

LUIZ ALBERTO PESSONI

**ESTRATÉGIAS DE ANÁLISE DA DIVERSIDADE EM GERMOPLASMA DE
CAJUEIRO (*Anacardium* spp. L.)**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2007

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

P475e
2007

Pessoni, Luiz Alberto, 1965-

Estratégias de análise da diversidade em germoplasma de
cajueiro (*Anacardium* spp. L.) / Luiz Alberto Pessoni.

– Viçosa : UFV, 2007.

xiii, 159f. : il. ; 29cm.

Orientador: Cosme Damião Cruz.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 148-159.

1. Caju - Genética. 2. Caju - Populações. 3. Caju -
Melhoramento genético. 4. Análise multivariada .

5. Marcadores genéticos. 6. *Anacardium occidentale*.

I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 634.5732

LUIZ ALBERTO PESSONI

**ESTRATÉGIAS DE ANÁLISE DA DIVERSIDADE EM GERMOPLASMA DE
CAJUEIRO (*Anacardium* spp. L.)**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 08 de março de 2007.

Prof. Everaldo Gonçalves de Barros
(Co-orientador)

Prof. Luiz Antônio dos Santos Dias
(Co-orientador)

Prof. Luiz Orlando de Oliveira

Dr. Levi de Moura Barros

Prof. Cosme Damião Cruz
(Orientador)

Dedico este trabalho à minha família, em especial aos meus pais Romilda e Otílio (*in memoriam*), como reconhecimento pelo esforço e incentivo dispensados para minha qualificação profissional, mas, acima de tudo, para formação e aperfeiçoamento pessoais.

AGRADECIMENTO

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento pela oportunidade de realização do curso.

À Universidade Federal de Roraima pela concessão de licença e apoio a esta importante etapa da minha capacitação profissional.

Aos órgãos financiadores de pesquisa, pelo apoio à manutenção da infraestrutura das instituições, em especial à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo e recursos financeiros para o trabalho de pesquisa, através do programa de qualificação institucional (PQI).

À Embrapa Agroindústria Tropical pela disponibilização do material genético de seu Banco Ativo de Germoplasma.

Ao Centro de Biotecnologia Aplicada à Agropecuária da UFV (Bioagro), nomeadamente aos Laboratórios de Bioinformática, Genética Molecular de Plantas (BIOMOL) e de Sequenciamento de DNA, imprescindíveis na execução deste trabalho.

Ao Professor Cosme Damião Cruz pelo incentivo, apoio, orientação e, sobretudo, pela amizade, momentos de convívio e exemplo de conduta.

Aos professores Everaldo Gonçalves e Luiz Orlando pela motivação, ensinamentos, disponibilização de espaço físico e recursos materiais para o desenvolvimento da pesquisa.

Ao Dr. João Rodrigues de Paiva, pela acolhida e viabilização do trabalho de campo na Embrapa Agroindústria Tropical.

Às servidoras da UFV, Maria Salvadora, Rita de Cássia e Rose Tomás, pelo pronto atendimento às minhas solicitações.

Aos membros da banca examinadora, professores Everaldo, Luiz Antonio, Luiz Orlando e doutor Levi de Barros, pelas críticas e contribuições para a melhoria do trabalho.

A todos os colegas da UFV, em especial aos de laboratório: Adésio, Ana, Andréia, Bia, Bruna, Carlos, Cassiana, Chico, Clever, Demerson, Gisele, Geovane, Glauco, João Paulo, Jorge, Luciano, Maíra, Márcia Flores, Márcia

Regina, Márcia Carvalho, Matheus, Nilton, Pedro, Polyana, Thiago, Valéria, Willian, pelo convívio, amizade e paciência nos momentos difíceis.

Aos Amigos de todas as horas, Leonarda, Marcos, Fábio, Sônia e Kelly, pelo companheirismo, momentos de alegria e descontração e, sobretudo, pelo indispensável suporte emocional.

À Cláudia Pombo, Flávio e Lívia pela atenção e acolhida dispensada durante o trabalho de coleta de dados em Campos RJ.

À Sr^a Cláudia Laranjeira pela simpática recepção e pronta autorização para realização do trabalho de campo nos cajuais nativos de sua propriedade.

À Tereza Cristina e Mônica pela amizade e acolhida em Fortaleza.

Ao Professor Fabrizio D'Ayala Valva, da Universidade Federal de Goiás, que foi sempre uma referência de dedicação e conduta profissional em minha vida.

Aos professores, pesquisadores, funcionários e estagiários das diversas instituições que, mesmo sem terem sido nomeados, contribuíram para minha formação profissional e para a realização deste trabalho.

BIOGRAFIA

Luiz Alberto Pessoni, filho de Otílio Pessoni e Romilda Montagnini Pessoni, nasceu em 18 de agosto de 1965, em Inhumas, Goiás.

Em dezembro de 1990, graduou-se em Agronomia pela Universidade Federal de Goiás.

Em fevereiro de 1995, concluiu o curso de Mestrado em Genética e Melhoramento de Plantas pela Universidade Federal de Goiás.

Foi Professor auxiliar, com contrato temporário, de bioestatística e genética do Departamento de Biologia Geral da UFG, no período de março de 1994 a fevereiro de 1995.

A partir de março de 1995, tornou-se professor assistente do Departamento de Biologia da Universidade Federal de Roraima, tendo sido aprovado em concurso público para provimento de cargo efetivo em agosto de 1995. Atualmente é professor adjunto II do Departamento de Biologia desta instituição.

Em março de 2003, iniciou o curso de Doutorado no Programa de Pós-Graduação em Genética e Melhoramento da UFV, submetendo-se à defesa de tese em 08 de março de 2007.

SUMÁRIO

	Página
RESUMO	x
ABSTRACT	xii
1 INTRODUÇÃO GERAL	1
2 REVISÃO DE LITERATURA	6
2.1 O gênero <i>Anacardium</i>	7
2.1.2 Distribuição geográfica e hábito	7
2.1.3 Morfologia floral	8
2.1.4 Biologia reprodutiva	9
2.1.5 Cariótipo	12
2.1.6 Análise cladística	12
2.1.7 Descrição de algumas espécies e/ou ecótipos de <i>Anacardium</i>	14
2.2 Análises da Diversidade de Recursos Genéticos Aplicadas a Caracteres Morfométricos	16
2.2.1 Componentes principais	17
2.2.2 Análise discriminante	19
2.2.2.1 Análise discriminante canônica	21
2.2.2.2 Análise discriminante baseada em distâncias de Mahalanobis	23
2.2.3 Análise de agrupamento	25
2.2.4 Aplicações de técnicas multivariadas na análise de diversidade baseada em caracteres fenotípicos	26
2.3 Marcadores Moleculares na Análise Genética Vegetal	29
2.3.1 Marcadores ISSR	31
2.3.2 Marcadores moleculares em cajueiro	34
2.4 Análises da Diversidade de Recursos Genéticos Aplicadas a Variáveis Binárias	36
2.4.1 Análise discriminante	36
2.4.2 Análise de variância molecular	37

CAPÍTULO 1

ANÁLISE DA DIVERSIDADE EM *Anacardium* spp. BASEADA EM DADOS MORFOMÉTRICOS

1 INTRODUÇÃO	43
2 MATERIAL E MÉTODOS	45
2.1 Material vegetal	45
2.2 Coleta de dados	48
2.3 Análise da divergência entre ecótipos e/ou espécies de <i>Anacardium</i>	49
2.4 Análise de divergência segundo a origem geográfica de acessos de <i>Anacardium</i>	49

2.5 Análises dos dados	50
2.5.1 Análise de variância	50
2.5.2 Análises discriminantes	51
2.5.3 Análise de agrupamento	52
3 RESULTADOS	53
3.1 Análise da divergência entre ecótipos e/ou espécies de <i>Anacardium</i>	53
3.2 Análise de divergência segundo a origem geográfica de acessos de <i>Anacardium</i>	59
3.3 Análises de agrupamento	65
4 DISCUSSÃO	69
4.1 Divergência entre ecótipos e/ou espécies de <i>Anacardium</i>	69
4.2 Divergência segundo a origem geográfica de acessos de <i>Anacardium</i>	71
4.3 Análises de agrupamento	72
5 CONCLUSÕES	74

CAPÍTULO 2

ANÁLISE DA DIVERSIDADE EM *Anacardium* spp. BASEADA EM MARCADORES MOLECULARES ISSR

1 INTRODUÇÃO	76
2 MATERIAL E MÉTODOS	80
2.1 Material vegetal	80
2.2 Extração do DNA	80
2.3 Quantificação e qualidade do DNA	84
2.4 Seleção de <i>primers</i>	84
2.5 Otimização das condições de amplificação	84
2.6 Resolução e visualização dos fragmentos amplificados	86
2.6.1 Sistema em gel desnaturante de poliacrilamida	86
2.6.2 Sistema em gel de agarose	88
2.7 Análise dos dados	89
2.7.1 Análises preliminares e descritivas	89
2.7.2 Estimação do número ótimo de marcadores	90
2.7.3 Análise discriminante	91
2.7.4 Projeção gráfica das distâncias	92

2.7.5 Análise de variância molecular	93
2.7.6 Análise de agrupamento entre populações e acessos	95
3 RESULTADOS	95
3.1 Padrões da variabilidade acessada	95
3.2 Estimação do número ótimo de marcadores	99
3.3 Análise discriminante	99
3.4 Projeção gráfica das distâncias	102
3.5 Análise de variância molecular	102
3.6 Análise de agrupamento	104
4 DISCUSSÃO	108
4.1 Aspectos da variabilidade acessada	108
4.2 Análises discriminantes	109
4.3 Análise de variância molecular e de agrupamento entre as populações	111
5 CONCLUSÕES	112

CAPÍTULO 3

ANÁLISE DA DIVERSIDADE EM *Anacardium* spp. BASEADA EM CARACTERES FENOTÍPICOS DISCRETOS

1 INTRODUÇÃO	115
2 MATERIAL E MÉTODOS	116
2.1 Material vegetal	116
2.2 Coleta de dados	116
2.3 Análises estatísticas	117
3 RESULTADOS	120
3.1 Distribuição e índices de diversidade dos caracteres	121
3.2 Análise discriminante e projeção das distâncias	121
3.3 Análise de agrupamento	123
4 DISCUSSÃO	124
5 CONCLUSÕES	128

CAPÍTULO 4

ANÁLISE COMPARATIVA DA DIVERSIDADE EM *Anacardium* spp. ACESSADA POR MEIO DE DADOS MORFOMÉTRICOS, FENOTÍPICOS DISCRETOS E MOLECULARES

1 INTRODUÇÃO	130
2 MATERIAL E MÉTODOS	133

2.1 Material vegetal	133
2.2 Coleta de dados	133
2.3 Análises estatísticas dos dados	134
3 RESULTADOS	135
3.1 Correlações entre medidas de dissimilaridades	135
3.1 Análises discriminantes	136
3.3 Análises de agrupamento dos acessos	138
4 DISCUSSÃO	142
5 CONCLUSÕES	144
3 CONCLUSÕES GERAIS	145
4 BIBLIOGRAFIA	148

RESUMO

PESSONI, Luiz Alberto, D.Sc., Universidade Federal de Viçosa, março de 2007. **Estratégias de avaliação da diversidade em germoplasma de cajueiro (*Anacardium* spp. L.)**. Orientador: Cosme Damião Cruz. Co-orientadores: Everaldo Gonçalves de Barros, João Rodrigues de Paiva e Luiz Antônio dos Santos Dias.

O cajueiro (*Anacardium occidentale*) é uma fruteira arbórea perene, atualmente cultivada em diversos países da faixa intertropical terrestre. O território brasileiro constitui o provável centro de origem e principal centro de diversidade desta e da maioria das espécies do gênero *Anacardium*. Por outro lado, estudos envolvendo a avaliação e caracterização do germoplasma disponível ainda são incipientes e restritos a acessos de *A. occidentale*. Neste trabalho foi avaliada a diversidade genética em *Anacardium* spp., a partir de acessos mantidos pelo Banco Ativo de Germoplasma do Cajueiro (BAG – Embrapa Agroindústria Tropical), localizado em Pacajus - CE e de indivíduos amostrados em uma população subespontânea de *Anacardium occidentale* L., localizada no estado do Rio de Janeiro. A caracterização envolveu a análise de dados de descritores fenotípicos discretos e morfométricos e de marcadores moleculares ISSR. Foram analisados 45 indivíduos da população do Rio de Janeiro e 91 acessos do BAG, de diferentes procedências, pertencentes às espécies *A. occidentale* (tipos comum e anão precoce), *A. othonianum*, *A. humile* e *A. microcarpum*. As informações morfométricas foram submetidas a técnicas multivariadas consistindo de análises discriminantes de Anderson e de componentes principais e análises de agrupamento por componentes principais, UPGMA e otimização de Tocher. Os dados moleculares e fenotípicos discretos foram investigados por meio de análises discriminantes baseadas no método não paramétrico de k vizinhos mais próximos, projeção gráfica das distâncias no espaço e agrupamento por UPGMA. Análise de variância molecular (AMOVA) foi aplicada aos dados moleculares e análises comparativas dos resultados, fornecidos pelas três classes de dados e pelas suas combinações, através da soma de matrizes de distâncias, também foram realizadas. Verificou-se que as técnicas de análise discriminante aplicadas às variáveis morfométricas foram adequadas para a caracterização dos grupos de

Anacardium avaliados. Entretanto, o método baseado em funções discriminantes de Anderson, proporcionou melhor discriminação dos grupos, além de possibilitar a classificação de acessos desconhecidos ou cujas informações foram consideradas duvidosas. O padrão de agrupamento UPGMA, baseado em dados morfométricos, exibido pelas populações analisadas refletiu adequadamente o grau de relacionamento esperado entre elas, considerando as evidências biológicas e históricas disponíveis. Em relação à análise discriminante, baseada em dados moleculares, observou-se que ela foi adequada apenas na classificação de acessos de diferentes populações de *A. occidentale*. A AMOVA indicou que 27,5% da variação acessada foi devido a diferenças entre as populações investigadas, enquanto o restante foi decorrente de diferenças entre acessos dentro das populações. Por outro lado, os valores estimados de ϕ_{ST} entre pares de populações demonstraram que, à exceção do par 'othonianum' x 'humile', todas as populações foram estatisticamente diferentes entre si. A maioria dos descritores fenotípicos discretos analisados apresentou ampla variabilidade, possibilitando discriminar, sem ambigüidades, todos os acessos de *Anacardium* avaliados. Por outro lado, os padrões de distribuição desta variabilidade foram "inadequados" para caracterizar grupos de acessos de origem comum. A análise comparativa das diferentes classes de dados mostrou que apenas os descritores morfométricos e os marcadores moleculares foram correlacionados, enquanto os descritores multicategóricos apresentaram correlações nulas com as classes de variáveis morfométricas e moleculares, respectivamente. Foi confirmado que a região Nordeste é um importante centro de diversidade de *A. occidentale*, mas o Norte Fluminense e o Estado de Roraima apresentam populações naturais que diferem daquelas encontradas nos Estados do Nordeste necessitando, por isso, serem mais investigadas através de prospecções e coletas de materiais. Em relação a *A. othonianum* e *A. humile*, concluiu-se que estudos adicionais envolvendo acessos e populações naturais destas espécies necessitam ser realizados, tanto para elucidar melhor o *status* taxonômico de cada uma delas, quanto para fins de utilização e conservação destes recursos genéticos.

ABSTRACT

PESSONI, Luiz Alberto, D.Sc., Universidade Federal de Viçosa, March of 2007.

Strategies for evaluating the diversity in germplasm of the cashew (*Anacardium* spp. L.). Adviser: Cosme Damião Cruz. Co-advisers: Everaldo Gonçalves de Barros, João Rodrigues de Paiva and Luiz Antônio dos Santos Dias.

The cashew (*Anacardium occidentale*) is a perennial arboreal fruit tree that is actually cropped in several tropical countries around the World. The Brazilian territory is probably the original and main diversity center of this tree fruit, as well as for most species of the genus *Anacardium*. On the other hand, studies concerning to the evaluation and characterization of the available germplasm are still incipient and restricted to accesses of *A. occidentale*. In this study, the genetic diversity in *Anacardium* spp. was evaluated from either the accesses kept by the Banco Ativo de Germoplasma do Cajueiro (BAG - Embrapa Agroindústria Tropical) located in Pacajus - CE and individuals sampled in a subspontaneous population of *Anacardium occidentale* L. located in Rio de Janeiro State. The characterization involved the analysis of the data for both morphometric and discrete phenotypic descriptors and the DNA molecular markers of the ISSR type. Forty five individuals of the Rio de Janeiro population and 91 accesses of BAG from different provenances, belonging to the species *A. occidentale* (common and precocious dwarf types), *A. othonianum*, *A. humile* and *A. microcarpum* were analyzed. The morphometric data were subjected to multivariate techniques consisting of discriminant analysis by Anderson's and principal components methods, and cluster analysis by hierarchical UPGMA method and optimizing Tocher's approach. Both molecular and discrete phenotypic data were investigated, by using discriminant analyses based on nonparametric *k*-nearest neighbor method, graphic projection of the distances in space and cluster analysis by UPGMA. Molecular variance analysis (AMOVA) was applied to the molecular data, as well as the comparative analyses of the results supplied by either three data classes and their combinations were also accomplished. The discriminant analysis techniques applied to the morphometric variables were appropriate for characterization of the *Anacardium* groups under evaluation. However, the

method based on Anderson's discriminant functions provided better discrimination of the groups, besides making possible the classification of unknown accesses or whose information were considered as doubtful. The cluster pattern UPGMA, based on morphometric data, exhibited by the populations under analysis rather appropriately reflected the level of the expected relationship among them, as considering both biological and historical evidences available. Relative to the discriminant analysis based on molecular data, it was only adequate for classification of the accesses from different *A. occidentale* populations. AMOVA showed that 27.5% of the accessed variation were due to the differences among the populations under investigation, whereas the remainder occurred because the differences among accesses within populations. On the other hand, the estimated values of ϕ_{ST} between the population pairs showed that all populations were statistically different from each others, except for the pair 'othonianum' x 'humile'. Most discrete phenotypic descriptors presented wide variability, making possible to discriminate without ambiguities all *Anacardium* accesses under evaluation. However, the distribution patterns of this variability were "inadequate" to characterizing access groups from common origin. The comparative analysis of the different data classes showed that only the morphometric descriptors and molecular markers were correlated, whereas the discrete phenotypic descriptors presented null correlations with the others data classes. The Nordeste region was confirmed to be an important diversity center of *A. occidentale*, but Norte Fluminense and the Roraima State show natural populations differing from those found on the Northeast States. Therefore, there is a need for a more accurate investigation and material collections of these populations. Relative to *A. othonianum* and *A. humile*, it was concluded that there is a need for additional studies involving accesses and natural populations of these species either for getting a better elucidation of the taxonomic *status* of each one and the use and conservation of these genetic resources.

1 INTRODUÇÃO GERAL

O estudo da diversidade visa elucidar relações genéticas, quantificar ou prever o nível de variabilidade total existente e a sua distribuição entre e/ou dentro de unidades taxonômicas, quer elas sejam indivíduos, acessos de bancos de germoplasma, linhagens, cultivares, populações (de sistemas controlados de acasalamento ou naturais) e espécies. Este conhecimento tem proporcionado importantes contribuições ao melhoramento genético, ao gerenciamento de bancos de germoplasma, a conservação de recursos genéticos e ao entendimento dos processos evolutivos das espécies.

A quantificação da diversidade genética auxilia no planejamento de estratégias mais eficazes que venham a maximizar os ganhos genéticos no melhoramento. Dentre as possibilidades de estudos da diversidade, destacam-se, entre outras (Mohammadi e Prasanna, 2003; Reif et al., 2005):

- a) determinação das inter-relações genéticas entre linhagens, cultivares e populações;
- b) identificação de combinações parentais adequadas à obtenção de híbridos altamente heteróticos e que possibilitem maior segregação em recombinações, com o aparecimento de transgressivos;
- c) introgressão de genes favoráveis provenientes dos acessos de bancos de germoplasma, componentes da base genética da espécie alvo;
- d) identificação de variedades derivadas no processo de proteção de cultivares.

Nos bancos de germoplasma, a análise da diversidade pode ajudar a classificar corretamente um acesso, identificar duplicatas e subgrupos de coleções núcleo e auxiliar na quantificação do nível de variabilidade presente em um *pool* gênico, bem como seu fluxo através do tempo (Reif et al., 2005).

Atualmente, existem muitos métodos disponíveis para o estudo da diversidade genética, seja em avaliações de acessos de bancos de germoplasma, cultivares melhoradas ou populações. Estes métodos podem ser aplicados na análise de dados de pedigree, dados morfológicos, dados de performance agrônômica, dados bioquímicos e dados moleculares baseados na análise de DNA. Entretanto, trabalhos envolvendo espécies perenes

tropicais, ainda são relativamente reduzidos e, normalmente, associados a algumas poucas espécies de grande expressão econômica mundial.

O cajueiro, (*Anacardium occidentale* L.), é uma importante fruteira arbórea tropical pertencente à família Anacardiaceae, cujos demais membros são também de ocorrência predominantemente tropical e subtropical. Devido, principalmente, à importância econômica que o comércio de sua amêndoa (castanha) apresenta no mercado internacional, esta espécie é explorada em diversos países tropicais (Barros e Crisóstomo, 1995). A área mundial ocupada com a cultura do cajueiro está estimada em 2,6 milhões de hectares (Pimentel et al. 2002) e os maiores produtores de castanha *in natura* em 2005 foram (em milhares de toneladas): Vietnã (960,8), Nigéria (594,0), Índia (460,0), Brasil (251,3) e Indonésia (122,0) (FAO, 2007). Entretanto, os países maiores produtores de castanha beneficiada são, respectivamente, Índia, Vietnã e Brasil, que respondem por 60%, 24% e 15,1% do volume de produção mundial (FAO, 2007).

No Brasil, mais de 90% do caju é produzido na região Nordeste, com áreas concentradas nos estados do Ceará, Piauí e Rio Grande do Norte (Barros et al. 2002). Nestes Estados, a agroindústria do caju exerce importante papel econômico e social, pelo número significativo de empregos gerados e expressiva participação na geração de divisas externas, com valores entre 140 e 160 milhões de dólares anuais (Barros, 2002). Outros produtos derivados, como o aproveitamento do pedúnculo na produção de suco concentrado, fabricação de doces e comercialização na forma de fruta fresca, apresentam enorme importância para mercado interno, ainda com grande potencial de expansão (Barros, 2002).

A baixa produtividade das lavouras é o problema mais significativo da cajucultura no Brasil, com média de 220 kg de castanhas *in natura* por hectare, razão pela qual o programa de pesquisa de melhoramento genético da Embrapa Agroindústria Tropical vem dando prioridade à obtenção de cultivares mais produtivas (Barros e Crisóstomo, 1995).

Diversos fatores contribuem para a baixa produtividade dos atuais plantios, dentre eles a propagação realizada por sementes é o principal, já que

o cajueiro é uma espécie predominantemente alógama, o que resulta na heterogeneidade de diversos caracteres da planta (Crisóstomo et al., 1992; Barros et al., 2002). Por outro lado, a propagação clonal, através de enxertia de genótipos de cajueiro do tipo anão precoce trouxe novas perspectivas para o desenvolvimento da cultura no Brasil, elevando o potencial produtivo, em regime de sequeiro, para níveis superiores a 1300 kg/ha de castanha, com a vantagem das plantas apresentarem redução de porte, em relação ao tipo comum (Barros e Crisóstomo, 1995).

O porte baixo das plantas é um caráter da maior importância na fruticultura moderna. A introdução de clones de cajueiro do tipo anão precoce proporcionou a uniformização da copa, facilitando as práticas de manejo, como poda e combate a pragas e doenças, assim como colheita do pseudofruto, atividades inviáveis em pomares de cajueiro do tipo comum (Barros et al., 2000). Entretanto, alguns caracteres como a qualidade e o peso da castanha e amêndoa, resistência a pragas e doenças, menor teor de tanino e conservação pós-colheita do pseudofruto, para consumo *in natura*, necessitam de avanços.

Estudos básicos de herança e de estimação de parâmetros genéticos são também fundamentais na orientação de um programa de melhoramento de qualquer espécie. No caso do cajueiro, os trabalhos desta natureza são relativamente recentes, podendo ser citados, entre outros, os de Paiva et al. (1998) e Holanda-Neto et al. (2001), que avaliaram os efeitos da endogamia sobre a cultura; Azevedo et al. (1998) e Cavalcanti et al. (2000a), que realizaram, respectivamente, estudos de correlações e de repetibilidade, para caracteres de produção e porte da planta; Cavalcanti et al. (2000b), que avaliaram a divergência genética e a heterose de híbridos interpopulacionais entre cajueiro comum e anão precoce; Cardoso et al. (1999), que realizaram estudos da genética da resistência do cajueiro anão à antracnose, mofo preto e mancha angular, três das principais doenças fúngicas no Brasil e Crisóstomo et al. (2002), que apresentaram resultados preliminares sobre incorporação de caracteres de melhoria da qualidade do pseudofruto, através da hibridização de *A. occidentale* e *A. microcarpum*.

Apesar de ser encontrada, na atualidade, em toda faixa intertropical da Terra, a maior variabilidade de *A. occidentale* ocorre no Brasil, particularmente nas zonas costeiras do Nordeste, onde populações espontâneas fazem parte

da vegetação de dunas e restingas (Lima, 1986; apud Barros et al. 2002). Por outro lado, Barros et al. (2002) destacam que, diferentemente de outras espécies do gênero, o cajueiro não suporta a concorrência com outras espécies arbóreas. Segundo Barros (1995) as teorias atuais sobre a origem do cajueiro fundamentam-se ainda em provas circunstanciais, as quais apontam o Brasil, ou pelo menos o norte da América do Sul e parte da América Central, como o mais provável centro de origem e domesticação da espécie. Dentre os fatos e argumentos apresentados pelo autor destacam-se: a distribuição geográfica e comportamento ecológico das espécies do gênero; os padrões de variação de *A. occidentale*, as mais antigas referências históricas e ilustrações sobre a planta; a origem etimológica da palavra caju e suas variações em outros idiomas, assim como a estreita associação, no território brasileiro, entre as atividades humanas (de populações indígenas e não indígenas) e a planta de cajueiro. Weiduschat (1999), a partir de uma revisão de documentos históricos, relatos de membros de comunidades indígenas que habitam as regiões de savana e análise de caracteres morfométricos de populações espontâneas de cajueiro, argumenta a favor do caráter “nativo” desta espécie no Estado de Roraima. Mitchell e Mori (1987), também reconhecem a existência de populações indígenas de *A. occidentale* nas savanas da Colômbia, Venezuela e Guianas.

De acordo com Mitchell e Mori (1987), *A. occidentale* é uma espécie polimórfica, constituída por dois ecótipos distintos: um ecótipo típico de zonas de restinga, da qual se originaram as formas cultivadas e um ecótipo típico dos cerrados do Brasil Central, não cultivado e associado à vegetação aberta de fisionomia savanóide desta região.

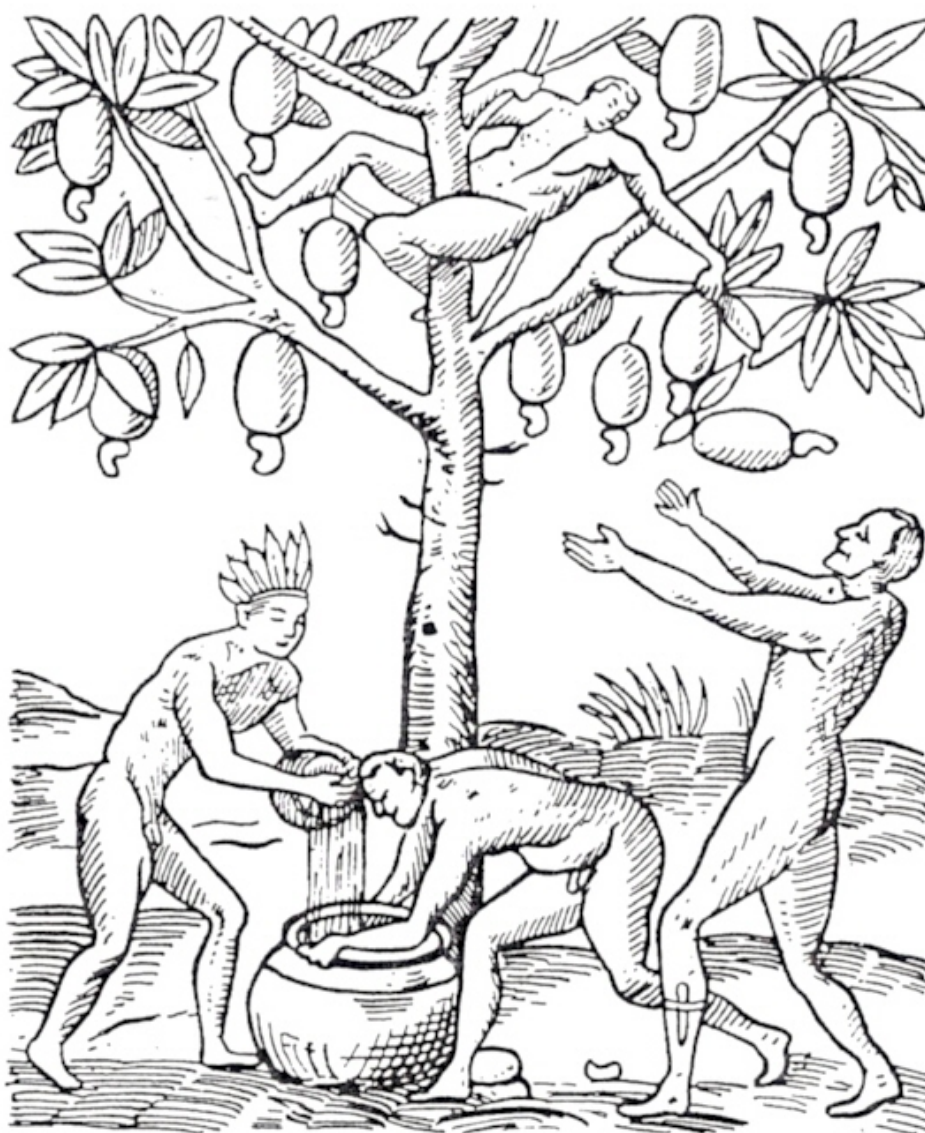
Apesar dos principais centros de diversidade do gênero *Anacardium* se localizarem na Amazônia e no Planalto Central do Brasil, trabalhos de avaliação e caracterização do germoplasma disponível ainda são incipientes, destacando-se apenas o estudo de Barros (1991), que fez a caracterização morfológica e isoenzimática, por meio de técnicas multivariadas, de parte do Banco Ativo Germoplasma do Cajueiro (BAG-Cajueiro), do Centro Nacional de Pesquisa de Caju (atual Embrapa Agroindústria Tropical). Trabalhos deste tipo fornecem subsídios fundamentais aos programas de melhoramento da espécie,

assim como ao processo de otimização nas atividades de coleta e de conservação dos recursos genéticos, tanto *in situ* quanto *ex situ*.

O BAG do cajueiro localiza-se no Campo Experimental de Pacajus, no litoral Leste do Ceará (4°10'S e 38°27'W) a 155 km de Fortaleza, pela BR 116. A coleção de germoplasma é formada por 621 acessos, sendo 565 de *A. occidentale* e 56 de outras espécies do gênero, originários da Região dos Cerrados (Paiva et al. 2003). Uma descrição mais detalhada do BAG, com informações relacionadas ao histórico de coletas, coleções disponíveis e atividades de caracterização já executadas, é encontrada no trabalho de Paiva et al. (2003).

Os objetivos deste trabalho foram avaliar estratégias de análise da diversidade genética em *Anacardium* spp., a partir de acessos mantidos pelo BAG do Cajueiro (Embrapa CNPAT – CE) e de indivíduos amostrados em uma população subespontânea de *Anacardium occidentale* L., localizada no estado do Rio de Janeiro, considerando a utilização de informações de descritores morfométricos e fenotípicos discretos e de marcadores de moleculares de DNA do tipo ISSR.

2 REVISÃO DE LITERATURA



A mais antiga ilustração conhecida de cajueiro, feita por Thevet (1558).
(extraído de Johnson (1974): O caju no Nordeste do Brasil – um estudo geográfico)

2.1 O gênero *Anacardium*

Anacardium (L.) é o gênero tipo da família Anacardiaceae. Segundo Ferreira (1973), esta família é composta por árvores ou arbustos de folhas alternas, coriáceas e sem estípulas, sempre providas de canais resiníferos em seus ramos. Vários gêneros são produtores de frutos comestíveis, como *Mangifera* (mangas), *Spondias* (cajás e umbu) e *Anacardium* (cajus), além de outros possuidores de madeira de ótima qualidade como *Astronium* (Gonçalo Alves, aroeiras) e *Schinopsis* (braúna e quebracho).

Mitchell e Mori (1987) realizaram a mais recente revisão taxonômica sobre *Anacardium*. De acordo com os autores, *Anacardium* é um pequeno gênero de árvores, arbustos e subarbustos nativos dos neotrópicos. Neste trabalho são reunidas informações sobre anatomia, morfologia, ecologia, fitogeografia, relações interespecíficas e botânica econômica sobre o gênero.

A primeira descrição e ilustração, baseadas em *Anacardium occidentale*, foi publicada pelo naturalista francês Andre de Thevet em 1558, a partir de observações de indivíduos ocorrentes na costa nordeste do Brasil. A este trabalho se seguiram outros, com descrições e ilustrações da mesma espécie, realizados por exploradores portugueses, cronistas e naturalistas europeus pré-Lineanos. Em 1735, Lineu descreveu *Anacardium* como um gênero e em 1853 a espécie *A. occidentale*. A partir de 1830, Lindley designou *Anacardium* como gênero tipo da família Anacardiaceae (Mitchell e Mori, 1987).

Aproximadamente, entre 21 e 22 espécies de *Anacardium* são reconhecidas pela taxonomia formal clássica (Ferreira, 1973; Barros, 1995; Ferrão, 1995). Por outro lado, na monografia de revisão do gênero, baseada em dados de coleções de diversos herbários e trabalhos de campo realizados no Panamá, Guiana Francesa, Amazônia, regiões Nordeste e Centro Oeste do Brasil, Mitchell e Mori (1987) reconheceram a existência de dez espécies apenas.

2.1.2 Distribuição geográfica e hábito

De acordo com Mitchell e Mori (1987), a distribuição natural do gênero *Anacardium* estende-se desde Honduras, na América Central, até o sul do

Paraná e leste do Paraguai. À exceção de *A. excelsum*, nenhuma espécie ocorre a oeste da cordilheira andina, na América do Sul. Segundo os autores, este gênero possui dois centros de diversidade: um centro localizado na Amazônia Central, ilustrado pela ocorrência de quatro espécies (*A. giganteum*, *A. spruceanum*, *A. parvifolium* e *A. microsepalum*) nas cercanias de Manaus e um centro localizado no Planalto Central do Brasil, onde três espécies correlatas (*A. humile*, *A. nanum* e *A. corymbosum*) ocupam o mesmo habitat.

O gênero *Anacardium* apresenta duas formas básicas, quanto ao hábito de crescimento: as espécies de floresta (*A. excelsum*, *A. giganteum*, *A. microsepalum*, *A. parvifolium* e *A. spruceanum*) são constituídas por árvores do dossel ou emergentes. As espécies *A. occidentale* e *A. fruticosum* são árvores de porte médio a baixo com grande envergadura da copa, adaptadas a ambientes de savana como os cerrados e as áreas de restinga do nordeste brasileiro. Por outro lado, *A. corymbosum*, *A. humile* e *A. nanum*, são espécies constituídas por subarbustos com tronco subterrâneo, confinadas ao cerrado do Planalto Central do Brasil (Mitchell e Mori, 1987).

2.1.3 Morfologia floral

As flores das espécies de *Anacardium* são reunidas em inflorescências terminais ou axilares, formando panículas esparsas ou densas, com flores estaminadas e hermafroditas. As flores bissexuadas apresentam cálice imbricado, com cinco sépalas lanceoladas ou ovais, corola pentâmera, campanulada ou cilíndrica, imbricada na pré-floração e reflexa na antese, geralmente apresentando cor avermelhada ou rósea na parte interna ou superior. Os estames são concrecidos na base, com número variando de 6 a 12, possuem comprimentos desiguais, sendo que um, (quatro no caso de *A. excelsum*) por ser muito maior que os demais, se projeta para além dos limites da corola. O ovário é oblíquo obovado, unilocular com um óvulo. O estilete é central ou lateral e geralmente é mais longo do que o maior dos estames. As flores estaminadas possuem formatos e tamanhos similares aos das flores bissexuadas. Contudo, o filete do estame mais longo geralmente é mais comprido do que seu equivalente nas flores bissexuadas, além do ovário e do estilete serem rudimentares (Ferreira, 1973; Mitchell e Mori, 1987).

Estudos histoquímicos, com gemas florais de *A. occidentale* de diferentes idades, revelaram que o tecido pistilar primordial das gemas florais presumivelmente hermafroditas acumulam maiores quantidades de proteína e taxa de crescimento, em relação às zonas pistilares das flores presumivelmente masculinas. Estas observações indicam que a acumulação protéica teria um importante papel na indução do desenvolvimento do gineceu e explicariam a maior preponderância de flores estaminadas na inflorescência, cujo crescimento é indeterminado. Uma vez que os pistilos são os órgãos de diferenciação mais tardia nas gemas primordiais é provável que a insuficiência de metabólicos seja a causa da baixa produção de flores hermafroditas (Pavithran et al., 1985).

Uma vez fertilizada, o pedicelo das flores bissexuadas de *Anacardium* aumentam de tamanho, formando o pedúnculo (hipocarpo) suculento denominado genericamente de caju. Esta combinação do desenvolvimento do fruto verdadeiro constituído por uma drupa rígida e o pedúnculo suculento constitui um processo complicado da chamada “frutificação dupla de *Anacardium*”. Os dois componentes desta frutificação se desenvolvem com diferentes taxas. A drupa alcança a maturação primeiro que o pedicelo. O desenvolvimento do pedicelo é caracterizado pelo maior crescimento inicial no comprimento, seguido pelo aumento no diâmetro, na medida em que se torna mais suculento. No processo de maturação o pedúnculo gradualmente perde seu sabor adstringente através da conversão de taninos livres em formas insolúveis. A cor vermelha brilhante, freqüente nos pedúnculos de várias espécies, é resultante da acumulação de antocianinas nas camadas subepidérmicas (Mitchell e Mori, 1987).

2.1.4 Biologia reprodutiva

A maior parte das informações a respeito da biologia reprodutiva de *Anacardium* é resultante de trabalhos realizados com *A. occidentale*. Freitas e Paxton (1998) analisaram a biologia reprodutiva em uma população espontânea de *A. occidentale*, localizada no litoral do Ceará. O estudo envolveu ainda a comparação da ação de duas espécies de abelhas na polinização: a indígena *Centris tarsata* e a exótica *Apis mellifera*. De acordo

com os autores, as panículas apresentaram uma proporção de 10% de flores hermafroditas para 90% de flores masculinas. A antese nas flores masculinas inicia-se por volta das 6h, encerrando-se por volta das 10h da manhã, enquanto nas flores hermafroditas a antese, a exposição e receptividade do estigma ocorrem no intervalo entre 10h e 12h. A deiscência das anteras ocorre depois das 9h, nas flores masculinas e por volta das 10h, nas flores hermafroditas. Por outro lado, as flores masculinas contribuem, proporcionalmente, com a maior quantidade de pólen viável.

As duas espécies de abelhas tocam os estames das flores masculinas e os estigmas das flores hermafroditas em suas visitas. Por outro lado, as operárias de *A. mellifera* e os machos de *C. tarsata* coletam apenas néctar, enquanto as fêmeas de *C. tarsata* coletam tanto pólen quanto néctar. Aparentemente o pólen é coletado somente da antera do estame mais longo das flores masculinas. Os horários de visitas ao longo do dia são similares para as duas espécies de abelhas, coincidindo com a liberação dos grãos de pólen pelas anteras e a maior receptividade dos estigmas, sugerindo que ambas as espécies são polinizadoras potenciais de *A. occidentale* (Freitas e Paxton, 1998).

Paulino (1992) também realizou trabalho de avaliação da polinização entomófila em cajueiro. O estudo foi conduzido em pomares do campo experimental da Embrapa CNPAT, localizado no estado do Ceará. Embora tenha observado a visita de insetos de diversas espécies, o autor não registrou visitas de *C. tarsata* e concluiu que *A. mellifera* é a espécie mais importante na polinização da cultura.

O ciclo de florescimento de *A. occidentale*, observado em populações naturais e em cultivo de sequeiro, na Região Nordeste, é de cinco a sete meses (Jul/Ago-Dez/Jan) para o cajueiro comum e de sete a nove meses (Jun/Jul-Jan/Fev) para o tipo anão precoce (Barros et al., 2002). Na panícula, o florescimento se inicia com a abertura de flores masculinas, passando por uma fase de florescimento misto, isto é, com abertura de flores masculinas e bissexuais, finalizando com uma fase de abertura de flores exclusivamente masculinas (Barros, 1995; Bueno, 1997). Por outro lado, estudos realizados na década de 1970 em Vittal (Índia), com 100 acessos de diversas procedências, mostraram que pode existir grande variação neste padrão (Kumaran et al.,

1985). De acordo com os autores, cerca de 50% dos acessos apresentaram florescimento bissexual inicialmente, isto é, apenas de flores hermafroditas, enquanto alguns outros acessos iniciaram o florescimento diretamente na fase mista.

De maneira geral, o florescimento das espécies de *Anacardium* coincide com a chegada da estação seca, sendo restrito a este período para as espécies de ambientes abertos ou de savana, onde esta estação é mais pronunciada (Mitchell e Mori, 1987).

Aparentemente, não há nenhum sistema de incompatibilidade que impeça a ocorrência de autogamia no cajueiro (Barros, 1995). Por outro lado, a espécie sofre depressão por endogamia e apresenta autoincompatibilidade parcial, manifestada no maior percentual de abortamento de frutos imaturos resultantes de autofecundações (Paulino, 1992; Holanda-Neto et al., 2002), assim como redução do percentual de germinação das sementes (Paiva et al., 1998) e do tamanho dos frutos (Paulino, 1992).

Estudos de polinização controlada, conduzidos por Holanda-Neto et al. (2002), mostraram que a taxa de frutificação inicial em *A. occidentale* não difere entre tratamentos de autopolinização e polinização cruzada. Confirmando resultados de outros trabalhos, de que o cajueiro aceita os dois tipos de polinização. Este comportamento talvez explique a pouca importância dada à polinização cruzada na produtividade da espécie. Por outro lado, a partir do sétimo dia após a polinização, houve uma taxa de abortamento diferenciada de frutos imaturos entre os dois tipos de tratamentos, de tal ordem que na maturação a taxa de frutos produzidos por polinização cruzada foi 240% maior do que pela autopolinização manual. Os autores ressaltam, contudo, que este significativo aumento no sucesso reprodutivo só ocorre se a transferência de pólen envolver indivíduos de genótipos diferentes e, portanto, esta deve ser uma preocupação adicional na implantação de novos pomares constituídos de clones melhorados.

De acordo com Mitchell e Mori (1987), três espécies simpátricas do planalto central brasileiro, *A. occidentale* (ou *A. othonianum*), *A. humile* e *A. nanum*, florescem no mesmo período. Suas flores são virtualmente idênticas em morfologia e são polinizadas pelos mesmos tipos de vetores, constituídos por borboletas e abelhas. Ainda, de acordo com os autores, parece haver

poucas barreiras extrínsecas ao cruzamento entre elas. Tais fatores explicariam a existência de indivíduos nas populações naturais que são intermediários entre *A. occidentale* e *A. humile*, assim como entre *A. humile* e *A. nanum*.

No que se refere à dispersão das sementes, sabe-se que castanha e pedúnculo juntos flutuam na água o que possibilita a dispersão da espécie pelos rios (Johnson, 1974). Entretanto, atribui-se a morcegos frugívoros o papel de principal agente dispersor das espécies de *Anacardium* que apresentam pedúnculo carnoso e suculento (Johnson, 1974; Mitchell e Mori, 1987). É importante destacar que a dispersão e distribuição atual de *A. occidentale* é resultante, em grande parte, de atividades antrópicas desde o período pré-colombiano (Mitchell e Mori, 1987).

2.1.5 Cariótipo

Estudos relacionados com a determinação do número de cromossomos foram realizados apenas para *A. occidentale*, e os resultados encontrados na literatura são bastante divergentes, com números de cromossomos relatados de $2n=24$, $2n=40$ e $2n=42$ (Goldblatt, 1981; Goldblatt, 1984). Thankamma-Pillai e Nambiar (1985) analisaram 223 células em metáfase I, de seis indivíduos diferentes, e verificaram que 64 a 68% delas apresentaram 21 bivalentes, o que corresponde a um cariótipo $2n=42$. As demais células apresentaram irregularidades cromossômicas que, segundo os autores, explicaria a elevada taxa de grãos de pólen estéreis.

Leão et al. (2001) investigaram células mitóticas, de pontas de raízes de plântulas, de *A. othonianum* e concluíram que o cariótipo da espécie é $2n=16$ cromossomos. Sugeriram ainda que a espécie seria diplóide, enquanto outras como *A. occidentale* seriam poliplóides. Entretanto, essas especulações carecem de mais estudos para serem confirmadas.

