGA, K.; BROWN, J. S.; KUHN, D. N.; SCHNELL, R. J.; MOTAMAYOR, J. C. Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L.. **Tree Genetics & Genomes**, v. 7, p. 1159-1168, 2011.

RUSSELL, J. R.; HOSEIN F.; JOHNSON, E.; WAUGH, R.; POWELL, W. Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. **Molecular Ecology**, v.2, p. 89-97, 1993.

SANSALONI, C. P.; PETROLI, C. D.; CARLING, J.; HUDSON, C. J.; STEANE, D. A.; MYBURG, A. A.; GRATTAPAGLIA, D.; VAILLANCOURT, R. E.; KILIAN, A. A. highdensity Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. **Plant Methods**, v. 6, p. 16, 2010.

SAYERS, E. R.; MURPHY, R. P. Seed set in alfalfa as related to pollen tube growth, fertilization frequency, and post-fertilization ovule abortion. **Crop Science**, v. 6, p. 365-368, 1966.

SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, v. 123, p. 218-223, 2006.

SEAVEY, S.R.; BAWA, K.S. Late-acting self-incompatibility in angiosperms. **The Botanical Review**, v. 52, p. 195-219, 1986.

SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Genomic selection using different marker types and densities. **Journal of Animal Science**, v. 86, p. 2447-2454, 2008.

SOTO-CERDA, B. J.; CLOUTIER, S. Association mapping in plant genomes. In: CALISKAN, M. (Org.). **Diversity in Plants**. Rijeka: University Campus STeP Ri, 2012. p. 29-54.

SOUZA, M. S.; VENTURIERI, G. A. Floral biology of cacaui (*Theobroma speciosum*-Malvaceae). **Brazilian Archives of Biology and Technology**, v. 53, n. 4, p. 861-872, 2010.

STICH, B.; MELCHINGER, A. E.; FRISCH, M.; MAURER, H. P.; HECKENBERGER, M.; REIF, J. C. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. **Theoretical and Applied Genetics**, v. 111, p. 723-730, 2005.

STOUT, A. B.; CHANDLET, C. Pollen-tube behavior in *Hemerocallis* with special reference to incompatibilities. **Bulletin of the Torrey Botanical Club**, v. 60, p. 397-416, 1933.

TAKAHASHI, Y.; SHOMURA, A.; SASAKI, T.; YANO, M. Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, p. 7922-7927, 2001.

TANKSLEY, S. D.; MEDINA-FILHO, H.; RICK, C. M. Use of naturally-occurring enzyme variations to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. **Heredity**, v. 49, p. 11-25, 1982.

TARODA, N.; GIBBS, P. E. Floral biology and breeding system of *Sterculia chicha* St. Hil (Sterculiaceae). **New Phytol**, v. 90, p. 735-743, 1982.

TERREROS, J. R.; CHAVARRO, G.; ROJAS, F. O. Determinacion de los genotipos de incompatibilidad o compatibilidad en varios cultivares de cacao. **Cacaotero Colombiano**, v. 24, p. 27-37, 1982.

THOMPSON, M. M. Genetics of incompatibility in *Corylus avellana* L. **Theoretical and Applied Genetics**, v. 54, p. 113-116, 1979.

THORNSBERRY, J. M.; GOODMAN, M. M.; DOEBLEY, J.; KRESOVICH, S.; NIELSEN, D.; BUCKLER, E. S. Dwarf8 polymorphisms associate with variation in flowering time. **Nature Genetics**, v. 28, p. 286-289, 2001.

TOVAR, G.; ORTIZ, M. Evaluación de algunos componentes del rendimiento en una plantación de híbridos de cacao de 5 años de edad en la región de Granada (Meta). **Agron Colombiana**, v. 8, p. 105-115, 1991.

TOXOPEUS, H. Cocoa breeding; a consequence of mating system, heterosis and population structure. In: Conference on Cocoa and Coconuts in Malaysia, **Proceedings...** Kuala Lumpur, 1972. p. 3-12.

VIRK, P. S.; FORD-LLOYD, B. V.; JACKSON, M. T.; POONI, H. S.; CLEMENO, T. P.; NEWBURY, H. J. Predicting quantitative variation within rice germplasm using molecular markers. **Heredity**, v. 76, p. 296-304, 1996.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, v. 23, p. 4407-4414, 1995.

WANG, J.; MCCLEAN, P.; LEE, R.; GOOS, R.; HELMS, T. Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. **Theoretical and Applied Genetics**, v. 116, p. 777-787, 2008.

WARREN, J.; MISIR, S.; KALAI. Isozyme markers for self-incompatibility and yield in *Theobroma cacao*. **Heredity**, v. 74, p. 354-356, 1995.

WEINER, M. P.; HUDSON, T. J.; Introduction to SNPs: Discovery of markers for disease. **Bio Techniques**, v. 32, p. 4-13, 2002.

WILLIAMS, J. G. K.; KUBELIK, A. R.; LIVAK, K. J.; RAFALSKI, J. A.; TINGEY, S. V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research**, v. 18, p. 6531- 6535, 1990.

WRAY, N. R.; VISSCHER, P. M. Population genetics and its relevance to gene mapping. In: NEALE, B. N. M.; FERREIRA, M. A. R.; MEDLAND, S. E.; POSTHUMA, D. (Org.).

Statistical genetics: Gene mapping through linkage and association. New York: Taylor & Francis Group, 2008. p. 87-112.

YAMADA, M. M.; BARTLEY, B. G. D.; CASTRO, G. C. T.; MELO, G. R. P. Herança do fator incompatibilidade em *Theobroma cacao* L.. I. Relações fenotípicas na família PA (Parinari). **Revista Theobroma**, v. 12, p. 163-167, 1982.

YAMADA, M. M.; FALEIRO, F. G.; CLEMENT, D.; LOPES, U. V.; PIRES, J. L.; MELO, G. R. P. Relationship between molecular markers and incompatibility in *Theobroma cacao* L.. **Agrotrópica**, v. 22, p. 71-74, 2010.

YAMADA, M. M.; FALEIRO, F. G.; LOPES, U. V.; BAHIA, R.; PIRES, J. L.; GOMES, L. M. C.; MELO, G. R. P. Genetic variability in cultivated cacao populations in Bahia, Brazil, using isozymes and RAPD markers. *Crop Breeding and Applied Biotechnology*, v. 1, p. 377-384, 2001.

YAMADA, M. M.; LOPES, U. V. Paternity analysis of cacao trees selected for resistance to witches' broom disease in plantations of Bahia, Brazil. **Agrotrópica**, v. 11, p. 83-88, 1999.

YANO, M.; KATAYOSE, Y.; ASHIKARI, M.; YAMANOUCHI, U.; MONNA, L.; FUSE, T.; BABA, T.; YAMAMOTO, K.; UMEHARA, Y.; NAGAMURA, Y.; SASAKI, T. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. **Plant Cell**, v. 12, p. 2473-2484, 2000.

YOUNG, A. M. **The Chocolate Tree:** A Natural History of Cacao, 1994, 200p.

YU, J.; PRESSOIR, G.; BRIGGS, W.H.; BI, I.V.; YAMASAKI, M.; DOEBLEY, J.F.; MCMULLEN, M.D.; GAUT, B.S.; NIELSEN, D.M.; HOLLAND, J.B.; KRESOVICH, S.; BUCKLER, E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, v. 38, p. 203-208, 2006.

ZHU, C.; GORE, M.; BUCKLER, E. S.; YU, J. Status and prospects of association mapping in plants. **Plant Genome**, v. 1, p. 5-20, 2008.