2.1.6 Análise cladística

A filogenia das espécies de *Anacardium*, reconhecidas por Mitchell e Mori (1987), foi estabelecida por Michell e Young (1987) através da aplicação

de um método de reconstrução filogenética baseada em parcimônia, utilizando 20 caracteres morfológicos qualitativos. O gênero *Fegimanra* foi utilizado como grupo externo no estabelecimento da polaridade dos estados de caráter e posicionamento da raiz do dendrograma (Figura 1). De acordo com os autores, várias árvores igualmente parcimoniosas foram obtidas a partir dos dados analisados, sendo possível a distinção de três clados:

- a) clado 1 – constituído por *A. excelsum*;
- b) clado 2 – abrangendo as espécies *A. occidentale*, *A. spruceanum*, *A. humile*, *A. nanum* e *A. corymbosum*;
- c) clado 3 – formado por *A. parvifolium*, *A. fruticosum*, *A. giganteum* e *A. microsepalum*.

Os resultados obtidos desta análise, entretanto, devem ser considerados apenas como uma aproximação das inter-relações filogenéticas das espécies de *Anacardium* em decorrência da precária resolução dos cladogramas obtidos (Michell e Young, 1987).

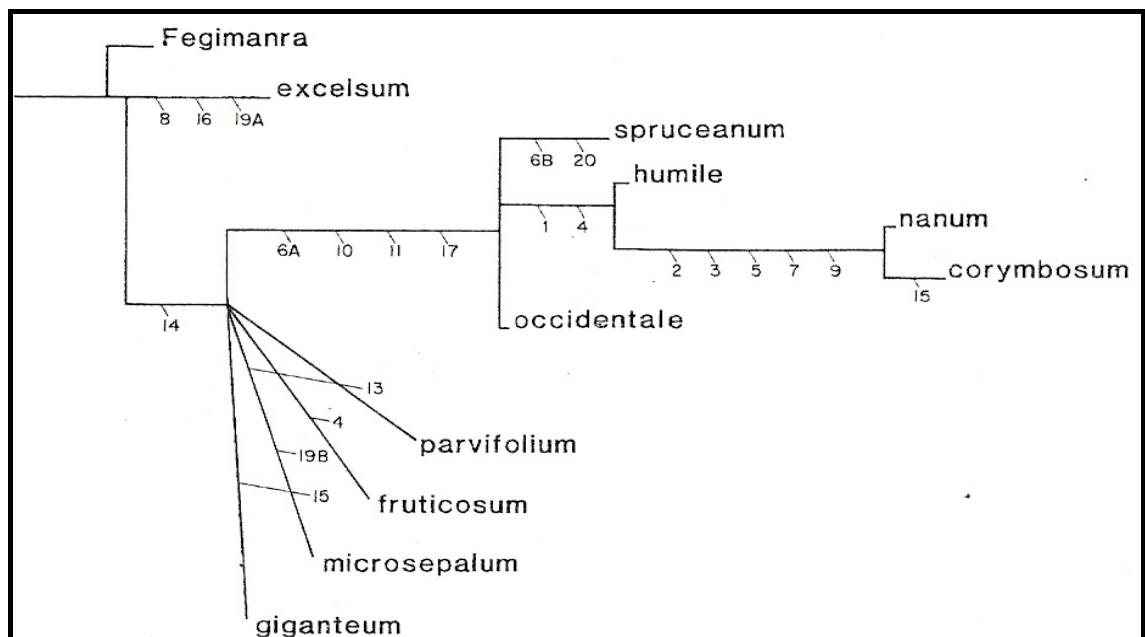


Figura 1 – Cladograma das espécies de *Anacardium* reconhecidas por Mitchell e Mori (1987), obtido por Mitchell e Young (1987) através da reconstrução filogenética, utilizando 20 caracteres morfológicos qualitativos.

2.1.7 Descrição de algumas espécies e/ou ecótipos de *Anacardium*

a) *Anacardium occidentale* L.

De acordo com Mitchell e Mori (1987), *A. occidentale* é uma espécie polimórfica, constituída por dois ecótipos distintos: um ecótipo típico de zonas de restinga, de qual se originaram as formas cultivadas e um ecótipo típico dos cerrados do Brasil Central, não cultivado e associado à vegetação aberta de fisionomia savanóide desta região.

O ecótipo de restinga consiste de árvores de folhas perenes, com ramificação baixa e cuja copa pode alcançar 10 a 15 metros de altura. Possui raiz pivotante muito desenvolvida e extensa rede de raízes fasciculadas. As folhas são simples alternas, arredondadas ou obovadas, muitas vezes indentadas no ápice e alongadas na base, com pecíolos curtos de 1 a 3 cm. Após a emergência, o limbo possui consistência delicada e apresenta coloração acobreada, avermelhada ou verde clara, passando progressivamente para o verde escuro e consistência cartácea ou coriácea, típicos das folhas adultas (Johnson, 1974; Ferrão, 1995). As flores estão reunidas em panículas esparsas a congestas (11-29 x 4,5-24,5 cm e são de dois tipos: bissexuais e estaminadas). As flores bissexuais com corola cilíndrica (3-5 mm de diâmetro), com pétalas reflexas de cor branca ou verde claro, apresentando linhas róseas ou vermelhas na antese, mudando para vermelho escuro após a fertilização. Apresentam de 6 a 10 estames, raramente 12, sendo um (raramente dois) muito maior que os demais, porém todos com anteras normais. Ovário com estilete central e estigma puntiforme. As flores estaminadas possuem pistilódio e estames de comprimentos desiguais como nas flores bissexuais (Mitchell e Mori, 1987). O fruto é uma drupa sub-reniforme (3-5 x 2-3,5 cm), que pesa de 3 a 20 gramas e apresenta cor cinza ou marrom na maturação. O pedúnculo é piriforme e sua cor, na maturação, varia do vermelho ao amarelo. Mede de 5 a 20 cm de comprimento e 2 a 8 cm de largura (Johnson, 1974).

Anacardium occidentale é nativo da América tropical, mas sua distribuição natural não é clara, em razão da longa e íntima associação desta espécie com atividades humanas no continente. Por outro lado, Mitchell e Mori (1987) acreditam que a espécie tenha evoluído originalmente nos

cerrados do Brasil Central e, posteriormente, colonizou as zonas de restingas da costa brasileira.

Alguns autores também diferenciam o ecótipo associado à restinga em duas formas, também denominadas de variedades ou ecótipos: cajueiro comum e cajueiro anão precoce (Almeida et al., 1993; Barros et al., 2002).

O Cajueiro comum possui porte elevado com altura variando de 8 a 15m e envergadura da copa com até 20 m. Apresenta grande variação na distribuição de ramos e formato da copa, que vai desde ereta e compacta até espaiada e grande variabilidade no peso do fruto e pseudofruto, atingindo valores de até 33 e 500 gramas, respectivamente. A idade mínima de estabilização da produção das plantas é superior a oito anos (Barros et al., 2002).

O cajueiro anão precoce caracteriza-se pelo porte baixo (inferior a 4 m), com copa compacta e homogênea e de maior envergadura que o porte (5 a 6,5 m). Nas populações naturais apresenta peso do fruto variando entre 3 e 10 g e pedúnculo com 20 a 160 g (Almeida et al., 1993; Barros et al., 2002). Especula-se que sua origem geográfica seja as savanas de Roraima, mas indivíduos isolados, cultivados ou espontâneos, são encontrados no Ceará e outro estados do nordeste (Almeida et al, 1993).

b) *Anacardium humile* (St. Hilaire)

Denominado de caju do campo ou caju rasteiro, é um subarbusto com tronco subterrâneo vigoroso e ramos ascendentes rígidos de 30 a 150 cm de altura. As folhas coriáceas são elípticas a oblanceoladas, de ápice obtuso, pecíolo muito curto ou séssil e nervura central proeminente. Apresenta inflorescência esparsa a congesta (9-27 x 6-24 cm), localizada na extremidade dos ramos. As flores bissexuadas apresentam corola cilíndrica com pétalas reflexas, de cor branca ou creme, com listas róseas ou vermelhas na antese, tornando-se vermelho escuro após a fertilização. Os estames, 5-9, apresentam comprimentos desiguais, sendo um, ou raramente dois, com filete muito maior que os demais, mas as anteras são normais nos dois tipos. O ovário apresenta estilete central e estigma puntiforme. As flores masculinas também apresentam um estame de maior comprimento e exserto. O pedicelo (pseudofruto) do fruto é tuberizado, desenvolvido, carnoso, suculento de sabor ácido, cônico ou

piriforme (1-3 x 1-2 cm) de cor avermelhada ou amarelada quando maduro. O fruto é uma drupa reniforme (1,3 x 12,3 x 1-1,17 cm) de cor cinza ou marrom na maturidade (Ferreira, 1973; Mitchell e Mori, 1987).

Distribuição geográfica: Santa Cruz, Bolívia, sul e leste do Paraguai e Brasil, do sudeste de Rondônia e norte de Goiás até sul do Paraná.

c) *Anacardium othonianum* (Rizzini)

Conhecida popularmente por cajuzinho de cerrado, esta espécie é uma árvore de porte mediano de 3 a 6 metros de altura e tronco de 20 a 40 cm de diâmetro, de córtex fusco e fissurado. Possui ramos crassos, sulcados e ferruginosos, com folhas elípticas, de ápice obtuso a emarginado, crassas, coriáceas, glabras com 13 a 18 nervuras laterais proeminentes e pecíolo curto. As flores são reunidas em panículas amplas, laxas mais ou menos corimbosas, com 15 a 25 cm de comprimento por 15 a 20 cm de largura. As flores bissexuadas têm pedicelos pilosos com 2 a 3 mm, pétalas estreitas alongadas e externamente pubérulas com 6 a 7 mm de comprimento. O ovário é oblíquo e o estigma é puntiforme. Os estames, em número de 8, são concrecidos na base. O pseudofruto apresenta cor avermelhada ou amarelada, suculento e ácido quando maduro. O fruto é reniforme, com 1,5-2,2 cm de comprimento e 1,2-1,5 cm de largura (Ferreira, 1973).

Esta espécie foi considerada por Mitchell e Mori (1987) como membro do ecótipo de *A. occidentale*, circunscrito à região dos cerrados e caracterizado por apresentar folhas mais coriáceas, onduladas e subsésseis, assim como pseudofruto menor e mais ácido que o ecótipo típico das áreas de restingas.

Distribuição Geográfica: Cerrados do Brasil Central, principalmente Goiás e Distrito Federal.

2.2 Análises da Diversidade de Recursos Genéticos Aplicadas a Caracteres Morfométricos

Técnicas analíticas multivariadas permitem a manipulação simultânea de múltiplas medidas, tomadas de cada indivíduo sob investigação e são amplamente utilizadas em estudos de diversidade, independente do tipo de dados analisados, sejam eles morfológicos, bioquímicos ou marcadores

moleculares (Mohammadi e Prasanna, 2003). Entre os algoritmos utilizados destacam-se as análises de componentes principais ou coordenadas principais, análise discriminante e análises de agrupamento.

2.2.1 Componentes principais

A análise dos componentes principais (CPs) é uma técnica multivariada cujo objetivo é partir de um conjunto de p variáveis, designadas por X_1, X_2, \dots, X_p , e encontrar combinações lineares que produzam índices Z_1, Z_2, \dots, Z_p ortogonais entre si. Esta ortogonalidade ou ausência de correlação entre os índices é uma propriedade desejável porque indica que eles estão medindo diferentes "dimensões" dos dados (Manly, 1998). De acordo Anderson (1984), o método dos CPs é, do ponto de vista estatístico, utilizado para encontrar combinações lineares com grande variância, considerando que em muitos estudos exploratórios o número de variáveis disponíveis é muito elevado e o principal interesse do pesquisador se concentra nos desvios das observações. Uma maneira de reduzir o número de variáveis a serem manipuladas é, segundo o autor, descartar combinações lineares que apresentem baixa variância, mantendo somente aquelas com variância elevada. Também é buscada uma ordenação decrescente da grandeza destes índices de tal forma que Z_1 , retenha o máximo da variabilidade disponível, Z_2 o máximo da variabilidade restante disponível e assim por diante (Manly, 1998; Cruz e Carneiro, 2003). Dessa forma $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \dots \geq \text{var}(Z_p)$, onde Z_i e $\text{var}(Z_i)$ indicam o componente principal em consideração e sua variância, respectivamente.

A expectativa de uma análise de componentes principais é que a maior parte da variação dos dados possa ser descrita adequadamente por poucos índices Z_s , enquanto as variâncias dos índices remanescentes sejam tão baixas que possam ser negligenciadas (Manly, 1998). Neste sentido, a análise busca algumas poucas combinações lineares que possam ser utilizadas para descrever os dados, de modo que a perda de informação seja a menor

possível. Esta tentativa de redução da dimensionalidade pode ser qualificada como “redução parcimoniosa” dos dados (Mardia et al., 1997).

Por outro lado, deve ser enfatizado que a técnica dos componentes principais nem sempre atua no sentido de reduzir um grande número de variáveis originais em pequeno número de variáveis transformadas. Os melhores resultados são obtidos quando as variáveis originais são altamente correlacionadas, seja positiva ou negativamente. Neste caso, a análise de CP será importante em estabelecer medidas relacionadas com as “dimensões” implícitas nos dados assim como apontar redundância de informações contidas nas variáveis originais, com muitas delas medindo coisas semelhantes (Manly, 1998).

De acordo com Manly (1998) a análise de componentes principais, de um conjunto de dados envolvendo p variáveis e n indivíduos, pode ser resumida na execução dos seguintes passos:

- a) Padronização das variáveis X_1, X_2, \dots, X_p , de modo que cada uma passe a exibir média zero e variância igual a unidade. Embora possa ser omitido em alguns casos, tal procedimento assegura que nenhuma das variáveis tenha uma influência excessiva sobre os componentes, decorrente de fatores como diferenças de escala ou de amplitude de variação;
- b) Cálculo da matriz de covariância C . No caso dos dados terem sido padronizados, esta matriz corresponde uma matriz de correlação;
- c) Determinação dos $\lambda_1, \lambda_2, \dots, \lambda_p$ autovalores e dos a_1, a_2, \dots, a_p autovetores correspondentes. Os coeficientes do i -ésimo componente principal são dados por a_i , enquanto λ_i é sua variância. Isto é, $Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$, sua variância é $\text{Var}(Z_i) = \lambda_i$ e as constantes $a_{i1}, a_{i2}, \dots, a_{ip}$ são os elementos correspondentes ao autovetor, com a restrição que $a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1$. Esta restrição é importante para que a $\text{Var}(Z_i)$ não aumente pela simples elevação dos valores de qualquer a_{ij} ;
- d) Descarte de todos os componentes que explicam apenas uma pequena porção da variabilidade total. Por exemplo, se os dados originais são representados por um conjunto de 20 variáveis e os dois primeiros

componentes principais explicam 80% ou mais da variabilidade total os outros 18 CP podem ser descartados ou ignorados.

Uma propriedade importante dos autovetores destacada por Manly (1998) é que eles são totalizados pela soma dos elementos diagonais (o traço) da matriz C. Isto é $\lambda_1 + \lambda_2 + \dots + \lambda_p = c_{11} + c_{22} + \dots + c_{pp}$. E, uma vez que c_{ii} é a variância de X_i e λ_i é a variância de Z_i , a soma das variâncias dos componentes principais é igual à soma das variâncias das variáveis originais padronizadas.

Outro importante aspecto da técnica de componentes principais é a possibilidade de avaliar a contribuição relativa de cada uma das variáveis originais na divergência das entidades sob avaliação. Conforme destacado por Cruz e Carneiro (2003) e Cruz et al. (2004), uma vez que a importância relativa dos CP decresce do primeiro para o último, a variável que apresenta maior coeficiente de ponderação ou elemento do autovetor no componente de menor autovalor é considerada de menor importância para explicar a variabilidade do material sob investigação, sendo possível o seu descarte sem comprometimento da qualidade das informações. De acordo estes autores a literatura recomenda, em estudos com variáveis padronizadas, descartar os caracteres de maiores coeficientes (em valores absolutos) nos autovalores menos importantes, partindo-se do último até àquele componente cujo autovetor não exceda a 0,70.

2.2.2 Análise discriminante

A análise discriminante procura determinar uma variável estatística, a partir da combinação linear de variáveis independentes, que fará a melhor distinção entre grupos definidos *a priori*. A discriminação é conseguida estabelecendo-se pesos às variáveis independentes de modo a maximizar a variância entre grupos em relação à variância dentro dos grupos (Hair et al., 2005). A combinação linear para uma análise discriminante, também conhecida como função discriminante, é determinada por uma equação que assume a seguinte forma:

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk}$$

Onde

Z_{jk} = escore Z discriminante da função discriminante J para o objeto K

a = constante

W_i = peso discriminante para a variável independente i

X_{ik} = variável independente i para o objeto k.

Calculando-se a média dos escores discriminantes para todos os indivíduos do grupo, obtém-se a média do grupo. Esta média de grupo é chamada centróide e existe uma para cada grupo envolvido na análise. Os centróides indicam o local mais típico de qualquer indivíduo de um grupo particular, e suas posições relativas ao longo da dimensão testada indicam o quão afastados ou divergentes são os grupos avaliados (Hair et al., 2005).

A alocação dos objetos ou unidades amostrais em uma das várias populações é um problema multivariado, por isso, é necessário produzir um índice ou um critério bem definido que possa ser utilizado como regra de classificação. Evidentemente, nenhuma regra de decisão será perfeita e, portanto, sempre haverá certa probabilidade de erro de classificação. Tal probabilidade, entretanto, pode ser controlada ou minimizada (Khattree e Naik, 2000).

Ao contrário do recomendado em uma análise de componentes principais ou análise de fatores, os dados que serão utilizados em uma análise discriminante não necessitam ser padronizados antes de serem submetidos aos cálculos. Isto porque o resultado de uma análise discriminante não é afetado por problemas de escala de variáveis individuais (Manly, 1998).

A análise discriminante pode ser considerada uma análise de perfil ou uma técnica preditiva analítica que permite ao pesquisador compreender diferenças de grupos ou classificar corretamente objetos em grupos ou classes. Em qualquer caso, esta técnica é mais apropriada onde só existe uma variável categórica dependente e diversas variáveis métricas independentes (Hair et al., 2005).

Por outro lado, segundo Khattree e Naik (2000), embora a distribuição normal multivariada seja uma pressuposição comum para populações, algumas análises discriminantes podem ser realizadas sob a pressuposição de outros tipos de distribuição também. Outras técnicas discriminantes, segundo os

autores, podem ser não paramétricas quando nenhuma pressuposição é feita sobre as distribuições das variáveis, ou ainda serem aplicadas quando algumas ou todas as variáveis são categóricas.

De acordo com Hair et al. (2005), a aplicação da análise discriminante pode ser vista como a construção de um modelo de seis estágios:

- 1) Estabelecimento dos objetivos, avaliando diferenças de grupos em um perfil multivariado e identificando dimensões de discriminação entre grupos;
- 2) Planejamento da pesquisa, selecionando as variáveis independentes, determinando o tamanho da amostra total e criando as amostras de análise e de teste;
- 3) Certificação das suposições inerentes ao modelo, como normalidade e ausência de multicolinearidade das variáveis independentes, linearidade das relações e igualdade das matrizes de dispersão;
- 4) Estimação das funções discriminantes e de suas significâncias estatísticas, avaliação da precisão preditiva através da aplicação de uma matriz de classificação;
- 5) Interpretação das funções discriminantes, determinando quantas funções serão interpretadas e quais variáveis independentes mais contribuem para distinguir os grupos;
- 6) Validação dos resultados discriminantes através da utilização de amostra de teste ou da validação cruzada.

Existem duas metodologias básicas de análise discriminante aplicadas a dados métricos multivariados: análise discriminante canônica e análise discriminante baseada em distâncias de Mahalanobis.

2.2.2.1 Análise discriminante canônica

A análise discriminante canônica ou de componentes principais é uma técnica de redução de dimensionalidade semelhante à análise de componentes principais e correlação canônica (Khattree e Naik, 2000). No contexto da análise discriminante, esta técnica fornece uma representação das várias populações sob investigação em um subespaço de dimensões reduzidas.

De maneira análoga aos componentes principais, o método visa obter combinações lineares das variáveis métricas independentes X , tal que

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

Estas funções “Z” são denominadas de funções discriminantes canônicas. O número máximo de funções estimadas, para um particular conjunto de dados, é valor mínimo entre $(p-1, v)$, em que p é o número de populações e v o número de variáveis (Manly, 1998; Cruz e Carneiro, 2003).

As funções são obtidas em ordem decrescente de importância, de tal forma que a primeira função explica o máximo da variância entre os grupos e a segunda, ortogonal à primeira, explica o máximo da variância remanescente e assim por diante (Manly, 1998; Cruz e Carneiro, 2003). Este processo é repetido até que toda a variância da matriz $W^{-1}B$ seja extraída, onde W é a matriz de variâncias e covariâncias dentro de grupos e B a matriz de variâncias e covariâncias entre grupos. Por outro lado, a expectativa do pesquisador é que apenas as duas ou três primeiras funções sejam suficientes para explicar a maior parte das diferenças entre os grupos (Manly, 1998).

Obtidas as funções que melhor discriminam os grupos sob análise, o passo seguinte é estimar os escores a partir dos valores dos indivíduos pertencentes a cada grupo, de forma que se possa visualizar as dissimilaridades entre estes indivíduos e seus respectivos grupos, por meio do exame da dispersão gráfica dos dados (Cruz e Carneiro, 2003).

Esta análise também possibilita avaliar a importância relativa dos caracteres avaliados na discriminação dos diferentes grupos. Utilizando o critério de peso das variáveis nos autovetores das funções canônicas estimadas, são consideradas de menor importância as variáveis de maiores pesos desde o último autovetor até aquele associado a um autovalor igual ou inferior a 0,70 (Cruz e Carneiro, 2003; Cruz et al. 2004).

2.2.2.2 Análise discriminante baseada em distâncias de Mahalanobis

Anderson (1958, 1984) descreve um método de análise discriminante quando existem $k(\geq 2)$ populações com distribuição normal multivariada, denominadas por $\pi_1, \pi_2, \dots, \pi_k$. O propósito, neste caso, é obter regras de discriminação para classificar quaisquer observações futuras em uma destas K populações (Khattree e Naik, 2000). Para isso, o espaço p -dimensional é dividido em k regiões disjuntas e mutuamente exclusivas, denominadas R_1, R_2, \dots, R_k e se uma observação cai dentro da região R_i assume-se que ela é proveniente da população π_i . Adicionalmente, são assumidas probabilidades a *priori* para cada uma das populações, de modo que $\pi_1 + \pi_2 + \dots + \pi_k = 1$, além de um custo de má classificação na construção das funções de discriminação (Anderson, 1958, 1984; Khattree e Naik, 2000).

Anderson (1958, 1984) mostrou que a regra de Bayes, para um dado conjunto de probabilidades a *priori* e uma matriz específica de má classificação, é baseada em razões de probabilidades para todos os pares de grupos, se vetores de médias (μ_k) e a matriz de covariância comum (Σ) ou suas estimativas são conhecidos.

Utilizando a notação de Cruz e Carneiro (2003), a forma genérica da função discriminante de Anderson é dada pela expressão:

$$D_j(\tilde{x}) = \ln(p_j) - \frac{1}{2}(2\tilde{x}'\Sigma^{-1}\mu_j + \mu_j'\Sigma^{-1}\mu_j)$$

onde

p_j : é a probabilidade a *priori* de indivíduo ou elemento pertencer a uma determinada população;

\tilde{x} : é o vetor de observações do indivíduo cuja população é desconhecida;

Σ : é a matriz de variâncias e covariâncias comum das populações;

μ_j : é o vetor de médias das populações.

$$\text{Ou } D_j(\tilde{x}) = \kappa_j + \alpha_{j1}x_1 + \alpha_{j2}x_2 + \dots + \alpha_{jv}x_v$$

Onde

$\kappa_j = \ln(p_j) - \frac{1}{2}\mu_j'\Sigma^{-1}\mu_j$: constante associada à função discriminante;

α_{jk} : coeficiente de ponderação da k-ésima variável ($k=1, 2, \dots, v$) na j-ésima função discriminante;

x_k : valor representativo do escore da k-ésima variável do elemento que se deseja classificar em uma das populações em estudo.

Depois de estimadas, é importante avaliar a eficácia das funções discriminantes obtidas. Esta eficácia depende do grau de divergência das populações, da quantidade e qualidade das variáveis utilizadas na discriminação (Cruz e Carneiro, 2003). Existem diversos métodos de avaliar a taxa de erro do esquema de discriminação adotado, entretanto, alguns utilizam os próprios dados para obter as estimativas de erro de classificação (Khattree e Naik, 2000).

Um destes métodos, denominado de taxa de erro aparente (TEA), consiste na reclassificação de todos os dados que foram utilizados para gerar as funções discriminantes. Baseado na suposição do conhecimento *a priori* da população a que cada elemento pertence, a classificação de um elemento da população π_j em outra população é indicativo da menor eficiência da função discriminante estimada (Cruz e Carneiro, 2003). A soma de todos os casos de reclassificação incorreta fornece a taxa de erro aparente (TEA), do conjunto de funções estimadas. Esta taxa é viesada para baixo, uma vez que o mesmo conjunto de dados utilizado para estimar as funções é empregado para avaliar suas eficiências (Khattree e Naik, 2000; Cruz e Carneiro, 2003).

Uma possibilidade alternativa de validação das funções consiste em dividir o conjunto de elementos de cada população em duas sub-amostras. Neste caso, um conjunto de sub-amostras é utilizado para estimar as funções discriminantes (amostras de análises) e o outro é utilizado para avaliar a eficiência de classificação (amostras de teste) (Khattree e Naik, 2000; Hair et al., 2005). Esta estratégia requer um grande número de observações por população para produzir estimativas satisfatórias, já que apenas uma parte dos dados é utilizada na construção das funções que serão utilizadas para classificar outro conjunto de dados, cujos coeficientes de discriminação possam ser comparáveis (Khattree e Naik, 2000).

O terceiro método é o da validação cruzada, em que a função é construída utilizando todas as observações disponíveis para a população

exceto uma, cuja alocação é posteriormente verificada através da função discriminante obtida. Este processo de construção de função e de classificação da observação excluída é repetido para todas as observações da população e para todas as populações. Posteriormente, é computado o número de más classificações individuais, por população, ou global, expressos em valores percentuais ou proporcionais. De acordo com Khattree e Naik (2000), este método conduz a estimativas com considerável redução no viés e as diversas estimativas de funções obtidas para uma única população, através da retirada de uma única observação do conjunto de dados praticamente não variam, a menos que haja algumas observações muito discrepantes.

2.2.3 Análise de agrupamento

“Análise de agrupamento” refere-se a um conjunto de técnicas multivariadas cujo principal propósito é agrupar indivíduos ou objetos baseados nas características que eles possuem. Assim, indivíduos com descrições similares são matematicamente alocados no mesmo grupo. Os grupos resultantes devem então exibir elevada homogeneidade interna e alta heterogeneidade externa (entre grupos). Por isso, se a classificação for bem sucedida, espera-se que os indivíduos dentro de um grupo formem aglomerados quando plotados geometricamente e, simultaneamente, se distanciem dos aglomerados formados pelos indivíduos de outros grupos.

De acordo com Mohammadi e Prasanna (2003), existem dois tipos básicos de métodos de agrupamento:

- a) **métodos baseados em distâncias**, nos quais uma matriz de distâncias entre pares de indivíduos é utilizada como entrada de dados, na análise realizada por meio de um algoritmo de agrupamento específico. Este tipo de método conduz a uma representação gráfica (tais como uma árvore ou dendrograma) que permite a identificação visual dos grupos;
- b) **métodos baseados em modelos**, nos quais se assume que as observações de cada grupo constituem amostras aleatórias de algum modelo paramétrico particular. Inferências a respeito dos parâmetros correspondentes de cada grupo, assim como dos membros dos grupos

são estimados conjuntamente, utilizando métodos estatísticos padrões tais como máxima verossimilhança ou métodos Bayesianos.

Os métodos baseados em distâncias são os mais freqüentemente utilizados e podem ser divididos em duas categorias: **hierárquicos** e **não hierárquicos**. A aplicação dos métodos hierárquicos envolve uma série de sucessivas “fusões” ou sucessivas divisões de um grupo de indivíduos. O primeiro caso, conhecido como método **hierárquico aglomerativo**, inicia-se com um único indivíduo. Assim há, inicialmente, tantos grupos quantos forem o número de indivíduos. Então os indivíduos mais similares são agrupados e a este grupo inicial são incorporados os demais indivíduos de acordo com suas similaridades (Mohammadi e Prasanna, 2003). Entre os vários métodos hierárquicos aglomerativos destacam-se, na análise de diversidade de espécies cultivadas, o UPGMA (Unweighted Paired Group Method using Arithmetic Averages) de Sneath e Sokal (1973) e o método da variância mínima de Ward (Ward, 1963).

Os procedimentos de agrupamento não hierárquicos não envolvem a construção de árvores ou dendrogramas. Tais procedimentos são também referidos como “agrupamento de k-médias”. A alocação dos indivíduos nos grupos é baseada em processos denominados de “limiar seqüencial”, “limiar paralelo”, ou “otimização”, uma vez que o número de grupos a serem formados é especificado. Este tipo de agrupamento raramente é utilizado em análises de diversidade intra-específica, já que dificilmente se dispõe de informação *a priori* sobre o número ótimo de grupos necessário para uma alocação acurada dos indivíduos (Mohammadi e Prasanna, 2003).

2.2.4 Aplicações de técnicas multivariadas na análise de diversidade baseada em caracteres fenotípicos

Métodos multivariados podem ser utilizados de diversas maneiras na análise de diversidade, a partir de dados morfométricos, por exemplo:

- a) Em sistemática e taxonomia, como técnica auxiliar na delimitação de táxons, caracterização de zonas híbridas e elucidação de complexos de espécies;

- b) Na caracterização de divergência genética entre populações naturais intra-específicas;
- c) Na caracterização da divergência entre genótipos ou acessos de plantas cultivadas, sejam eles resultantes de material melhorado ou de uso tradicional, como meio de identificar grupos divergentes para fins de conservação e/ou hibridação.

Nielsen et al. (2003) trabalharam com populações “típicas” e “atípicas” de duas espécies de Asteraceae arbustivas, endêmicas das ilhas Galápagos, *Scalesia divisa* e *S. incisa*. Os autores avaliaram dados morfométricos vegetativos e reprodutivos das populações, assim como dados moleculares. Análises multivariadas, baseadas em componentes principais e função discriminante de Anderson, mostraram que as populações desviantes são resultantes de uma zona de hibridação formada entre as populações “típicas” das duas espécies. Estas espécies estiveram geograficamente isoladas no passado, em decorrência de atividade vulcânica, mas agora se encontram em processo de expansão sobre o solo vulcânico recém formado, se hibridando na zona de contato, uma vez que as barreiras do isolamento reprodutivo ainda não haviam sido estabelecidas por completo.

Em um outro trabalho, também envolvendo uma espécie de Asteraceae, *Senecio aquaticus*, Pelsner e Houchin (2004) avaliaram, por componentes principais e análise discriminante, um total inicial de 60 caracteres morfométricos diagnósticos potenciais de dois subgrupos da espécie: *Aquaticus* e *Barbareifolius*. A aplicação de análise discriminante, através de processo *stepwise*, possibilitou a identificação de sete caracteres diagnósticos mais confiáveis para os dois subgrupos. Os autores atribuíram o *status* de variedades para ambos os subgrupos uma vez que, embora distintos, há sobreposição dos agrupamentos de *Aquaticus* e *Barbareifolius* no espaço morfométrico multivariado e não há uma distinção clara referente às suas distribuições geográficas.

Em um estudo envolvendo o gênero *Vicia* (Fabaceae), van de Woune et al. (2003) avaliaram 53 caracteres morfológicos da série *Vicia* (uma subdivisão do gênero), empregando análises de coordenadas principais e agrupamento por UPGMA. Baseados nos resultados obtidos, os autores justificaram a delimitação de quatro espécies distintas dentro da série e de seis subespécies

dentro do agregado *Vicia sativa*, que constitui a espécie morfológicamente mais diversa e de maior distribuição geográfica do gênero.

A análise de características morfométricas através de técnicas multivariadas possibilitou a caracterização de seis espécies de crustáceos do gênero *Artemia*. A utilização da análise discriminante, baseada em variáveis canônicas, mostrou que 92,8% da variância, contida nas 12 variáveis morfométricas originalmente computadas, poderiam ser condensadas nas três primeiras funções canônicas. A primeira destas funções foi relacionada, principalmente, com as variáveis morfométricas associadas a características da furca, a segunda relacionou-se com variáveis associadas ao comprimento do animal e a terceira com variáveis associadas à largura do animal (Baxevanis et al., 2005).

Diversos trabalhos de avaliação de acessos de plantas cultivadas através de técnicas multivariadas estão disponíveis na literatura. Por exemplo, Ferreira et al. (1995) aplicaram análise discriminante em genótipos de arroz para classificá-los quanto à tolerância à toxidez de alumínio. Vidigal et al. (1997) avaliaram a divergência entre cultivares de mandioca, adaptadas ao cultivo no norte do Paraná, através de variáveis canônicas e análise de agrupamento UPGMA, identificando grupos heteróticos para possíveis hibridações futuras. Cruz et al. (2004) também empregaram técnicas multivariadas a dados morfométricos de trigo com o intuito de identificar variabilidade e possíveis grupos heteróticos para resistência ao acamamento. Fonseca et al. (2004) avaliaram a adequação da composição de genótipos de três cultivares clonais de café robusta (*Coffea canephora*). Os autores empregaram a análise discriminante para classificar os clones em uma das três categorias: a) maturação precoce, b) maturação intermediária e c) maturação tardia.

O principal trabalho de análise de dados morfométricos, através de técnicas multivariadas, envolvendo acessos de cajueiro foi realizado por Barros (1991). O autor estudou 67 acessos de *Anacardium occidentale* do BAG do cajueiro da Embrapa Agroindústria Tropical, sendo 10 acessos do tipo anão precoce, 27 do tipo comum, coletados em diferentes municípios do Ceará, 9 introduzidos da Índia, 11 da Venezuela e 10 do Estado de São Paulo. Foram avaliados 30 caracteres morfológicos, entre características quantitativas e

qualitativas ou multicategóricas. As estimativas de divergência foram obtidas por meio da técnica de componentes principais e agrupamento pelo método de otimização de Tocher, baseado em distâncias Euclidianas médias padronizadas. De acordo com o autor, os dois primeiros componentes principais explicaram apenas 44,6% da variância contida nos dados originais, enquanto a inclusão do terceiro componente ampliou este valor para 57,09%. Dessa forma, a projeção gráfica dos acessos nos espaços bi ou tridimensionais apresentam elevadas taxas de distorção, prejudicando a confiabilidade nos resultados e nas interpretações subseqüentes. Por outro lado, a análise agrupamento pelo método de Tocher alocou os 67 acessos em 12 grupos diferentes. Sete dos 10 acessos de cajueiro do tipo anão foram alocados juntos, constituindo um grupo distinto em relação os demais acessos. Enquanto a maior parte dos acessos coletados no Ceará foi distribuída em 10 grupos diferentes, todos os acessos introduzidos da Índia, Venezuela e Valinhos foram alocados em um único grupo. Barros (1991) concluiu que estes resultados reforçam a hipótese de que o Nordeste brasileiro seja em um importante centro de diversidade de *A. occidentale* e que o cajueiro anão precoce constitui uma forma ou ecótipo da espécie *A. occidentale*.

Samal et al. (2003) avaliaram 20 variedades de cajueiro através de oito variáveis morfológicas relacionadas com o padrão de ramificação, florescimento, razão sexual entre flores na panícula e componentes de produção de castanha, por um período de três anos consecutivos. Análises de divergência, baseada na distância de Gower, e de agrupamento, por UPGMA, evidenciaram a formação de quatro grupos distintos com nível de similaridade igual ou superior a 70%. Entretanto, o maior grupo foi formado por 12 genótipos, enquanto dois grupos foram formados, cada um, por um genótipo apenas.

2.3 Marcadores Moleculares na Análise Genética Vegetal

A utilização de marcadores moleculares de diversos tipos tornou possível a identificação e caracterização de germoplasma, a construção de mapas genéticos e a estimativa da distância genética entre indivíduos e/ou populações de várias espécies vegetais. Tais tecnologias podem contribuir

para uma aceleração de ganhos genéticos em plantas de interesse, especialmente em culturas cujas informações ainda limitadas, como é o caso da maioria das espécies tropicais arbóreas (Federezzi, 1998; Ferreira & Grattapaglia, 1998; Picoli et al., 2001). De modo semelhante, as informações de diversidade e distância genética têm sido avaliadas para a identificação das melhores combinações híbridas e na organização de bancos de germoplasma; indicando redundâncias e deficiências das coleções, além de gerar dados sobre a eficiência do processo de coleta, manutenção e ampliação de um banco de germoplasma (Barbosa-Neto & Bered, 1998; Ferreira & Grattapaglia, 1998).

O desenvolvimento de marcadores bioquímicos e moleculares proporcionou um salto qualitativo e quantitativo em estudos da estrutura populacional e do sistema reprodutivo de diversas espécies. Vários parâmetros populacionais, como grau de endogamia e sistema reprodutivo predominante, entre outros, podem ser obtidos e são de grande importância na determinação de estratégias de conservação da variabilidade genética na natureza, bem como para sua melhor utilização em programas de melhoramento genético das espécies domesticadas (Carlini-Garcia et al., 2001; Reif et al., 2005).

Barbosa-Neto e Bered (1998) argumentam que não há sérias restrições quanto aos marcadores moleculares preferenciais para utilização em estudos de diversidade genética. De modo geral, os marcadores devem ser confiáveis em termos de repetibilidade, ter baixo custo e serem fáceis e rápidos de analisar. Neste sentido, os marcadores baseados em técnicas de PCR (reação em cadeia da polimerase) são mais atrativos, principalmente aqueles que apresentam grande polimorfismo como microssatélites e AFLP.

Em um trabalho de compilação de dados, de 307 estudos de avaliação de diversidade inter e intrapopulacional de espécies de plantas selvagens, baseados em marcadores de DNA nuclear, Nybom (2004) observou que as estimativas derivadas de marcadores de herança dominante (RAPD, AFLP e ISSR) são muito similares e podem ser comparáveis diretamente. Por outro lado, marcadores de herança co-dominante (microssatélites) produziram valores de diversidade intrapopulacional quase três vezes superiores, enquanto que as estimativas de diversidade interpopulacionais foram similares às aquelas obtidas pelos marcadores de herança dominante. A distância geográfica

máxima entre populações amostradas, produziu acentuado efeito positivo na diversidade interpopulacional e, para espécies perenes, alógamas e/ou de sucessão tardia, a maior parte da variabilidade é de natureza intrapopulacional. A autora também observou que o número de plantas amostradas por população e o número de alelos de microssatélite detectados foram correlacionados com alguns parâmetros genéticos populacionais.

2.3.1 Marcadores ISSR

Os marcadores ISSR (Inter Simple Sequence Repeat) são produtos de uma técnica baseada no método de amplificação de DNA pela reação em cadeia da polimerase (PCR) (Reddy et al., 2002). Esta metodologia foi originalmente descrita, com pequenas variações e diferentes denominações, pelos trabalhos de Zietkiewicz et al. (1994), Gupta, et al. (1994) e Wu et al. (1994). Basicamente, a técnica envolve a amplificação de regiões entre seqüências de microssatélites adjacentes, inversamente orientadas. Os *primers* são baseados nas próprias seqüências repetitivas (com motivos di, tri, tetra ou penta nucleotídeos) proporcionando amplo arranjo de possíveis produtos amplificáveis (Zietkiewicz et al., 1994). Nenhum conhecimento *a priori* do genoma é necessário e *primers* baseados numa seqüência repetitiva, tal como $(CA)_n$, podem ser desenhados considerando apenas o motivo existente na própria seqüência tais como $(CA)_8$ ou $(CA)_9$ (Gupta et al., 1994; Bornet e Branchard, 2001) ou com uma ancoragem, formada por um pequeno segmento de bases arbitrárias e/ou degeneradas, localizado em uma das extremidades (3' ou 5') da seqüência repetitiva, tais como $(CA)_8RG$ e $(CA)_8YT$ ou $BDB(CA)_7$ (onde R=purinas, Y=pirimidinas, B=C, G ou T e D=A, G, ou T) (Zietkiewicz et al., 1994; Godwin et al., 1997, Fang et al., 1997, Reddy et al., 2002).

Considerando a ubiquidade das seqüências microssatélites nos genomas dos eucariotos, o resultado da amplificação por *primers* ISSR é um sistema de marcadores multilocos útil para fingerprinting, análises de filogenia e diversidade, mapeamento e etiquetagem gênica e análises de biologia evolutiva (Godwin, et al., 1997; Reddy et al., 2002). Para que a amplificação de locos ISSR tenha sucesso, os pares de seqüências repetitivas devem ocorrer dentro de uma distância suficientemente curta, que possa ser amplificada pela

reação de PCR e os seus produtos separados em géis padrões de agarose ou poliacrilamida (Zietkiewicz et al., 1994, Gupta, et al., 1994; Godwin, et al., 1997; Reddy et al., 2002).

De acordo com Reddy et al. (2002), as fontes de variabilidade e o nível de polimorfismo detectado através dos marcadores ISSR, são decorrentes dos seguintes fatores: a) elevada taxa de mutação das seqüências microssatélites, alvos do anelamento do *primer* empregado; b) natureza do *primer* utilizado, isto é, seqüência repetitiva ou motivo, existência ou não de região de ancoragem, assim como localização (5' ou 3'), extensão e composição da seqüência de ancoragem; d) método de resolução utilizado, isto é, gel desnaturante de poliacrilamida (PAGE) em combinação com marcação radiativa ou coloração com nitrato de prata ou gel de agarose, corado com brometo de etídeo.

Segundo Blair et al. (1999), se a distribuição das diferentes seqüências microssatélites é aleatória, o número de bandas produzidas por um *primer* ISSR, de um dado motivo, deveria ser o reflexo deste motivo em um dado genoma. Dessa forma, tal situação forneceria uma estimativa da abundância de diferentes motivos, a partir de uma biblioteca de hibridização, e auxiliaria na seleção de motivos-alvos para o desenvolvimento futuro de marcadores de microssatélites.

O número de trabalhos utilizando marcadores ISSR tem sido crescente desde as primeiras publicações sobre a técnica. Esta classe de marcadores combina a maioria das vantagens dos marcadores AFLP, microssatélite, e RAPD, como elevado grau de reprodutibilidade dos dois primeiros e baixo custo, simplicidade e universalidade, do último (Godwin, et al., 1997; Fang et al., 1997, Bornet e Branchard, 2001; Reddy et al., 2002). A alta reprodutibilidade é resultante do comprimento dos *primers* empregados (16 a 25 pb) e de condições de anelamento mais estridentes (com temperaturas entre 45°C e 60°C), enquanto o baixo custo e a simplicidade são em decorrência dos *primers* serem universais e livres de patenteamento, assim como da possibilidade de utilização sistemas de resolução que dispensam o uso de radiatividade, como o gel agarose corado com brometo de etídeo. Reddy et al. (2002) discutem as aplicações, relacionando diversos exemplos, desta classe de marcadores no melhoramento de plantas cultivadas. Por outro lado, uma parcela significativa dos trabalhos encontrados na literatura atual

envolve a aplicação dos marcadores ISSR na análise genética de espécies silvestres, sejam elas aparentadas ou não de espécies de plantas cultivadas. Estes trabalhos envolvem estudos de diversidade genética intra e interespecíficas de acessos ou genótipos, assim como análise da estrutura genética entre e dentro de populações naturais.

Huang e Sun (2000) avaliaram a diversidade e as inter-relações de acessos de batata doce (*Ipomoea batatas*) e de outras nove espécies silvestres relacionadas, utilizando marcadores ISSR e DNA do cloroplasto. De acordo com os autores, as análises filogenéticas, conduzidas com os dois tipos de marcadores, produziram resultados semelhantes, entretanto o elevado nível de polimorfismo exibido pelos marcadores ISSR resultou em melhor separação dos acessos ao nível infra-específico. Análise de diversidade e relações filogenéticas, utilizando marcadores ISSR, foi realizada no gênero *Oryza* por Joshi et al. (2000). O trabalho envolveu 17 espécies de arroz selvagem e duas cultivadas (*O. sativa* e *O. glaberrima*), além de espécies de outros três gêneros relacionados, perfazendo um total de 42 genótipos avaliados. O emprego de 11 *primers* polimórficos e informativos possibilitou o agrupamento dos genótipos de acordo com seus respectivos genomas, além de revelar a existência de 87 marcas espécies e/ou genoma específicos para oito dos nove genomas existentes em *Oryza* e sugerir um padrão evolutivo polifilético para o gênero.

Estudos com marcadores ISSR envolvendo a caracterização de acessos pertencentes a uma única espécie foram realizados por Fang et al. (1997), que avaliaram 48 acessos de *Poncirus trifoliata* da coleção de Citrus da Universidade da Califórnia; Blair et al. (1999), que analisaram 58 acessos representativos da diversidade de arroz cultivado (*Oryza sativa*); Eiadthong et al. (1999), que investigaram a divergência entre 22 cultivares de manga (*Mangifera indica*), plantados na Tailândia; Assefa et al. (2003), que estudaram a divergência genética de 92 genótipos do cereal *Eragrostis tef*, originários de oito regiões diferentes da Etiópia. Em todos estes trabalhos, o nível de polimorfismo acessado foi considerado satisfatório pelos autores, possibilitando a distinção entre genótipos, individualmente e por grupo de afinidade, na quase totalidade dos casos. Os dados observados também foram compatíveis com os resultados obtidos de outros estudos, quando os mesmos materiais foram

investigados através de outros tipos de marcadores moleculares, como isoenzimas, RAPD, RFLP e AFPL.

Os marcadores ISSR também têm sido utilizados com sucesso em estudos de diversidade e estrutura genética de populações naturais de espécies silvestres, tais como mangue (*Aegiceras corniculatum*) da costa da China (Ge e Sun, 1999); amora (*Morus serrata*) na Índia (Vijayan et al., 2004); café (*Coffea arabica*) na Etiópia, (Aga et al., 2005); leguminosas do gênero *Ammopiptanthus* do deserto do Noroeste da China (Ge et al., 2005) e de três espécies de montana (*Symplocos lauriana*, *Gaultheria fragrantissima* e *Eurya nitida*), endêmicas da região de Western Ghats da Índia (Deshpande et al., 2001).

2.3.2 Marcadores moleculares em cajueiro

O primeiro trabalho envolvendo a utilização de marcadores moleculares de DNA em cajueiro foi descrito por Silva-Neto et al. (1995). Neste estudo os autores avaliaram a viabilidade da utilização de marcadores RAPD na identificação de clones de cajueiro anão precoce e concluíram que a técnica era apropriada para ser empregada em análises de germoplasma de *A. occidentale*. Por outro lado, ainda hoje a utilização de marcadores moleculares na caracterização de recursos genéticos de cajueiro é bastante restrita, se limitando a um pequeno número de trabalhos conduzidos com coleções de bancos de germoplasma da Tanzânia e da Índia. O primeiro destes estudos foi realizado por Mneney et al. (2001), que avaliaram a divergência genética, baseada em marcadores RAPD, de acessos da Tanzânia, Sri Lanca, Moçambique, Índia, Brasil e Cook Islands. Os autores verificaram que os acessos brasileiros formaram um grupo separado e mais distinto em relação aos das demais procedências. Posteriormente, Dhanaraj et al (2002) estimaram a diversidade genética entre 90 acessos do Banco Nacional de Genética do Cajueiro da Índia, também utilizando marcadores RAPD. De acordo com os autores, este estudo permitiu confirmar que diversidade do banco indiano de germoplasma de caju pode ser considerada de moderada a alta. Os resultados também permitiram estabelecer uma coleção nuclear (core

colletion) menor, com representação da mesma diversidade da população inteira.

Archak et al. (2003a) compararam a eficiência e utilidade de três tipos de marcadores moleculares (RAPD, ISSR e AFLP) e de descritores morfológicos na medida de variação e discriminação genotípica de germoplasma de caju. Os autores verificaram que, a partir de cada um dos três tipos de marcadores utilizados, foi possível discriminar os 19 acessos avaliados, com maior eficiência para os marcadores AFLPs. Por outro lado, não foi observada a ocorrência de correlação entre o conjunto de 27 descritores morfométricos utilizados e os marcadores moleculares. Apenas os marcadores RAPD e ISSR apresentaram correlação significativa ($r=0,63$ $p<0,01$). Archak et al. (2003b) também utilizaram marcadores RAPD e ISSR na caracterização fingerprinting de 35 variedades elites, sendo 24 seleções e 11 híbridos, desenvolvidas por diferentes centros de pesquisa da Índia. A partir de 94 locos amplificados foi possível diferenciar, com alto grau de certeza, todos os genótipos. Por lado, os resultados evidenciaram que a base genética entre eles é muito estreita, apesar da grande amplitude geográfica entre os locais de obtenção dos cultivares, fato atribuído ao intenso intercâmbio de material genético entre os centros de pesquisa e ao número reduzido de parentais empregados nos programas de produção de híbridos.

Trabalhos utilizando marcadores moleculares específicos para o cajueiro, como microssatélites, ainda não estão disponíveis na literatura até a presente data. Entretanto, Croxford et al. (2006) relataram o desenvolvimento de 21 pares de *primers* microssatélites, a partir de uma biblioteca genômica não enriquecida de *A. occidentale*. A viabilidade destes *primers* foi testada em uma população de 49 indivíduos de cajueiro e a maioria amplificou dois ou três alelos distintos, com um deles produzindo um máximo de cinco alelos. Syed et al. (2005) descreveram o desenvolvimento de marcadores baseados no sistema de polimorfismo de amplificação de seqüência específica (SSAP) da região LTR (Long Terminal Repeat) do retrotransposon Ty1- *copia*, a partir do sequenciamento de duas LTRs isoladas do genoma de *A. occidentale*. De acordo com os autores, esta classe de marcadores apresentou nível de polimorfismo superior ao obtido por marcadores AFLP em cajueiro, sendo de grande potencial na construção mapas saturados e seleção assistida.

2.4 Análises da Diversidade de Recursos Genéticos Aplicadas a Variáveis Binárias

2.4.1 Análise Discriminante

Diversos procedimentos de análises discriminantes, aplicados a dados binários, já foram sugeridos por diferentes autores (Moore, 1973; Hand, 1983; Beharav e Nevo, 2003). Nos últimos anos, as estimativas de variação genética têm sido, de forma crescente, baseadas na informação extraída ao nível de seqüências de DNA, através de diferentes classes de marcadores moleculares. Este tipo de informação, frequentemente, é convertida em um sistema binário e utilizada em análises discriminantes, com o propósito de identificar os locos que sejam os melhores fatores de diferenciação, isto é, aqueles locos que apresentem elevado nível de diferenciação nas frequências alélicas para diferentes populações e, por isso, são úteis na classificação correta dos indivíduos nas suas respectivas populações (Beharav e Nevo, 2003).

Os métodos de análises discriminantes normalmente empregados na classificação de genótipos, a partir de dados moleculares binários, são ditos não paramétricos, uma vez que não é assumida nenhuma forma paramétrica de distribuição para a função de densidade das variáveis analisadas (Khattree e Naik, 2000). Uma das técnicas frequentemente empregadas é o método dos k vizinhos mais próximos. Este procedimento é baseado em certo critério envolvendo distâncias entre indivíduos imediatos. Ou seja, dado um grupo de observações a serem classificadas, um algoritmo procura, para uma particular observação, aquelas que lhe sejam mais próximas. Assim, cada observação é alocada na classe que contenha a maior proporção de k vizinhos mais próximos (Khattree e Naik, 2000; Beharav e Nevo, 2003).

Supondo que se deseje alocar um indivíduo, cujo valor observado seja igual a x , em uma de duas ou mais populações, e que k seja um número inteiro positivo definido pré-especificado. A partir de todas as distâncias genéticas entre pares de indivíduos, define-se os k indivíduos mais próximos daquele que se deseja classificar. Supondo que dentre esses k indivíduos, k_t são provenientes da população t , cuja probabilidade *a priori* é π_t . Então, a

probabilidade a *posteriori* de que este indivíduo pertença a t -ésima população é estimada por:

$$P = \frac{\pi_t(k_t / n_t)}{\sum_{s=1}^p \pi_s(k_s / n_s)}$$

Onde

s : é número de populações avaliadas ($s = 1, \dots, p$);

n : é o número de indivíduos de cada população.

A observação x será classificada na t -ésima população se a probabilidade posterior para esta população for a maior entre todas. Ou seja, x será classificado na t -ésima população se $\pi_t k_t / n_t$ for maior que $\pi_s k_s / n_s$ para todo s (Khattree e Naik, 2000). Geralmente, no caso de ocorrência de empate entre probabilidades, x é alocado uma população designada de “desconhecida”.

2.4.2 Análise de Variância Molecular

Técnicas multilocos de *fingerprinting* de DNA, baseadas na reação em cadeia da polimerase (PCR), tais como RAPD (*random amplified polymorphic DNA*), AFLP (*amplified fragment length polymorphism*) e ISSR (*inter-simple sequence repeats*), apresentam enorme valor nos estudos de variação genética de populações naturais, uma vez que elas proporcionam a geração de grande quantidade de dados de maneira rápida e relativamente fácil (Krauss, 2000), além de prescindir de qualquer conhecimento prévio do genoma da espécie ou populações analisadas. Por outro lado, apesar das vantagens citadas a análise desse tipo de marcador apresenta alguns problemas de ordem prática, sendo o principal deles o fato de não ser possível distinguir, diretamente, os genótipos heterozigotos dos homozigotos dominantes, ou seja, se um haplótipo ou “alelo” de um sítio de RAPD é amplificado, não é possível distinguir entre o loco homozigoto marcador/marcador (situação em que são amplificados os fragmentos dos dois haplótipos do indivíduo) e o heterozigoto marcador/ loco

nulo (situação em que ocorre a amplificação de apenas um dos haplótipos do indivíduo) (Lynch e Milligan, 1994).

Uma estratégia alternativa é a análise de progênies ou de tecido haplóide (como o macrogametófito de coníferas), de cada indivíduo que apresenta o alelo dominante. Entretanto tal estratégia amplia significativamente o volume de trabalho laboratorial e, mais importante ainda, em muitos casos não se dispõe de tais progênies e/ou tecidos haplóides (Zhivotovsky, 1999). Por isso, segundo Krauss (2000), três soluções, de natureza estatística, têm sido propostas para contornar este problema:

- a) A frequência do alelo nulo, para um dado loco, pode ser estimada através da raiz quadrada da frequência do genótipo homozigoto nulo (banda ausente no gel), assumindo equilíbrio de Hardy-Weinberg na população e/ou loco;
- b) Aplicação da metodologia proposta por Lynch e Milligan (1994), cujo propósito principal é reduzir o viés das estimativas obtidas, pelo método da raiz quadrada. Para isto, a análise ignora os locos cujas frequências de indivíduos homozigotos nulos não excedam a três e utiliza estimadores de parâmetros populacionais obtidos pela expansão, de segunda ordem, da série de Taylor;
- c) Utilização de uma abordagem Bayesiana, proposta por Zhivotovsky (1999), na estimação da frequência dos alelos nulos.

As três metodologias relacionadas assumem, de algum modo, que o coeficiente de endogamia intrapopulacional é conhecido (Holsinger et al., 2002). Uma outra metodologia, amplamente utilizada atualmente, é AMOVA (*analysis of molecular variance*) desenvolvida por Excoffier et al. (1992). Neste caso, os dados fenotípicos multilocos, de cada indivíduo, são tratados como haplótipos e utilizados na construção de matrizes que descrevem a distância entre haplótipos na análise de variância molecular (Holsinger et al., 2002).

A análise de variância molecular (AMOVA) foi concebida como uma extensão dos trabalhos de Cockerham (1973) e de Long (1986), para a correlação alélica entre demes, para dados de diversidade entre haplótipos. Dessa forma, partindo do pressuposto que uma soma de quadrados convencional pode ser escrita como uma soma de quadrados das diferenças entre todos os pares de observações, Excoffier et al. (1992) elaboraram uma

análise hierárquica da variância molecular diretamente da matriz de quadrados de distâncias entre todos os pares de haplótipos, produzindo estimativas de componentes de variância análogas às estatísticas F de Wright, denominadas de “estatísticas Φ ”. O método é flexível para acomodar diversos tipos de matrizes, correspondentes a diferentes tipos de dados moleculares e/ou pressuposições evolutivas, sem necessidade de mudança na estrutura básica da análise.

O termo haplótipo, empregado neste contexto, refere-se à combinação de marcadores moleculares presentes em alguma área definida do genoma. Assim, o haplótipo de um indivíduo é representado por um vetor p , com valores 1, se a banda homóloga estiver presente e 0, caso contrário, enquanto um marcador, pode ser um fragmento de restrição/amplificação, um nucleotídeo ou um evento mutacional (Excoffier et al., 1992).

Considerando o caso em que os indivíduos amostrados são oriundos de diferentes populações, arranjadas em grupos, definidos *a priori* por um critério não genético, os haplótipos podem ser representados pelo modelo linear:

$$x_{ijk} = x + a_k + b_{jk} + c_{ijk}$$

onde x_{ijk} é a freqüência do vetor do i -ésimo haplótipo, da j -ésima população no k -ésimo grupo, x é o valor esperado desconhecido de x_{ijk} , \mathbf{a} é efeito de grupos, \mathbf{b} é efeito de populações e \mathbf{c} de indivíduos ou haplótipos. Tais efeitos são assumidos como sendo aditivos, aleatórios, não correlacionados, associados aos componentes de variância σ_a^2 , σ_b^2 e σ_c^2 , respectivamente, que são utilizados na obtenção das estatísticas Φ

$$\Phi_{ST} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma^2}, \quad \Phi_{CT} = \frac{\sigma_a^2}{\sigma^2}, \quad \Phi_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}$$

onde $\sigma^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2$; Φ_{ST} é correlação entre pares de haplótipos aleatórios dentro de populações, em relação à correlação de pares de haplótipos tomados ao acaso da espécie como um todo; Φ_{CT} é a correlação entre haplótipos dentro de cada grupo de populações, em relação à de pares de haplótipos tomados ao acaso da espécie como um todo e Φ_{SC} é correlação da diversidade molecular entre haplótipos aleatórios dentro das populações,

relativa à correlação entre pares de haplótipos tomados ao acaso dentro de uma região (Excoffier et al., 1992).

Originalmente, a metodologia da AMOVA foi desenvolvida para aplicação em sistemas haplóides (mtDNA) constituídos de um único grupo de ligação. Posteriormente, Michalakis e Excoffier (1996), descreveram estimadores para dados multilocos de microssatélites em sistemas diplóides. Os autores sugeriram manipular os genótipos dos indivíduos como dois haplótipos distintos, por isso, seria necessário conhecer a fase gamética e/ou que os marcadores fossem codominantes. Entretanto, Huff et al. (1993) aplicaram a metodologia original de Excoffier et al. (1992), na análise de diversidade de populações de uma espécie alógama obrigatória de gramínea (*Buchoë dactyloides*), genotipadas com marcadores RAPD, decompondo a variação total observada em componentes devido às diferenças entre indivíduos dentro de populações, entre populações dentro de regiões adaptativas e entre regiões. A partir daí este método tornou-se um dos mais utilizados, para partição da variabilidade genética, em trabalhos envolvendo populações naturais avaliadas com marcadores moleculares de herança dominante (Nybom, 2004). Posteriormente, modelos com decomposição da variação total, específicos para diversos tipos de dados, incluindo dados genotípicos com alelos recessivos e fase gamética desconhecida, foram incorporados ao aplicativo computacional empregado na execução da AMOVA (Schneider et al., 2000).

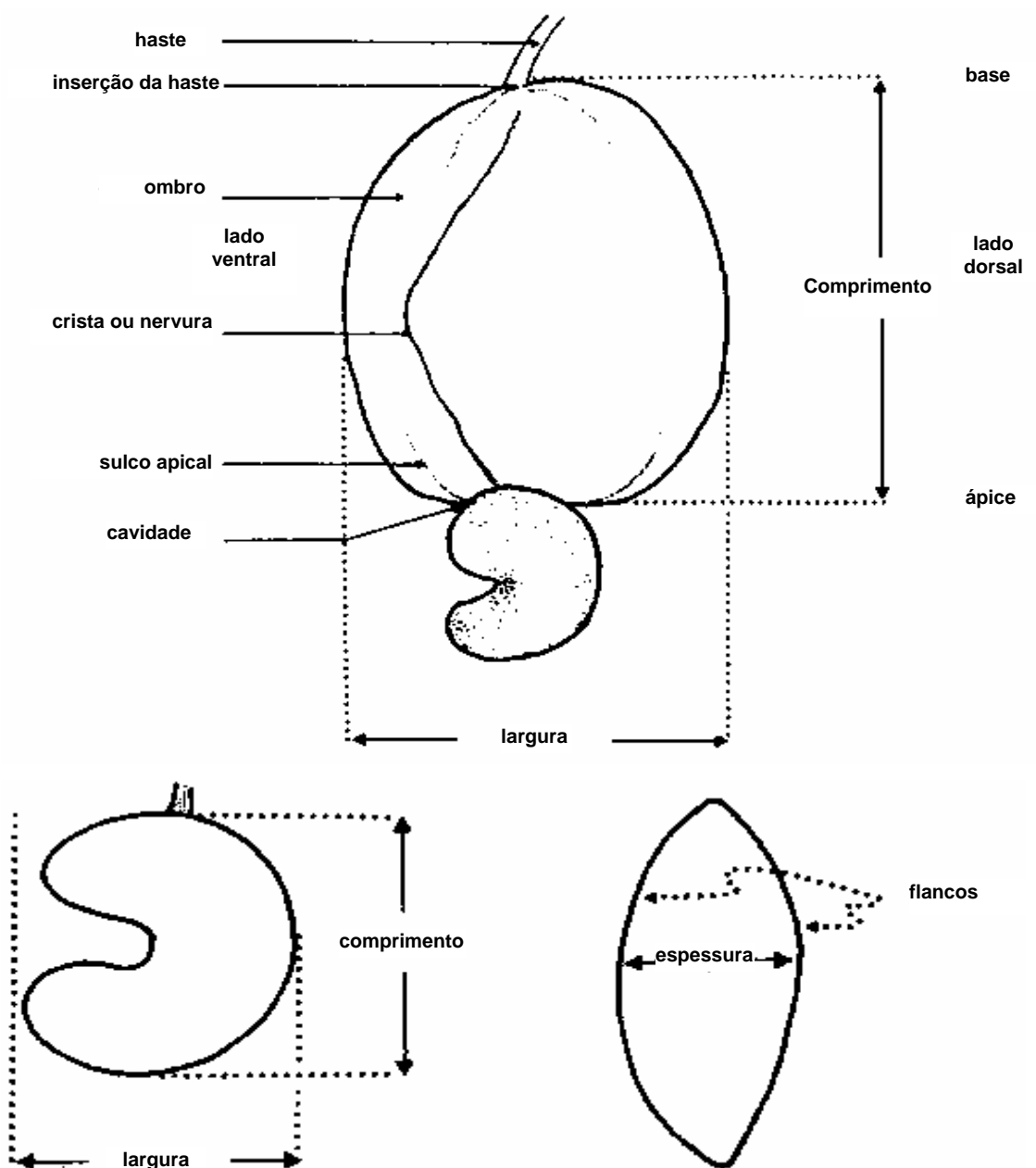
Diferentes métodos para obtenção da matriz de distância têm sido utilizados, destacando-se o quadrado da distância euclidiana, distância de Nei e Li e distância de Jaccard, mas todos eles produzem resultados praticamente idênticos na análise de variância molecular, quando aplicados ao mesmo conjunto de dados (Nybom, 2004). A significância dos componentes de variância e das estatísticas Φ são testadas através de análise de permutação (Excoffier et al., 1992). Neste caso, uma distribuição nula é computada por reamostragem dos dados, de modo que, a cada permutação, cada indivíduo é alocado aleatoriamente dentro da população, mantendo-a de tamanho constante. A repetição do processo diversas vezes possibilita a construção de uma distribuição nula, que será utilizada na estimação do grau de significância da estatística obtida.

O método da AMOVA apresenta uma relação direta com as estatísticas F , estando sujeito às mesmas limitações das pressuposições assumidas pelo modelo proposto por Cockerham (1973). É assumido que os acasalamentos ocorram completamente ao acaso e que não há endogamia. Caso contrário, o nível de heterozigose será inferior e a variabilidade poderá não ser convenientemente estimada. Em razão da deriva genética, fica difícil assumir que qualquer haplótipo seja representativo da variação completa do genoma. Também é importante que os dados sejam derivados de um número adequado de marcadores moleculares. Os efeitos de seleção também não são completamente considerados nesta metodologia de análise. Se as subpopulações estão sob diferentes pressões seletivas, a seleção poderá ter efeitos muito diferentes sobre os alelos e sobre as combinações genotípicas.

Por outro lado, conforme destaca Excoffier et al.(1992), o propósito principal da AMOVA é delinear a extensão da diferenciação genética entre e dentro de populações, através de componentes de variância. Por isso, os resultados obtidos deverão ser interpretados com a devida cautela.

CAPÍTULO 1

ANÁLISE DA DIVERSIDADE EM *Anacardium* spp. BASEADA EM DADOS MORFOMÉTRICOS



Desenhos esquemáticos extraídos de Cashew Descriptors (IBPGR, 1986)

1 INTRODUÇÃO

Anacardium é um pequeno gênero da família Anacardiaceae constituído de árvores, arbustos e subarbustos nativos dos neotrópicos. Segundo Ferreira (1973), esta família é composta por árvores ou arbustos de folhas alternas, coriáceas e sem estípulas, sempre providas de canais resiníferos em seus ramos. Vários gêneros são produtores de frutos comestíveis, como *Mangifera* (mangas), *Spondias* (cajás e umbu) e *Anacardium* (cajus), além de outros possuidores de madeira de ótima qualidade como *Astronium* (Gonçalo Alves, aroeiras) e *Schinopsis* (braúna e quebracho).

A única espécie cultivada é *A. occidentale*, devido, principalmente, à importância econômica que o comércio de sua amêndoa (castanha) apresenta no mercado internacional. No Brasil, a baixa produtividade das lavouras é o problema mais significativo desta cultura, com produtividade média de 220 kg de castanhas *in natura* por hectare, razão pela qual o programa de pesquisa de melhoramento genético vem dando prioridade à obtenção de cultivares mais produtivas (Barros e Crisóstomo, 1995). Diversos fatores contribuem para a baixa produtividade dos atuais plantios, dentre eles a propagação realizada por sementes é o principal, já que o cajueiro é uma espécie predominantemente alógama. Isto resulta na heterogeneidade de diversos caracteres da planta como: altura, formato e expansão da copa, época de florescimento e de produção, tamanho do fruto, peso total, peso da amêndoa, despeliculagem, relação peso da amêndoa/peso do fruto e do pseudofruto, coloração, formato e composição química do suco (Crisóstomo et al., 1992; Barros et al., 2002).

A propagação clonal, através de enxertia de genótipos de cajueiro anão precoce, trouxe novas perspectivas para o desenvolvimento da cajucultura brasileira, elevando o potencial produtivo, em regime de sequeiro, para níveis superiores a 1300 kg/ha de castanha, com a vantagem das plantas apresentarem redução de porte, em relação ao tipo comum (Barros e Crisóstomo, 1995). Por outro lado, existem poucos clones do tipo anão recomendados para plantio, todos resultantes de um programa de melhoramento de base genética muito estreita, constituindo situação de vulnerabilidade à cultura (Cavalcanti et al., 2000b; Barros et al., 2000). Esta realidade pode ser modificada com uma maior disponibilização de diversidade

genética, caracterizada e quantificada, para uso imediato nos programas de melhoramento para obtenção de cultivares mais produtivos e adaptados a diferentes ecossistemas tropicais, como os cerrados e o semi-árido brasileiros (Barros et al. 2000).

No que se refere à origem geográfica, o Brasil é considerado o local de origem do cajueiro, uma vez que a espécie ocorre em praticamente todo o território nacional e já era amplamente utilizada, pelas populações indígenas, à época da chegada dos europeus (Lima, 1986; apud Barros et al. 2002). Além de *A. occidentale*, outras vinte espécies de *Anacardium* foram descritas pela taxonomia clássica, número este reduzido a dez espécies por Mitchell e Mori (1987). Estas espécies são tipicamente tropicais, sendo encontradas na Amazônia (florestas úmidas, matas de galeria e cerrado), planalto central e Nordeste (cerrado) do Brasil, embora *A. humile* seja encontrada até o sul do trópico de Capricórnio, na América do Sul (Mitchell e Mori, 1987).

De acordo com Mitchell e Mori (1987), *A. occidentale* é uma espécie polimórfica, constituída por dois ecótipos distintos: um ecótipo típico de zonas de restinga, de qual se originaram as formas cultivadas e um ecótipo típico dos cerrados do Brasil Central, não cultivado e associado à vegetação aberta de fisionomia savanóide desta região. Por outro lado, os autores acreditam que a espécie tenha evoluído originalmente nos cerrados do Brasil Central e, posteriormente, colonizou as zonas de restingas, mais recentes, da costa brasileira.

Embora seja considerado o principal centro de diversidade do gênero *Anacardium*, poucos trabalhos de avaliação e caracterização do germoplasma foram realizados no Brasil, destacando-se apenas estudo de Barros (1991), que fez a caracterização morfológica e isoenzimática, por meio de técnicas multivariadas, de parte do Banco Ativo Germoplasma do Cajueiro, do Centro Nacional de Pesquisa de Caju (atual Embrapa Agroindústria tropical). Trabalhos desta natureza fornecem subsídios fundamentais aos programas de melhoramento da espécie, assim como ao processo de otimização nas atividades de coleta e de conservação dos recursos genéticos, tanto *in situ* quanto *ex situ*.

Este trabalho teve por finalidade avaliar a diversidade genética de parte dos acessos de *Anacardium* spp., mantidos pelo Banco Ativo de Germoplasma

do Cajueiro (BAG – Embrapa CNPAT – CE) e de indivíduos amostrados em uma população subespontânea de cajueiro (*Anacardium occidentale*), localizada no estado do Rio de Janeiro, através de descritores morfométricos. Especificamente, os objetivos foram:

- Caracterizar e quantificar a variabilidade genética existente entre acessos e dentro de populações de cajueiro, através de variáveis morfométricas;
- Estimar as distâncias genéticas entre as populações e avaliar suas relações com a origem geográfica;
- Comparar diferentes estratégias de análises baseadas em técnicas multivariadas aplicadas a caracteres de variação contínua.

2 MATERIAL E MÉTODOS

2.1 Material vegetal

As informações utilizadas para o estudo foram obtidas de 69 acessos do Banco Ativo de Germoplasma do Cajueiro (Quadro 01) e de 23 indivíduos de uma população subespontânea de uma área de restinga do litoral norte do Estado do Rio de Janeiro. Cada acesso foi representado por um indivíduo apenas, perfazendo um total de 92 genótipos analisados, os quais serão referidos daqui em diante simplesmente como acessos, independentemente da origem.

O BAG do Cajueiro pertence a Embrapa Agroindústria Tropical e localiza-se na Microrregião Litoral de Pacajus, Município de Pacajus, Estado do Ceará, no km 5 da Estrada Pacajus-Itaipaba, a partir da BR-116, a aproximadamente 55 km da cidade de Fortaleza e cerca de 40 km, em linha reta, do oceano. Esta localização corresponde às coordenadas 4°10' S e 38°27' W, com altitude de 60 m acima do nível do mar. O clima da Região enquadra-se no tipo seco/sub úmido (c2) da classificação de Thornthwaite. A precipitação média anual é de 1100 mm com duas estações definidas: uma de chuvas que vai, normalmente, de janeiro a junho com cerca de 90% do total da precipitação anual, caracterizada por ser extremamente irregular, e outra seca, de julho a dezembro, na qual ocorrem normalmente o florescimento e a

frutificação do cajueiro (Barros, 1991, Paiva et al., 2003). A área é composta de três unidades pedogenéticas: a) podzólico vermelho-amarelo eutrófico; b) podzólico vermelho-amarelo distrófico e, 3) areias quartzosas distróficas. O relevo é plano e as características físicas de maior importância são associadas à textura. A condição arenosa resulta em fraca agregação do solo, baixa retenção de umidade, lixiviação de fertilizantes aplicados à superfície e drenagem acentuada, forte ou excessiva. O pH está entre 5,0 e 6,0 (moderadamente ácido) e a fertilidade natural é baixa (Barros, 1991).

Quadro 1 – Informações das coleções e dos acessos do Banco Ativo de Germoplasma do Cajueiro empregados na análise de diversidade baseada em dados morfométricos.

Nome da coleção, procedência e data da introdução	Número da planta no BAG	Espécie	Local da Coleta
Coleção C-02, "Cerrado" Goiás, Mato Grosso, Pará e Piauí. Introdução por sementes em 1975	29	<i>A. othonianum</i>	Campos Belos (GO)
	100	<i>A. othonianum</i>	São Domingos (GO)
	138	<i>A. othonianum</i>	Arraias (GO)
	197	<i>Anacardium</i> sp.	Cristalina (PI)
	205	<i>A. occidentale</i>	Itaituba (PA)
	209	<i>Anacardium</i> sp.	Eliseu Martins (PI)
	226	<i>A. othonianum</i>	Barra do Garças (MT)
	241	<i>A. occidentale</i>	Itaituba (PA)
	245	<i>A. occidentale</i>	Itaituba (PA)
	281	<i>A. othonianum</i> (?)	Água Boa (MT)
Coleção C-03 Venezuela, Índia e Valinhos Introdução por sementes em 1973	05	<i>A. occidentale</i>	Venezuela - El Tigre
	09	<i>A. occidentale</i>	Venezuela - El Tigre
	10	<i>A. occidentale</i>	Índia - Ullal
	12	<i>A. occidentale</i>	Índia - Ullal
	16	<i>A. occidentale</i>	Índia - Ullal
	19	<i>A. occidentale</i>	Índia - Ullal
	25	<i>A. occidentale</i>	Valinhos (SP)
	30	<i>A. occidentale</i>	Valinhos (SP)
Coleção C-04 Roraima Introdução por sementes em 1983	01	<i>Anacardium</i> sp.	Rio Tacutu / Faz. Caju
	18	<i>Anacardium</i> sp.	Est. Tabalascada - Cantá
	19	<i>Anacardium</i> sp.	Est. Normandia Km 25
	24	<i>Anacardium</i> sp.	Est. Normandia Km 25
	30	<i>Anacardium</i> sp.	Maloca do Sucuba
	38	<i>Anacardium</i> sp.	Est. Normandia Km 25
	46	<i>Anacardium</i> sp.	Est. Normandia Km 25
54	<i>Anacardium</i> sp.	Rio Tacutu / Faz. Caju	
Coleção C-05 Paraíba Introdução por sementes em 1990	40	<i>A. occidentale</i>	-
	11	<i>A. occidentale</i>	Pacajus

Coleção C-06 Ceará Introdução por enxertia em 1979	28	<i>A. occidentale</i>	Aracati
	34	<i>A. occidentale</i>	Trairi
	47	<i>A. occidentale</i>	Trairi
	55	<i>A. occidentale</i>	Aracati
	65	<i>A. occidentale</i>	Aracati
	75	<i>A. occidentale</i>	Trairi
	78	<i>A. occidentale</i>	Aracati
	102	<i>A. occidentale</i>	Aracati
	118	<i>A. occidentale</i>	Trairi
	174	<i>A. occidentale</i>	Cascavel
	184	<i>A. occidentale</i>	Russas
	187	<i>A. occidentale</i>	Russas
	195	<i>A. occidentale</i>	Aracoiaba
	196	<i>A. occidentale</i>	Aracoiaba
	223	<i>A. occidentale</i>	Russas
	236	<i>A. occidentale</i>	-
	261	<i>A. occidentale</i>	Russas
275	<i>A. occidentale</i>	Aracoiaba	
C-07 Camocin -1979	01	<i>Anacardium</i> sp.	Camocin (CE)
<i>A. microcarpum</i> sementes - 1964	-	<i>A. microcarpum</i>	Ceará
	-	<i>A. microcarpum</i>	Ceará
Coleção C-09 Goiás Introdução por sementes em 1975	08	<i>A. othonianum</i>	-
	14	<i>A. othonianum</i>	-
	30	<i>A. occidentale</i>	-
Coleção Pacajus – CE Introdução por sementes entre 1956 e 1959	CP12	<i>A. occidentale</i>	Pacajus
	64	<i>A. occidentale</i>	Pacajus
	63 (CP77)	<i>A. occidentale</i>	Pacajus
	67 (CP137)	<i>A. occidentale</i>	Pacajus
	09 (CP138)	<i>A. occidentale</i>	Pacajus
	54 (CP96)	<i>A. occidentale</i>	Pacajus
	CP 136	<i>A. occidentale</i>	Pacajus
Coleção Maranguape – CE Introdução por sementes em 1956	14	<i>A. occidentale</i>	Maranguape
	15	<i>A. occidentale</i>	Maranguape
	16 (CP76)	<i>A. occidentale</i>	Maranguape
	22 (CP06)	<i>A. occidentale</i>	Maranguape
	23 (CP09)	<i>A. occidentale</i>	Maranguape
	25(CP1001)	<i>A. occidentale</i>	Maranguape
	28	<i>A. occidentale</i>	Maranguape
	29	<i>A. occidentale</i>	Maranguape
	30	<i>A. occidentale</i>	Maranguape
35	<i>A. occidentale</i>	Maranguape	

A população subespontânea amostrada de *A. occidentale* localiza-se na Fazenda Pontinha (21°43'77"S e 41°05'68"O), distrito do Cajueiro, município de São João da Barra – RJ. De acordo com Dias (1984, apud Assumpção e Nascimento, 1998) as planícies quaternárias próximas à desembocadura do rio Paraíba do Sul classificam-se como cristas de praia ou simplesmente restingas. O clima da região é quente e úmido, com estação chuvosa no verão e

pluviosidade anual atingindo cerca de 1000 mm (Assumpção e Nascimento, 1998).

2.2 Coleta de dados

Os caracteres botânico-agronômicos foram avaliados durante a fase reprodutiva, nos meses de novembro e dezembro de 2004, para os acessos do Rio de Janeiro e outubro e novembro de 2005, para os acessos do BAG.

Treze características quantitativas relacionadas na lista de descritores do cajueiro, estabelecida pelo IBPGR (1986), foram empregadas no estudo. As variáveis utilizadas foram referentes à inflorescência, pedúnculo, castanha e porte da planta, conforme relação a seguir:

- a) Comprimento da inflorescência (CIN), em centímetros;
- b) Largura máxima da inflorescência (LIN), em centímetros;
- c) Número total de ramificações da inflorescência (NRIN), em relação ao eixo principal;
- d) Comprimento do pedúnculo (CPE), em centímetros;
- e) Diâmetro da base do pedúnculo (DBPE), em centímetros;
- f) Diâmetro do ápice do pedúnculo (DAPE), em centímetros;
- g) Peso do pedúnculo (PPE), em gramas;
- h) Comprimento da castanha (CCA), em centímetros;
- i) Largura da castanha (LCA), em centímetros;
- j) Espessura da castanha (ECA), em centímetros;
- k) Peso da castanha (PCA), em gramas;
- l) Altura máxima da planta (AMP), em metros;
- m) Diâmetro médio da copa (DMC), considerando os eixos N-S e W-E, em metros.

À exceção das variáveis relativas ao porte da planta, foram realizadas 10 medições, por indivíduo, para cada uma das características. Este número foi considerado suficiente para obter médias com confiabilidade igual ou superior a 90 por cento, a partir de um estudo preliminar de repetibilidade realizado por Personi e Cruz (2005).

Dados referentes ao porte das plantas não foram tomados dos acessos do Rio de Janeiro, em decorrência da ausência de uniformidade quanto à idade e espaçamento entre os indivíduos.

2.3 Análise da divergência entre ecótipos e/ou espécies de *Anacardium*

Este estudo envolveu apenas os 69 acessos do BAG do cajueiro, para os quais se dispunha de informações relativas a 13 variáveis morfométricas e dos registros individuais de identificação parcial ou total, quanto à espécie e/ou ecótipo.

A caracterização da divergência genética foi realizada por meio do emprego de análise discriminante, baseadas em componentes principais e em funções de classificação de Anderson, considerando 13 variáveis relativas à inflorescência, pedúnculo, castanha e porte da planta.

Os acessos foram classificados segundo a espécie e/ou ecótipo em um dos três grupos: *A. occidentale* comum, *A. occidentale* anão precoce e *A. othonianum*. Quatorze acessos não foram pré-allocados por estarem identificados incorretamente ou de modo incompleto, ou ainda por não pertencerem a nenhum dos três grupos. Nestes casos tais acessos foram alocados *a posteriori*, através das funções de classificação de Anderson, para averiguar suas similaridades com cada um dos grupos pré-estabelecidos.

2.4 Análise de divergência segundo a origem geográfica de acessos de *Anacardium*

Para esta caracterização, foi considerada, além dos diferentes ecótipos e/ou espécies, a origem geográfica dos acessos avaliados. O trabalho foi conduzido com todos os acessos com dados morfológicos disponíveis, ou seja, 69 acessos do BAG e 23 indivíduos da população do Rio de Janeiro. Foram estabelecidos seis grupos *a priori*, envolvendo 80 dos acessos sob investigação. A definição dos grupos foi realizada utilizando como critérios a origem geográfica dos acessos e tamanho mínimo do grupo igual ou superior a quatro acessos.

As funções discriminantes foram construídas a partir de 11 variáveis morfométricas comuns a todos os acessos avaliados, referentes à inflorescência, pedúnculo e castanha.

Os grupos pré-estabelecidos apresentaram as seguintes denominações e constituições:

- a) Grupo Nordeste (NE): 28 acessos de *A. occidentale*, 27 do estado do Ceará e 1 da Paraíba;
- b) Grupo Anão Precoce (AP): 10 acessos de *A. occidentale* anão precoce, coletados no Município de Maranguape, estado do Ceará;
- c) Grupo Rio de Janeiro (RJ): 23 acessos de *A. occidentale* do norte fluminense;
- d) Grupo Roraima (RR): 08 acessos de *A. occidentale* oriundos do Estado de Roraima;
- e) Grupo Índia (IN): 04 acessos de *A. occidentale* provenientes da Índia;
- f) Grupo *othonianum* (OT): 06 acessos de *A. othonianum* do Cerrado do Brasil Central.

Os 12 acessos remanescentes não foram agrupados por não atenderem aos critérios de agrupamento pré-estabelecidos. A alocação *a posteriori* destes acessos, dentro dos grupos pré-estabelecidos, foi averiguada através da análise discriminante de Anderson.

2.5 Análises dos dados

2.5.1 Análise de variância

As medidas individuais das variáveis relativas à inflorescência, pedúnculo e castanha foram submetidas, inicialmente, a análises de variâncias considerando um fator de classificação, à exceção das variáveis peso do pedúnculo e peso da castanha, para os indivíduos do BAG, cujos valores foram obtidos pela média aritmética do peso conjunto de 10 frutos ou 10 castanhas.

2.5.2 Análises discriminantes

A divergência ao nível de grupo foi avaliada através do emprego de análises discriminantes baseada em componentes principais (CP) e função discriminante de Anderson, utilizado as médias de variáveis dos indivíduos amostrados.

Na análise discriminante baseada em CP as funções são obtidas em ordem decrescente de importância, de tal forma que a primeira função explica o máximo da variância entre os grupos e a segunda, ortogonal à primeira, explica o máximo da variância remanescente e assim por diante (Manly, 1998; Cruz e Carneiro, 2003). O modelo matemático destas funções possui a seguinte composição:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{iv}X_v$$

onde

Z_i : função discriminante de ordem i ;

$a_{i1}, a_{i2}, \dots, a_{iv}$: pesos discriminantes para variáveis independentes v ($i = 1, 2, \dots, v$), na função discriminante Z_i ;

X_1, X_2, \dots, X_v : variáveis independentes.

O número máximo de funções estimadas, para um particular conjunto de dados, é o valor mínimo entre $(p-1, v)$, em que p é o número de populações e v o número de variáveis (Manly, 1998; Cruz e Carneiro, 2003).

A partir das funções discriminantes de grupos foram estimados os escores dos indivíduos pertencentes a cada grupo. Em seguida, as dissimilaridades entre indivíduos e seus respectivos grupos foram visualizadas por meio de dispersão gráfica, no espaço bi ou tridimensional (Cruz e Carneiro, 2003).

A importância relativa dos caracteres empregados na discriminação dos grupos foi avaliada utilizando o critério de peso das variáveis nos autovetores. Foram consideradas de menor importância, as variáveis com maiores pesos, desde o último autovetor até aquele associado a um autovalor igual ou inferior a 0,70 (Cruz e Carneiro, 2003; Cruz et al. 2004).

Na análise discriminante de Anderson (1958, 1984) parte-se de um conjunto de observações pertencentes a $p(\geq 2)$ populações conhecidas com

distribuição normal multivariada, denominadas por $\pi_1, \pi_2, \dots, \pi_p$, para obter regras de discriminação de futuras observações de origem desconhecida (Khattree e Naik, 2000). Para isso, o espaço p-dimensional é dividido em p regiões disjuntas e mutuamente exclusivas, denominadas R_1, R_2, \dots, R_p e se uma observação cai dentro da região R_i assume-se que ela é proveniente da população π_i . Adicionalmente, são assumidas probabilidades *a priori* para cada uma das populações, de modo que $\pi_1 + \pi_2 + \dots + \pi_p = 1$, além de um custo de má classificação na construção das funções de discriminação (Anderson, 1958, 1984; Khattree e Naik, 2000).

De acordo com a notação de Cruz e Carneiro (2003), a forma genérica da função discriminante de Anderson é dada pela expressão:

$$D_i(\tilde{x}) = \kappa_i + \alpha_{i1}x_1 + \alpha_{i2}x_2 + \dots + \alpha_{iv}x_v$$

Onde

κ_i : constante associada à função discriminante;

α_{ij} : coeficiente de ponderação da j -ésima variável ($j = 1, 2, \dots, v$) na i -ésima função discriminante;

x_j : valor representativo do escore da j -ésima variável do elemento que se deseja classificar em uma das populações em estudo.

A eficácia das funções discriminantes estimadas foi testada através do cálculo da taxa de erro aparente (TEA). Esta taxa é obtida através da soma de todos os casos de alocação incorreta, decorrentes da reclassificação dos dados utilizados na geração das funções, dividido pelo número total de classificações realizadas.

2.5.3 Análise de agrupamento

As inter-relações dos grupos, descritos anteriormente, foi avaliada através de uma árvore não enraizada, construída pelo método UPGMA, a partir das distâncias de Mahalanobis.

A divergência entre todos os acessos foi avaliada através da dispersão baseada em componentes principais. Os autovetores e autovalores foram estimados a partir da matriz de correlação dos dados padronizados das médias

originais. Adicionalmente, foi comparado o padrão de agrupamento submetido à análise discriminante de Anderson com o padrão de agrupamento obtido pela técnica de otimização de Tocher.

3 RESULTADOS

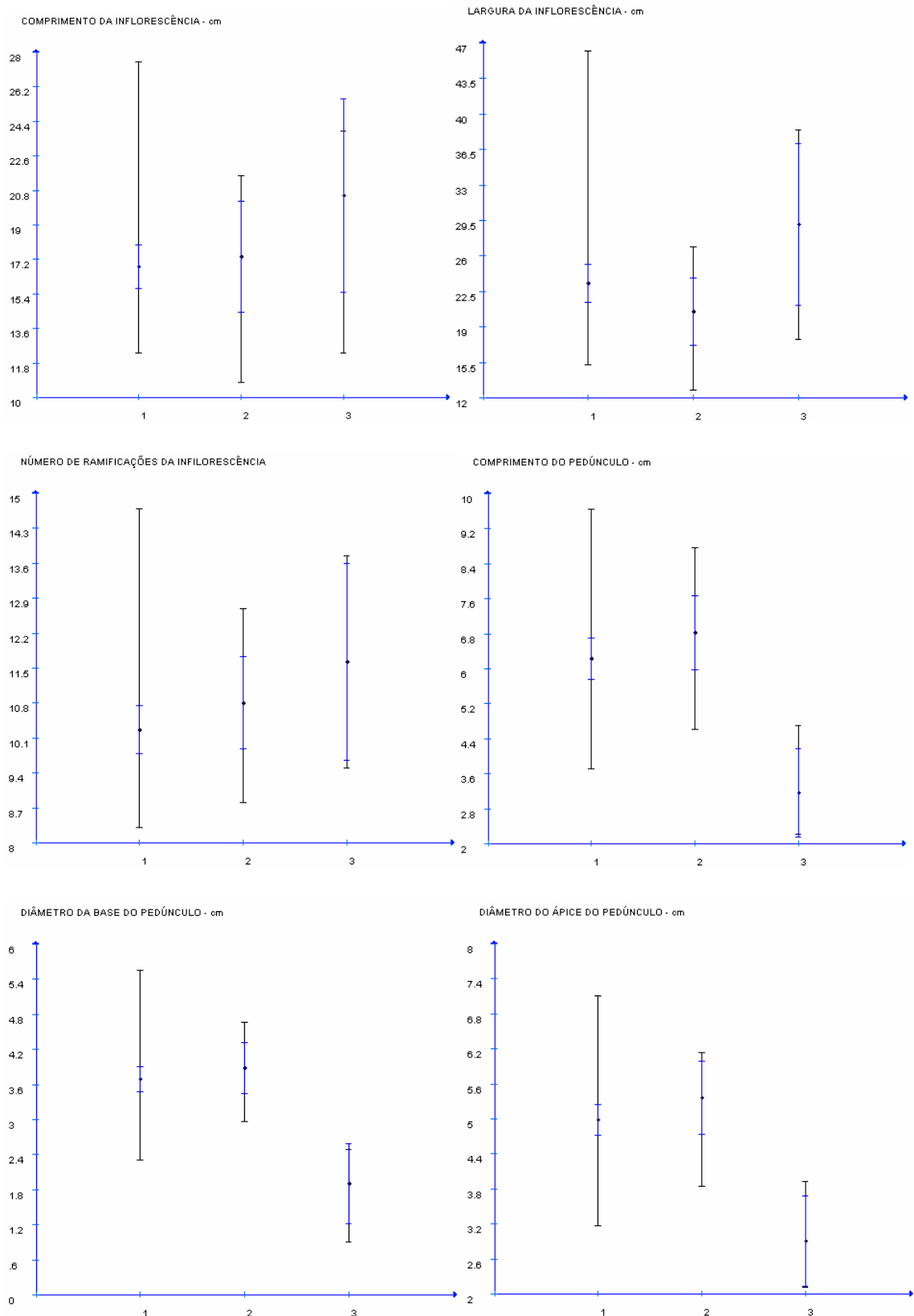
3.1 Análise da divergência entre ecótipos e/ou espécies de *Anacardium*

As médias, amplitudes totais e intervalos de confiança (95%) das 13 características morfométricas avaliadas, segundo os grupos ou ecótipos, são apresentadas graficamente na Figura 1. De modo geral, observa-se que o grupo 1 (*A. occidentale* comum) apresentou as maiores amplitudes em todas as variáveis, mas com a maior parte das observações concentradas em torno das médias do grupo, conforme destacado pelos intervalos de confiança. O grupo 3 (*othonianum*) apresentou as maiores médias para os caracteres relativos à inflorescência, mas estas diferenças foram estatisticamente significativas (teste t a 0,05 de probabilidade) apenas para variável 'largura da inflorescência'. Por outro lado, este grupo apresentou valores significativamente inferiores aos dos demais para as variáveis relativas ao pedúnculo, castanha e porte da planta.

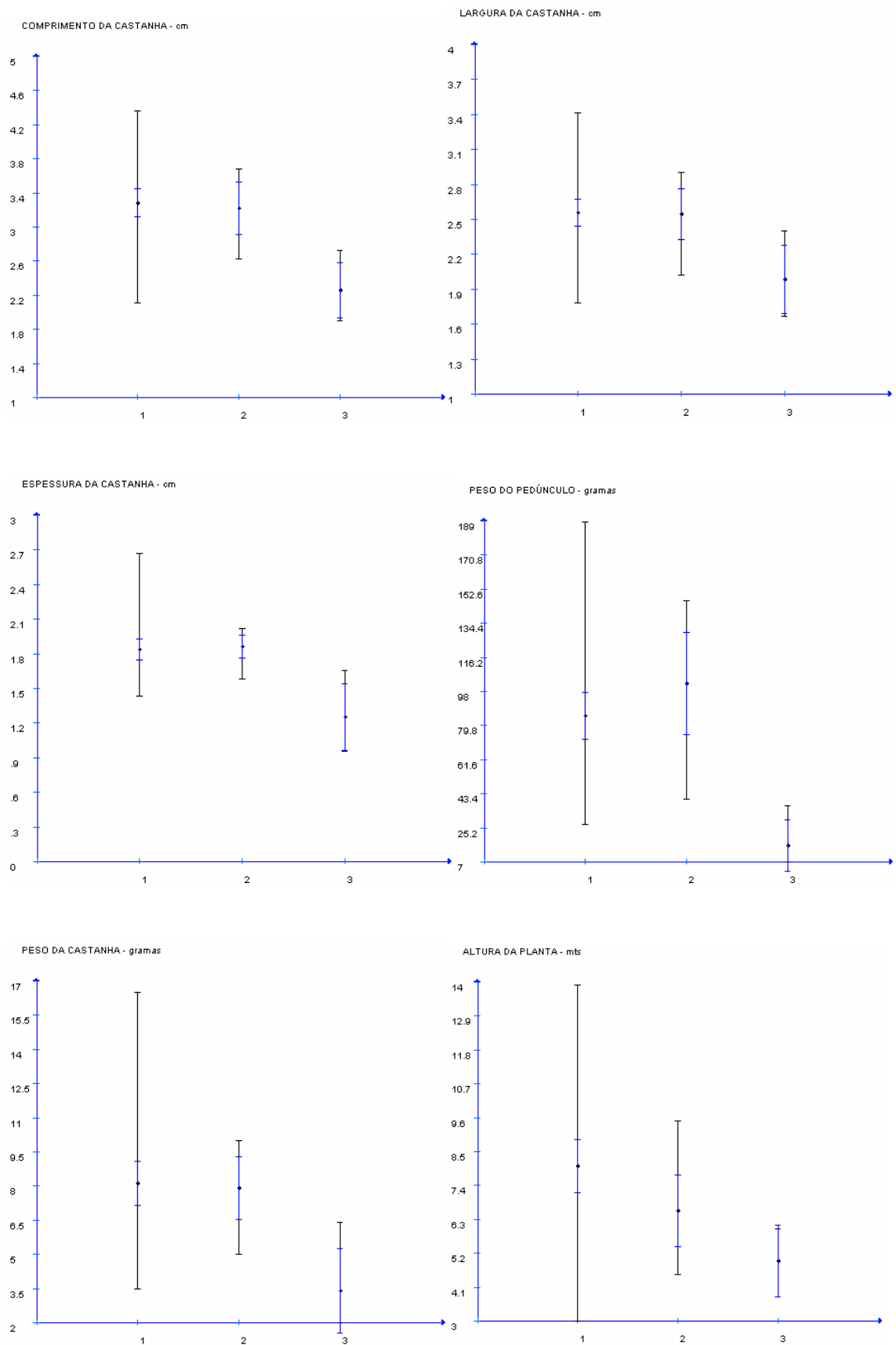
Nenhuma das variáveis analisadas apresentou médias que diferissem estatisticamente ($\alpha = 0,05$) entre os grupos de *A. occidentale* comum (1) e anão precoce (2).

Os resultados da análise discriminante baseada em componentes principais estão resumidos na Figura 2. Observa-se que o centróide e a dispersão dos acessos de *A. othonianum* são claramente divergentes das dispersões apresentadas pelos outros dois grupos. Por outro lado, é possível notar também uma tendência de diferenciação entre grupos na dispersão dos acessos 'comum' e 'anão precoce', em torno de seus centróides.

Figura 1 – Médias (ponto), amplitudes (barra escura) e intervalos de confiança 95% (barra azul) de 13 variáveis morfométricas, de acessos de *Anacardium* agrupados em ecótipos e/ou espécies (1 - *A. occidentale* comum; 2 - *A. occidentale* anão precoce; 3 - *A. othonianum*).



Continuação Figura 1



Continuação Figura 1

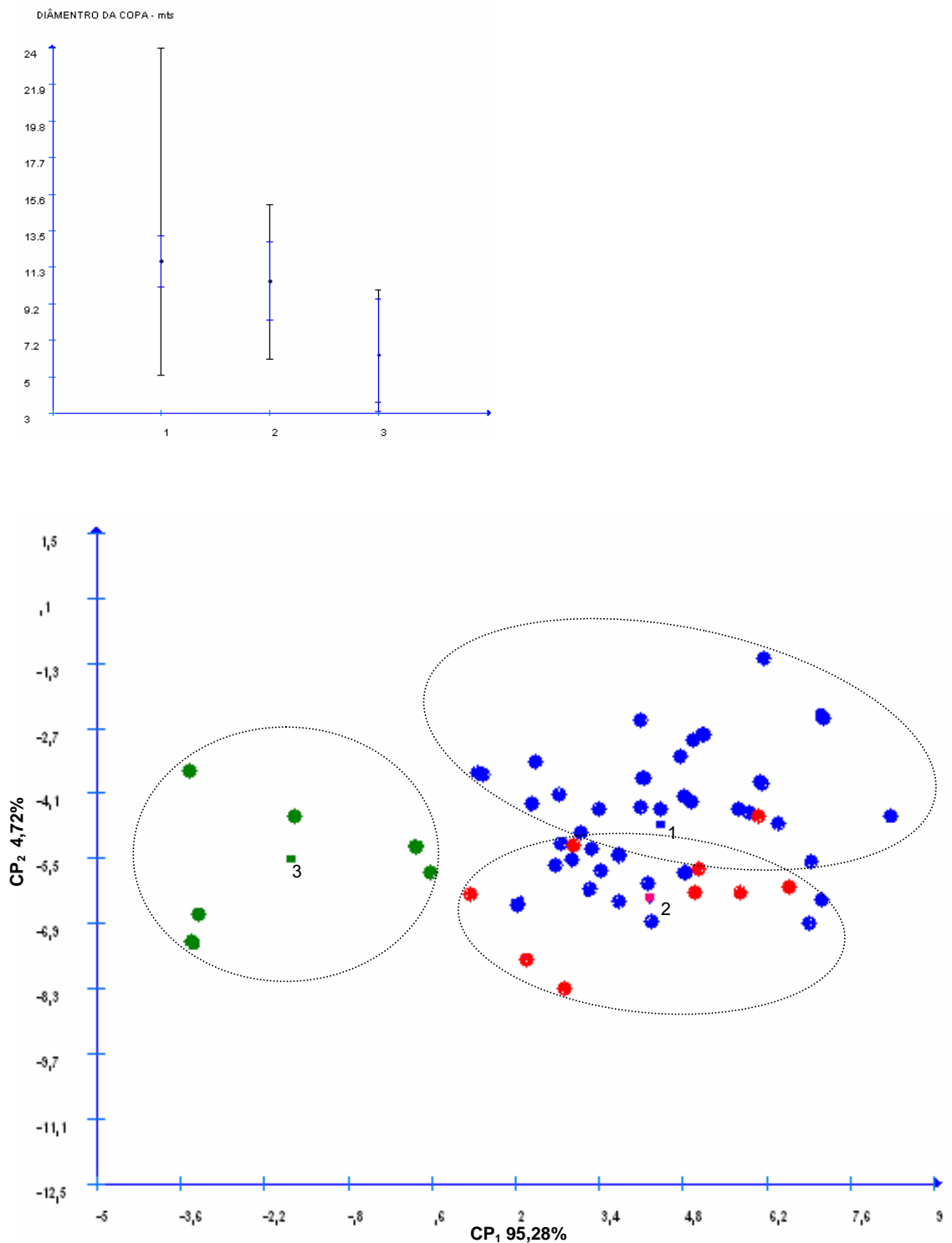


Figura 2 - Dispersão gráfica dos centróides (números) e acessos de três ecótipos de *Anacardium*, em relação funções discriminante estimadas por componentes principais, a partir da combinação linear entre 13 características morfométricas. (1-azul: *A. occidentale* comum, 2-vermelho: *A. occidentale* anão precoce, 3-verde: *A. othonianum*).

O primeiro componente principal é um vetor que concentra quase toda a variação disponível, respondendo por 95,3% da variação total dos dados. Todas as variáveis analisadas apresentaram pesos semelhantes para este componente. A única exceção foi variável altura da planta, que teve um efeito um pouco menor no primeiro componente ao mesmo tempo em que foi a de maior efeito no segundo componente principal, indicando com isso que é a variável de menor importância na discriminação dos grupos (Tabela 1).

Na Tabela 1 são apresentadas as funções discriminantes de Anderson, estimadas para os três grupos investigados. A classificação dos acessos obtida a partir destas funções, apresentada na Tabela 2, confirma o maior grau de diferenciação do grupo 'othonianum', com todos os seus representantes alocados corretamente no grupo respectivo, ao mesmo tempo em que este grupo não recebeu a alocação de nenhum acesso pertencente a outro grupo. Por sua vez, os percentuais de classificações incorretas nos grupos 'comum' e 'anão precoce' também foram relativamente baixos, com maior frequência de acessos do grupo 'comum' sendo alocados incorretamente no grupo 'anão precoce' do que o contrário. Dessa forma, dos 55 acessos pré-classificados apenas oito foram alocados incorretamente, o que resultou em uma taxa de erro aparente total de 14,5%.

Entre os 14 acessos alocados *a posteriori*, oito foram alocados no grupo 'comum', quatro no grupo 'anão precoce' e dois no grupo 'othonianum'. Dos oito acessos provenientes do estado de Roraima e identificados no BAG apenas como *Anacardium* sp, quatro foram alocados no grupo 'anão precoce' e quatro grupo 'comum'. Entre os dois acessos alocados no grupo 'othonianum' um foi coletado no estado do Piauí e encontra-se identificado apenas como *Anacardium* sp. O outro acesso está identificado como *A. microcarpum* e sua origem é desconhecida. Os quatro acessos restantes foram alocados no grupo 'comum', sendo três identificados apenas como *Anacardium* sp. e um como *A. othonianum*.

Tabela 1 - Constantes, coeficientes de ponderação e estimativas de autovetores referentes às funções discriminantes por componentes principais (1) e de Anderson (2), considerando três ecótipos de *Anacardium* e 13 variáveis métricas.

Variáveis	¹ Coeficientes de ponderação e estimativas de autovetores (λ_j)		² Constantes e coeficientes de ponderação das funções discriminantes de Anderson		
	CP ₁	CP ₂	-205,1661	-203,4227	-190,7639
CIN	-0,2827	-0,1292	-1,3352	-0,9138	-1,1182
LIN	-0,2673	0,4331	1,4140	1,1066	1,3391
NRIN	-0,2644	-0,4672	6,1883	6,3746	6,3922
CPE	0,2797	-0,2254	-1,7381	-2,1354	-2,8950
DBPE	0,2823	-0,1437	-17,3110	-17,8891	-20,0715
DAPE	0,2800	-0,2174	37,6960	37,9786	36,7747
CCA	0,2840	0,0369	42,3910	44,4063	41,0549
LCA	0,2841	-0,0170	65,4769	63,4634	66,7939
ECA	0,2826	-0,0816	55,2291	54,9172	52,5099
PPE	0,2776	-0,2729	-0,6168	-0,5623	-0,5419
PCA	0,2841	0,0180	-19,1260	-19,5991	-18,8419
AMP	0,2543	0,5692	2,2889	1,5162	2,1589
DMC	0,2796	0,2272	0,2218	0,4585	0,0460
λ_j	13,3861	0,6138	-	-	-
λ_j (%)	95,28	4,72	-	-	-

Tabela 2 – Análise discriminante de ecótipos de *Anacardium* baseada em funções discriminantes de Anderson, estimadas a partir da combinação linear de 13 características morfométricas. Os valores destacados em negrito referem-se aos percentuais de alocação dos acessos nos seus respectivos grupos de origem.

Grupo de origem	Número de Indivíduos	Percentuais de alocação nos diferentes grupos		
		<i>A. occidentale</i> comum	<i>A. occidentale</i> anão precoce	<i>A. othonianum</i>
<i>A. occidentale</i> comum	39	82,05	17,95	0,00
<i>A. occidentale</i> anão precoce	10	10,00	90,00	0,00
<i>A. othonianum</i>	06	0,00	0,00	100
Alocados a <i>posteriori</i>	14	57,14	28,57	14,28
Total de indiv.	69	41	20	08

Taxa de erro aparente = 14,5%.

3.2 Análise de divergência segundo a origem geográfica de acessos de *Anacardium*

As médias, amplitudes e intervalos de confiança (95%) das 11 características avaliadas, dos seis grupos de *Anacardium* estabelecidos a priori, são apresentadas na Figura 3. Novamente o grupo 'othonianum' (6) se destaca por apresentar os maiores valores médios para os caracteres relativos à inflorescência e as menores médias para os demais caracteres. Entre os grupos constituídos por acessos de *A. occidentale*, o do 'Rio de Janeiro' (3) apresentou as menores médias para 10 das 11 características, a única exceção foi a variável número de ramificações da inflorescência. Entretanto estas diferenças foram estatisticamente significativas apenas para os caracteres relativos ao pedúnculo e ao peso da castanha.

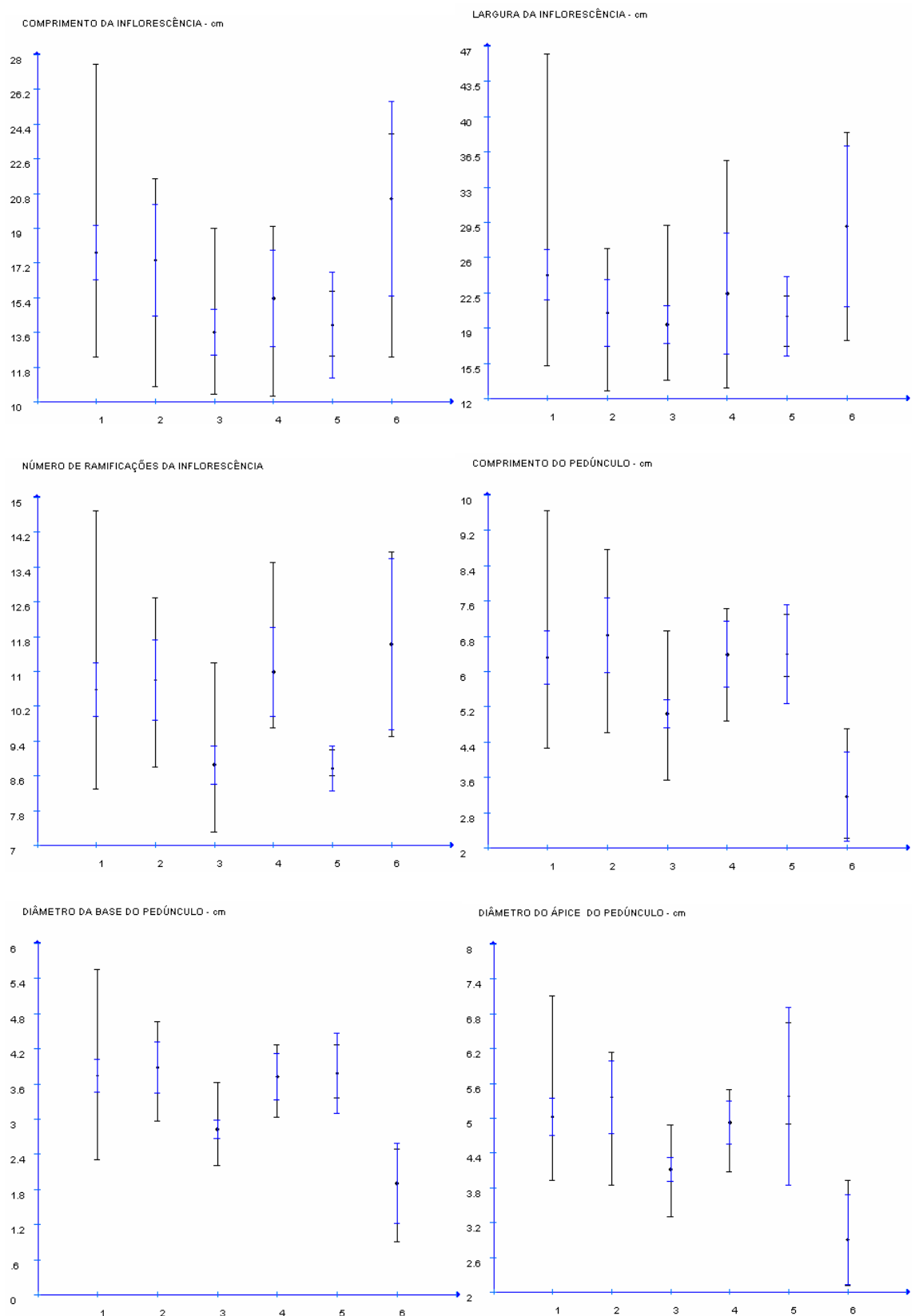
Embora os grupos 3 e 5 (Índia) não tenham diferido para caráter número de ramificações da inflorescência, esta variável foi importante na discriminação destes dois grupos em relação aos demais envolvidos no estudo.

A Tabela 3 e a Figura 4 apresentam os resultados da análise discriminante baseada em componentes principais, considerando as seis populações pré-estabelecidas, segundo os critérios de ecótipo e/ou espécie e origem geográfica dos acessos.

Verifica-se que a maior parte da variabilidade foi retida nos primeiros componentes principais, com o primeiro e o segundo componentes explicando, respectivamente, 77,58% e 17,84% da variação total. Dentro destas circunstâncias, os dois primeiros componentes retêm cerca de 95% da variância dos dados.

Examinando a Tabela 3 também é possível identificar as variáveis de maiores pesos nos últimos autovetores e, portanto, com menor contribuição na discriminação dos grupos. Estas variáveis foram, em ordem crescente de importância, comprimento da castanha, comprimento da inflorescência e número de ramificações da inflorescência, conforme valores destacados em negrito na Tabela.

Figura 3 – Médias (ponto), amplitudes (barra escura) e intervalos de confiança 95% (barra azul) de 11 variáveis morfométricas de acessos de *Anacardium*, agrupados segundo a origem geográfica, ecótipos ou espécie. Grupos: 1 – Nordeste ; 2 – Anão Precoce; 3 – Rio de Janeiro; 4 – Roraima; 5 – Índia; 6 - *A. othonianum*.



Continuação Figura 3

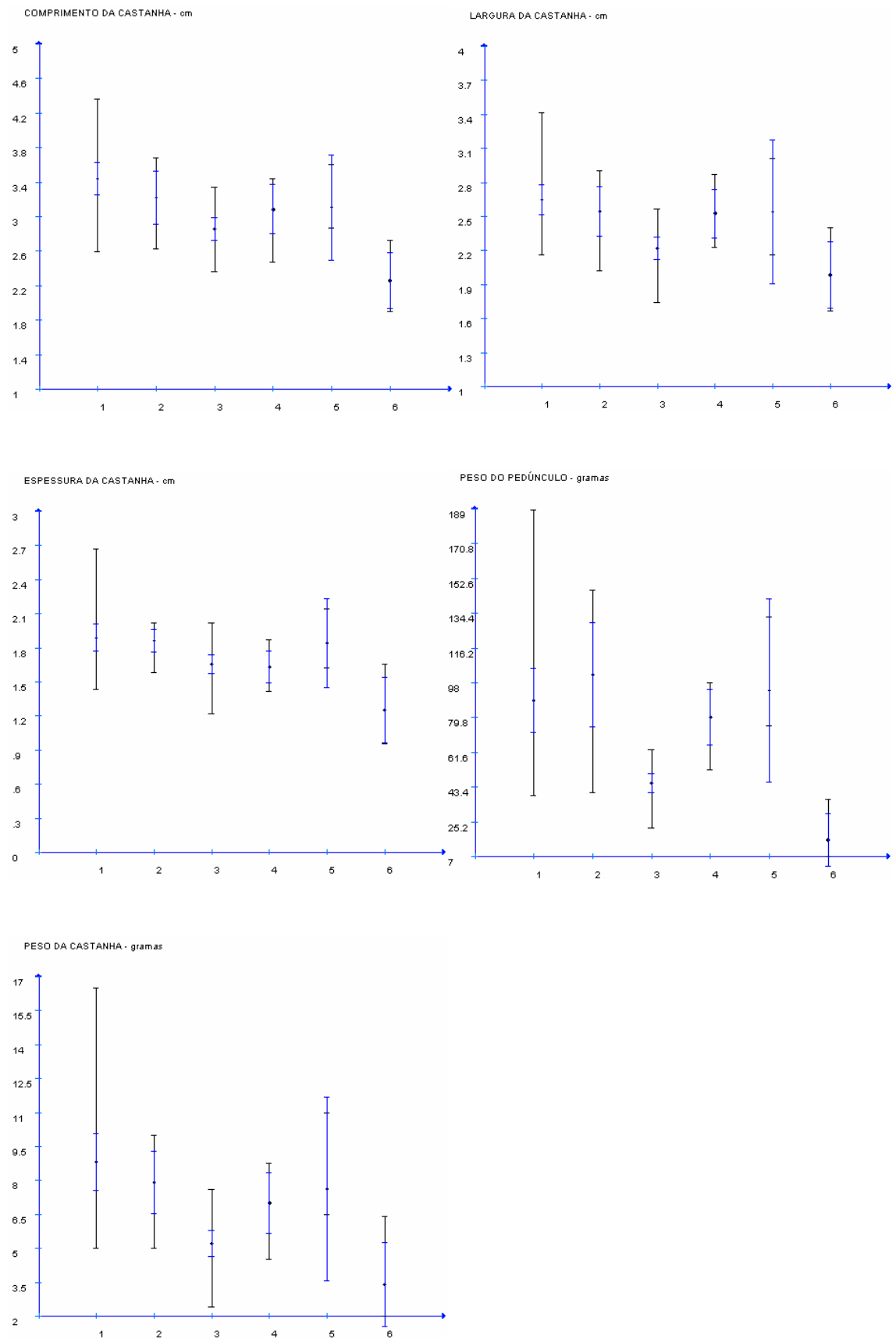


Tabela 3 – Estimativas de autovalores e autovetores associados aos componentes principais, obtidos da matriz de correlação entre 11 variáveis morfométricas de *Anacardium* spp. Valores expressos em: ¹centímetros, ²números e ³gramas.

Variável	Autovetores (λ_j) associados aos Componentes Principais				
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅
CIN ¹	-0,1999	0,5499	0,2804	-0,5679	-0,0866
LIN ¹	-0,2566	0,4260	0,3423	0,5208	-0,3762
NRIN ²	-0,1462	0,6070	-0,5435	-0,0057	0,3833
CPE ¹	0,3360	0,0797	-0,2957	-0,0269	0,0385
DBPE ¹	0,3349	0,1129	-0,2264	0,1301	-0,1348
DAPE ¹	0,3379	0,0491	-0,1537	-0,0963	-0,3704
CCA ¹	0,3263	0,1302	0,2699	0,1877	0,5854
LCA ¹	0,3251	0,2000	0,0560	0,3684	-0,0488
ECA ¹	0,3305	-0,0004	0,4164	-0,3664	0,1727
PPE ³	0,3291	0,1528	-0,1318	-0,2455	-0,4149
PCA ³	0,3223	0,2144	0,2824	0,1183	0,0119
λ_j	8,5334	1,9628	0,2711	0,1344	0,0978
λ_j (%)	77,58	17,84	2,46	1,22	0,89

* valores em negrito destacam a variável de maior peso no respectivo autovetor.

A dispersão gráfica dos centróides e acessos, apresentadas na Figura 4, mostra ser possível delimitar os grupos '*othonianum*' e 'Rio de Janeiro' com relativa facilidade. Por outro lado, nos demais grupos há uma acentuada proximidade dos centróides e sobreposição na dispersão de seus respectivos acessos, impossibilitando a distinção destes grupos.

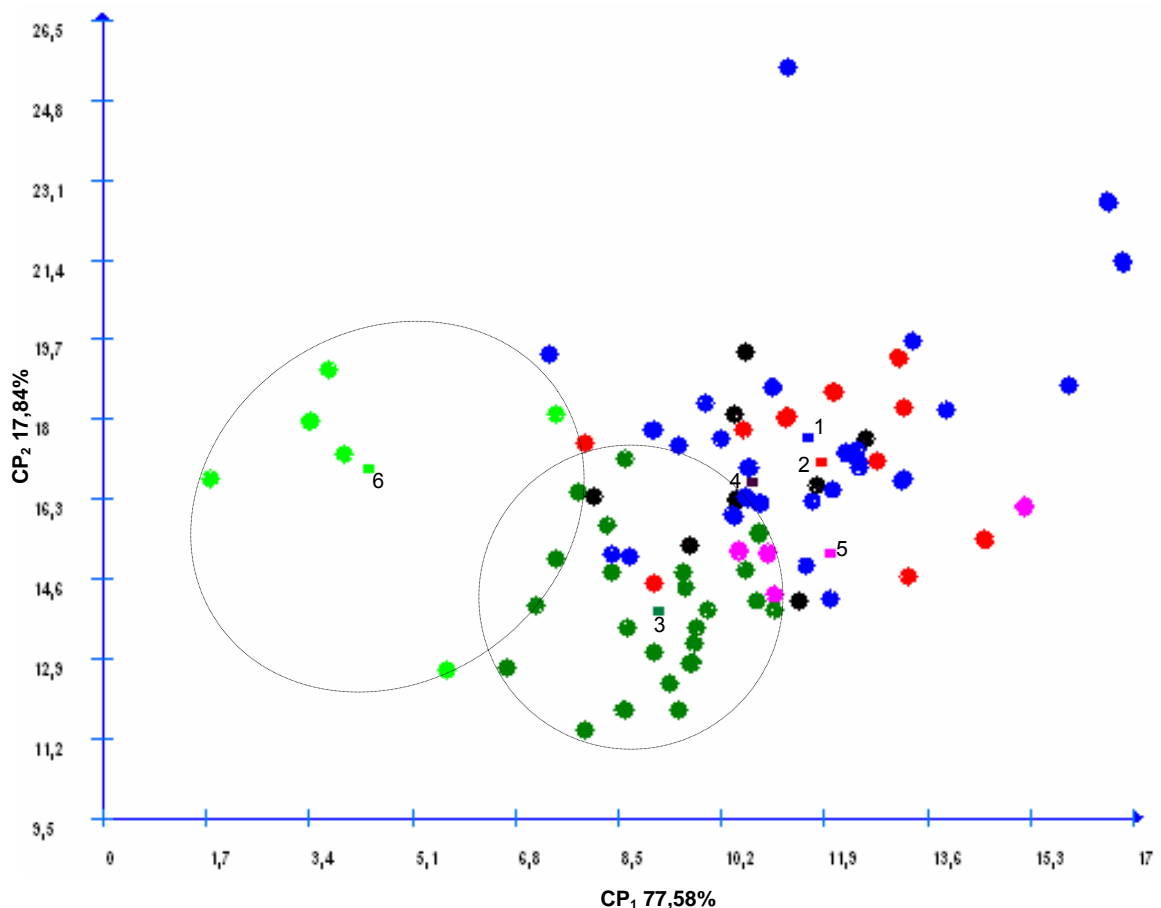


Figura 4 - Dispersão gráfica dos centróides (números) e acessos de seis populações de *Anacardium*, em relação a funções discriminantes estimadas por componentes principais, a partir da combinação linear entre 11 características morfométricas. Grupos: 1-azul – Nordeste; 2-vermelho – Anão Precoce; 3-verde escuro – Rio de Janeiro; 4-preto – Roraima; 5-rosa – Índia; 6-verde claro – *A. othonianum*.

A Tabela 5 apresenta a alocação dos acessos nos diferentes grupos, obtida de acordo com as seis funções discriminantes de Anderson estimadas, apresentadas na Tabela 4. Uma vez mais a singularidade do grupo ‘*othonianum*’ é destacada. Observa-se também que todos os acessos do grupo da ‘Índia’ foram alocados corretamente, embora diversos acessos pertencentes a outros grupos também tenham sido alocados neste grupo. O grupo ‘Rio de Janeiro’ também apresentou elevado grau de discriminação, com a ocorrência de uma permuta recíproca apenas entre um de seus acessos com outro do grupo ‘Nordeste’.

Por outro lado, a discriminação dos demais grupos foi menos eficiente, com muitos acessos sendo alocados em grupos diferentes dos estabelecidos *a priori*. O grupo ‘Nordeste’ foi o que apresentou o maior percentual de erro de

alocação, com 37% de seus acessos sendo alocados em outros grupos. Considerando as 78 classificações realizadas 16 foram incorretas, o que resultou em uma taxa de erro aparente de 20,51%.

Tabela 4 – Funções discriminantes de Anderson, referentes a seis populações de *Anacardium* spp., obtidas pela combinação linear de 11 características métricas.

Componente	Nome das Populações					
	Nordeste	Anão precoce	Rio de Janeiro	Roraima	Índia	<i>othonianum</i>
Constante	-244,0509	-243,9762	-241,5911	-241,6326	-237,9737	-222,9650
CIN	1,2668	1,6745	1,1383	0,6773	1,2671	1,3475
LIN	0,7340	0,4159	0,7476	0,8384	0,7523	0,6938
NRIN	4,2782	4,3513	4,0415	5,1833	3,4931	4,6913
CPE	-1,03557	-0,9620	-1,5422	0,5584	-1,0907	-2,5812
DBPE	-19,0902	-19,5064	-20,8918	-17,2405	-20,3531	-22,6102
DAPE	43,8324	44,8787	45,1468	44,2412	47,0456	43,0151
CCA	66,2065	63,7227	67,3847	60,8981	61,3708	60,9453
LCA	70,7953	72,3593	70,9039	73,3362	74,1993	73,8511
ECA	60,5603	60,9390	63,4861	52,9955	59,2953	56,7851
PPE	-0,6869	-0,6512	-0,6748	-0,7449	-0,6844	-0,5963
PCA	-24,0431	-24,6695	-24,8457	-23,4100	-24,2122	-23,4764

Entre as classificações realizadas *a posteriori*, três acessos foram alocados no grupo ‘Nordeste’, sendo um originário de Venezuela e os outros dois da Região Nordeste. Estes últimos identificados apenas como *Anacardium* sp. O grupo ‘Anão precoce’ recebeu a alocação de dois acessos, um originário do Pará e outro de São Paulo, ambos identificados como *A. occidentale*. No grupo ‘Roraima’ foram alocados cinco acessos, dois do Pará, um de Mato Grosso, um da Venezuela e um do Ceará. Este último identificado apenas como *Anacardium* sp. Apenas um acesso originário de São Paulo foi alocado no grupo ‘Índia’, enquanto no grupo ‘othonianum’ foram alocados três acessos. Dois destes acessos já haviam sido alocados neste grupo, na análise discriminante de ecótipos (Tabela 1). O terceiro acesso incluído no grupo ‘othonianum’ é originário de Goiás e está identificado como *A. occidentale*.

Tabela 5 – Análise discriminante de seis populações de *Anacardium* baseada em funções discriminantes de Anderson, estimadas a partir da combinação linear de 11 características morfométricas. Os Valores destacados em negrito referem-se aos percentuais de alocação dos acessos nos seus respectivos grupos de origem.

Grupo de origem	Nº de Indiv.	Percentuais de alocação nas diferentes populações					
		NE	AP	RJ	RR	ID	OT
Nordeste (NE -01)	27	62,96	7,41	3,70	11,11	14,81	0,00
Anão Precoce (AP – 02)	10	10,00	70,00	0,00	10,00	10,00	0,00
Rio de Janeiro (RJ – 03)	23	4,35	0,00	95,65	0,00	0,00	0,00
Roraima (RR - 04)	08	0,00	12,50	0,00	75,00	12,50	0,00
Índia (ID - 05)	04	0,00	0,00	0,00	0,00	100	0,00
<i>othonianum</i> (OT - 06)	06	0,00	0,00	0,00	0,00	0,00	100
Alocados a <i>posteriori</i>	14	21,43	14,28	0,00	37,71	7,14	21,43
Total de Indiv. alocados	92	22	12	23	14	12	09

Taxa de erro aparente = 20,51%.

3.3 Análises de agrupamento

As similaridades entre as seis populações, delimitadas segundo a origem geográfica, ecótipo e/ou espécie, são apresentadas no dendrograma da Figura 05. As duas populações mais similares foram Nordeste (NE) e Anão precoce (AP), consideradas ecótipos diferentes, embora de mesma origem geográfica (região Nordeste). A este grupo inicial as populações mais próximas, em ordem decrescente de similaridade, foram Índia (ID), Roraima (RR) e Rio de Janeiro (RJ), todas pertencentes a *A. occidentale*. Novamente a população '*othonianum*' (OT) mostrou-se mais divergente em relação a todas as demais avaliadas.

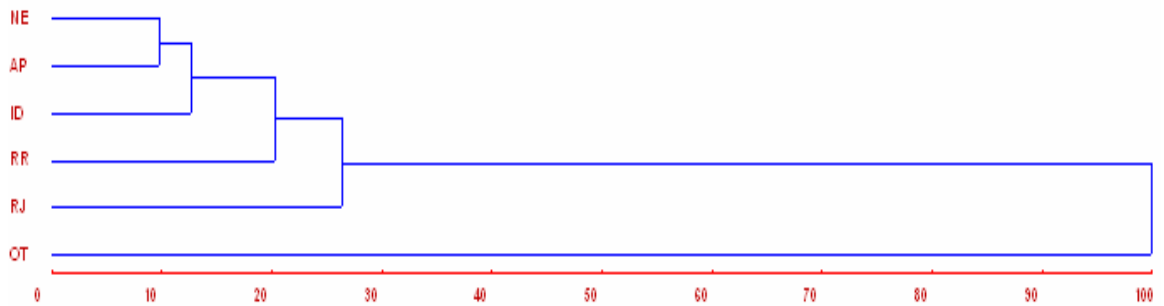


Figura 5 - Dendrograma de seis populações de *Anacardium*, obtido por agrupamento UPGMA a partir de distâncias generalizadas de Mahalanobis. Populações: NE - nordeste, AP - anão precoce, ID - Índia, RR – Roraima, RJ – Rio de Janeiro, OT – *othonianum*.

Os resultados da avaliação da diversidade entre os 92 acessos baseada em análise multivariada por componentes principais são apresentados na Figura 6 e Tabela 6.

A análise da Tabela 6 indica que o primeiro e segundo componentes respondem, respectivamente, por 59,83 e 21,54 por cento da variabilidade total dos dados. Portanto, a combinação destes dois componentes explica mais de 80% da variação disponível nos dados originais e, dessa forma, a dispersão gráfica dos acessos no espaço bidimensional traduz bem suas similaridades relativas.

A partir dos dados apresentados na Tabela 6 é possível também identificar a importância relativa das variáveis sobre a diversidade genética dos acessos. Variáveis de maior peso nos menores autovetores, são de menor importância na discriminação dos acessos. Partindo-se do último até o terceiro autovetor, as variáveis de maiores pesos foram, respectivamente: peso da castanha (PCA), diâmetro do ápice do pedúnculo (DAPE), largura da castanha (LCA), peso do pedúnculo (PPE), comprimento da inflorescência (CIN), espessura da castanha (ECA), diâmetro da base do pedúnculo (DBPE), número de ramificações da inflorescência (NRIN) e comprimento do pedúnculo (CPE).

Na Figura 6 além das similaridades entre os acessos, evidenciada pela dispersão gráfica dos pontos, procurou-se demonstrar os agrupamentos obtidos segundo dois critérios diferentes.

O primeiro tipo de agrupamento destaca os grupos formados pela análise discriminante de Anderson, considerando seis populações previamente

estabelecidas segundo a origem geográfica, ecótipo e/ou espécie dos acessos. Estes agrupamentos são diferenciados pelas cores dos pontos e os acessos de cada grupo incluem aqueles determinados *a priori*, assim como aqueles alocados *a posteriori*, a partir das funções discriminantes obtidas (Tabela 4).

O segundo tipo de agrupamento foi obtido através do método de otimização de Tocher. Neste caso, foram produzidos sete grupos de acessos, delimitados pelas figuras geométricas (elipses).

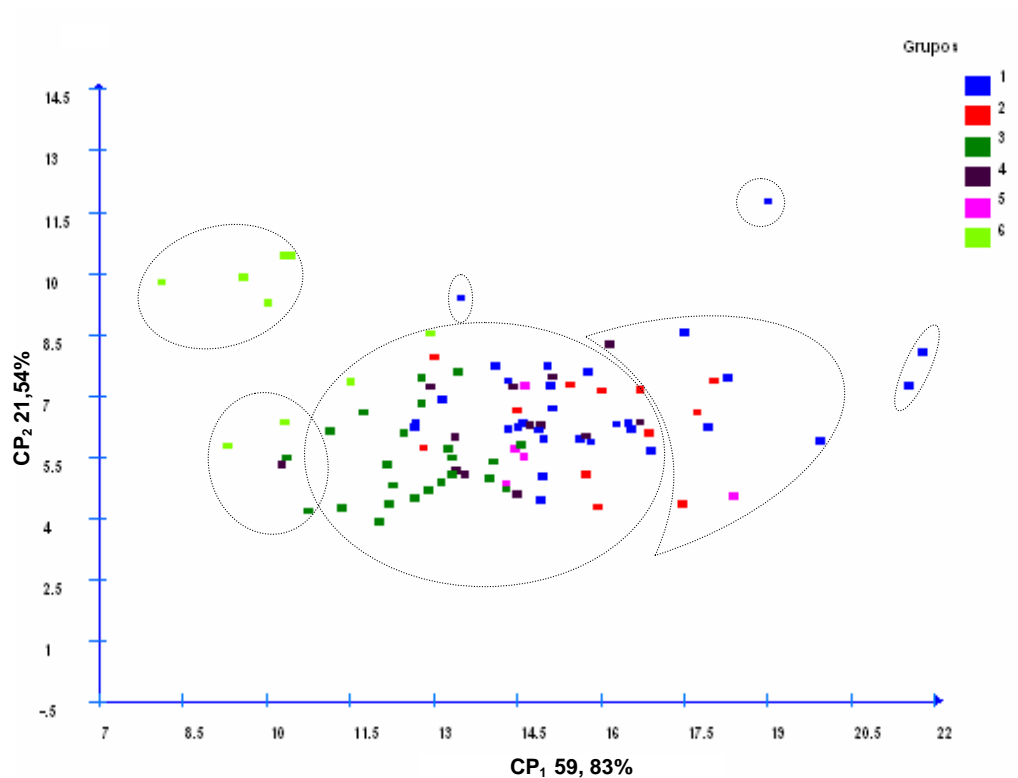


Figura 6 - Dispersão gráfica de 92 acessos de *Anacardium*, em relação ao primeiro e segundo componentes principais, estabelecidos pela combinação linear de 11 variáveis morfométricas. As cores diferenciam os acessos alocados nas seis populações pela análise discriminante de Anderson, enquanto as figuras geométricas delimitam os grupos obtidos pela análise de agrupamento de Tocher. Grupos: 1 – Nordeste; 2 – Anão Precoce; 3 – Rio de Janeiro; 4 – Roraima; 5 – Índia; 6 – *A. othonianum*.

Tabela 6 – Estimativas de autovalores e autovetores associados aos componentes principais, obtidos da matriz de correlação entre 11 variáveis morfométricas de 92 acessos de *Anacardium* spp. CIN – comprimento da inflorescência; LIN – largura da inflorescência; NRIN – número de ramificações da inflorescência; CPE – comprimento do pedúnculo; DBPE – diâmetro da base do pedúnculo; DAPE – diâmetro do ápice do pedúnculo; CCA – comprimento da castanha; LCA – largura da castanha; ECA – espessura da castanha; PPE – peso do pedúnculo; PCA – peso da castanha.

λ_j	$\lambda_j(\%)$ Acumulada	Elementos de autovetores associados										
		CIN	LIN	NRIN	CPE	DBPE	DAPE	CCA	LCA	ECA	PPE	PCA
6,5814	59,83	0,0846	0,0478	0,0909	0,3085	0,3465	0,3635	0,3566	0,3591	0,3284	0,3659	0,3703
2,3699	81,37	0,6043	0,5810	0,5119	-0,0569	-0,0631	-0,0938	-0,0229	0,0127	-0,1359	-0,0083	-0,0018
⊗ 0,5873	86,71	0,0462	0,1835	-0,2548	0,6927	-0,0918	0,1679	-0,1293	-0,2755	-0,3534	0,3253	-0,2440
0,5128	91,38	-0,1306	-0,4157	0,6751	0,0234	0,3153	0,2323	-0,2967	-0,2045	-0,1076	0,1486	-0,1922
0,3366	94,44	0,0151	-0,3568	0,3534	0,4451	-0,5659	-0,3292	0,2880	0,0408	0,1364	-0,0944	0,0932
0,2485	96,69	0,3134	-0,0382	-0,1286	0,0510	-0,1032	0,0674	-0,3861	-0,3051	0,7794	0,0510	-0,1223
0,1179	97,77	-0,5455	0,4196	0,1908	0,1768	-0,1433	-0,0147	-0,3967	0,4593	0,2059	-0,0746	-0,1473
0,0937	98,62	0,1048	0,0040	-0,0314	0,3349	0,5217	-0,1879	0,2208	0,0683	0,1211	-0,5587	-0,4337
0,0848	99,40	-0,4438	0,3855	0,1751	-0,0384	0,0379	0,0188	0,3967	-0,6441	0,1932	-0,0397	0,0970
0,0406	99,76	-0,0569	0,0151	-0,0291	0,1351	0,3779	-0,7372	-0,2704	-0,0952	-0,0058	0,2625	0,3742
0,0265	100	-0,0360	0,0044	0,0078	-0,2396	-0,0161	-0,2909	0,3144	0,1376	0,1290	0,5817	-0,6193

* Valores em negrito destacam a variável de maior peso nos respectivos autovalores.

4 DISCUSSÃO

Estudos de diversidade baseados na análise de vários caracteres morfológicos, tomados de cada indivíduo ou acesso, podem ser realizados por meio do emprego de métodos multivariados. A confiabilidade dos resultados é maior quando os dados são obtidos a partir de um delineamento experimental com repetições. Entretanto, esta estratégia é muito difícil quando se trabalha com plantas perenes arbóreas, como no caso do cajueiro, e impraticável quando se avaliam indivíduos de populações naturais. Neste caso, a obtenção de medidas repetidas para uma mesma variável em cada acesso e, quando possível, a avaliação dos acessos em um mesmo ambiente, pode minimizar os erros aleatórios que incidem sobre os resultados.

No presente trabalho, a avaliação dos acessos foi realizada através da média de observações por variável, valor este considerado satisfatório a partir de um estudo preliminar de repetibilidade conduzido com as variáveis sob investigação. Além do mais, a maior parte dos acessos avaliados é cultivada em um mesmo local, que é o BAG do Cajueiro, o que reduz as diferenças decorrentes da variação ambiental. Adicionalmente, os dados são analisados por diferentes métodos multivariados, considerando diferentes arranjos, com o propósito de avaliar a consistência dos resultados obtidos.

4.1 Divergência entre ecótipos e/ou espécies de *Anacardium*

As análises discriminantes, baseadas em componentes principais e funções discriminantes de Anderson, evidenciaram a singularidade do grupo '*othonianum*' em relação aos grupos 'comum' e 'anão precoce'. Os acessos deste grupo foram coletados na região dos cerrados do Brasil Central. De acordo com Ferreira (1973), *A. othonianum* foi descrita como uma espécie distinta de *Anacardium* por Rizzini em 1969. Conhecida popularmente por cajuzinho de cerrado, esta espécie é caracterizada como uma árvore de porte mediano de 3 a 6 metros de altura, com tronco de 20 a 40 cm de diâmetro, de córtex fusco e fissurado. Entretanto, Mitchell e Mori (1987), consideraram esta e outras espécies ocorrentes na região do Brasil Central (i.e. *A. curatellaefolium*, *A. rondonianum*, *A. amilcarianum* e *A. kuhlmannianum*) como membros do ecótipo de *A. occidentale* circunscrito à região dos cerrados, sendo

caracterizado por apresentar folhas mais coriáceas, onduladas e subsésseis, assim como pseudofruto menor e mais ácido que o ecótipo típico das áreas de restingas. Por outro lado, a diferenciação entre os grupos 'comum' e 'anão precoce' foi bem menos pronunciada, especialmente se considerarmos a dispersão gráfica obtida pela análise discriminante baseada em componentes principais. Observa-se grande proximidade entre os centróides dos dois grupos, assim como elevada sobreposição na dispersão entre os acessos de cada grupo.

Entretanto, a utilização das funções discriminantes de Anderson na classificação dos acessos, segundo os grupos de alocação *a priori*, mostrou-se bastante eficiente, com nenhum erro de classificação associado ao grupo 'othonianum' e com a alocação correta de nove dos dez acessos de 'anão precoce' e 32 dos 39 acessos de 'comum'. A utilização destas funções na classificação dos acessos alocados *a posteriori* também produziu resultados satisfatórios. Quase todos os acessos coletados no Estado de Roraima apresentam porte reduzido, com as demais características típicas de *A. occidentale*, apesar de serem identificados apenas como *Anacardium* sp. Dessa forma, a alocação de metade destes acessos no grupo 'comum' e metade no grupo 'anão precoce' é plenamente justificável. Estes resultados, de certa forma, corroboram a hipótese de que o tipo 'anão precoce' teria sua origem nas savanas de Roraima (Almeida et al., 1993). Entre os acessos alocados no grupo 'othonianum', aquele identificado apenas como *Anacardium* sp, coletado no Estado do Piauí, apresenta muitas características fenotípicas próprias de *A. othonianum*, além das avaliadas neste trabalho, como tronco de casca muito fissurada e folhas coriáceas e subsésseis. O outro acesso, identificado como *A. microcarpum*, apresentou castanha e pseudofruto com dimensões semelhantes às de *A. othonianum*, mas suas características vegetativas como porte da planta e tipo foliar são mais associadas a *A. occidentale*. De fato, *A. microcarpum* é considerada, por Mitchell e Mori (1987), uma sinonímia de *A. occidentale* (ecótipo de restinga). Os demais acessos alocados no grupo 'comum' apresentam características típicas de *A. occidentale*. Embora um destes acessos esteja identificado como *A. othonianum*, todos os seus indivíduos indiscutivelmente são de *A. occidentale*

do tipo comum e, portanto, trata-se de um erro de identificação do acesso no BAG.

Portanto, as funções discriminantes de Anderson obtidas permitiram uma boa discriminação dos grupos pré-estabelecidos, assim como uma classificação adequada de novos acessos cujos grupos sejam desconhecidos, mas assumindo que eles pertençam a um destes grupos definidos.

4.2 Divergência segundo a origem geográfica de acessos de *Anacardium*

A análise discriminante baseada em componentes principais, considerando as seis populações, definidas *a priori*, evidenciou a maior divergência do grupo 'othonianum' em relação aos demais tanto pela posição mais distanciada de seu centróide como pela relativa facilidade de delimitação da dispersão de seus acessos. Situação semelhante também foi verificada para a população 'Rio de Janeiro'. Por outro lado, não é possível distinguir as demais populações entre si, considerando os padrões de dispersão que seus centróides e acessos apresentaram no plano bi-dimensional. Deve-se ressaltar que essa dispersão representa bem as interrelações entre os elementos investigados, uma vez que os dois primeiros componentes explicam mais de 95% da variação total, contida nos dados originais. Por isso, parece que o grau de diferenciação entre estas populações é baixo. Mesmo a população 'anão precoce' que havia se mostrado razoavelmente distinta quando confrontada com o grupo 'comum', na análise discriminante de ecótipos, agora apresenta seus acessos com um padrão de dispersão que não mais possibilita sua delimitação. Outro aspecto que deve ser destacado são as grandes distâncias relativas, apresentadas pelos acessos da população 'othonianum' comparadas com dispersão dos acessos da população 'Rio de Janeiro'. Estas distâncias refletem o grau de divergência intrapopulacional. No caso do grupo 'othonianum' esta divergência reflete, ainda que parcialmente, o grau de diferenciação de várias populações naturais da espécie ou ecótipo, uma vez que os acessos avaliados são de procedências diversas. Por outro lado, os acessos da população 'Rio de Janeiro' são todos de uma mesma população natural, cujos membros apresentam níveis variados de parentesco e, portanto, maior similaridade relativa.

As funções discriminantes de Anderson possibilitaram melhor distinção entre as populações em comparação com discriminação obtida por componentes principais. A taxa global de erro aparente foi de pouco mais de 20%, mas as populações 'othonianum' e 'Rio de Janeiro' novamente foram as que apresentaram os melhores níveis de discriminação de seus acessos. Embora nenhum acesso da 'Índia' tenha sido alocado incorretamente, vários acessos de outras populações foram classificados como pertencentes a este grupo. Por outro lado, o grupo 'Nordeste' teve muitos de seus acessos alocados em outros grupos, embora tenha recebido apenas dois acessos de outras populações. Estes resultados e a amplitude de dispersão de seus acessos, observada na Figura 04, apontam para a maior diversidade relativa do grupo 'Nordeste', confirmando a tese de que as zonas costeiras do Nordeste Brasileiro constituem um centro de diversidade de *A. occidentale* (Lima, 1986; apud Barros et al., 2002).

A utilização das funções discriminantes de Anderson na alocação de acessos "desconhecidos" gerou alguns resultados consistentes. Por exemplo, dos três acessos alocados no grupo 'othonianum' dois já haviam recebido esta mesma alocação quando foram classificados por meio das funções discriminantes de ecótipos (Tabela 1) e dos três acessos coletados na região Nordeste, mas não incluídos *a priori* por se encontrarem indeterminados ao nível de espécie, dois foram alocados no grupo 'Nordeste'. Por outro lado, resultados aparentemente contraditórios foram obtidos na alocação de acessos oriundos da Venezuela, Pará e São Paulo. Nestes casos, acessos de um mesmo local de origem foram alocados em grupos diferentes.

4.3 Análises de agrupamento

O padrão de agrupamento das populações observado na Figura 5, parece representar bem o grau de similaridade esperado entre as populações, considerando as evidências históricas e biológicas disponíveis. As populações com o maior grau de similaridade, 'Nordeste' e 'anão precoce', embora distintas possuem a mesma origem geográfica, que é a região Nordeste do Brasil, principalmente o Estado do Ceará. O terceiro grupo mais similar foi a 'Índia' que é o geograficamente mais distante, mas onde *A. occidentale* é uma

espécie introduzida pelos portugueses a partir da segunda metade do século XVI, provavelmente a partir de sementes coletadas no nordeste brasileiro (Jonhson, 1974; Ferrão, 1995). A seguir, são agrupadas as populações de 'Roraima' e do 'Rio de Janeiro', localizações onde *A. occidentale* é considerada uma espécie indígena ou nativa, ou seja, de ocorrência anterior à colonização européia da América (Mitchell e Mori, 1987). O conjunto destas populações difere substancialmente da população 'othonianum', que consiste uma espécie distinta ou, de acordo Mitchell e Mori, (1987), em ecótipo distinto de *A. occidentale*, associado à região dos cerrados.

A análise de diversidade, por componentes principais, entre os 92 acessos, mostrou que as variáveis de maior peso no primeiro componente principal (Tabela 6) são aquelas relativas ao pedúnculo e à castanha, portanto, este componente exprime um fator relativo às dimensões do pedúnculo e da castanha, enquanto as variáveis relativas à inflorescência foram mais importantes no segundo componente que, neste caso, exprime um fator relativo às dimensões da inflorescência.

No que se refere aos agrupamentos apresentados na Figura 6, a discordância entre os dois critérios utilizados é evidente. Os acessos do grupo 1 (NE) foram distribuídos em quatro grupos segundo o método de Tocher. Por outro lado, a maioria dos acessos deste grupo e dos grupos 2 (AP), 3 (RJ), 4 (RR) e 5 (ID) formaram um único grupo (grupo 1 de Tocher). Já os acessos do grupo 6 (OT) foram distribuídos em três grupos. Dois destes grupos foram formados por cinco indivíduos cada, sendo um constituído exclusivamente por membros de *A. othoianum*, e outro por membros das populações 3, 4 e 6. Os dois acessos restantes foram alocados no grupo 1 de Tocher. Apesar de todas estas discrepâncias, a tendência de agrupamento entre os acessos da população 'othonianum' e da população 'Rio de Janeiro' ainda é percebida na Figura 6.

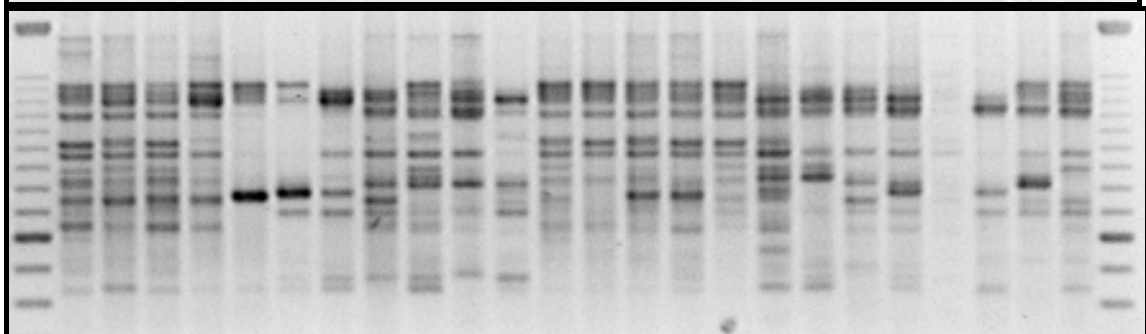
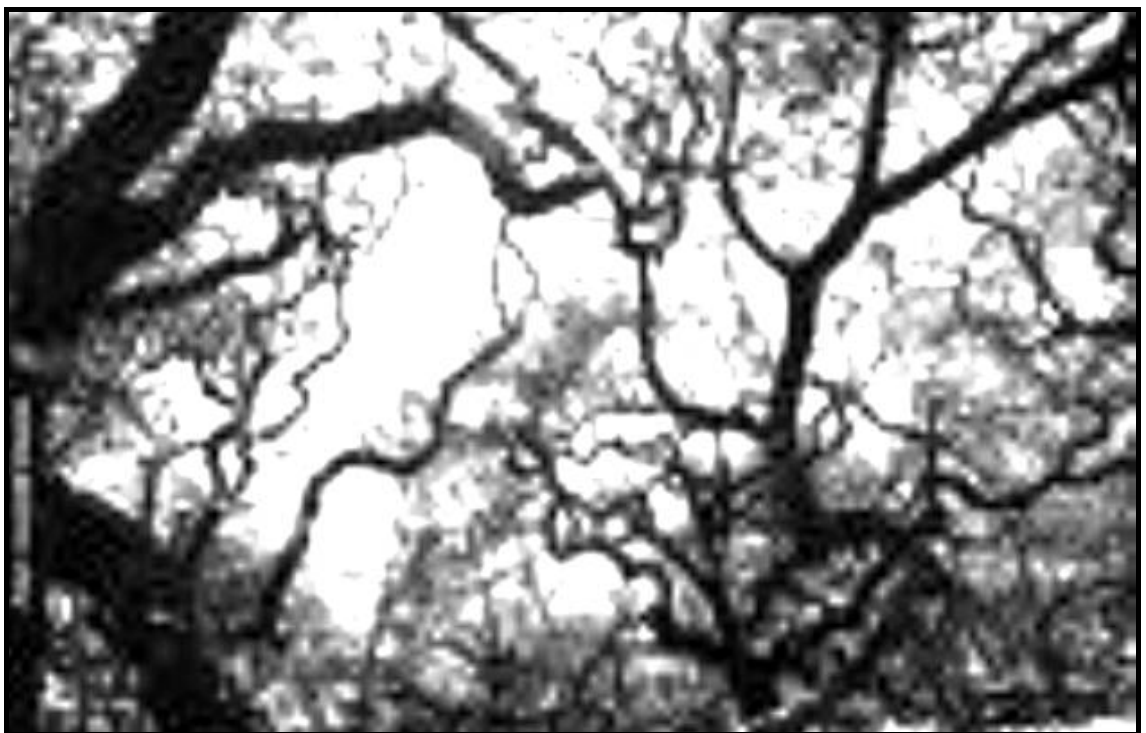
5 CONCLUSÕES

- a) Análises discriminantes, aplicadas a variáveis morfométricas, são ferramentas eficientes na caracterização de grupos de *Anacardium*, sejam eles espécies, ecótipos ou populações de origens diferentes.

- b) A utilização de funções discriminantes de Anderson, em comparação com a técnica baseada em componentes principais, possibilitou melhor discriminação dos grupos, além de permitir a classificação de acessos desconhecidos ou cujas informações foram consideradas duvidosas.
- c) Foi confirmado que a região Nordeste é um centro de diversidade de *A. occidentale*.
- d) A alocação da metade dos acessos de Roraima no grupo 'anão' é mais uma evidência da origem amazônica desse ecótipo ou variedade, destacando a necessidade da realização de novas coletas na região, com o propósito de ampliar a base genética deste grupo.
- e) *A. othonianum*, apesar de ser considerado por alguns autores apenas um ecótipo de *A. occidentale*, constituiu-se em um grupo morfológicamente muito distinto dos demais. Considerando que divergência foi observada entre indivíduos cultivados em um mesmo ambiente, ou seja, o Banco Ativo de Germoplasma do Cajueiro, pode-se assumir que ela é majoritariamente de natureza genética.
- f) O padrão de agrupamento exibido pelas populações analisadas refletiu adequadamente o grau de relacionamento esperado entre elas, considerando as evidências biológicas e históricas disponíveis.
- g) Por outro lado, o agrupamento dos acessos pelo método de otimização de Tocher produziu resultados muito diferente dos grupos estabelecidos *a priori*, nas análises discriminantes. Possivelmente, a maior proporção da divergência acessada não está relacionada com a origem geográfica dos acessos.

CAPÍTULO 2

ANÁLISE DA DIVERSIDADE EM *Anacardium* spp. BASEADA EM MARCADORES MOLECULARES ISSR



1 INTRODUÇÃO

A utilização de marcadores moleculares de diversos tipos tornou possível a identificação e caracterização de germoplasma, a construção de mapas genéticos e a obtenção de estimativa da distância genética entre indivíduos e/ou populações de várias espécies vegetais. Tais tecnologias podem permitir uma aceleração de ganhos genéticos para plantas de interesse, especialmente em culturas com informações genéticas limitadas, como é caso da maioria das espécies arbóreas tropicais (Federezzi, 1998; Ferreira e Grattapaglia, 1998; Picoli et al., 2001). De modo semelhante, as informações de diversidade e distância genética têm sido avaliadas para a identificação das melhores combinações híbridas e na organização de bancos de germoplasma; indicando redundâncias e deficiências das coleções, além de gerar dados sobre a eficiência do processo de coleta, manutenção e ampliação destes recursos genéticos (Barbosa-Neto e Bered, 1998; Ferreira e Grattapaglia, 1998).

O desenvolvimento de marcadores bioquímicos e moleculares proporcionou um salto qualitativo e quantitativo em estudos da estrutura populacional e do sistema reprodutivo de diversas espécies. Vários parâmetros populacionais, como grau de endogamia e sistema reprodutivo predominante, entre outros, podem ser obtidos e são de grande importância na determinação de estratégias de conservação da variabilidade genética na natureza, bem como para sua melhor utilização em programas de melhoramento genético das espécies domesticadas (Carlini-Garcia et al., 2001; Reif et al., 2005).

Barbosa-Neto e Bered (1998) argumentam que não há sérias restrições quanto aos marcadores moleculares preferenciais para utilização em estudos de diversidade genética. De modo geral, os marcadores devem ser confiáveis em termos de repetibilidade, ter baixo custo e serem fáceis e rápidos de analisar. Neste sentido, os marcadores baseados em técnicas de PCR (reação em cadeia da polimerase) são mais atrativos, principalmente aqueles que apresentam grande polimorfismo como microssatélites e AFLP.

Em um trabalho de compilação de dados, de 307 estudos de avaliação de diversidade inter e intrapopulacional de espécies de plantas selvagens, baseados em marcadores de DNA nuclear, Nybom (2004) observou que as

estimativas de diversidade derivadas de marcadores de herança dominante (RAPD, AFLP e ISSR) são muito similares e podem ser comparáveis diretamente. Por outro lado, marcadores de herança co-dominante (microsatélites) produziram valores de diversidade intrapopulacional quase três vezes superiores, enquanto que as estimativas de diversidade interpopulacionais foram similares àquelas obtidas pelos marcadores de herança dominante. A distância geográfica máxima entre populações amostradas, produziu acentuado efeito positivo na diversidade interpopulacional e, para espécies perenes, alógamas e/ou de sucessão tardia, a maior parte da variabilidade é de natureza intrapopulacional. A autora também observou que o número de plantas amostradas por população e o número de alelos de microsatélite detectados foram correlacionados com alguns parâmetros genéticos populacionais.

Os marcadores ISSR (Zietkiewicz et al. 1994; Gupta, et al., 1994; Wu et al. 1994) são produtos de uma técnica baseada no método de amplificação de DNA pela reação em cadeia da polimerase (PCR) (Reddy et al., 2002) e, basicamente, envolve a amplificação de regiões entre seqüências de microsatélites adjacentes, inversamente orientadas. Nenhum conhecimento *a priori* do genoma é necessário e *primers* baseados numa seqüência repetitiva, tal como $(CA)_n$, podem ser desenhados considerando apenas o motivo existente na própria seqüência tais como $(CA)_8$ ou $(CA)_9$ (Gupta et al., 1994; Bornet e Branchard, 2001) ou com uma ancoragem, formada por um pequeno segmento de bases arbitrárias e/ou degeneradas, localizado em uma das extremidades (3' ou 5') da seqüência repetitiva, tais como $(CA)_8RG$ e $(CA)_8YT$ ou $BDB(CA)_7$ (onde R=purinas, Y=pirimidinas, B=C, G ou T e D=A, G, ou T) (Zietkiewicz et al., 1994; Godwin et al., 1997, Fang et al., 1997, Reddy et al., 2002).

De acordo com Reddy et al. (2002), as fontes de variabilidade e o nível de polimorfismo obtido pelos marcadores ISSR são decorrentes dos seguintes fatores: a) elevada taxa de mutação das seqüências microsatélites, alvos do anelamento do *primer* empregado; b) natureza do *primer* utilizado, isto é, seqüência repetitiva ou motivo, existência ou não de região de ancoragem, assim como localização (5' ou 3'), extensão e composição da seqüência de ancoragem; d) método de resolução utilizado, isto é, gel desnaturante de

poliacrilamida (PAGE) em combinação com marcação radiativa ou coloração com nitrato de prata ou gel de agarose, corado com brometo de etídio.

A utilização de marcadores moleculares na caracterização de recursos genéticos de cajueiro ainda é bastante restrita, se limitando a um pequeno número de trabalhos conduzidos em coleções de bancos de germoplasma da Tanzânia e da Índia. O primeiro destes estudos foi realizado por Mneney et al. (2001), que avaliaram a divergência genética, baseada em marcadores RAPD, de parte da coleção do banco de germoplasma do Cashew Development Centre (Tanzânia), constituído por acessos provenientes da Tanzânia, Sri Lanka, Moçambique, Índia, Brasil e Cook Islands. Os autores verificaram que os acessos brasileiros formaram um grupo separado e mais distinto em relação aos das demais procedências.

Dhanaraj et al (2002) estimaram a diversidade genética entre 90 acessos do Banco Nacional de Genética do Cajueiro da Índia, também utilizando marcadores RAPD. De acordo com os autores, este estudo demonstrou que diversidade do banco indiano de germoplasma de caju pode ser considerada de moderada a alta. Os resultados também permitiram estabelecer uma coleção nuclear (core collection) menor, com representação da mesma diversidade da população inteira.

Archak et al. (2003a) compararam a eficiência e utilidade de três tipos marcadores moleculares (RAPD, ISSR e AFLP) e de descritores morfológicos na medida de variação e discriminação genotípica de germoplasma de caju. Os autores verificaram que, a partir de cada um dos três marcadores moleculares utilizados, foi possível discriminar os 19 acessos avaliados, com maior eficiência para os marcadores AFLPs. Por outro lado, não foi observada a ocorrência de correlação entre o conjunto de 27 descritores morfométricos utilizados e os marcadores moleculares. Apenas os marcadores RAPD e ISSR apresentaram correlação significativa ($r=0,63$; $p<0,01$). Archak et al. (2003b) também utilizaram marcadores RAPD e ISSR na caracterização fingerprinting de 35 variedades elites, sendo 24 seleções e 11 híbridos, desenvolvidas por diferentes centros de pesquisa da Índia. A partir de 94 locos amplificados foi possível diferenciar, com alto grau de certeza, todos os genótipos. Por lado, os resultados evidenciaram que a base genética entre eles é muito estreita, apesar da grande amplitude geográfica entre os locais de obtenção dos

cultivares, fato atribuído ao intenso intercâmbio de material genético entre os centros de pesquisa e ao número reduzido de parentais empregados nos programas de produção de híbridos.

Este trabalho teve por finalidade avaliar a diversidade genética de *Anacardium* spp. a partir de acessos mantidos pelo Banco Ativo de Germoplasma do Cajueiro (BAG – Embrapa CNPAT) e de indivíduos amostrados em uma população subespontânea de cajueiro (*Anacardium occidentale* L.), localizada no estado do Rio de Janeiro, utilizando marcadores moleculares ISSR. Especificamente, os objetivos foram:

- Caracterizar e quantificar a variabilidade genética existente entre acessos e dentro de populações de cajueiro, através de marcadores moleculares ISSR;
- Estimar as distâncias genéticas entre as populações e avaliar suas relações com a origem geográfica;
- Estabelecer relações fenéticas entre as espécies e/ou ecótipos estudados;
- Comparar diferentes estratégias de análises multivariadas aplicadas a dados moleculares binários.

2 MATERIAL E MÉTODOS

2.1 Material vegetal

As informações utilizadas para este estudo foram obtidas de 91 acessos do Banco Ativo de Germoplasma do cajueiro e de 45 indivíduos de uma população subespontânea, de uma área de restinga do litoral norte do estado do Rio de Janeiro. Cada acesso foi representado por um indivíduo apenas, perfazendo um total de 136 genótipos analisados, os quais serão referidos daqui em diante simplesmente como acessos, independentemente da origem. No Quadro 1 estão descritos a procedência, espécie/ecótipo e local de coleta dos acessos do BAG do cajueiro.

O material foi coletado no final da fase reprodutiva, durante a emissão de folhagem nova, nos meses de novembro e dezembro de 2004, para os acessos do Rio de Janeiro e novembro de 2005, para os acessos do BAG.

A coleta consistiu de folhas jovens completamente expandidas, que foram identificadas e envolvidas em papel de filtro umedecido, acondicionadas em sacos plásticos e mantidas resfriadas até a chegada ao Laboratório de Genética Molecular de Plantas (Biomol), do Instituto de Biotecnologia Aplicada à Agropecuária (Bioagro) da UFV.

No Biomol, as folhas foram submetidas à extração imediata do DNA total, sendo mantidas por um período máximo de cinco dias sob refrigeração a 4°C. Alternativamente, parte das amostras foi armazenada à temperatura de 80°C negativos, até o momento apropriado para a operação de extração do DNA.

2.2 Extração do DNA

Diversos protocolos de extração de DNA, baseados no método CTAB, foram testados com o objetivo de obter DNA em quantidade e qualidade adequados para reações de PCR. O protocolo adotado foi baseado, com modificações, no ajustado por Cavalcanti (2004) para *A. occidentale*, o qual é descrito a seguir.

Quadro 1 – Informações das coleções e dos acessos do Banco Ativo de Germoplasma do Cajueiro empregados na análise de diversidade

Nome da coleção, procedência e data da introdução	Número da planta no BAG	Espécie	Local da Coleta
Coleção C-01, Índia, Introdução por sementes em 1973	08	<i>A. occidentale</i>	-
	37	<i>A. occidentale</i>	-
	70	<i>A. occidentale</i>	-
	72	<i>A. occidentale</i>	-
Coleção C-02, “Cerrado” Goiás, Mato Grosso, Pará e Piauí. Introdução por sementes em 1975	29	<i>A. othonianum</i>	Campos Belos (GO)
	67	<i>A. humile</i>	Cavalcanti (GO)
	100	<i>A. othonianum</i>	São Domingos (GO)
	109	<i>A. humile</i> (?)	Cavalcanti (GO)
	122	<i>A. humile</i>	São João d’Aliança (GO)
	131	<i>A. humile</i>	São João d’Aliança (GO)
	138	<i>A. othonianum</i>	Arraias (GO)
	148	<i>A. othonianum</i>	Galheiros (GO)
	151	<i>A. othonianum</i>	Arraias (GO)
	163	<i>A. humile</i>	Alto Paraíso (GO)
	197	<i>Anacardium</i> sp.	Cristalina (PI)
	205	<i>A. occidentale</i>	Itaituba (PA)
	209	<i>Anacardium</i> sp.	Eliseu Martins (PI)
	219	<i>A. othonianum</i>	Cavalcanti (GO)
	226	<i>A. othonianum</i>	Barra do Garças (MT)
	229	<i>A. humile</i>	Jaciara (GO)
	241	<i>A. occidentale</i>	Itaituba (PA)
245	<i>A. occidentale</i>	Itaituba (PA)	
281	<i>A. othonianum</i> (?)	Água Boa (MT)	
Coleção C-03 Venezuela, Índia e Valinhos Introdução por sementes em 1973	05	<i>A. occidentale</i>	Venezuela - El Tigre
	09	<i>A. occidentale</i>	Venezuela – El Tigre
	10	<i>A. occidentale</i>	Índia - Ullal
	12	<i>A. occidentale</i>	Índia - Ullal
	16	<i>A. occidentale</i>	Índia - Ullal
	19	<i>A. occidentale</i>	Índia - Ullal
	25	<i>A. occidentale</i>	Valinhos (SP)
	30	<i>A. occidentale</i>	Valinhos (SP)
Coleção C-04 Roraima Introdução por sementes em 1983	01	<i>Anacardium</i> sp.	Rio Tacutu / Faz. Caju
	06	<i>Anacardium</i> sp.	Rio Tacutu / Faz. Caju
	18	<i>Anacardium</i> sp.	Est. Tabalascada - Cantá
	19	<i>Anacardium</i> sp.	Est. Normandia Km 25
	24	<i>Anacardium</i> sp.	Est. Normandia Km 25
	30	<i>Anacardium</i> sp.	Maloca do Sucuba
	38	<i>Anacardium</i> sp.	Est. Normandia Km 25
	46	<i>Anacardium</i> sp.	Est. Normandia Km 25
	51	<i>Anacardium</i> sp.	Est. Taiano km 20
	54	<i>Anacardium</i> sp.	Rio Tacutu / Faz. Caju
Coleção C-05 Paraíba	03	<i>A. occidentale</i>	-
	22	<i>A. occidentale</i>	-

Introdução por sementes em 1990	29	<i>A. occidentale</i>	-
	40	<i>A. occidentale</i>	-
	53	<i>A. occidentale</i>	-
Coleção C-06 Ceará Introdução por enxertia em 1979	11	<i>A. occidentale</i>	Pacajus
	28	<i>A. occidentale</i>	Aracati
	34	<i>A. occidentale</i>	Trairi
	47	<i>A. occidentale</i>	Trairi
	55	<i>A. occidentale</i>	Aracati
	65	<i>A. occidentale</i>	Aracati
	75	<i>A. occidentale</i>	Trairi
	78	<i>A. occidentale</i>	Aracati
	102	<i>A. occidentale</i>	Aracati
	118	<i>A. occidentale</i>	Trairi
	174	<i>A. occidentale</i>	Cascavel
	184	<i>A. occidentale</i>	Russas
	187	<i>A. occidentale</i>	Russas
	195	<i>A. occidentale</i>	Aracoiaba
	196	<i>A. occidentale</i>	Aracoiaba
	223	<i>A. occidentale</i>	Russas
236	<i>A. occidentale</i>	-	
261	<i>A. occidentale</i>	Russas	
275	<i>A. occidentale</i>	Aracoiaba	
C-07 Camocin -1979	01	<i>Anacardium</i> sp.	Camocin (CE)
<i>A. microcarpum</i> sementes - 1964	-	<i>A. microcarpum</i>	Ceará
	-	<i>A. microcarpum</i>	Ceará
Coleção C-09 Goiás Introdução por sementes em 1975	08	<i>A. othonianum</i>	-
	14	<i>A. othonianum</i>	-
	18	<i>A. othonianum</i>	-
	24	<i>A. occidentale</i>	-
	30	<i>A. occidentale</i>	-
	38	<i>A. humile</i>	-
Coleção Pacajus – CE Introdução por sementes entre 1956 e 1959	CP12	<i>A. occidentale</i>	Pacajus
	64	<i>A. occidentale</i>	Pacajus
	63 (CP77)	<i>A. occidentale</i>	Pacajus
	67 (CP137)	<i>A. occidentale</i>	Pacajus
	09 (CP138)	<i>A. occidentale</i>	Pacajus
	54 (CP96)	<i>A. occidentale</i>	Pacajus
	CP 136	<i>A. occidentale</i>	Pacajus
Coleção Maranguape – CE Introdução por sementes em 1956	14	<i>A. occidentale</i>	Maranguape
	15	<i>A. occidentale</i>	Maranguape
	16 (CP76)	<i>A. occidentale</i>	Maranguape
	22 (CP06)	<i>A. occidentale</i>	Maranguape
	23 (CP09)	<i>A. occidentale</i>	Maranguape
	25(CP1001)	<i>A. occidentale</i>	Maranguape
	28	<i>A. occidentale</i>	Maranguape
	29	<i>A. occidentale</i>	Maranguape
	30	<i>A. occidentale</i>	Maranguape
35	<i>A. occidentale</i>	Maranguape	

Folhas frescas ou congeladas foram maceradas em nitrogênio líquido, utilizando almofariz de porcelana. Aproximadamente 50 mg de tecido pulverizado foi transferido para um tubo Eppendorf de 2 mL, previamente identificado e também congelado em nitrogênio líquido. Em seguida, foi adicionado 1 mL do tampão de extração, pré-aquecido a 65°C, composto da seguinte mistura: 100 mM de Tris-HCl pH 8,0, 20 mM de EDTA (*ethylenediaminetetracetate*), 1,4 M de NaCl, 2% de CTAB (*Cationic Hexadecyl Trimethyl Ammonium Bromide*), 2% de PVP 40 (Polivinilpirrolidona) e 0,1% de β-Mercaptoetanol. O material foi homogeneizado em vortex e incubado em banho-maria por aproximadamente cinco minutos, com uma a duas inversões dos tubos durante este período. Depois de retirados do banho-maria, os tubos foram mantidos em repouso à temperatura ambiente por cerca de 10 minutos. Em seguida adicionou-se 900 µL de clorofórmio/álcool isoamílico (24:1) em cada tubo e procedeu-se a mistura das soluções por inversões suaves durante dois a três minutos.

Posteriormente, as amostras foram centrifugadas em microcentrífuga Eppendorf, modelo 5415C, a 13.500 rpm por 10 minutos e 800 µL da fase aquosa superior foi transferido para um novo tubo de 1,5 mL. O DNA foi precipitado pela adição de 600 µL de isopropanol gelado, seguido da inversão suave dos tubos e seu armazenamento a 20°C por duas horas, no mínimo. O precipitado foi recuperado através de nova centrifugação dos tubos a 13.600 rpm durante cinco minutos. Depois do descarte do sobrenadante, o precipitado e as paredes internas foram lavados com 500 µL etanol 95%, através da inversão cuidadosa, por duas vezes, de cada tubo. Para evitar a perda do precipitado procedeu-se a uma nova centrifugação de 13.600 rpm durante um minuto. A solução de lavagem foi descartada e os precipitados foram secos à temperatura ambiente por uma hora, aproximadamente.

Os precipitados foram resuspenso em 30 µL de TE (10 mM de Tris-HCl pH 8,0; 1 mM de EDTA pH 8,0) contendo RNase A na concentração final igual a 30 µg/mL e incubados a 37°C por 30 minutos, com agitações leves até o desaparecimento completo de qualquer partícula sólida na suspensão.

2.3 Quantificação e qualidade do DNA

A avaliação da concentração e da qualidade do DNA das amostras foi realizada em géis de agarose, na concentração de 0,8% (p/v), corados com 1 µL/50mL de solução de brometo de etídio (10 mg/mL). Alíquotas de cada amostra de DNA foram aplicadas no gel ao lado de concentrações conhecidas de DNA do fago λ (10, 20, 30, 50 e 100 ng). As bandas de DNA foram visualizadas sob luz ultravioleta e fotodocumentadas pelo sistema de captura de imagens 'Eagle Eye' (Stratagene). Amostras de trabalho foram diluídas para a concentração de 10 ng/µL. Tanto o DNA estoque quanto as amostras de trabalho foram mantidas a -20°C.

2.4 Seleção de *primers*

Testes preliminares de amplificação foram conduzidos com 11 *primers* não ancorados, utilizados por Archak et al. (2003a) na análise de diversidade de germoplasma de cajueiro, e 72 *primers* ancorados desenvolvidos pelo Laboratório de Biotecnologia da Universidade da Columbia Britânica (coleção nº 9). Um total de cinco *primers* não ancorados e 16 ancorados (Tabela 1), que produziram melhores amplificações em termos de quantidade e nitidez de bandas, foram pré-selecionados para determinação da melhor temperatura de anelamento e nível de polimorfismo produzidos. A amplitude de temperatura de anelamento testada foi entre 45 e 60°C, utilizando termociclador de gradiente modelo 'Robocycler Gradient 96' (Stratagene). Apenas nove *primers* (6 ancorados e 3 não ancorados) foram utilizados no trabalho final de amplificação de todas as amostras.

2.5 Otimização das condições de amplificação

Diversas condições de amplificação relacionadas na literatura foram testadas, como diferentes concentrações de *primer* e DNA molde, número de ciclos e duração de cada etapa dos ciclos, assim como diferentes concentrações de reagentes.

Tabela 1 – Lista de *primers* de marcadores ISSR avaliados em ampliações de DNA genômico de *Anacardium* spp.

<i>Primers</i> não ancorados	Seqüência completa (5'-3')	Nº. de nucleotídeos	T _A (°C)
*(GACA) ₄	GACAGACAGACAGACA	16	45-60
*(GAAGTGGG) ₂	GAAGTGGGGAAGTGGG	16	50-60
(AGG) ₆	AGGAGGAGGAGGAGGAGG	18	49-60
*(GTG) ₆	GTGGTGGTGGTGGTGGTG	18	50-60
(ACTG) ₄	ACTGACTGACTGACTG	16	45-60
<i>Primers</i> UBC (ancorados)	Seqüência completa (5'-3')	Nº. de nucleotídeos	T _A (°C)
P808 - (AG) ₈ C	AGA GAG AGA GAG AGA GC	17	49-55
P810 - (GA) ₈ T	GAG AGA GAG AGA GAG AT	17	49-55
*P812 - (GA) ₈ A	GAG AGA GAG AGA GAG AA	17	49-55
*P840 - (GA) ₈ YT	GAG AGA GAG AGA GAG AYT	18	49-55
*P841 - (GA) ₈ YC	GAG AGA GAG AGA GAG AYC	18	49-55
*P842 - (GA) ₈ YG	GAG AGA GAG AGA GAG AYG	18	49-55
*P834 - (AG) ₈ YT	AGA GAG AGA GAG AGA GYT	18	49-55
P835 - (AG) ₈ YC	AGA GAG AGA GAG AGA GYC	18	49-55
P836 - (AG) ₈ YA	AGA GAG AGA GAG AGA GYA	18	49-55
P827 - (AC) ₈ G	ACA CAC ACA CAC ACA CG	18	49-55
P855 - (AC) ₈ YT	ACA CAC ACA CAC ACA CYT	18	49-55
P884 - HBH(AG) ₇	HBH AGA GAG AGA GAG AG	17	49-55
P885 - BHB(GA) ₇	BHB GAG AGA GAG AGA GA	17	49-55
*P886 - VDV(CT) ₇	VDV CTC TCT CTC TCT CT	17	49-55
P888 - BDB (CA) ₇	DVD TCT CTC TCT CTC TC	17	49-55
P889 - BDB (AC) ₇	DBD ACA CAC ACA CAC AC	17	49-55

**Primers* efetivamente utilizados no trabalho

T_A = Intervalo de temperaturas de anelamento testado

Y = (C ou T); B = (C, G ou T); D = (A, G ou T); H = (A, C, ou T); V = (A, C ou G)

As concentrações adotadas para um volume final de reação de 15 µl, consistiram em: 10 mM de Tris-HCl pH 8,3; 50 mM de KCl; 1,5 mM de MgCl₂; 0,2 mM de cada dNTP; 0,5 µM de *primer*; 15 ng de DNA molde; 0,6 unidades de Taq polimerase e 0,5% de BSA (albumina de soro bovino). No caso dos *primers* não ancorados, a concentração utilizada foi de 1 µM de *primer*, mantendo-se inalteradas as concentrações dos demais reagentes. Algumas reações foram realizadas com um volume final de 20 µl, neste caso foram utilizados 20 ng DNA molde e 1 unidade de Taq polimerase, ajustando-se, proporcionalmente, as concentrações dos demais reagentes para este volume.

A reação de PCR consistiu de uma fase inicial de desnaturação a 94°C por três minutos, seguida por 35 ciclos de desnaturação (94°C/1 minuto), anelamento (1 minuto, com temperatura dependente do *primer*) e extensão

(72°C/2 minutos) e uma fase de extensão final de 72°C por sete minutos. Todas as reações foram realizadas em termocicladores modelo 'GeneAmp PCR System 9600' (Applied Biosystem 9600).

2.6 Resolução e visualização dos fragmentos amplificados

Os produtos das reações de PCR foram separados e visualizados utilizando um de dois sistemas distintos: gel desnaturante de poliacrilamida, corado com nitrato de prata ou gel de agarose, corado com brometo de etídio.

2.6.1 Sistema em gel desnaturante de poliacrilamida

Os géis desnaturantes de poliacrilamida foram confeccionados de acordo com as recomendações de Creste et al. (2001). A solução estoque de poliacrilamida a 30% (29:1) foi preparada pela adição de 290 g de acrilamida e 30 g de bisacrilamida em água ultrapura, com um volume final de 1 litro. Na confecção de 500 mL de solução matriz de poliacrilamida a 6% e 7 M de uréia, foram utilizados 100 mL da solução estoque de poliacrilamida a 30%, 210 g de uréia ultra pura, 100 mL de TBE 5X (445 mM de Tris, 445 mM de ácido bórico, 10 mM de EDTA) e água ultrapura até completar o volume final de 500 mL. A solução foi filtrada em papel de filtro e transferida para frasco âmbar, envolvido por papel alumínio e armazenado em geladeira.

Para montagem do gel foram utilizadas placas de vidro temperado de 6 mm de espessura. As placas foram lavadas com detergente neutro a 2% e enxaguadas em água corrente de torneira e água destilada e deixadas na posição vertical para secagem. Posteriormente, cada placa foi limpa por três vezes com etanol a 95% esfregando-se lenços de papel, com movimentos horizontais e verticais. Depois de seca à temperatura ambiente, a placa maior foi tratada com uma solução composta de 1,1 mL de etanol a 95% e ácido acético glacial a 0,5% mais 2,2 µl de Bind Silane. Antes da aplicação, os dois reagentes foram colocados em um tubo Eppendorf e homogeneizados em um vortex. O produto foi aplicado à placa e depois de 10 minutos o excesso foi removido com um lenço de papel umedecido em álcool a 95%. A placa menor foi tratada com 200 µl, aproximadamente, de uma solução anti-chuva utilizada

em vidros de carros. Após a secagem por 10 minutos o excesso foi removido pela limpeza da placa, por três vezes, com etanol a 95% e lenços de papel. As placas foram montadas utilizando espaçadores de 0,5 mm e vaselina como agente de vedação.

O gel foi preparado utilizando 80 mL de solução matriz, 400 µl de persulfato de amônio a 10% e 168 µl de TEMED (Promega). A solução foi homogeneizada em um béquer e vertida diretamente no aparato das placas, posicionado com uma inclinação de 30°, aproximadamente. Em seguida, o pente com 60 poços foi posicionado e o aparato de placas foi mantido na posição horizontal por uma hora para permitir a polimerização do gel. Após a polimerização, as placas foram montadas na cuba de eletroforese e os poços foram limpos com auxílio de uma pipeta automática.

Antes da aplicação das amostras, o gel foi submetido a uma pré-corrída, sob potência constante de 50W, por uma hora ou até a placa atingir uma temperatura de 50°C. Na parte superior da cuba foi utilizado TBE 1X (89 mM de Tris, 89 mM de ácido bórico, 2 mM de EDTA) e na parte inferior TBE 1X e acetato de sódio a 0,375 M (Resende, 2004). Antes da aplicação das amostras procedeu-se a uma nova limpeza dos poços, através da injeção de tampão de corrida com auxílio de uma pipeta automática.

À reação de PCR com volume igual a 15 µl foi adicionado 6 µl tampão de carregamento (98% de formamida, 10mM EDTA, azul de bromofenol 1mg/mL e xileno cianol 1mg/mL). Em seguida, as amostras foram desnaturadas a 94°C por 5 minutos e transferidas para um recipiente com gelo, uma alíquota de 8 µl foi aplicada em cada poço do gel. Em todos os géis também foram aplicados 5 µl de marcador 100 pb (Invitrogen), preparado na concentração 25 ng/µl em solução tampão de carregamento e água ultrapura (1:2,33). As alíquotas do marcador foram submetidas às mesmas condições de desnaturação dos produtos de PCR. Após aplicação de todas as amostras, o sistema foi submetido à eletroforese sob potência constante de 50 W por 3 horas, aproximadamente, ou até a chegada do xileno cianol na borda inferior do gel.

A coloração dos géis foi realizada com nitrato de prata, utilizando o kit *Silver SequenceTM DNA staining reagents* (Q 4132) da Promega. Após a

separação das placas, o gel (aderido na placa maior) foi imerso em dois litros de solução fixadora gelada (ácido acético glacial a 10%, em água ultra pura) por 40 minutos, sob agitação orbital leve. Em seguida a solução fixadora foi removida e novamente armazenada em geladeira. O gel foi lavado três vezes, por três minutos cada vez, em dois litros de água destilada, sempre com o auxílio de agitador orbital. Após as lavagens o gel foi imerso em dois litros solução de nitrato de prata 1 g/l e formaldeído 37% 1,5mL/l, mantido sob agitação por 20 minutos. Após a remoção da solução de nitrato de prata, procedeu-se a uma lavagem em água destilada por 10 segundos. A revelação foi realizada pela imersão em um litro de solução gelada contendo 3% (p/v) de carbonato de cálcio anidro, 1,5ml/l de formaldeído a 37% e 400 µl/l de tiosulfato de sódio a 10%. A placa foi agitada manualmente até o aparecimento das primeiras bandas, acrescentando-se mais um litro da solução reveladora e continuando a agitação até se atingir o padrão de coloração desejado. Em seguida, um litro da solução fixadora foi adicionado para bloquear o processo de revelação. A placa foi novamente lavada por duas vezes com água destilada e armazenada na posição vertical em local arejado para secagem do gel.

Após a secagem os géis foram fotodocumentados com auxílio de um scanner. A anotação dos dados foi realizada a partir do exame direto do gel sobre transiluminador de luz branca, auxiliada pela imagem digital obtida.

2.6.2 Sistema em gel de agarose

Neste sistema os géis foram confeccionados com 1,5% (p/v) de agarose ultra pura (Invitrogen) em tampão TBE 1X. Os produtos de reação de PCR (volume de 20µl) foram misturados com 3 µl do corante tipo IV (Sambrook et al., 1989) e o volume resultante foi aplicado a cada um dos poços. Em seguida, o sistema foi submetido à eletroforese sob voltagem constante de 4V/cm, por um período aproximado de 4 horas. Em todos os géis foram aplicados 4 µl de marcador de 100 pares de bases (Invitrogen), na concentração de 0,1µg/µl. Após a corrida, os géis foram imersos em uma solução de brometo de etídio (0,6 ng/ml) e mantidos sob agitação leve, em agitador orbital, por 30 minutos.

Após a retirada da solução de brometo, os géis foram enxaguados em água de torneira, visualizados e fotodocumentados sob luz ultravioleta pelo sistema de captura de imagens 'Eagle Eye' (Stratagene).

2.7 Análise dos dados

A interpretação dos padrões de bandas exibidos pelos géis, para ambos os sistemas de resolução, levou em consideração o princípio em que bandas geradas por um mesmo *primer* e que ocupam a mesma posição relativa referem-se à amplificação do mesmo fragmento de DNA, ou seja, pertencem ao mesmo loco gênico, enquanto bandas ocupando posições relativas diferentes são de locos distintos. Dessa forma, foi construída uma matriz de dados envolvendo todos os acessos, atribuindo-se valor igual a 1, se banda homóloga estivesse presente e 0, caso contrário.

A não ser quando explicitamente mencionado, todas as análises estatísticas foram realizadas com auxílio do aplicativo computacional GENES (Cruz, 2001, 2006a, 2006b).

2.7.1 Análises preliminares e descritivas

A matriz binária de dados, obtida a partir da leitura dos géis, foi utilizada para o cálculo do coeficiente de similaridade de Sorenso ou Nei e Li entre pares de acessos, obtido pela expressão (Cruz e Carneiro, 2003):

$$S_{ii'} = \frac{2a}{2a + b + c}$$

Onde:

$S_{ii'}$: similaridade entre os acessos i e i' ;

a : valor que quantifica o número de coincidências do tipo 1-1 para cada par de acessos;

b : valor que quantifica o número de coincidências do tipo 1-0 para cada par de acessos;

c : valor que quantifica o número de coincidências do tipo 0-1 para cada par de acessos.

Este coeficiente ($S_{ii'}$) ou seu complemento ($1 - S_{ii'}$) foi empregado na construção de matrizes de similaridade ou de dissimilaridades, respectivamente, entre os acessos. Tais matrizes foram utilizadas na avaliação do número ótimo de marcadores, necessário para discriminar adequadamente os acessos avaliados; na análise discriminante, baseada em k vizinhos mais próximos; e, na análise da projeção gráfica das distâncias entre acessos.

A partir da matriz de dados binários também foram obtidas as seguintes estatísticas descritivas:

a) Porcentagem de polimorfismo de cada *primer*, obtida pela equação:

$$P = \frac{N_{pb}}{N_{bt}}$$

Onde:

P : porcentagem de locos polimórficos;

N_{bp} : número de bandas polimórficas;

N_{bt} : número de bandas totais.

b) Número de locos ou bandas exclusivas

N_{be} : número total de bandas restritas a um grupo ou população;

2.7.2 Estimação do número ótimo de marcadores

O número mínimo de marcadores, necessário para a caracterização da dissimilaridade genética dos acessos avaliados, foi estimado a partir de correlações entre matrizes de dissimilaridades. Foram obtidas, por amostragem, diversas matrizes de dissimilaridades a partir de subconjuntos, de diferentes tamanhos, de marcadores moleculares. Os dados de cada uma destas matrizes foram correlacionados com dados da matriz de dissimilaridade obtida a partir de todos os locos analisados no estudo.

O procedimento foi iniciado com matrizes de dissimilaridade estimadas a partir de amostras de 10 locos, sendo repetido com incrementos de cinco locos marcadores, no intervalo entre 10 e 220 locos. Para cada tamanho amostral foram realizadas 50 réplicas e os resultados obtidos foram apresentados graficamente, considerando as correlações entre as diversas matrizes com m_i marcas amostradas e a matriz original com m marcas ($m_i < m$).

2.7.3 Análise discriminante

Análises discriminantes, baseadas no método não paramétrico de k vizinhos mais próximos, foram aplicadas considerando dois critérios de classificação *a priori*: espécies ou ecótipos apenas ou espécies, ecótipos e origem geográfica das populações.

No método dos k vizinhos mais próximos, o valor de k é definido *a priori* de maneira arbitrária. Supondo que se deseja alocar um indivíduo, cujo valor observado seja igual a x , em uma de duas ou mais populações, a partir de todas as distâncias genéticas entre pares de indivíduos, define-se os k indivíduos mais próximos daquele que se deseja classificar. Supondo que dentre esses k indivíduos, k_t são provenientes da população t , cuja probabilidade *a priori* é π_t . Então, a probabilidade *a posteriori* de que este indivíduo pertença a t -ésima população é estimada por:

$$P = \frac{\pi_t (k_t / n_t)}{\sum_{s=1}^p \pi_s (k_s / n_s)}$$

Onde

s : é número de populações avaliadas ($s = 1, \dots, p$);

n : é o número de indivíduos de cada população.

A observação x será classificada na t -ésima população se $\pi_t k_t / n_t$ for maior que $\pi_s k_s / n_s$ para todo s (Khattree e Naik, 2000).

Neste trabalho foi pré-estabelecido um valor $k=3$, no caso de ocorrência de empate entre probabilidades, a observação x foi alocada uma população designada de “desconhecida”.

A classificação *a priori* dos acessos estudados, de acordo com o critério baseado em espécies ou ecótipos, foi assim estabelecida:

- a) *A. occidentale* ‘Comum’ – 93 acessos;
- b) *A. occidentale* ‘Anão Precoce – 10 acessos;
- c) *A. othonianum* – 10 acessos;
- d) *A. humile* – 06 acessos.

Os 17 acessos restantes foram alocados a *posteriori*, por não atenderem aos critérios de inclusão nas populações pré-definidas.

A classificação dos acessos pelo critério baseado na origem geográfica, além do critério de espécie e/ou ecótipo, possibilitou a alocação prévia dos acessos em sete populações diferentes:

- a) Nordeste (NE) – 31 acessos;
- b) Ano Precoce (AP) – 10 acessos;
- c) Rio de Janeiro (RJ) – 45 acessos;
- d) Roraima (RR) – 10 acessos;
- e) Índia (ID) – 08 acessos;
- f) othonianum (OT) – 10 acessos;
- g) humile (HU) – 06 acessos.

Os 16 acessos remanescentes foram alocados a *posteriori*, uma vez que não atenderam aos critérios para inclusão em uma das populações pré-definidas.

2.7.4 Projeção gráfica das distâncias

A partir da matriz de distâncias, obtida do complemento do coeficiente de similaridade de Nei e Li, e da alocação de todos os genótipos nos grupos definidos pela análise discriminante descrita anteriormente, foi realizada a projeção gráfica dos acessos no espaço tridimensional. De acordo com Cruz (2006a), neste procedimento as distâncias entre genótipos são convertidas em escores relativos a três variáveis (X, Y e Z), que, quando representadas em gráficos de dispersão tridimensional, irão refletir as distâncias originalmente obtidas a partir do espaço n-dimensional (n=número de caracteres avaliados).

As coordenadas dos genótipos ou acessos são estabelecidas considerando a ordem decrescente da divergência média de cada um. As coordenadas dos dois primeiros acessos (i e j) são estabelecidas arbitrariamente, enquanto as coordenadas do quarto e quinto acessos são estabelecidas matematicamente, baseadas nas propriedades de um triângulo. Para os demais acessos, as coordenadas são estimadas estatisticamente, a partir de um sistema de equações, visando minimizar a distorção entre as distâncias originais e as distâncias gráficas (Cruz e Carneiro, 2003; Cruz, 2006a).

A eficiência da projeção gráfica dos dados foi estimada através das seguintes estatísticas (Cruz, 2006a):

a) Coeficiente de correlação entre as distâncias originais e as que foram representadas no gráfico de dispersão;

b) Grau de distorção (1- α), considerando que:

$$\alpha = \frac{\sum_{i < i'} \sum d_{gi'}^2}{\sum_{i < i'} \sum d_{oi'}^2}$$

Onde $d_{gi'}$ e $d_{oi'}$ são as distâncias gráficas (espaço tridimensional) e originais, respectivamente, de todos os pares de indivíduos i e i' .

c) Coeficiente de estresse, definido por Kruskal (1964), dado por:

$$s = \sqrt{\frac{\sum_{i < i'} \sum (d_{oi'} - d_{gi'})^2}{\sum_{i < i'} \sum d_{oi'}^2}}$$

2.7.5 Análise de variância molecular

A análise de variância molecular (AMOVA) foi delineada por Excoffier et al. (1992) e é baseada em um esquema de análise de variância hierarquizada produzindo estimativas de componentes de variância análogas às estatísticas F de Wright, denominadas de “estatísticas Φ ”. O método parte de uma matriz com os quadrados das distâncias entre todos os pares de haplótipos que, neste contexto, referem-se à combinação de marcadores moleculares presentes em alguma área definida do genoma. Assim, o haplótipo de um indivíduo é representado por um vetor p , com valores 1, se a banda homóloga estiver presente e 0, caso contrário (Excoffier et al., 1992).

Considerando o caso presente, em que é assumido que os indivíduos amostrados são oriundos de diferentes populações independentes, os haplótipos podem ser representados pelo modelo linear:

$$x_{ij} = x + a_j + b_{ij}$$

Onde:

x_{ij} : é a frequência do vetor do i -ésimo haplótipo, da j -ésima população;

x : é o valor esperado desconhecido de x_{ij} ;

a: é o efeito de populações e

c: é o efeito de indivíduos ou haplótipos.

Tais efeitos são assumidos como sendo aditivos, aleatórios, não correlacionados, associados aos componentes de variância σ_a^2 e σ_b^2 , respectivamente, que são utilizados na obtenção da estatística ϕ_{ST} . O esquema de análise de variância baseada neste modelo é apresentado na Tabela 2.

Tabela 2 – Esquema de Análise de Variância Molecular (AMOVA), para o caso de um grupo de populações independentes.

Fonte de Variação	GL	SQ	QM	E(QM)	Componentes de variância
Entre populações	P-1	SQa	QMa	$\sigma_b^2 + n\sigma_a^2$	$\sigma_T^2 = \sigma_a^2 + \sigma_b^2$
Dentro de Populações	N-P	SQb	QMb	σ_b^2	$\phi_{ST} = \frac{\sigma_a^2}{\sigma_T^2}$
Total	N-1	-	-	-	-

$$n = \frac{N - \sum_P \frac{N_P^2}{N}}{P - 1}$$

Onde:

P : é número de populações;

N : é número total de indivíduos ou acessos;

Np : é o número de indivíduos ou acessos na população P .

Neste trabalho a AMOVA foi realizada considerando os 120 acessos alocados *a priori* em sete populações de *Anacardium* spp.: 'Nordeste', 'Anão Precoce', 'Rio de Janeiro', 'Roraima', 'Índia' 'othonianum' e 'humile'. As análises foram realizadas com auxílio dos aplicativos Arlequin (Schneider et al., 2000) e Genes (Cruz, 2006a, 2006b).

A significância de ϕ_{ST} global foi testada por 10.000 permutações de haplótipos entre as populações, enquanto as significâncias dos ϕ_{ST} entre pares de populações foram testadas através de 1000 permutações de haplótipos entre as populações.

2.7.6 Análise de agrupamento entre populações e acessos

Os valores de ϕ_{ST} entre pares de populações foram utilizados para construir um dendrograma que evidenciasse melhor as similaridades relativas entre os grupos. Foram utilizados os métodos de agrupamento UPGMA e o baseado na variância mínima de Ward (Cruz et al., 2004). A adequação dos agrupamentos foi testada através do cálculo do coeficiente de correlação cofenética.

O agrupamento entre os 91 acessos de *Anacardium*, do BAG do Cajueiro, foi obtido pelo método UPGMA, a partir do complemento dos coeficientes de similaridade de Nei e Li (Cruz e Carneiro, 2003).

3 RESULTADOS

3.1 Padrões da variabilidade acessada

A relação dos *primers* ISSR selecionados para a análise da diversidade é apresentada na Tabela 3, onde também são informados as respectivas temperaturas de anelamento, o sistema de resolução dos produtos amplificados, o número de locos amplificados, o percentual de polimorfismo e a amplitude aproximada dos fragmentos amplificados.

O número de locos distinguíveis com o uso gel desnaturante de poliacrilamida, corado com nitrato de prata, foi três vezes superior, em média, que no sistema de gel de agarose, corado com brometo de etídio, muito embora esta comparação seja feita entre *primers* de seqüências distintas.

O nível de polimorfismo detectado foi elevado e, de maneira geral, independente do sistema de resolução empregado. O sistema de resolução também não apresentou influência relevante na amplitude dos fragmentos amplificados (Tabela 3).

Outro aspecto a ser destacado é a prevalência do motivo 'GA' nas seqüências dos *primers* selecionados. Esta prevalência já era evidente a partir dos *primers* relacionados na Tabela 1, que também mostra uma freqüência igualmente elevada do motivo 'AG'. Entretanto, conforme ilustrado pela Figura

1, *primers* com motivos iguais, mas diferindo apenas nas seqüências de ancoragem, produzem perfis de amplificação muito diferentes.

A Tabela 4 mostra, de maneira detalhada, o número de locos fixados ($f = 1$ ou $f = 0$) ou com freqüências $f \geq 0,9$ e $f \leq 0,1$ para os produtos da amplificação dos nove *primers* ISSRs, considerado os quatro grupos de acessos de *Anacardium* spp., definidos segundo a espécie e / ou ecótipo. Verifica-se que o grupo 'Anão Precoces' é o que apresenta o maior número locos fixados e o grupo 'Comum' o menor, mas com grande prevalência de locos com $f \geq 0,9$ e $f \leq 0,1$. Comparativamente, o grupo 'othonianum' apresentou maior proporção de locos com freqüências intermediárias, isto é, $0,9 \geq f \geq 0,1$, além de poucos locos fixados.

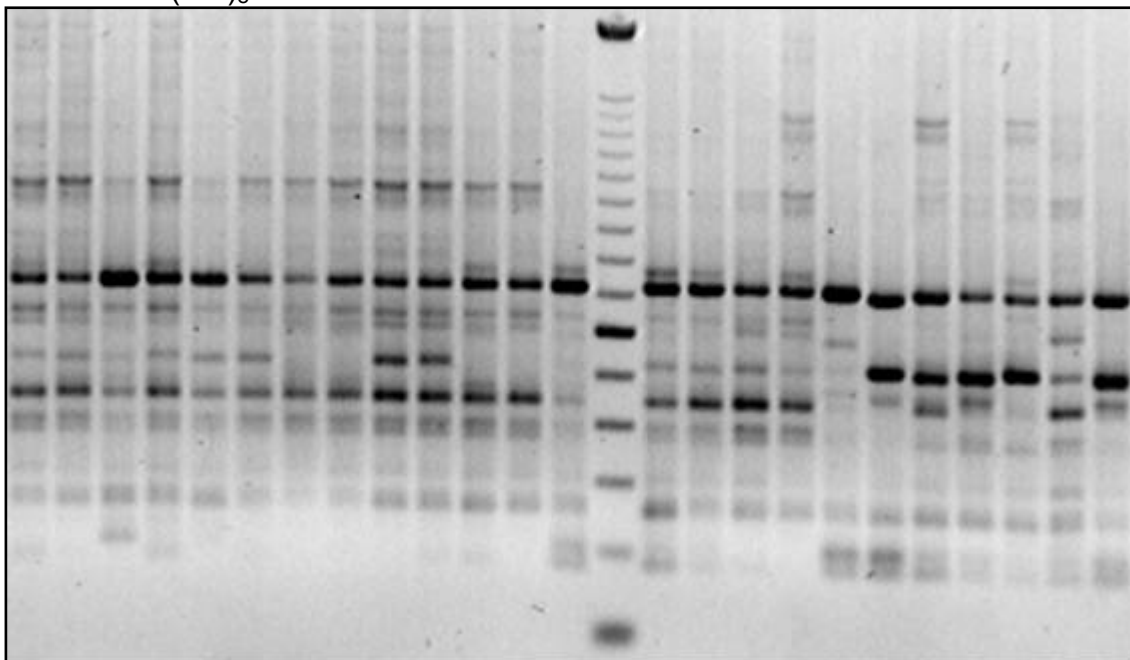
Tabela 3 – Relação de *primers* utilizados com respectivas temperaturas de anelamento (T_A), sistema de resolução dos produtos amplificados, número de bandas registradas, porcentagem de polimorfismo e amplitudes aproximadas, em pares de bases (pb), dos fragmentos amplificados.

<i>Primer</i>	T_A (°C)	Sistema de Resolução	Nº de marcas registradas	% de Polimorfismo	Amplitude dos fragmentos (pb)
(GACA) ₄	48	Agarose	16	87,50	370-2000
(GAAGTGGG) ₂	52	Agarose	16	100	230-2000
(GTG) ₆	60	Agarose	11	72,73	280-1200
812 - (GA) ₈ A	50	Poliacrilamida	26	92,30	250-2200
840 - (GA) ₈ YT	50	Agarose	14	100	380-1500
841 - (GA) ₈ YC	50	Poliacrilamida	78	100	180-2200
842 - (GA) ₈ YG	50	Agarose	14	92,28	190-1100
834 - (AG) ₈ YT	50	Poliacrilamida	30	100	270-2100
886 - VDV(CT) ₇	50	Agarose	18	100	450-1500
Total			223	96,40	180-2200

Na Tabela 4 verifica-se também que o *primer* (GTG)₆ amplificou o menor número de locos (n=11), além de ter apresentado o maior número de locos fixados para o conjunto de acessos. Por outro lado, este *primer* amplificou três locos exclusivos ou privados, um para o grupo 'Comum' ($f = 0,05$), um para o grupo 'othonianum' ($f = 0,1$) e um para *A. occidentale*

($f = 0,88$), considerando 'Comum' e 'Anão 'Precoce', em conjunto. Bandas exclusivas também obtidas com os *primers* GACA para 'humile' ($f = 0,17$), GAA para 'othonianum' $f = 0,20$ e 841 para 'Comum' ($f = 0,11$).

Primer 842 (GA)₈YG



Primer 840 (GA)₈YT

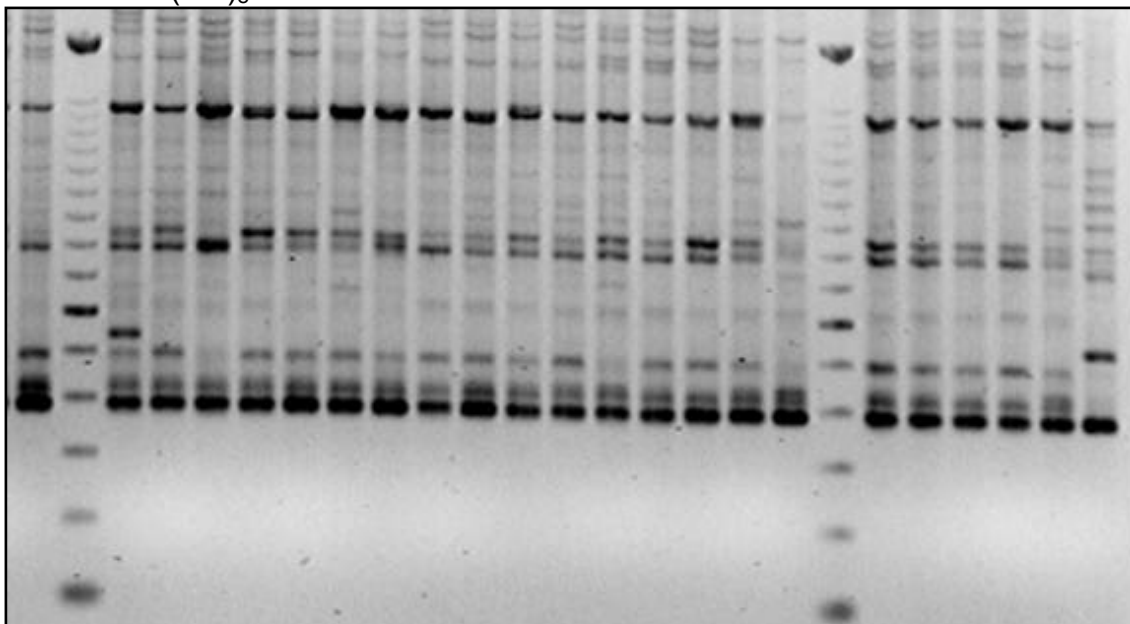


Figura 1 – Diferenças nos perfis de bandas amplificadas pelos *primers* (GA)₈YG e (GA)₈YT, em DNA nuclear de *Anacardium* spp, com resolução em gel de agarose (1,5%), corado com brometo de etídio.

Tabela 4 – Número de locos fixados ($f = 1$ ou $f = 0$) e com frequências $f \geq 0,9$ e $f \leq 0,1$ para os produtos da amplificação de nove primers ISSR, considerado quatro grupos de acessos de *Anacardium* spp.

Primer	Freq. de bandas	<i>Anacardium occidentale</i> Comum (n=93)	<i>Anacardium occidentale</i> Anão (n=10)	<i>Anacardium othonianum</i> (n=10)	<i>Anacardium humile</i> (n=6)	Conjunto acessos (N=119)
(GACA) ₄ n=16	$f = 1$	2	4	2	2	2
	$f = 0$	1	3	3	0	-
	$f \geq 0,9$	1	1	1	-	0
	$f \leq 0,1$	1	2	0	-	0
(GAAGT GGG) ₂ n=16	$f = 1$	1	7	2	2	0
	$f = 0$	2	3	0	2	-
	$f \geq 0,9$	6	3	3	-	1*
	$f \leq 0,1$	2	1	2	-	1
(GTG) ₆ n=11	$f = 1$	3	4	4	4	3
	$f = 0$	0	2	2	2	-
	$f \geq 0,9$	3	0	0	-	1
	$f \leq 0,1$	3	1	1	-	2
812 (GA) ₈ A n=26	$f = 1$	2	11	5	8	2
	$f = 0$	0	4	2	1	-
	$f \geq 0,9$	10	4	1	-	3
	$f \leq 0,1$	1	0	1	-	0
840 (GA) ₈ YT n=14	$f = 1$	0	2	1	1	0
	$f = 0$	1	4	1	1	-
	$f \geq 0,9$	3	2	1	-	1
	$f \leq 0,1$	1	1	7	-	1
841 (GA) ₈ YC n=78	$f = 1$	0	4	2	5	0
	$f = 0$	0	9	8	6	-
	$f \geq 0,9$	9	8	2	-	2*
	$f \leq 0,1$	6	13	8	-	1
842 (GA) ₈ YG n=14	$f = 1$	2	5	3	2	1
	$f = 0$	0	2	0	4	-
	$f \geq 0,9$	1	0	0	-	1*
	$f \leq 0,1$	1	0	2	-	0
834 (AG) ₈ YT n=30	$f = 1$	1	14	5	4	0
	$f = 0$	0	4	0	1	-
	$f \geq 0,9$	12	3	1	-	4
	$f \leq 0,1$	1	0	1	-	0
886 VDV(CT) ₇ n=18	$f = 1$	1	3	1	0	0
	$f = 0$	0	3	1	3	-
	$f \geq 0,9$	2	2	0	-	0
	$f \leq 0,1$	4	1	2	-	0
Totais N=223	$f = 1$	12	54	25	28	8
	$f = 0$	4	34	17	20	-
	$f \geq 0,9$	47	23	9	-	13
	$f \leq 0,1$	20	19	24	-	5
	Geral (%)	83 (37,2)	130 (58,3)	75 (33,6)	48 (21,5)	26 (111,7)

(-) valor não estimado, (*) incluindo casos com $f = 0,83$ no grupo *A. humile*.

3.2 Estimação do número ótimo de marcadores

A relação entre o número de locos marcadores investigados e o número mínimo necessário para acessar o mesmo nível de variabilidade é apresentada na Figura 2. A estabilização do coeficiente de correlação de Pearson (r) em patamares iguais ou superiores a 0,90 e 0,95 ocorreu, respectivamente, com tamanhos amostrais de 115 e 155 locos. Os coeficientes médios de estresse para matrizes de distâncias obtidas a partir de amostras com 115 e 155 locos foram, respectivamente, 4,81% e 2,88%.

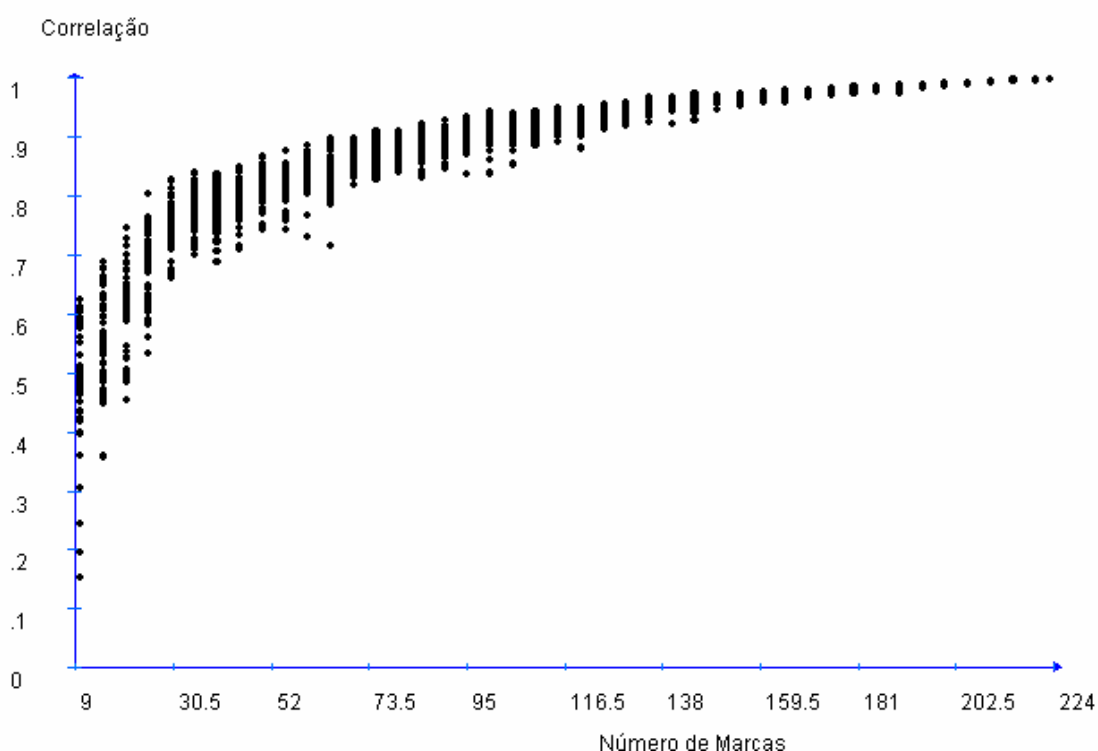


Figura 2 – Representação gráfica do número ótimo de marcadores para acessar a diversidade entre 136 acessos de *Anacardium* spp. Pontos referem-se à correlação entre matrizes com m_i marcas amostradas ($10 \leq m_i \leq 223$) e a matriz original com m marcas ($m = 223$).

3.3 Análise discriminante

Os resultados das análises discriminantes, baseadas em k vizinhos mais próximos são apresentados nas Tabelas 5 e 6. A discriminação dos acessos segundo a espécie e/ou ecótipo, apresentada na Tabela 5 evidencia que os

acessos dos grupos 'Comum' e 'Anão Precoce' foram eficientemente alocados nos seus grupos de origem. Entretanto, o método foi ineficiente para discriminar os acessos dos grupos 'othonianum' e 'humile'. A taxa de erro aparente foi de 12,6%. Embora relativamente baixo, este valor é resultante, quase que exclusivamente, da má classificação dos acessos dos grupos 'othonianum' e 'humile'.

A maior parte (13) dos 17 acessos classificados *a posteriori* foi alocada no grupo 'Comum'. Dez destes acessos são originários do Estado de Roraima e, embora estejam identificados apenas como *Anacardium* sp., apresentam fenótipo típico de *A. occidentale*, mas com porte reduzido e copa compacta como do grupo 'Anão Precoce'. Entre os acessos alocados no grupo 'Anão Precoce', dois estão identificados como *A. microcarpum*, cujas origens são desconhecidas, e o terceiro como *Anacardium* sp. coletado em Camocim-CE. O acesso alocado no grupo 'othonianum' está identificado como *A. humile*, mas seu fenótipo é de *A. othonianum*, tratando-se possivelmente de um erro de identificação.

Tabela 5 – Análise discriminante de quatro populações de *Anacardium* spp., baseada na metodologia de k vizinhos mais próximos (k=3), estimados a partir de 223 marcadores moleculares binários. Os Valores destacados em negrito referem-se aos percentuais de alocação dos acessos nos seus respectivos grupos de origem.

Grupo de origem	Nº de acessos	Percentuais de alocação nos diferentes grupos			
		'Comum'	'Anão'	'othonianum'	'humile'
<i>A. occidentale</i> 'Comum'	93	97,85	1,08	0,00	1,08
<i>A. occidentale</i> 'Anão Precoce'	10	0,00	100	0,00	0,00
<i>A. othonianum</i> 'othonianum'	10	30,00	0,00	30,00	40,00
<i>A. humile</i> 'humile'	06	50,00	0,00	50,00	0,00
Alocados a <i>posteriori</i>	17	76,47	17,65	5,88	0,00
Total de acessos alocados	136	110	14	07	05

A discriminação dos acessos considerando, além da espécie e/ou ecótipo, o local de origem das populações é apresentada na Tabela 6. À exceção do grupo 'Nordeste', todas as demais populações de *A. occidentale*

foram bem discriminadas, em termos de alocação correta de seus acessos. A população mais distinta foi 'Rio de Janeiro', uma vez que 100% de seus acessos foram alocados no próprio grupo, além de não ter recebido a alocação de nenhum acesso de outra população. Todos os acessos das populações 'Anão Precoce' e 'Índia' também foram corretamente alocados. Por outro lado, vários acessos de outras populações foram incorretamente alocados no grupo 'Índia', principalmente acessos do grupo 'Nordeste'. A maior parte dos acessos classificados a *posteriori* também foi alocada no grupo 'Índia', embora quase todos eles sejam originários do Brasil.

Assim como na situação anterior, a análise não proporcionou uma discriminação adequada das populações 'othonianum' e 'humile'. Nos dois casos nenhum acesso de 'humile' foi alocado nesta população, mas apenas acessos do grupo 'othonianum'. Entretanto, a maior parte dos acessos de 'humile' foram alocados no grupo 'othonianum'. Além disso, muitos acessos destes dois grupos foram alocados nas populações 'Nordeste', 'Roraima' e 'Índia'.

Tabela 6 – Análise discriminante de sete populações de *Anacardium* spp., baseada na metodologia de k vizinhos mais próximos (k=3), estimados a partir de 223 marcadores moleculares binários. Os Valores destacados em negrito referem-se aos percentuais de alocação dos acessos nos seus respectivos grupos de origem.

Grupo de origem	Nº de Acessos	Percentuais de alocação nos diferentes grupos						
		NE	AP	RJ	RR	ID	OT	HU
Nordeste (NE)	31	54,84	3,23	0,00	9,68	32,26	0,00	0,00
Anão Precoce (AP)	10	0,00	100	0,00	0,00	0,00	0,00	0,00
Rio de Janeiro (RJ)	45	0,00	0,00	100	0,00	0,00	0,00	0,00
Roraima (RR)	10	0,00	0,00	0,00	90,00	10,00	0,00	0,00
Índia (ID)	8	0,00	0,00	0,00	0,00	100	0,00	0,00
<i>othonianum</i> (OT)	10	30,00	0,00	0,00	0,00	20,00	30,00	20,00
<i>humile</i> (HU)	6	16,67	0,00	0,00	16,67	0,00	66,67	0,00
Alocados a <i>posteriori</i>	16	0,00	18,75	0,00	0,00	68,75	6,25	6,25
Total de acessos	136	21	14	45	13	32	8	3

Por outro lado, a classificação a *posteriori* dos acessos alocados nos grupos 'Anão Precoce' e 'othonianum' mostrou-se consistente. Uma vez que nas duas análises discriminantes estes acessos foram alocados nos mesmos grupos.

3.4 Projeção gráfica das distâncias

A projeção das dissimilaridades entre os 136 acessos é apresentada na Figura 3. As cores evidenciam os grupos de classificação dos acessos, considerando as alocações a *priori* e a *posteriori*. As elipses destacam a dispersão dos grupos 'Rio de Janeiro'(3), 'Anão Precoce' (2) e a região onde se concentra a dispersão da maior parte dos acessos avaliados, mas, principalmente, daqueles pertencentes aos grupos 1, 2, 4 e 5. Dessa forma, fica claro a maior divergência do grupo 'Rio de Janeiro' e a sobreposição dos grupos 'Nordeste', 'Anão Precoce', 'Roraima' e 'Índia. Os acessos, dos grupos 'othonianum' e 'humile', por outro lado, apresentaram padrões de dispersão que impossibilita o delineamento de qualquer tipo de grupo.

O coeficiente de correlação entre as distâncias originais e as estimadas para o espaço tridimensional foi de 0,73 enquanto a grau de distorção e o coeficiente de estresse foram, respectivamente, 47,6% e 50,85%. Os valores elevados destes dois últimos coeficientes implicam em considerar com cautela os resultados da Figura 3.

3.5 Análise de variância molecular

Os resultados da AMOVA, apresentados na Tabela 7, indicam que 27,5% da variação, em média, foi devido às diferenças entre populações ($\phi_{ST} = 0,2749$), enquanto 72,5% resultou de diferenças entre acessos dentro das populações.

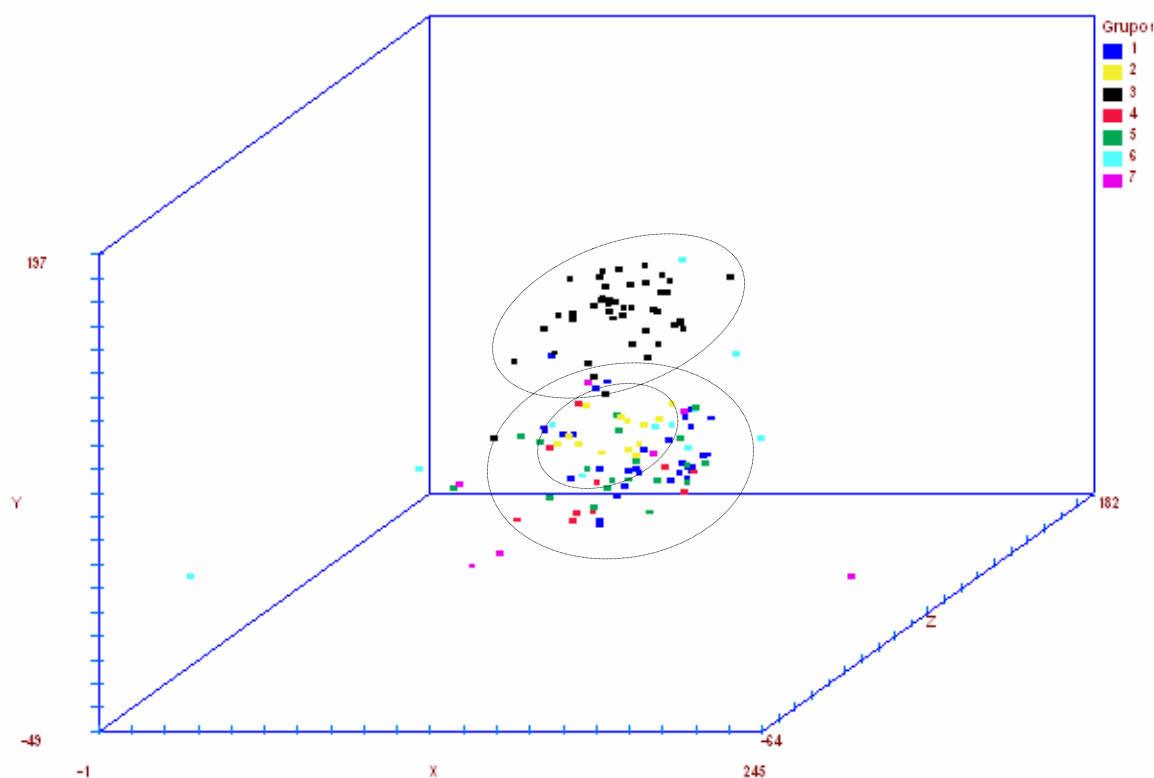


Figura 3 – Projeção gráfica de 136 acessos de *Anacardium* spp., alocados em sete populações diferentes através de análise discriminante não paramétrica, baseada em k vizinhos mais próximos. Populações ou grupos: 1 Nordeste, 2 Anão Precoce, 3 Rio de Janeiro, 4 Roraima, 5 Índia, 6 othonianum e 7 humile.

Tabela 7 - Análise de variância molecular (AMOVA), de 120 acessos amostrados de sete populações *Anacardium* spp., utilizando 221 locos marcadores ISSR.

Fonte de Variação	GL	SQ	QM	Componente de Variância	Porcentagem (total)	P
Entre Populações	6	965,57	160,93	8,97	27,49	(<0,001)
Dentro de Populações	113	2674,40	23,67	23,67	72,51	(<0,001)
Total	119	3639,97	30,59	-	100	-

Na Tabela 08 são apresentados, na diagonal principal, os índices ϕ_{ST} específicos de cada população, que é uma medida da variabilidade intrapopulacional. No restante da tabela são apresentados os índices ϕ_{ST} estimados entre pares de populações. Estes valores variaram de 0,004 a 0,4289 e todos eles foram significativos com probabilidades inferiores a 0,1%, com exceção dos pares de populações 'Nordeste' x 'Índia', cuja probabilidade

foi $0,001 < P < 0,01$, e ‘othonianum’ x ‘humile’, que apresentou ϕ_{ST} não significativo.

Portanto, à exceção do par ‘othonianum’ x ‘humile’ todas as populações podem ser consideradas diferentes entre si, com ‘Rio de Janeiro’ apresentado o maior grau de divergência médio relativo.

Tabela 08 – Índices ϕ_{ST} específicos (diagonal principal) e entre pares de populações de sete populações de *Anacardium* spp., estimados a partir de 221 locos marcadores ISSR.

Pop.	NE	AP	RJ	RR	ID	OT	HU
NE	0,2700						
AP	0,1290***	0,2740					
RJ	0,2916***	0,2950***	0,2826				
RR	0,1146***	0,2471***	0,4289***	0,2815			
ID	0,0600**	0,2476***	0,4011***	0,1907***	0,2873		
OT	0,2010***	0,1625***	0,3624***	0,2610***	0,2673***	0,2518	
HU	0,2472***	0,2080***	0,4052***	0,3217***	0,3225***	0,0040 ^{ns}	0,2552

** $P < 0,01$; *** $P < 0,001$; ^{ns} P não significativo.

3.6 Análise de agrupamento

Os dendrogramas obtidos pelos métodos de agrupamento baseados na variância mínima de Ward e UPGMA são apresentados, respectivamente, nas Figuras 4 e 5. Os dois dendrogramas foram construídos a partir dos valores ϕ_{ST} entre pares de populações, apresentados na Tabela 8. Embora exista congruência nas associações entre as populações ‘othonianum’, ‘humile’ e ‘Anão Precoce’ de um lado, ‘Nordeste’, ‘Índia’ e ‘Roraima’ de outro, os dois dendrogramas diferem no posicionamento da população ‘Rio de Janeiro’. Por outro lado, os resultados dos dois agrupamentos são concordantes com aqueles obtidos na análise discriminante e projeção das distâncias entre os acessos, porque apresentam a população ‘Rio de Janeiro’ como a mais divergente, seguida de ‘Anão Precoce’. Enquanto ‘Nordeste’, ‘Índia’ e ‘Roraima’ são menos diferenciadas entre si e ‘othonianum’ e ‘humile’, são praticamente indistintas, mas diferentes de todas as demais.

Os coeficientes de correlação cofenética foram iguais a 65,16%, para o agrupamento de Ward, e 89,56% para o agrupamento obtido por UPGMA, ambos significativos a 1% de probabilidade pelos testes t e de Mantel.

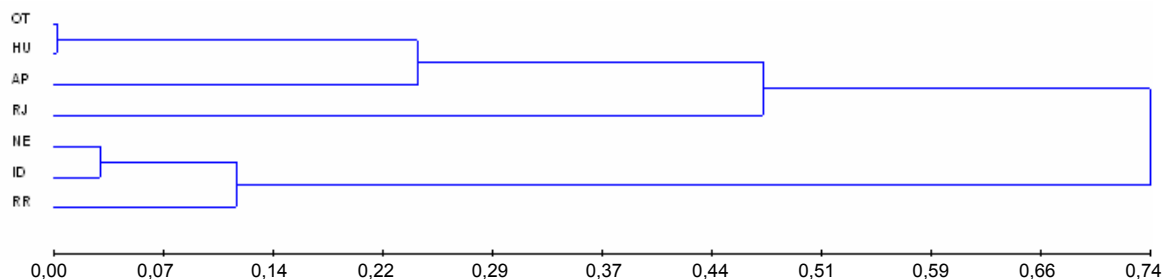


Figura 4 – Dendrograma de sete populações de *Anacardium* spp., obtido pelo método de agrupamento baseado na variância mínima de Ward, a partir da matriz de valores de ϕ_{ST} entre pares de populações.

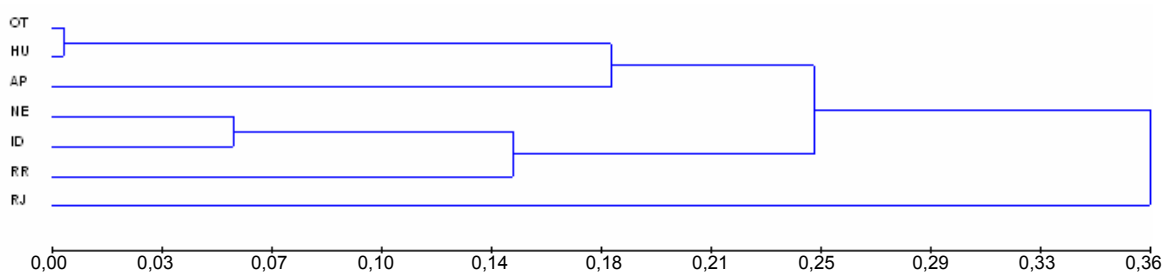


Figura 5 – Dendrograma de sete populações de *Anacardium* spp., obtido pelo método de agrupamento UPGMA, a partir da matriz de valores de ϕ_{ST} entre pares de populações.

A Figura 6 apresenta o dendrograma formado pelos 91 acessos de *Anacardium* avaliados do BAG do Cajueiro. Este agrupamento foi obtido pelo método UPGMA, a partir do complemento dos coeficientes de similaridade de Nei e Li, estimados entre pares de acessos. Embora não tenha ocorrido uma separação clara dos acessos, segundo a origem ou espécie, é percebida uma tendência ao agrupamento dos acessos de algumas populações. Conforme destacado na Figura, a maior parte dos acessos do ‘Nordeste’ se concentrou em dois subgrupos distintos, assim como os acessos da ‘Índia’. Os acessos de ‘Anão Precoce’ formaram apenas um grupo principal, enquanto os acessos de ‘Roraima’ não formaram nenhum subgrupo principal, embora a maior parte deles se encontrem posicionados na parte superior do dendrograma, juntos com um subgrupo do ‘Nordeste’ e os dois subgrupos da ‘Índia’. Por outro lado, os acessos de várias procedências, não agrupados *a priori*, e cuja maioria foi

alocada no grupo da 'Índia' na análise discriminante, apresentam distribuição dispersa ao longo do dendrograma. Estes acessos estão codificados por iniciais em letras maiúsculas: SP – São Paulo; MT – Mato Grosso; PI – Piauí; PA – Pará, CE – Ceará; GO – Goiás; VNZ – Venezuela.

Os acessos de 'othonianum' e 'humile', posicionados na parte inferior da Figura, não apresentaram nenhuma tendência à formação de grupos entre si, além de serem mais divergentes, de um modo geral, em relação aos acessos das populações de *A. occidentale*.

Os acessos do 'Rio de Janeiro', quando foram incluídos na análise de agrupamento, formaram um grupo único, posicionado na extremidade superior do dendrograma.

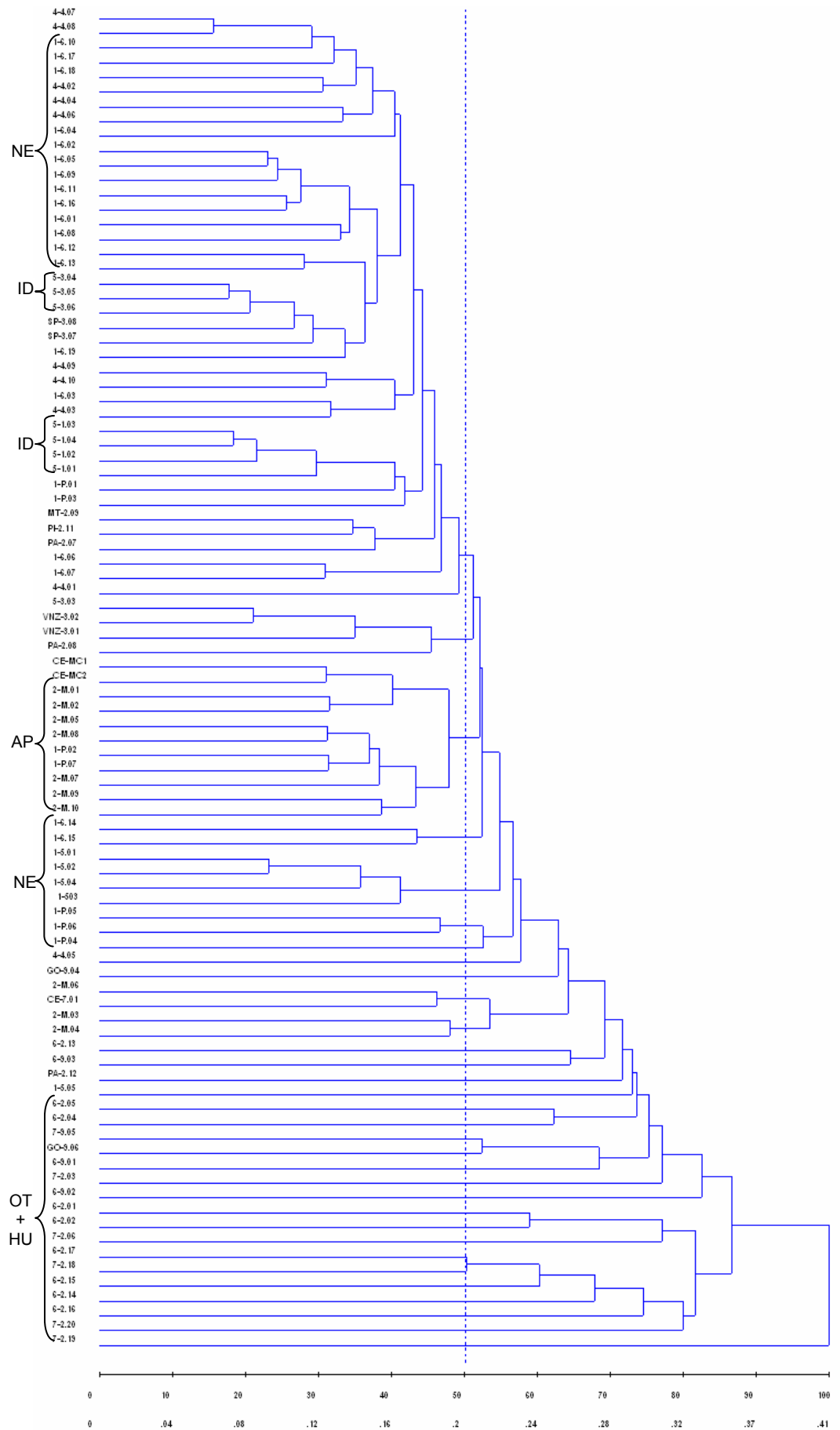


Figura 6 - Dendrograma UPGMA, baseado no coeficiente de Nei e Li, de 91 acessos do BAG do Cajueiro, Embrapa-CNPAT-CE. Grupos em destaque: NE- Nordeste, AP- Anão Precoce, ID – Índia, OT – ‘othonianum’, HU – ‘humile’.

4 DISCUSSÃO

4.1 Aspectos da variabilidade acessada

Os marcadores ISSR têm sido utilizados com sucesso em diversos estudos envolvendo a caracterização molecular em plantas, tais como *fingerprinting* de cultivares (Eiadthong et al., 1999; Prevost e Wilkinson, 1999; Charters e Wilkinson, 2000; Ruas et al., 2003); divergência entre acessos de bancos de germoplasma (Fang et al., 1997; Blair et al., 1999; Huang e Sun, 2000; Joshi et al., 2000; Assefa et al., 2003); diversidade entre e dentro de populações co-específicas ou interespecíficas (Ge e Sun, 1999; Deshpande et al., 2001; Mattioni et al. 2002; Vijayan et al., 2004; Aga, et al., 2005; Ge et al., 2005).

No presente trabalho também foi confirmada a utilidade desta classe de marcadores na caracterização de germoplasma de cajueiro. Do conjunto de nove *primers* utilizados, três amplificaram 134 locos (resolução em PAGE, corado com nitrato de prata) e os outros seis amplificaram 89 locos (resolução em gel de agarose, corado com brometo de etídio). A proporção de locos polimórficos foi bastante elevada (96,4%) considerando o conjunto de acessos avaliados. Por outro lado, o percentual de locos fixados dentro de cada ecótipo ou espécie variou de 7,17%, no grupo 'Comum', a 39,46%, no grupo 'Anão Precoce'. O número de locos com baixo conteúdo médio de informação polimórfica, ou seja, locos com frequências $f \geq 0,9$ e $f \leq 0,1$ também foi elevado dentro dos grupos. Entretanto, a existência de locos privados ou exclusivos, além de ter sido pouco freqüente, foi restrita a locos raros dentro dos grupos, com exceção de um loco amplificado pelo *primer* (GTG)₆, exclusivo para *A occidentale*. Mesmo locos comuns, mas com grande divergência de frequências entre os grupos não foram observados. Tais resultados apontam para certas restrições na capacidade discriminatória dos grupos proporcionada pelos locos amplificados. Muito embora o aumento no número de locos amostrados não exerça impacto significativo nos resultados obtidos, conforme ficou demonstrado através da avaliação do número ótimo de marcadores.

4.2 Análises discriminantes

Os resultados obtidos pelas análises discriminantes, baseada no método de k vizinhos mais próximos, mostraram que os locos marcadores ISSR obtidos foram apropriados para diferenciar populações da espécie *A. occidentale*, mas inadequados para discriminar acessos das espécies *A. othonianum* e *A. humile*, tanto entre si quanto em relação a *A. occidentale*. Estes resultados foram, de certa forma, surpreendentes, uma vez que os grupos 'othonianum' e 'humile', são morfologicamente mais divergentes, tanto entre si quanto em relação aos demais grupos.

Muitos trabalhos envolvendo a caracterização de indivíduos de diferentes espécies, através de marcadores ISSR têm produzidos resultados satisfatórios e sem ambigüidades. Por exemplo, Mattioni et al. (2002) conseguiram agrupar corretamente 125 indivíduos de diferentes populações, pertencentes a três espécies arbóreas produtoras de madeira do gênero *Nothofagus*, utilizando apenas 6 *primers* ISSR da coleção UBC, que produziram um total de 63 bandas. Huang e Sun (2000), utilizando 15 *primers* ISSR, descreveram a divergência genética de 40 acessos de germoplasma de batata doce, pertencentes a 10 espécies diferentes. A análise de agrupamento dos dados demonstrou que os acessos de uma mesma espécie foram mais similares entre si do que acessos de espécies diferentes. Resultados semelhantes foram obtidos por Joshi et al. (2000), trabalhando com acessos de germoplasma de arroz de diferentes espécies selvagens e cultivadas do gênero *Oriza*.

Archak et al. (2003a) compararam a eficiência de três tipos de marcadores moleculares (RAPD, ISSR e AFLP) na medida de variação e discriminação genotípica de germoplasma de caju. Os autores verificaram que, a partir de cada um dos três tipos de marcadores utilizados, foi possível discriminar os 19 acessos de *A. occidentale* avaliados, com maior eficiência para os marcadores AFLPs. Foram empregados 12 *primers* ISSR não ancorados, sendo três deles com seqüências idênticas aos utilizados neste trabalho, que produziram 171 bandas (73,7% polimórficas). Em um outro estudo, conduzido por Archak et al. (2003b), cinco *primers* RAPD e quatro ISSR não ancorados foram utilizados na caracterização *fingerprinting* de 35

variedades elites de cajueiro, sendo 24 seleções e 11 híbridos, desenvolvidas por diferentes centros de pesquisa da Índia. A partir de 94 locos amplificados foi possível diferenciar, com alto grau de certeza, todos os genótipos. Por lado, os resultados evidenciaram que base genética entre eles é muito estreita, apesar da grande amplitude geográfica entre os locais de obtenção dos cultivares.

No trabalho de revisão taxonômica do gênero *Anacardium*, Mitchell e Mori (1987) consideram *A. othonianum* e *A. occidentale* como ecótipos da mesma espécie, com o primeiro circunscrito à região dos cerrados do Brasil Central e o último tendo evoluído e sido domesticado nas áreas de dunas e restingas das zonas costeiras da América do Sul. Por outro lado, os autores reconhecem *A. humile* como uma espécie válida de *Anacardium*, simpátrica a *A. othonianum*, embora com distribuição geográfica mais ampla. Mitchell e Mori (1987) relatam ainda a ocorrência ocasional de híbridos entre estas duas espécies, uma vez que elas apresentam morfologia floral muito similar e sobreposição nas épocas de florescimento. Entretanto, as duas espécies são facilmente distinguíveis pelo hábito de crescimento: *A. humile* é sempre um subarbusto com tronco subterrâneo e ramos ascendentes rígidos de 30 a 150 cm de altura, enquanto *A. othonianum* é uma árvore de porte mediano de 3 a 6 metros de altura, com tronco fissurado de 20 a 40 cm de diâmetro e folhas coriáceas.

Dentre os acessos mantidos pelo BAG, e utilizados no presente trabalho, todos os indivíduos de *A. humile* e *A. othonianum* apresentam os fenótipos típicos descritos para as respectivas espécies. Entretanto, esses indivíduos foram introduzidos por sementes coletadas de populações naturais de diferentes localidades, conforme descrito no Quadro 1. Este último aspecto poderia explicar a grande divergência relativa apresentada pelos acessos destas duas espécies, evidenciada no gráfico de projeção das distâncias da Figura 2. Esta maior divergência relativa, associada a um número reduzido de indivíduos amostrados, pode ter contribuído para má classificação destes acessos, nas análises discriminantes baseada no método de *k* vizinhos mais próximos. Também é importante destacar que, apesar da grande proporção de classificações errôneas de acessos de ‘humile’ e ‘othonianum’ em grupos de *A.*

occidentale, apenas um acesso de *A. occidentale* foi classificado erradamente como ‘humile’, na análise discriminante envolvendo ecótipos e / ou espécies.

4.3 Análise de variância molecular e de agrupamento entre as populações

A maior parte da variabilidade genética acessada se encontra dentro das populações de *Anacardium*, enquanto 27,5% da variação é resultante da diferença entre populações ($\phi_{ST} = 0,2749$). Este valor é compatível com dados médios de ϕ_{ST} estimados por Nybom (2003), para populações de espécies perenes, alógamas e de sucessão tardia, a partir de diversos estudos conduzidos com marcadores RAPD em espécies de plantas silvestres. A média de ϕ_{ST} , obtida a partir de trabalhos conduzidos com marcadores ISSR, mas sem levar em conta diferenças de histórias de vida das espécies, foi igual a $0,35 \pm 0,25$ (Nybom, 2003). Este valor não diferiu das estimativas de ϕ_{ST} compiladas pela autora a partir de trabalhos conduzidos com marcadores RAPD ou AFLP. Portanto, segundo a autora, o tipo de marcador dominante seria de pouca importância nos valores estimados de ϕ_{ST} .

Por outro lado, em um trabalho realizado com populações de pimenta longa (*Piper hispidinervum*), uma espécie arbustiva pioneira da região amazônica, Wadt e Kageyama (2004) obtiveram um valor de $\phi_{ST} = 0,281$, a partir de marcadores RAPD. Por sua vez, Lacerda et al. (2001) estimaram $\phi_{ST} = 0,123$, para populações de vinhático (*Plathymeria reticulata*), uma leguminosa arbórea comum do cerrado do sudeste brasileiro. Enquanto Cavers et al. (2003) obtiveram estimativas de $\phi_{ST} = 0,20$ e $\phi_{ST} = 0,47$ entre populações de cedro (*Cedrela odorata*) pertencentes à ecótipos de regiões áridas e úmidas, respectivamente, da Costa Rica. Embora todos estes trabalhos envolvam a avaliação de espécies arbóreas tropicais e predominantemente alógamas, os resultados obtidos ilustram o quão variáveis podem ser as estimativas de ϕ_{ST} .

As “populações” de *Anacardium* do presente trabalho não constituem, necessariamente, amostras de unidades biológicas intercruzantes, mas conjuntos de genótipos que compartilham uma origem geográfica comum, não

obstante muito amplas e delimitadas de maneira arbitrária. A única exceção é a população 'Rio de Janeiro'. Por isso, era esperado que a maior parte da variabilidade acessada se concentrasse mesmo dentro das populações. Por outro lado, o fato do trabalho incluir "populações" de espécies diferentes contribuiria para a obtenção de um valor de ϕ_{ST} superestimado, comparado com o valor que seria obtido se o estudo envolvesse apenas grupos intra-específicos. Entretanto, a comparação estatística entre os pares de populações analisadas mostrou que a população mais divergente foi 'Rio de Janeiro', enquanto as populações constituídas por acessos de *A. humile* e *A. othonianum* não diferiram entre si.

Anacardium humile se constitui em uma espécie válida de *Anacardium*, segundo todos os autores que estudaram o gênero. Enquanto *A. othonianum* seria outra espécie ou, no mínimo, um ecótipo de *A. occidentale*, isolado evolutivamente do ecótipo de restinga, ao qual pertenceriam as demais populações avaliadas, inclusive 'Rio de Janeiro'. Sob estas circunstâncias seria esperado que os maiores níveis de divergência, em ordem decrescente de grandeza, fossem observados entre populações pertencentes a espécies e ecótipos diferentes, respectivamente.

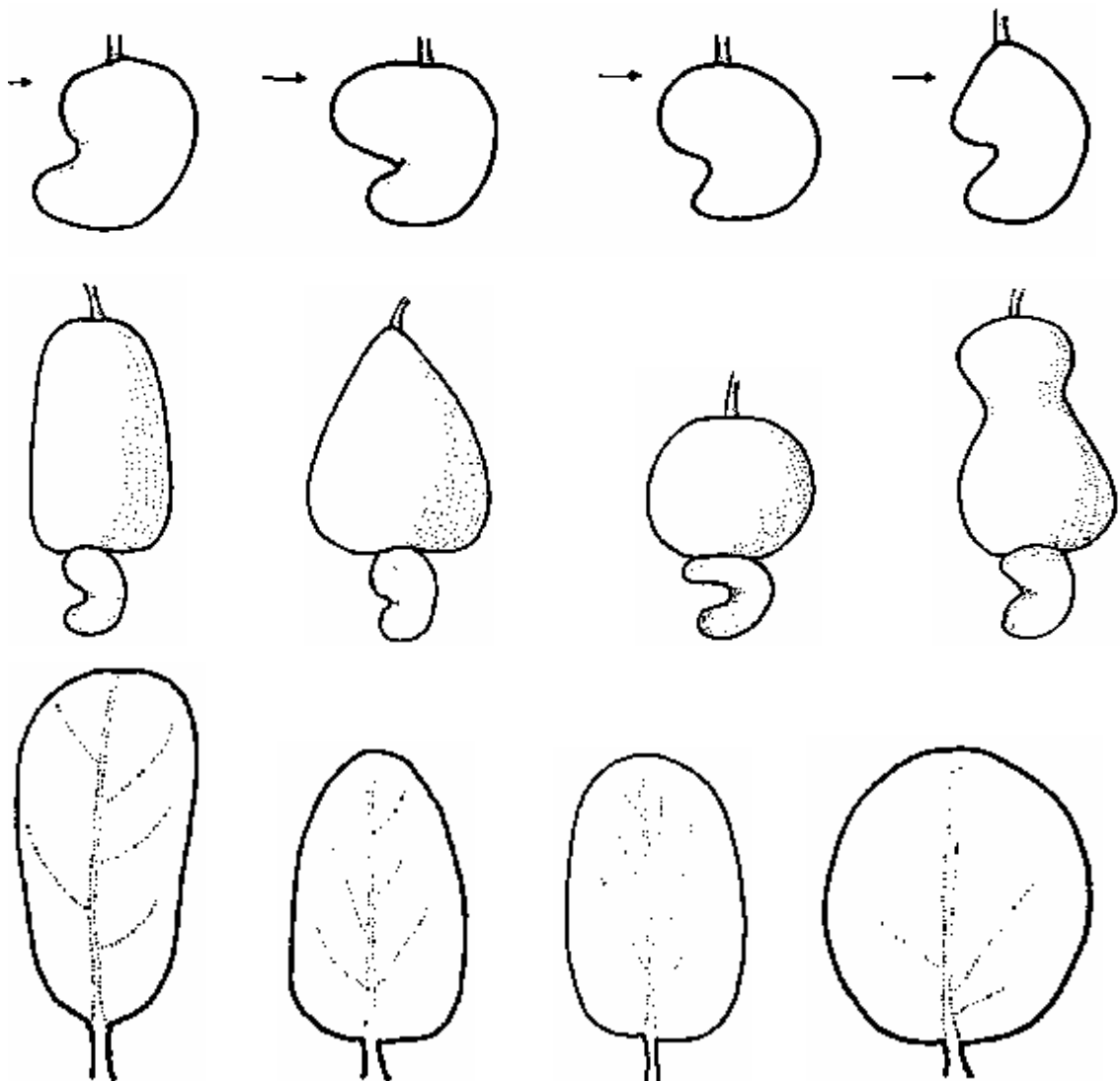
5 CONCLUSÕES

- a) Os dois sistemas de resolução empregados, gel de poliacrilamida e gel de agarose, foram adequados para detectar níveis elevados de polimorfismos, gerados pelos diferentes *primers* utilizados;
- b) O número mínimo de locos marcadores, necessário para acessar o mesmo nível de variabilidade, foi estimado em 155, considerando um coeficiente de correlação de Pearson (r) igual a 95%;
- c) A análise discriminante pelo método de k vizinhos mais próximos mostrou-se adequada apenas na classificação de acessos de diferentes populações de *A. occidentale*;

- d) Os padrões de divergência observados entre os acessos de *A. othonianum* e de *A. humile* não foram consistentes com os esperados para espécies distintas;
- e) Apesar da ampla diversidade observada entre os acessos do BAG do cajueiro, verificou-se que a variabilidade exibida pela população 'Rio de Janeiro' não está representada nesta coleção;
- f) Os percentuais da variação entre e dentro de populações, obtidos pela decomposição da variação total por meio da AMOVA, foram compatíveis com o esperado para espécies perenes de polinização mista;
- g) O padrão de agrupamento dos acessos pelo método UPGMA não refletiu, inequivocamente, a similaridade inferida na definição *a priori* das populações analisadas, à exceção da população 'Rio de Janeiro'.

CAPÍTULO 3

ANÁLISE DA DIVERSIDADE EM *Anacardium* spp. BASEADA EM CARACTERES FENOTÍPICOS DISCRETOS



Desenhos esquemáticos extraídos de Cashew Descriptors (IBPGR, 1986)

1 INTRODUÇÃO

Em taxonomia de plantas, características não métricas tais como forma, cor e textura são utilizadas com frequência. Este tipo de descritor fenotípico apresenta, em geral, estados de caráter que são mutuamente exclusivos, isto é, se expressam de modo descontínuo ou semidescontínuo. Por isso, tais descritores desempenham papel importante na classificação de unidades taxonômicas, uma vez que a probabilidade de erro é pequena (Engels, 1983a). Esta classe de caracteres também é muito importante na caracterização de recursos genéticos vegetais uma vez que são, em geral, altamente herdáveis, se expressam de modo consistente em todos os ambientes e podem ser avaliadas de forma rápida e direta à vista desarmada. Por isso, listas de descritores têm sido elaboradas por órgãos internacionais como o IPGRI (International Plant Genetic Resources Institute) para várias espécies cultivadas e suas afins silvestres, envolvendo características tais como cor da flor, formato e tamanho dos frutos, formato das folhas, época de florescimento etc. (Perry e Bettencourt, 1995).

Além da facilidade de avaliação, características fenotípicas discretas apresentam, em muitos casos, importância agrônômica, industrial ou mercadológica, relacionadas com aspectos tais como facilidade no manejo da cultura, processamento da produção e aceitação comercial do produto. Para a cultura do cajueiro (*Anacardium occidentale* L.), uma lista de descritores foi disponibilizada pelo IBPGR (atual IPGRI) em 1986 (IBPGR, 1986). Estes descritores são divididos nas seguintes categorias: dados de passaporte, caracterização e avaliação preliminar e caracterização e avaliação adicional.

Neste trabalho foi seccionado um subconjunto de descritores fenotípicos discretos da lista elaborada pelo IBPGR (1986) para a cultura do cajueiro, considerando a facilidade de observação e possível importância econômica dos caracteres, com objetivo de avaliar sua eficiência na discriminação de acessos *Anacardium* spp. de bancos de germoplasma.

2 MATERIAL E MÉTODOS

2.1 Material vegetal

As informações utilizadas para o estudo foram obtidas de 69 acessos do Banco Ativo de Germoplasma do Cajueiro da Embrapa Agroindústria Tropical, localizado no município de Pacajus – CE. Cada acesso foi representado por um indivíduo apenas, cujas denominações e origens geográficas são relacionadas a seguir:

- a) *A. occidentale* tipo comum – 49 acessos: Ceará (26), Roraima (8), Pará (3), São Paulo (2), Goiás (1), Mato Grosso (1), Piauí (1), Paraíba (1), Índia (4) e Venezuela (2);
- b) *A. occidentale* tipo anão precoce - 10 acessos coletados no Ceará;
- c) *A. othonianum* (Rizzini) – 7 acessos: Goiás (5), Mato Grosso (1) e Piauí (1);
- d) *A. microcarpum* (Ducke) – 2 acessos coletados no Ceará;
- e) *Anacardium* sp. – 1 acesso coletado no Ceará.

2.2 Coleta de dados

As avaliações foram realizadas durante a fase reprodutiva, nos meses de outubro e novembro de 2005. As características avaliadas foram selecionadas a partir da lista de descritores do cajueiro, definida pelo IBPGR (1986), totalizando 18 variáveis referentes à folha, inflorescência, pedúnculo, castanha e hábito de crescimento, conforme descrito a seguir:

- Folha

- a) Formato do limbo: 1-obovado, 2-ovado, 3-oblongo, 4-circular;
- b) Formato do ápice do limbo: 1-pontiagudo, 2-arredondado, 3-indentado;
- c) Ângulo de inserção no caule: 1-agudo, 2-obtuso;
- d) Plano de secção do limbo: 1-plano, 2-reflexo, 3-encurvado, 4-torcido;

- Pedúnculo

- a) Formato geral: 1-cilíndrico, 2-cônico a obovado, 3-redondo, 4-piriforme;
- b) Formato da base: 1-angular, 2-arredondado, 3-aplanado, 4-obliquamente aplanado;

- c) Formato do ápice: 1-nivelado, 2-oblíquo;
- d) Cavidade do ápice: 1-ausente, 2-superficial, 3-profunda;
- e) Cor do pedúnculo maduro: 1-amarelo, 2-amarelo-laranja, 3-laranja, 4-vermelho-laranja, 5-vermelho;

-Castanha

- a) Formato da base: 1-arredondado, 2-apanado, 3-obliquamente aplanado, 4-angular;
- b) Sutura: 1-arredondada, 2-angular;
- c) Posição relativa sutura x ápice: 1-sutura à frente do ápice, 2-sutura alinhada com ápice, 3-ápice à frente da sutura;
- d) Flancos: 1-apanado, 2-arredondado, 3-alargado;
- e) Formato do ápice: 1-arredondado, 2-intermediário, 3-pontiagudo;

-Inflorescência

- a) Formato geral: 1-cônica, 2-deltóide, 3-piramidal;
- b) Grau de compactação: 1-compacta, 2-frouxa;
- c) Tipo de ramificação: 1-ao redor de todo o eixo principal, 2-ao longo de dois lados em um plano;

-Hábito de crescimento

- a) Formato da copa: 1-ereta e compacta, 2-ereta e aberta, 3-aberta e projetada horizontalmente.

A determinação das categorias de cada acesso, para cada variável considerada, foi realizada com o auxílio de desenhos esquemáticos da lista de descritores do cajueiro (IBPGR, 1986).

2.3 Análises estatísticas

Os dados foram codificados para um sistema binário, de modo que as múltiplas categorias de uma variável foram representadas por diferentes colunas da matriz, utilizando valores iguais a 1 na coluna correspondente ao estado de caráter do acesso em questão e zeros nas demais (Cruz, 2006a).

Para cada variável, foram calculadas as freqüências relativas dos diferentes estados de caráter e estimados os índices de diversidade de Shannon-Wiener normalizados, obtidos pela expressão (Bekele e Bekele, 1996):

$$H' = \frac{-\sum_{i=1}^n P_i \log(P_i)}{\log(n)}, \text{ onde}$$

n = número de classes de estados de caráter da variável;

P_i = freqüência do estado de caráter i da variável em análise;

\log = logaritmo (podendo ser de base 2, base e ou base 10).

A partir da matriz de dados codificados foram calculadas as dissimilaridades entre todos os pares de acessos, por meio das seguintes informações (Cruz, 2006a):

a_j : número de concordância do tipo 1-1 para a j -ésima variável;

b_j : número de discordância do tipo 1-0 para a j -ésima variável;

c_j : número de discordância do tipo 0-1 para a j -ésima variável.

O índice de dissimilaridade (d_{ij}) foi obtido pela expressão:

$$d_{ij} = \sum_{j=1}^v \frac{b_j + c_j}{a_j + b_j + c_j}$$

A partir da matriz de dissimilaridade entre os acessos foi realizada a análise discriminante, baseada no método não paramétrico de k vizinhos mais próximos, considerado como critério de classificação *a priori* a espécie dos acessos, ou seja:

a) *A. occidentale* – 58 acessos;

b) *A. othonianum* – 7 acessos.

Três acessos não foram incluídos nesta análise, dois identificados como *A. microcarpum* e um como *Anacardium sp.*

O valor de k foi pré-estabelecido, arbitrariamente, em $k=3$ e a probabilidade *a posteriori*, de que um dado indivíduo pertença a t -ésima população, é estimada por:

$$P = \frac{\pi_t(k_t / n_t)}{\sum_{s=1}^p \pi_s(k_s / n_s)}$$

Onde

s : é número de populações avaliadas ($s = 1, \dots, p$);

n : é o número de indivíduos de cada população

π_t : é a probabilidade *a priori* de um indivíduo pertencer à população t

A observação x é classificada na t -ésima população se $\pi_t k_t / n_t$ for maior que $\pi_s k_s / n_s$ para todo s (Khattree e Naik, 2000).

A relação entre os acessos foi também avaliada através da projeção gráfica das distâncias no espaço tridimensional e por métodos de agrupamento hierárquicos.

Na análise de projeção, as distâncias entre genótipos foram convertidas em escores relativos a três variáveis (X, Y e Z), de modo que, na representação gráfica de dispersão tridimensional, refletissem as distâncias originalmente obtidas a partir do espaço n -dimensional, sendo n o número de caracteres avaliados (Cruz e Carneiro, 2004; Cruz, 2006a). Neste estudo foram considerados os acessos e respectivos grupos envolvidos na análise discriminante e a eficiência da projeção foi estimada através das seguintes estatísticas (Cruz, 2006a):

- a) Coeficiente de correlação entre as distâncias originais e as que foram representadas no gráfico de dispersão;
- b) Grau de distorção $(1-\alpha)$, considerando que:

$$\alpha = \frac{\sum_{i < i'} \sum d_{gii'}^2}{\sum_{i < i'} \sum d_{oii'}^2}$$

Onde $d_{gii'}^2$ e $d_{oii'}^2$ são as distâncias gráficas (espaço tridimensional) e originais, respectivamente, de todos os pares de indivíduos i e i' .

- c) Coeficiente de estresse, de Kruskal (1964), dado por:

$$s = \frac{\sum_{i < i'} \sum (d_{oij} - d_{gij})^2}{\sqrt{\sum_{i < i'} \sum d_{oii'}^2}}$$

A análise de agrupamento por métodos hierárquicos foi realizada considerando os 69 acessos avaliados. Foram utilizados os métodos UPGMA (médias aritméticas entre todos os pares de acessos) e da variância mínima de Ward (Ward, 1963).

3 RESULTADOS

3.1 Distribuição e Índices de Diversidade dos Caracteres

Houve ampla variação para a grande maioria das características avaliadas. Apenas um descritor, 'formato do ápice do pedúnculo', foi monomórfico, enquanto o caráter 'tipo de ramificação da inflorescência', apresentou o estado alternativo em apenas um acesso e o caráter 'formato do ápice da castanha' apresentou apenas duas das três classes possíveis. As demais características, além de variáveis, apresentaram frequências de estados de caráter relativamente equitativas, conforme atestam os índices de diversidade de Shannon-Wiener que, na maioria dos casos, foi superior a 0,7 (Tabela 1).

3.2 Análise discriminante e projeção das distâncias

Os resultados da análise discriminante (Tabela 2) demonstram que a discriminação pelo método do k vizinhos mais próximos produziu proporção de alocações incorretas maior entre os acessos de *A. othonianum*, enquanto na alocação pelo método do vizinho médio todos os acessos de *A. othonianum* foram classificados de maneira apropriada, mas grande parte dos acessos de *A. occidentale* foi alocada incorretamente no grupo *othonianum*. Por outro lado, a taxa erro aparente (TEA) global entre dois métodos mostra que o método de k vizinhos mais próximos foi mais eficiente na alocação dos acessos que o método do vizinho médio, com um valor de TEA quase três vezes menor.

A projeção gráfica das distâncias, apresentada na Figura 1, evidencia sobreposição na dispersão dos acessos das duas populações, o que dificulta uma discriminação adequada dos grupos. Por outro lado, a correlação entre distâncias originais e estimadas foi apenas de 0,46, enquanto o grau de distorção foi 40,01% e o coeficiente de stress igual a 46,3%. Por isso, a análise destas relações deve ser realizada com a devida cautela, considerando as restrições de confiabilidade que elas apresentam.

Tabela 1 - Freqüências relativas e índices de diversidade Shannon-Wiener normalizados de 18 variáveis fenotípicas discretas, avaliadas em 69 acessos de *Anacardium* sp. do BAG do cajueiro, Embrapa-CNPAT

Caracteres Avaliados	Freqüências de Estados de Caráter					Índice Shannon-Wiener
	1	2	3	4	5	
Folha						
Formato do limbo	0,41	0,10	0,41	0,08	-	0,8486
Formato do ápice	0,03	0,59	0,38	-	-	0,7097
Ângulo de inserção	0,64	0,36	-	-	-	0,9446
Plano de secção	0,25	0,42	0,07	0,26	-	0,9018
Pedúnculo						
Formato geral	0,10	0,59	0,28	0,03	-	0,7207
Formato da base	0,39	0,28	0,17	0,16	-	0,9516
Formato do ápice	1,00	0,00	-	-	-	Indet.
Cavidade do ápice	0,16	0,68	0,16	-	-	0,7710
Cor do pedúnculo	0,35	0,12	0,17	0,23	0,13	0,9481
Castanha						
Formato da base	0,36	0,27	0,25	0,12	-	0,9507
Formato da sutura	0,49	0,51	-	-	-	0,9998
Posição sutura x ápice	0,54	0,43	0,03	-	-	0,7272
Angulação dos flancos	0,14	0,41	0,45	-	-	0,9151
Formato do ápice	0,65	0,35	0,00	-	-	0,5893
Inflorescência						
Formato geral	0,01	0,35	0,64	-	-	0,6514
Grau de compactação	0,29	0,71	-	-	-	0,8685
Tipo de ramificação	0,99	0,01	-	-	-	0,1093
Hábito de Crescimento						
Formato da copa	0,09	0,68	0,23	-	-	0,7399

(-) estados não aplicáveis ao caráter em questão.

Tabela 2 – Análise discriminante de acessos de duas espécies de *Anacardium*, baseada nos métodos de *k* vizinhos mais próximos (*k*=3) e vizinho médio, considerando a matriz de distâncias obtida da avaliação de 17 variáveis fenotípicas discretas.

Grupo de Origem	Nº de Indivíduos	Percentuais de Alocação nos grupos*			
		Método <i>k</i> vizinhos mais próximos		Método do vizinho médio	
		<i>A. occidentale</i>	<i>A. othonianum</i>	<i>A. occidentale</i>	<i>A. othonianum</i>
<i>A. occidentale</i>	59	91,53	8,47	66,10	33,90
<i>A. othonianum</i>	7	28,57	71,43	0,00	100
Total de Indiv. alocados	66	56	10	39	27
Taxa de erro aparente (%)		10,60		30,3	

* valores em negrito destacam os percentuais de alocação de acessos nos respectivos grupos de origem.

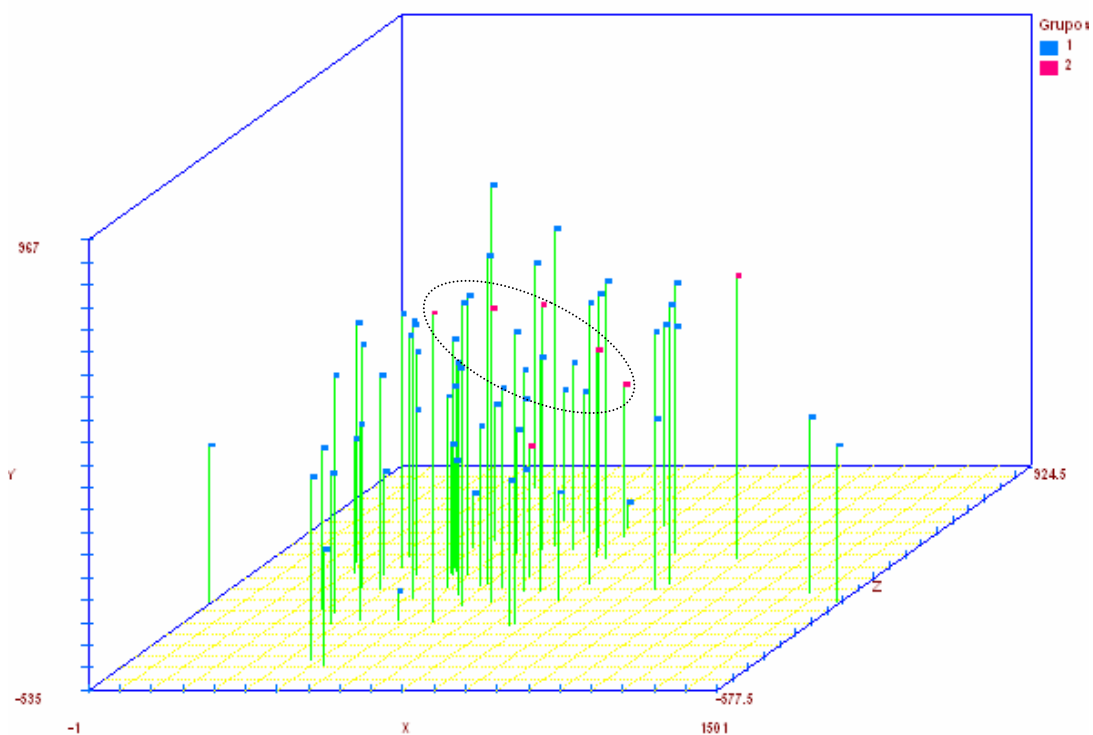


Figura 1 – Projeção gráfica das distâncias entre 66 acessos de *Anacardium* spp., discriminados em duas populações diferentes. Populações ou grupos: 1 - *A. occidentale*; 2 - *A. othonianum*. Nota: elipse delimita a região de dispersão da maioria dos acessos de *A. othonianum*.

3.3 Análise de agrupamento

Os dendrogramas de agrupamento dos 69 acessos obtidos pelos métodos UPGMA e Ward são apresentados, respectivamente, nas Figuras 2 e 3. Conforme evidenciado pelos dois métodos, as medidas de dissimilaridades entre pares de acessos, obtidas pela combinação das 17 variáveis consideradas, possibilitou a distinção de todos os indivíduos avaliados sem ambigüidades.

Uma linha de corte traçada a 75% da variação total, evidencia a formação de seis subgrupos maiores, isto é, com cinco ou mais acessos (destacados pelas barras verticais à esquerda) e vários subgrupos menores, no agrupamento obtido por UPGMA. Por outro lado, o mesmo ponto de corte no dendrograma de Ward produz apenas sete subgrupos que abrangem todos os acessos.

Nos dois dendrogramas não houve a formação de subgrupos que envolvessem a maioria dos acessos de uma espécie ou ecótipo. Pelo contrário, os acessos de cajueiro anão precoce (AP) ficaram dispersos em vários subgrupos, assim como os acessos de *A. othonianum* (OT). Embora, neste último caso, tenha ocorrido uma tendência de agrupamento dos acessos OT nos dois subgrupos localizados na parte superior do dendrograma.

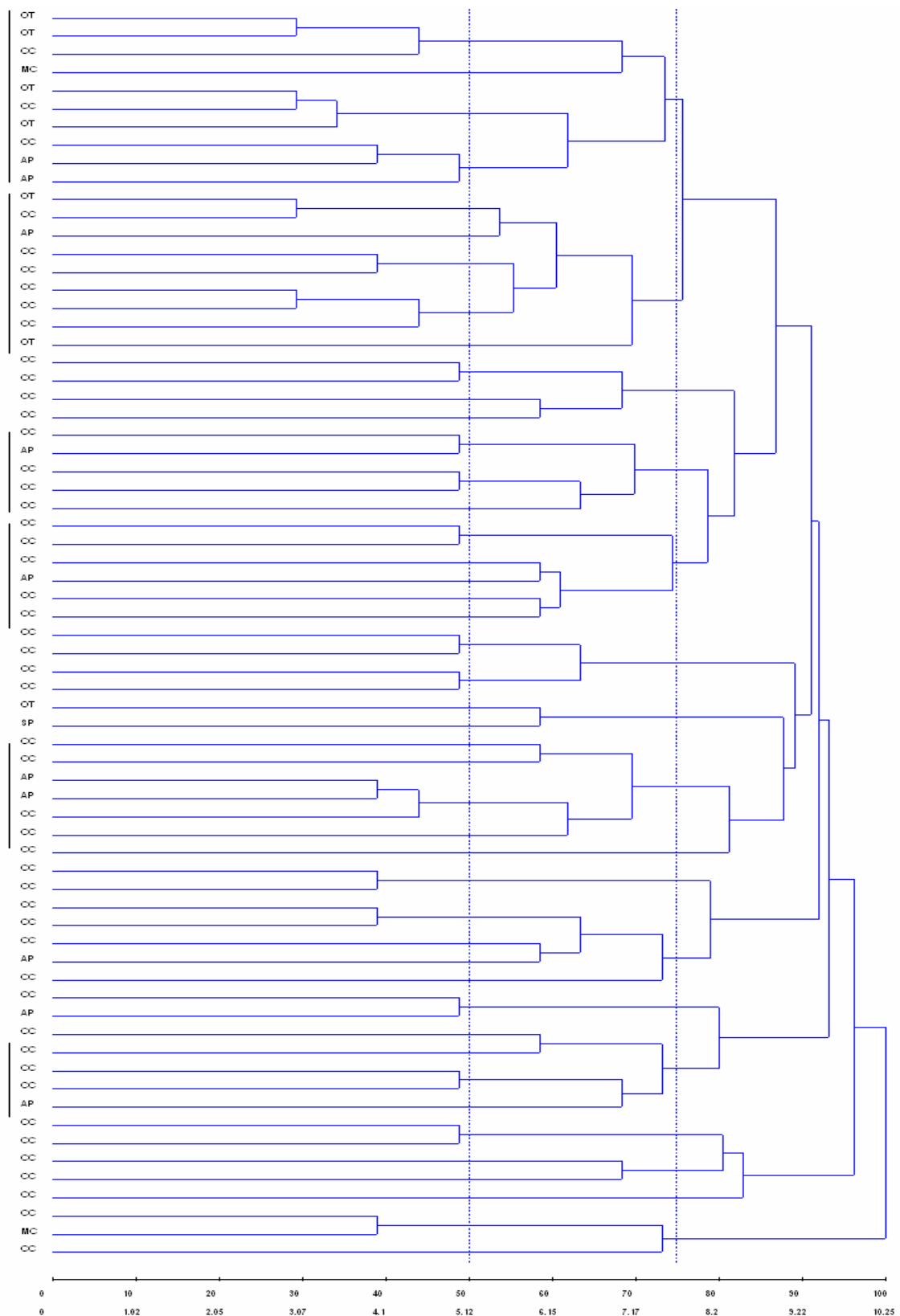


Figura 2 - Agrupamento UPGMA de 69 acessos do BAG do Cajueiro, Embrapa-CNPAT-CE, gerado a partir da matriz de distâncias obtida da avaliação de 17 variáveis fenotípicas discretas. Códigos: OT – *A. othonianum*, CC cajueiro comum (*A. occidentale*), AP – cajueiro anão precoce (*A. occidentale*), MC – *A. microcarpum*, SP – *Anacardium* sp.

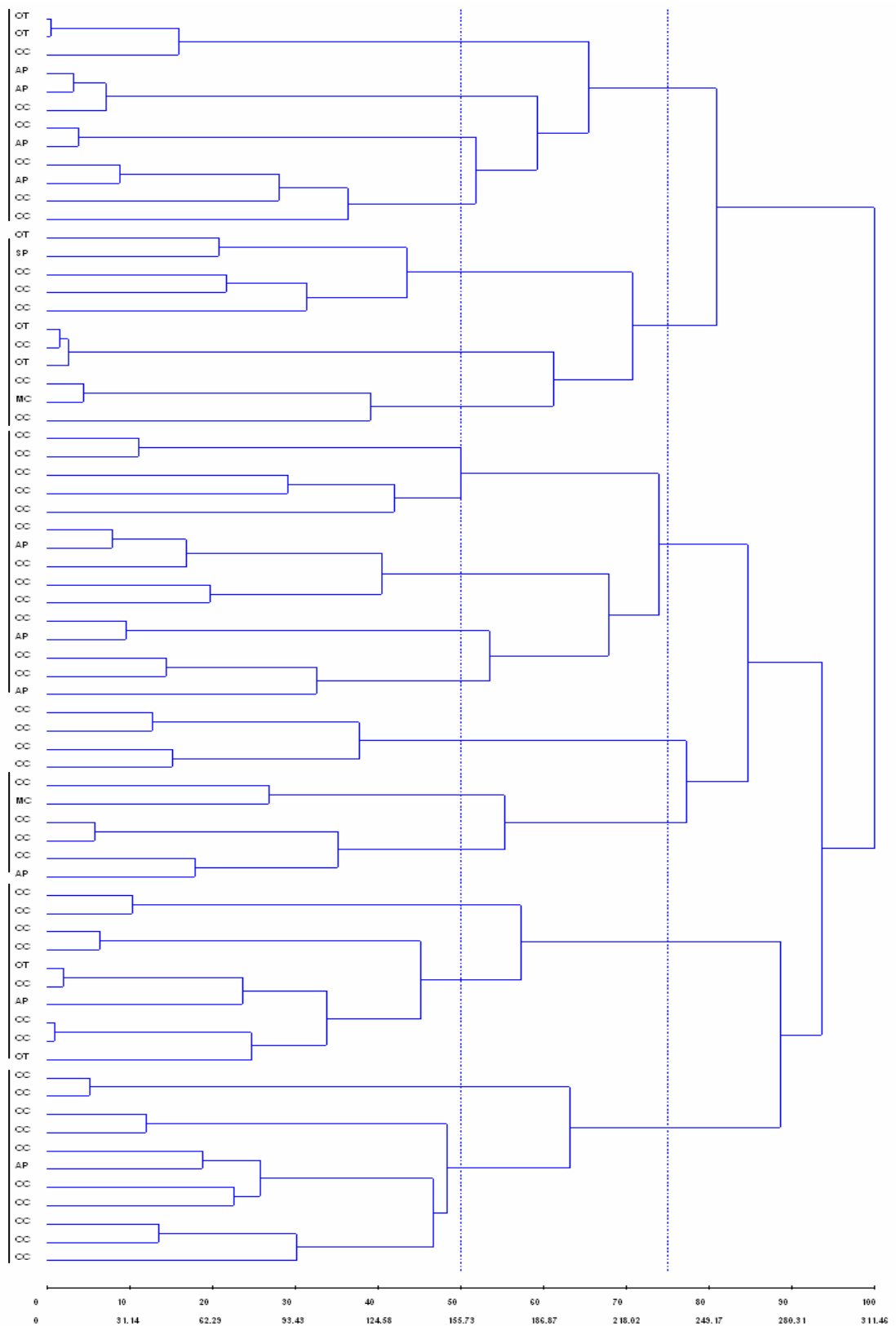


Figura 3 – Dendrograma de agrupamento, baseado no método da variância mínima de Ward, de 69 acessos do BAG do Cajueiro, Embrapa-CNPAT-CE, gerado a partir da matriz de distâncias obtida da avaliação de 17 variáveis fenotípicas discretas. Códigos: OT – *A. othonianum*, CC cajueiro comum (*A. occidentale*), AP – cajueiro anão precoce (*A. occidentale*), MC – *A. microcarpum*, SP – *Anacardium* sp.

4 DISCUSSÃO

Estudos de diversidade genética utilizando descritores morfológicos não métricos são comuns em trabalhos envolvendo a caracterização de recursos genéticos vegetais (Engels, 1983b; Bekele e Bekele, 1996; Colombo et al., 1998; Bajracharya et al., 2005; Duran et al., 2005). Por outro lado, os resultados obtidos nem sempre são satisfatórios no que se refere à diferenciação de todos os acessos analisados (Colombo et al., 1998; Bajracharya et al., 2005). Engels (1983a) afirma que a utilidade de um descritor qualitativo depende do número de estados de caráter que esse descritor pode assumir, assim como da quantidade de informação que ele compartilha com outros descritores utilizados no mesmo estudo.

Neste trabalho foram utilizados 18 descritores fenotípicos discretos selecionados da lista de descritores elaborada pelo IBPGR (1986) para o cajueiro. A seleção teve como principais critérios a facilidade de determinação do estado do caráter do descritor e sua associação aos produtos de importância econômica da planta, ou seja, castanha e pedúnculo. Das 56 classes fenotípicas esperadas, para o conjunto dos 18 descritores avaliados, foram observadas 54. Apenas um descritor, formato do ápice do pedúnculo, foi monomórfico, enquanto o caráter 'formato do ápice da castanha' apresentou apenas duas das três classes possíveis.

Os índices normalizados de diversidade de Shannon-Wiener apresentaram valores acima de 0,50 para 16 descritores e a acima de 0,8 para nove dos 18 descritores. Estes valores indicam que, além de variáveis, a maioria dos descritores apresentou freqüências de classes relativamente eqüitativas, atributos que lhes conferem elevado poder de discriminação. O poder discriminatório de descritores com estas características foi observado por Bekele e Bekele (1996) ao avaliarem 100 acessos do banco internacional de germoplasma do cacaueteiro, utilizando 28 variáveis quantitativas e 26 qualitativas. Entre os descritores fenotípicos discretos, 69% apresentaram índices de diversidade maior que 0,5. Por isso, de acordo com os autores, este conjunto de variáveis teve um peso muito maior que os descritores morfométricos na discriminação dos acessos avaliados. Por outro lado, Bajracharya et al. (2005) não conseguiram discriminar adequadamente 147

raças locais de arroz do Nepal, utilizando 27 descritores fenotípicos discretos, uma vez que apenas 12 apresentaram alguma variação, mas com índice médio de diversidade Shannon-Wiener igual a 0,23.

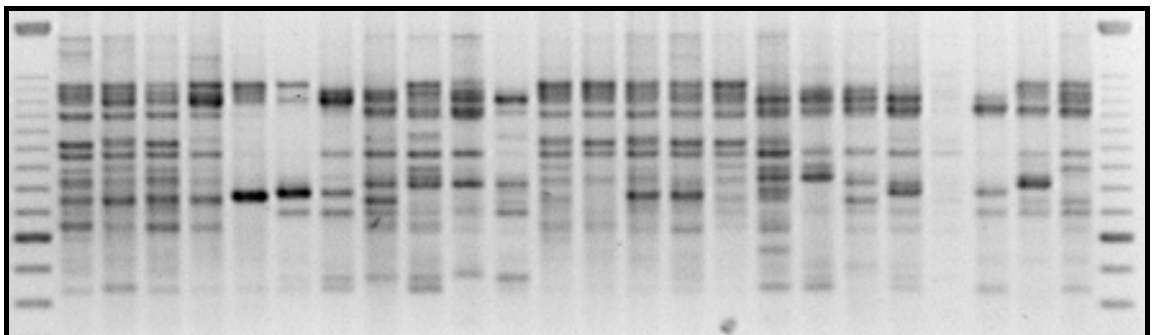
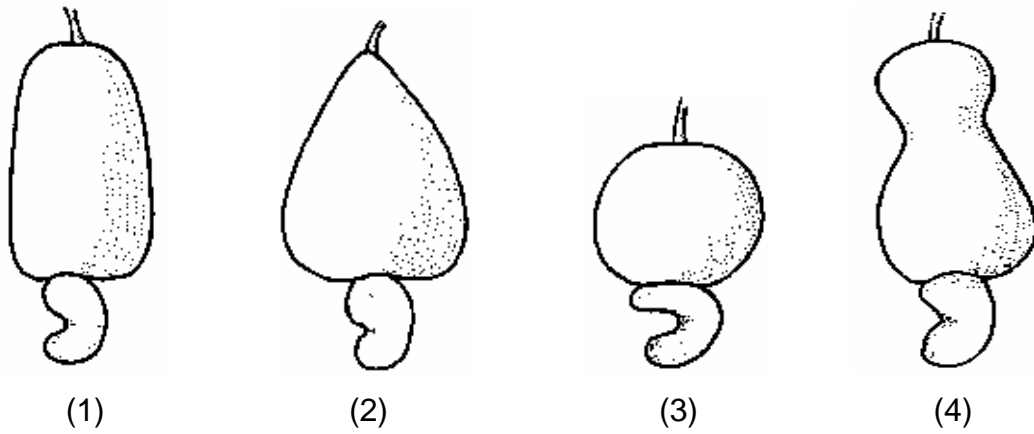
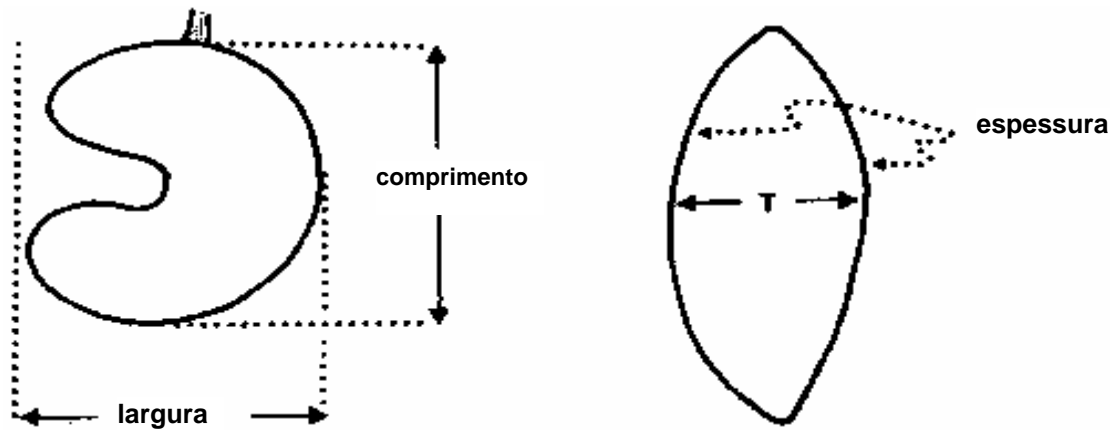
Os descritores utilizados no presente estudo apresentaram variação suficiente para discriminar adequadamente os 69 acessos de *Anacardium* avaliados, conforme evidenciado pelas Figuras 2 e 3. De acordo com os resultados da Tabela 2, estes descritores também foram razoavelmente eficientes na diferenciação de acessos pertencentes a *A. occidentale* e *A. othonianum*, quando empregados na análise discriminante pelo método não paramétrico baseado em k vizinhos mais próximos. Em um trabalho de caracterização de 31 cultivares de mandioca, através de 27 descritores fenotípicos discretos, realizado por Colombo et al. (1998), foi possível diferenciar um grupo de cultivares plantados para consumo *in natura* de outro grupo constituído por cultivares destinados à fabricação de farinha. Duran et al. (2005) ao avaliarem 65 raças locais de feijão comum da região do Caribe, utilizando sete descritores fenotípicos discretos e seis morfométricos, conseguiram distinguir os acessos de origem Andina daqueles de origem Mesoamericana, resultados que foram confirmados pela análise molecular dos tipos proteína faseolina dos acessos e por marcadores RAPD. Os padrões de agrupamento exibidos pelos dendrogramas das Figuras 2 e 3, por outro lado, demonstram que não houve a formação de subgrupos associados com a espécie, ecótipo ou origem geográfica dos acessos. Estes resultados indicam que tais descritores são adequados para discriminar acessos, mas inapropriados para discriminar grupos. Isto ocorreu porque, aparentemente, não houve classes fenotípicas mutuamente exclusivas dos descritores entre os grupos avaliados. Dessa forma, a variabilidade acessada encontra-se dispersa entre todos os acessos avaliados, indicando que a divergência genética não está associada com grupo de origem.

5 CONCLUSÕES

- a) A maioria dos descritores analisados apresentou ampla variabilidade entre os acessos avaliados.
- b) As informações geradas pelo conjunto de descritores foram adequadas para discriminar, sem ambigüidades, todos os acessos de *Anacardium* avaliados.
- c) A utilização de um número menor de descritores, do que o inicialmente recomendado para avaliar os recursos genéticos do cajueiro, parece ser apropriada, resultando em economia de tempo e esforços no processo de caracterização preliminar dos acessos.
- d) Por outro lado, os padrões de distribuição da variabilidade acessada foram inadequados para caracterizar grupos de acessos de origem comum, exceto no caso dos grupos acessos de *A. occidentale* x *A. othonianum*, que puderam ser razoavelmente distinguidos através da análise discriminante pelo método não paramétrico de *k* vizinhos mais próximos.

CAPÍTULO 4

ANÁLISE COMPARATIVA DA DIVERSIDADE EM *Anacardium* spp. ACESSADA POR MEIO DE DADOS MORFOMÉTRICOS, FENOTÍPICOS DISCRETOS E MOLECULARES



1 INTRODUÇÃO

O estudo da diversidade genética é um processo através do qual a variação entre indivíduos ou entre populações é examinada empregando um método específico ou combinação de métodos (Mohammadi e Prasanna, 2003). Atualmente, existe grande número de métodos disponíveis para análise da diversidade genética, seja em avaliações de acessos de bancos de germoplasma, cultivares melhoradas ou populações. De acordo com Franco et al. (2001), para alcançar os objetivos almejados, vários tipos de atributos são medidos em cada genótipo: (1) variáveis fenotípicas contínuas, tais como características morfoagronômicas; (2) variáveis fenotípicas discretas ou fenotípicas discretas e (3) marcadores moleculares discretos binários, tais como RFLPs, AFLPs, RAPDs, entre outros. Assim, para um total de p atributos, cada genótipo pode ser visualizado como sendo localizado em um espaço p multidimensional no qual cada dimensão é representada por um atributo.

De um modo geral, de acordo com Franco et al. (2001), quando se dispõe de dados fenotípicos morfológicos e de marcadores moleculares para um conjunto de genótipos, os estudos de diversidade e de classificação hierárquica são realizados de forma independente. Uma análise é baseada nos dados morfológicos, através do cálculo de uma distância métrica padrão (tal como o quadrado da distância Euclidiana) e da aplicação de uma estratégia de agrupamento como UPGMA ou Ward. Outra classificação é obtida baseada nos atributos moleculares, através da determinação da similaridade (ou dissimilaridade) genética entre os indivíduos, também seguido da aplicação de uma estratégia de agrupamento dos genótipos (p. ex. Colombo et al., 1998; Nielsen et al., 2003; Baxevanis et al., 2005).

Por outro lado, em muitos casos em que são realizadas avaliações de caracteres morfométricos e fenotípicos discretos, os dados são analisados como um único conjunto de atributos, assumindo que a ordenação arbitrária atribuída às variáveis discretas constitui diferenças de grau entre os acessos (p. ex. Barros, 1991; Bekele e Bekele, 1996; Duran et al., 2005). Apesar da atribuição de valores numéricos tais como 1, 2, 3 etc. para cada categoria de uma variável discreta não é apropriada a utilização de uma medida de distância

padrão, como as aplicadas a caracteres de variação contínua, uma vez que não é possível assumir que um par de genótipos com valores 1 e 4 sejam mais distantes que outro com valores 1 e 2, por exemplo (Cruz e Carneiro, 2003).

Alguns autores têm proposto métodos para combinar diferentes classes de dados, de modo que se possa obter uma única matriz de distâncias ou de dissimilaridades, que será empregada na classificação e agrupamento dos genótipos investigados em um único procedimento. Engels (1983a) propôs um método para combinar descritores quantitativos e qualitativos de germoplasma de cacau, baseado no cálculo do valor discriminatório de cada atributo avaliado. Cole-Rodgers et al. (1997) também propuseram uma forma de combinar dados fenotípicos qualitativos e quantitativos de modo que pudessem ser submetidos a análises estatísticas. Os autores utilizaram caracteres de germoplasma de *Capsicum* para ilustrar a aplicabilidade do método, mas afirmaram que sua utilidade é extensível a qualquer trabalho envolvendo a análise de recursos genéticos. Uma estratégia que possibilita a combinação de informações moleculares binárias e atributos fenotípicos foi proposta por Franco et al. (2001). Neste método, inicialmente é buscado um número mínimo de fragmentos relevantes na diferenciação dos genótipos analisados que são, posteriormente, combinados com atributos morfoagronômicos disponíveis.

De acordo com Mohammadi e Prasanna (2003), duas questões assumem considerável importância na decisão de analisar diferentes grupos de dados individualmente ou de forma combinada: (a) se a análise e a interpretação dos resultados deveriam ser baseadas nos dados de grupos individuais ou nos dados combinados dos diversos grupos, quando múltiplos grupos de dados estão disponíveis; (b) como combinar eficientemente diferentes grupos de dados.

O ponto mais importante a ser considerado, antes da combinação de grupos de dados, é a congruência ou correspondência entre os resultados derivados dos diferentes grupos de dados (Mohammadi e Prasanna, 2003). Entretanto, Burstin e Charcosset (1997) concluíram que as distâncias fenotípicas e moleculares entre genótipos não apresentam relação linear, mas “triangular” para caracteres de herança poligênica, que são normalmente utilizados para computar as distâncias fenotípicas. Assim, distâncias pequenas estimadas por marcadores moleculares são, sistematicamente, associadas

com pequenas distâncias fenotípicas, enquanto distâncias moleculares elevadas correspondem tanto a distâncias fenotípicas baixas quanto elevadas. Burst e Charcosset (1997) demonstram que a magnitude do coeficiente de correlação entre distâncias fenotípicas e moleculares depende da associação entre locos marcadores e caracteres de herança quantitativa (QTLs). Por isso, quando a associação entre locos marcadores e QTLs é baixa, o coeficiente de correlação linear entre estas duas distâncias decresce com aumento do número de QTLs envolvidos no controle das características avaliadas.

Diversos estudos analisaram as correlações entre matrizes de distâncias ou de similaridades genéticas, derivadas da aplicação de diferentes tipos de marcadores moleculares. Por outro lado, poucos trabalhos envolvendo análises de diversidade tentaram comparar os resultados derivados de grupos individuais vs combinados de diferentes classes de dados (Mohammadi e Prasanna, 2003). Pejic et al. (1998) realizaram análise comparativa da similaridade genética entre linhas endogâmicas de milho, estimadas a partir de quatro sistemas diferentes de marcadores moleculares (RAPD, RFLP, SSR e AFLP). Os coeficientes de correlação entre as diferentes matrizes de similaridade foram todos significativos variando entre 0,51 e 0,70. Trabalho semelhante foi realizado por Russell et al. (1997) em cevada. Zucchi (2002) observou a existência correlações significativas e elevadas entre distâncias genéticas e congruência entre estimativas de diversos parâmetros populacionais de *Eugenia dysenterica*, uma frutífera nativa do cerrado, estimados a partir de dados isoenzimáticos, marcadores SSR e RAPD. Por outro lado, em uma análise de genótipos de cajueiro através de variáveis morfométricas e de marcadores AFLP, RAPD e ISSR, Archak et al. (2003a) verificaram a existência de correlação significativa apenas entre as matrizes de similaridade derivadas de dados de RAPD e ISSR. Kamada (2006) comparou matrizes de dissimilaridades e padrões de agrupamento obtidos de dados morfométricos, fenotípicos discretos e de RAPD, avaliados em quatro populações de fáfia (*Pfaffia glomerata*). O autor verificou que as correlações entre as diferentes matrizes foram fracas, com o maior valor para matrizes de distâncias entre dados fenotípicos discretos e de RAPD. Correlações nulas entre dados moleculares e fenotípicos foram também observadas por Roldán-

Ruiz et al (2001) em estudo envolvendo a caracterização de variedades de centeio forrageiro.

De acordo com Mohammadi e Prasanna (2003), diante de tantos resultados contraditórios, as opiniões divergem a respeito da utilidade de se combinar grupos de dados de diferentes tipos de variáveis, com o propósito de analisar a diversidade genética. Os autores ainda chamam a atenção para o cuidado que deve ser tomado ao combinar dados de medidas quantitativas e qualitativas, em razão de possíveis vieses nas medidas de distância, baseadas em caracteres quantitativos e da alta correlação entre os caracteres qualitativos.

Neste trabalho foram comparadas as matrizes de distâncias ou dissimilaridades entre acessos de *Anacardium*, obtidas a partir de dados morfométricos, fenotípicos discretos e de marcadores ISSR. Análises discriminantes e de agrupamento dos acessos também foram realizadas, considerando as diferentes classes de dados e de suas combinações, através da soma de matrizes de divergência padronizadas.

2 MATERIAL E MÉTODOS

2.1 Material vegetal

As informações utilizadas para o estudo foram obtidas de 69 acessos do Banco Ativo de Germoplasma do Cajueiro da Embrapa Agroindústria Tropical, localizado no município de Pacajus – CE. Cada acesso foi representado por um indivíduo apenas, cujas denominações e origens geográficas foram descritas no item 2.1 do Capítulo 3.

2.2 Coleta de dados

As informações referentes a características morfométricas e fenotípicas discretas foram obtidas durante a fase reprodutiva, nos meses de outubro e novembro de 2005. As características avaliadas foram selecionadas a partir da lista de descritores do cajueiro, definida pelo IBPGR (1986), totalizando 13 variáveis morfométricas, relativas à inflorescência (3), pedúnculo (4), castanha

(4) e hábito de crescimento (2) e 18 variáveis fenotípicas discretas referentes à folha (4), inflorescência (3), pedúnculo (5), castanha (5) e hábito de crescimento (1), conforme descrito nos itens 2.2 dos Capítulos 1 e 3, respectivamente.

Os dados moleculares de marcadores ISSR foram produzidos no Laboratório de Genética Molecular de Plantas (Biomol), do Instituto de Biotecnologia Aplicada à Agropecuária (Bioagro) da UFV. O material foi coletado no final da fase reprodutiva, durante a emissão de folhagem nova, no mês de novembro de 2005. Os métodos referentes aos processos de extração DNA, seleção de *primers*, otimização das condições de amplificação e interpretação dos produtos amplificados foram descritos nos itens 2.2 a 2.7 do Capítulo 2.

2.3 Análises estatísticas dos dados

Matrizes de dissimilaridades entre todos os pares de acessos foram obtidas para cada uma das classes de dados. A matriz de dissimilaridade dos dados morfométricos foi gerada a partir do cálculo das distâncias Euclidianas médias padronizadas (Cruz e Carneiro, 2003; Cruz, 2006a). A matriz de dissimilaridade dos dados fenotípicos discretos foi estimada a partir do modelo binário qualitativo descrito no item 2.3 do Capítulo 3 e a matriz de dissimilaridade dos dados moleculares foi obtida a partir do complemento do coeficiente de similaridade de Sorenso ou Nei e Li, conforme descrito no item 2.7.1 do Capítulo 2.

As diferentes medidas de distâncias entre pares de acessos foram combinadas através da soma de matrizes, considerando todas as combinações de pares de matrizes assim como matriz soma dos três conjuntos de dados.

As associações entre as medidas de dissimilaridades obtidas das três classes de variáveis foram verificadas através da correlação entre matrizes, cujas significâncias foram testadas através do teste Z de Mantel, descrito por Manly (1986, 1998) para matrizes envolvendo dados biológicos. O valor de Z é dado por:

$$Z = \sum_{i,j=1}^n X_{ij} Y_{ij}$$

onde X_{ij} e Y_{ij} são elementos das matrizes X e Y a serem comparadas. A significância de Z é testada comparando-se o valor obtido a partir dos dados observados com uma distribuição nula, construída a partir do cálculo de Z de diversas correlações entre uma das matrizes na forma original com outra sendo aleatorizada a cada cálculo do novo coeficiente de correlação. Neste trabalho, a significância de Z foi avaliada com base em 1000 simulações aleatórias.

A utilidade na discriminação de grupos de acessos de *Anacardium*, constituídos por espécies ou ecótipos distintos, foi avaliada para cada uma das três classes de variáveis isoladamente, assim como pela suas combinações em matrizes de dissimilaridades. Os grupos definidos a priori foram constituídos por 66 dos 69 acessos avaliados assim distribuídos:

- a) 49 acessos de *A. occidentale* tipo comum, com diversas origens geográficas;
- b) 10 acessos de *A. occidentale* tipo anão precoce, coletados no município de Maranguape – CE;
- c) Sete acessos de *A. othonianum*, coletados em diferentes locais do Cerrado do Brasil Central.

Em todos os casos foi aplicado o método não paramétrico de análise discriminante baseado em k vizinhos mais próximos ($k=3$), considerando os dados da matriz de distâncias genéticas, conforme metodologia descrita no item 2.3 do Capítulo 3.

As diferentes medidas de dissimilaridade também foram comparadas quanto aos padrões de agrupamento dos acessos, produzidos pelo método de otimização de Tocher.

Todas as análises foram realizadas com auxílio do aplicativo computacional Genes (Cruz, 2001; 2006a).

3 RESULTADOS

3.1 Correlações entre medidas de dissimilaridades

As correlações entre as medidas de dissimilaridades são apresentadas na Tabela 1. Apenas as medidas de dados morfométricos e moleculares demonstraram ser, de alguma forma, correlacionadas, exibindo um coeficiente r

relativamente baixo ($r=0,365$), mas significativo pelo teste de Mantel ($p<0,01$). A matriz de dissimilaridades dos dados fenotípicos discretos apresentou correlações nulas com as outras matrizes de dissimilaridades.

Tabela 1 – Coeficientes de correlação e intervalos de confiança do teste de Mantel (1000 simulações) entre matrizes de dissimilaridade de caracteres morfométricos (matriz 1), fenotípicos discretos (matriz 2), moleculares (matriz 3), obtidas da avaliação de 69 acessos de *Anacardium* spp. do Banco Ativo de Germoplasma do Cajueiro.

Matrizes	Coeficientes de correlação	Teste de Mantel - níveis críticos	
		5%	1%
1 X 2	0,0769	0,0163 - 0,0817	0,0221 - 0,0718
1 X 3	0,3650	0,2457 - 0,3204	0,2558 - 0,3071
2 X 3	0,0043	-0,0299 - 0,0154	-0,0404 - 0,0313

A Figura 1 apresenta a relação entre as distâncias fenotípicas (matriz 1) e moleculares (matriz 3). A dispersão gráfica dos pontos não apresenta uma configuração “triangular”, diferentemente do que foi preconizado por Burstin e Charcosset (1997).

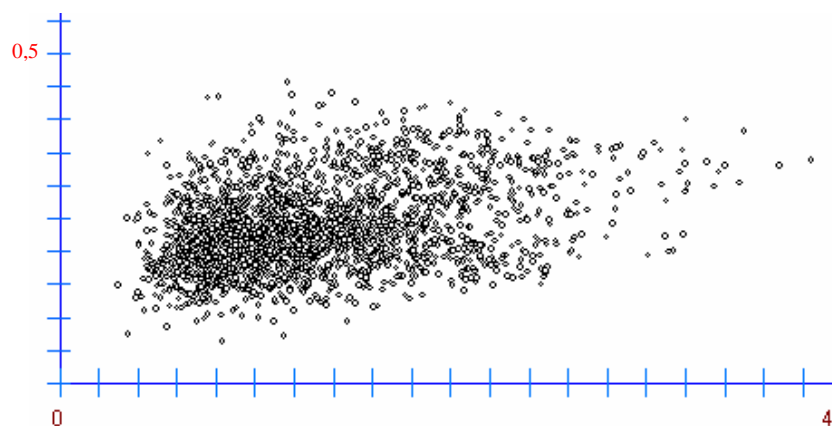


Figura 1 – Dispersão dos valores de distâncias estimadas a partir de dados morfométricos (abscissa) e moleculares (ordenada) de 69 acessos de *Anacardium* spp.

3.1 Análises discriminantes

As classificações dos acessos através de análise discriminante pelo método dos k vizinhos mais próximos, considerando as diferentes matrizes de dissimilaridades estimadas, são apresentadas na Tabela 2. As matrizes que proporcionaram menores taxas de erros aparentes foram, respectivamente, a

de dados moleculares (1) e a matriz soma entre dados moleculares e morfométricos (1+3). Enquanto a matriz de dados fenotípicos discretos (2) e a matriz soma entre dados fenotípicos discretos e morfométricos (1+2) resultaram nas maiores proporções de classificações incorretas, com TEA de 37,88% e 40,91%, respectivamente. Por outro lado, a matriz soma (1+2) foi a única que proporcionou a discriminação correta de todos os acessos de *A. othonianum*, enquanto a matriz soma (1+3) possibilitou melhor discriminação média entre todos os grupos investigados.

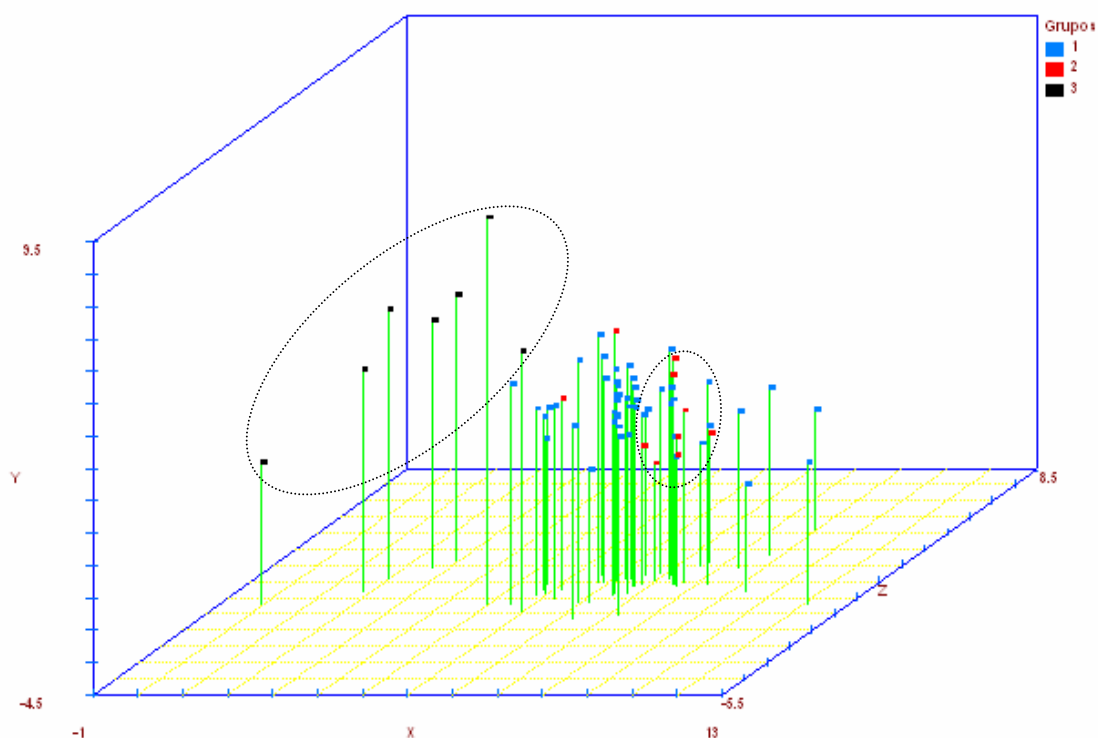
Tabela 2 – Análises discriminantes de três populações de *Anacardium* spp. obtidas pelo método não paramétrico de *k* vizinhos mais próximos (*k*=3), a partir de matrizes de distâncias de dados morfométricos (1), fenotípicos discretos (2), moleculares (3) e da combinação de medidas de distâncias, através da soma de matrizes.

Dados da matriz de distâncias	Grupo de origem	Nº de indivíduos	Percentuais de alocação nos grupos ¹			Taxa de erro aparente (%)
			CC	AP	OT	
Métricos (1)	Cajueiro comum (CC)	49	67,35	32,65	0,00	28,79
	Anão precoce (AP)	10	10,00	80,00	10,00	
	<i>A. othonianum</i> (OT)	07	0,00	14,29	85,71	
Categóricos (2)	Cajueiro comum (CC)	49	67,35	22,45	10,20	37,88
	Anão precoce (AP)	10	70,00	30,00	0,00	
	<i>A. othonianum</i> (OT)	07	28,57	0,00	71,43	
Moleculares (3)	Cajueiro comum (CC)	49	95,92	4,08	0,00	9,09
	Anão precoce (AP)	10	0,00	100	0,00	
	<i>A. othonianum</i> (OT)	07	42,86	14,29	42,86	
Soma (1+2)	Cajueiro comum (CC)	49	57,14	34,69	8,16	40,91
	Anão precoce (AP)	10	50,00	40,00	10,00	
	<i>A. othonianum</i> (OT)	07	0,00	0,00	100	
Soma (1+3)	Cajueiro comum (CC)	49	87,76	12,24	0,00	10,60
	Anão precoce (AP)	10	0,00	100	0,00	
	<i>A. othonianum</i> (OT)	07	0,00	14,29	87,71	
Soma (2+3)	Cajueiro comum (CC)	49	81,63	18,37	0,00	18,18
	Anão precoce (AP)	10	10,00	90,00	0,00	
	<i>A. othonianum</i> (OT)	07	28,57	0,00	71,43	
Soma (1+2+3)	Cajueiro comum (CC)	49	85,71	14,29	0,00	15,15
	Anão precoce (AP)	10	10,00	90,00	0,00	
	<i>A. othonianum</i> (OT)	07	28,57	0,00	71,43	

¹Valores destacados em negrito referem-se aos percentuais de alocação de acessos nos seus respectivos grupos de origem.

Na Figura 2 são apresentadas as projeções das distâncias no espaço tridimensional das distâncias genéticas estimadas pela matriz soma (1+3). Nesta figura é possível delimitar claramente o agrupamento formado pelos acessos de *A. othonianum*, apesar da grande amplitude de dispersão entre eles. Por outro lado, os acessos de cajueiro anão precoce se encontram circunscritos dentro da área de dispersão dos acessos de cajueiro comum não formando, portanto, grupos visualmente distintos no espaço tridimensional.

A correlação entre as distâncias originais e as projetadas na Figura 1 foi a igual a 85,6%, enquanto o grau de distorção e o coeficiente de stresse foram, respectivamente, iguais a 49% e 50,3%.



3.3 Análises de agrupamento dos acessos

Diferentes padrões de agrupamento dos acessos, pelo método de otimização de Tocher, foram obtidos para cada uma das matrizes de dissimilaridades empregadas. A matriz de dissimilaridades de dados

moleculares proporcionou o agrupamento com o menor número de grupos e com o grupo que abrangeu o maior número de acessos. Por outro lado, os dados fenotípicos discretos proporcionaram o maior número de grupos e o menor entre os maiores grupos de cada padrão de agrupamento (Tabelas 2 e 3).

Tabela 2 – Agrupamentos pelo método de otimização de Tocher, de 69 acessos de *Anacardium* spp., obtidos por diferentes matrizes de dissimilaridades. Acessos: 1 a 49 – cajueiro comum; 50 a 59 – cajueiro anão precoce; 60 a 66 *A. othonianum*; 67 e 68 *A. microcarpum* e 69 *Anacardium* sp.

Grupos	Natureza dos Caracteres da Matriz de Dissimilaridade																					
	Métricos (1)					Fenotípicos discretos (2)						Moleculares (3)										
1	13	39	40	44	19	9	33	19	35	46	63	32	33	10	17	11	2	5	9			
	48	16	28	18	29	61	66	50	62	11	16	40	38	39	12	42	41					
	12	41	46	42	11	40	6	31	58	57	2	19	18	1	31	3	44	29				
	68	8	32	3	4	35						13	37	4	28	36	7	8	21			
	58	7	17	1	55	59						26	46	20	34	6	22	27				
	6	51	34	36	37	2						56	54	50	68	15	14	57				
	45	30	27	54	53						47	67	51	24	59	23	25					
	15	31	33	56	66						43	58	30	69	35	45	53					
	57	43	52	9	50						55	49	62	52	64	48						
2	49	69	67	61	12	68	32	60	61													
3	60	64	65	62	63	20	22	10	28	24	63											
4	24	25	23	14	26	36	43	55	26	66												
	47																					
5	20	22				3	34	44	65													
6	21					5	48	59	14	29												
7	38					18	45	54	23													
8	5					1	4	52	49													
9	10					8	25															
10						13	60															
11						17	65															
12						21	30															
13						27	37	39	53													
14						41	42															
15						16	67															
16						47	64															
17						7																
18						15																
19						56																
20						69																
21						38																
22						51																

Os agrupamentos com dados combinados (Tabela 3), através da soma de matrizes de dissimilaridade, apresentaram padrões que foram dependes do tipo de dado envolvido, ou seja, agrupamentos que levaram em consideração matrizes de distâncias de informações fenotípicas discretas, formaram muitos

grupos (14 a 16 grupos), enquanto que o agrupamento obtido a partir dos dados morfométricos e moleculares foram apenas oito grupos, com mais de 80% dos acessos alocados em um único grupo. Por outro lado, tanto os dados morfométricos quanto os moleculares discriminaram melhor os acessos de *A. othonianum*, o que pode ser visualizado melhor pelo exame da Figura 3, que apresenta o agrupamento pelo método UPGMA, a partir da matriz soma entre as distâncias estimadas por estas duas classes de dados.

Tabela 3 – Agrupamentos pelo método de otimização de Tocher, de 69 acessos de *Anacardium* spp., obtidos de matrizes de dissimilaridade resultantes da soma matrizes de distâncias de dados morfométricos (1), fenotípicos discretos (2) e moleculares (3). Acessos: 1 a 49 – cajueiro comum; 50 a 59 – cajueiro anão precoce; 60 a 66 *A. othonianum*; 67 e 68 *A. microcarpum* e 69 *Anacardium* sp.

Grupos	Natureza dos Caracteres da Matriz de Dissimilaridade																			
	Matriz 1+2				Matriz 1+3				Matriz 2+3				Matriz 1+2+3							
1	44	57	28	58	39	40	13	12	2	11	8	19	9	2	11	8	19	40		
	42	31	11	2	40	19	42	41	16	33	31	40	41	41	42	46	9	31		
	19	8	41	6	46	11	2	44	18	29	46	42	28	39	28	33	39	44		
	66	33	35	9	48	32	28	3	1	46	44	10	5	57	34	29	32	12	3	34
	18	43	29	59	8	17	4	31	37	3	50	6	38	29	18	16	1	57	6	
	53	55	56	36	7	36	33	34	9	32	27	20	37	1	58	4	36	68	56	
					6	68	5	38	27	21	58	26	16	17	13	43	35			
					54	15	10	51	56					37	10	5	50	38		
					56	35	59	50						59	54	27				
					57	58	30	14												
					67	47	43	45												
					55	24	26	53												
					48	25	23													
2	12	68	32	39	20	22	21		12	13	68	4	22	20	22	26	23			
	13	27	37						47					21						
3	20	22	26		49	69			36	43	35	45		14	15	7	47			
									18	17	59	23								
									54											
4	50	52	23	47	62	63	64		14	15	67	7		60	61					
5	14	15	7	3	16	60	61		60	61				24	25	53				
6	60	61	62	63	66				24	25				45	51					
	64																			
7	1	4	30	49	69	65			62	63				62	63					
8	45	51	34	54	52				52	69				49	69					
9	10	24	5						53	55				52	55					
10	17	65							51					66	67					
11	67								30					30						
12	38								49					48						
13	21								48					65						
14	25								66					64						
15									64											
16									65											

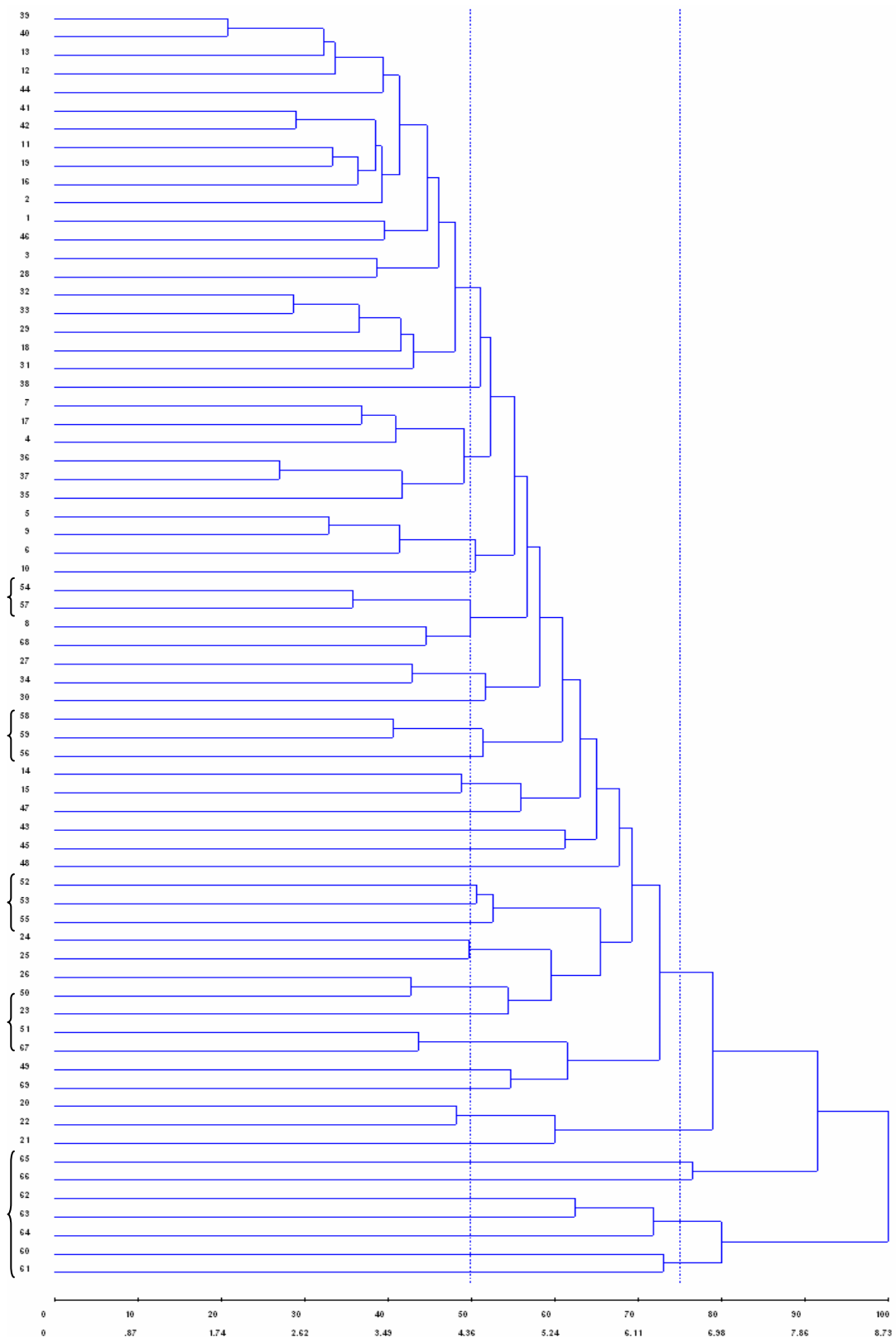


Figura 3 – Agrupamento de 69 acessos de *Anacardium* spp. , pelo método UPGMA, a partir da matriz soma de distâncias baseadas em dados morfométricos e divergência baseada em marcadores moleculares ISSR. Acessos: 1 a 49 – cajueiro comum; 50 a 59 – cajueiro anão precoce; 60 a 66 *A. othonianum*; 67 e 68 *A. microcarpum* e 69 *Anacardium* sp.

4 DISCUSSÃO

A incongruência entre as matrizes de distâncias de dados morfométricos e dados fenotípicos discretos, assim como entre dados fenotípicos discretos e dados moleculares é praticamente absoluta, conforme foi evidenciado pelos coeficientes nulos de correlação (Tabela 1). Por outro lado, as matrizes de distâncias de dados morfométricos e moleculares foram positiva e significativamente correlacionadas, mas com um valor de r relativamente baixo ($r=0,365$). Estes resultados diferem, em certa medida, dos encontrados por Archak et al. (2003a) que não observaram correlações significativas entre dados morfométricos e diferentes tipos de matrizes de dados moleculares (AFLP, ISSR e RAPD), ao analisarem a divergência genética de 19 genótipos de cajueiros, consistindo de cultivares comercial e linhas melhoradas, selecionados aleatoriamente de um banco de germoplasma com mais de 300 acessos. Ausência de correlações significativas entre divergência genética molecular e divergência fenotípica é relatada em diversos trabalhos na literatura, tanto em estudos plantas silvestres (Nielsen et al., 2003; Kamada, 2006) quanto de plantas cultivadas (Roldán-Ruiz et al., 2001).

A dispersão de pontos da Figura 1 não apresenta uma relação linear, mas também não remete a uma configuração “triangular”, nos moldes propostos por Burstin e Charcosset (1997). Esta relação “triangular” foi verificada por Dillmann et al., que correlacionaram distâncias morfológicas e moleculares em linhas endogâmicas de milho e também por Lefebvre et al. (2001), ao avaliarem acessos de pimenta (*Capsicum annum*) através de caracteres fenotípicos e moleculares, constituídos por marcadores AFLP e RAPD. Todos estes casos, contudo, referem-se a análises de genótipos homozigotos ou altamente endogâmicos, diferentemente de que se espera em *Anacardium*, conforme destacado por Archak et al. (2003a).

As análises discriminantes conduzidas com as diferentes classes de dados demonstraram que a maior proporção de classificações corretas foi obtida com a matriz de dissimilaridade construída com os marcadores ISSR. Enquanto a matriz de distâncias construída com os dados fenotípicos discretos proporcionou o maior percentual de erros de classificação geral dos acessos. Por outro lado, a classificação com base nos dados moleculares resultou na

menor proporção de alocações corretas para o grupo *A. othonianum*, indicando incongruência entre divergência fenotípica e molecular dentro do grupo.

A combinação, através da soma de matrizes, dos dados moleculares com os fenotípicos morfométricos ou fenotípicos discretos, resultou na melhoria significativa da discriminação dos grupos, em comparação ao uso de matrizes de dados fenotípicos isoladamente. No caso da matriz soma (1+3), que combinou dados morfométricos e moleculares, houve uma melhoria na classificação dos acessos de *A. othonianum* com a manutenção do baixo percentual da taxa de erro aparente (TEA), proporcionado pela classificação baseada nos dados moleculares isoladamente. No caso da combinação dos dados morfométricos com dados fenotípicos discretos, entretanto, a TEA foi superior à observada para qualquer uma das classificações realizadas com as diferentes matrizes isoladamente. Ainda que este tenha sido o único caso que discriminou todos os acessos de *A. othonianum* de maneira correta. Nielsen et al. (2002) aplicaram análise discriminante para classificar indivíduos de seis populações naturais pertencentes ao gênero *Scalesia*, sendo duas de *S. divisa*, duas de *S. incisa* e duas de indivíduos com características intermediárias. Os autores observaram aumento significativo na discriminação correta dos indivíduos, em cinco das seis populações, quando a análise foi realizada a partir de marcadores AFLPs, em comparação com a análise realizada com dados morfométricos. A caracterização da diversidade genética e diferenciação de cultivares de mandioca, por marcadores moleculares RAPD, também foi mais eficiente, em comparação com aquela obtida através de descritores botânicos fenotípicos discretos (Colombo et al., 1998).

Nas análises de agrupamento por Tocher observou-se, de modo geral, que os acessos de *A. othonianum* (acessos 60 a 66) foram suficientemente distintos para não serem incluídos no grupo principal (ou grupo 1) dos diferentes agrupamentos, considerando as dissimilaridades dos dados morfométricos e moleculares, isoladamente ou combinados. Estes acessos formaram pequenos grupos, isolados ou associados com acessos de cajueiro comum. Por outro lado, a maioria dos acessos de cajueiro anão precoce (acessos 50 a 59) foi sempre alocado no primeiro grupo, juntamente com a maior parte dos acessos de cajueiro comum, evidenciando pouca divergência

fenotípica e molecular entre os dois tipos. Este padrão de agrupamento dos acessos também é nitidamente percebido no dendrograma da Figura 3.

Apesar das correlações fracas ou ausentes entre medidas de divergência, Kamada (2006) verificou que os padrões de agrupamento dos acessos de fáfia, pelo método de otimização de Tocher, foram aproximadamente correspondentes, quando gerados a partir das matrizes de distâncias obtidas de dados morfométricos, moleculares ou da matriz soma, envolvendo as três classes de dados. Neste trabalho, por outro lado, comportamento semelhante foi observado apenas entre os agrupamentos produzidos a partir das matrizes 1 e 3 e da matriz soma (1+3), ou seja, entre as matrizes de dados que foram correlacionados.

Dessa forma, fica evidente que a combinação de classes de dados não correlacionados em uma única matriz de distâncias não é recomendável, uma vez que a cada uma destas classes parece acessar porções da variabilidade total com padrões de distribuição completamente diversos.

5 CONCLUSÕES

- a) Apenas os dados morfométricos e moleculares apresentaram correlação significativa, ainda que em um patamar relativamente baixo.
- b) As correlações baixas ou nulas entre as diferentes classes de dados indicam que eles acessam informações de diferentes partes do genoma.
- c) A discriminação dos acessos por marcadores moleculares foi superior às verificadas com base em dados fenotípicos, embora a combinação entre dados moleculares e morfométricos tenha proporcionado uma melhor discriminação média entre os grupos.
- d) As análises de agrupamento pelo método de otimização de Tocher apresentaram resultados discrepantes para as diferentes matrizes de distâncias empregadas, relativa congruência foi verificada apenas entre agrupamentos com base em dados morfométricos e / ou moleculares.

CONCLUSÕES GERAIS

- Técnicas de análise discriminante por componentes principais e por função discriminante de Anderson mostraram-se adequadas na caracterização dos grupos de *Anacardium* avaliados, quando aplicadas a variáveis morfométricas.
- Apesar de ser considerado por alguns autores apenas um ecótipo de *A. occidentale*, os acessos de *A. othonianum* constituíram um grupo morfologicamente muito distinto dos demais e, considerando que esta divergência foi observada entre indivíduos cultivados em um mesmo ambiente, pode-se assumir que ela é majoritariamente de natureza genética.
- O padrão de agrupamento UPGMA, baseado em dados morfométricos, exibido pelas populações analisadas refletiu adequadamente o grau de relacionamento esperado entre elas, considerando as evidências biológicas e históricas disponíveis.
- A análise discriminante, baseada em dados moleculares, pelo método de k vizinhos mais próximos, mostrou-se adequada apenas na classificação de acessos de diferentes populações de *A. occidentale*, uma vez que os padrões de divergência observados entre os acessos de *A. othonianum* e de *A. humile* não foram consistentes com os esperados para espécies distintas.
- Os percentuais da variação entre e dentro de populações, obtidos pela decomposição da variação total por meio da AMOVA, foram compatíveis com o esperado para espécies perenes de polinização mista.
- No agrupamento pelo método UPGMA, baseado em dados moleculares, a população 'Rio de Janeiro' apresentou a maior

divergência relativa, situação discordante do padrão obtido com base nas informações morfométricas.

- A maioria dos descritores fenotípicos discretos analisados apresentou ampla variabilidade, possibilitando a discriminação de todos os acessos de *Anacardium* avaliados, sem ambigüidades. Por outro lado, os padrões de distribuição desta foram inadequados para caracterizar grupos de acessos de origem geográfica comum.
- A análise comparativa das diferentes classes de dados mostrou que apenas os descritores morfométricos e os marcadores moleculares foram correlacionados, enquanto os descritores fenotípicos discretos apresentaram correlações nulas com os dados morfométricos e moleculares.
- A discriminação dos acessos por marcadores moleculares, pelo método de k vizinhos mais próximos, foi superior às verificadas com base em dados fenotípicos, embora a combinação entre dados moleculares e morfométricos tenha proporcionado uma melhor discriminação média entre os grupos.
- A combinação de diferentes classes de dados, através da soma de matrizes de distâncias, parece ser pertinente apenas quando há congruência entre as estimativas de divergência.
- Foi confirmado que a Região Nordeste é um importante centro de diversidade para *A. occidentale*. Por outro lado, o Norte Fluminense e o Estado de Roraima apresentam populações naturais que diferem daquelas encontradas nos Estados do Nordeste, o que aponta para a necessidade da realização trabalhos adicionais com vistas à coleta, caracterização e conservação destes recursos genéticos.

- Mais estudos envolvendo acessos e populações naturais de *A. othonianum* e *A. humile* também necessitam ser realizados, tanto para elucidar melhor o *status* taxonômico destas espécies, quanto para fins de utilização e conservação destes recursos genéticos, considerando o ritmo acelerado de destruição dos Cerrados.

REFERÊNCIAS

- AGA, E.; BEKELE, E.; BRYNGELSSON, T. Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. **Genetica**, v. 124, p. 213-221. 2005.
- ALMEIDA, J.I.L. de; ARAÚJO, F. DE; LOPES, J.G.V. **Evolução do cajueiro anão precoce na estação experimental de Pacajus, Ceará**. Fortaleza: EPACE, 1993. 17p.
- ANDERSON, T.W. **An introduction to multivariate statistical analysis**. Nova York: John Wiley & Sons, 1958, 242 p.
- ANDERSON, T.W. **An introduction to multivariate statistical analysis**. 2ª ed. Nova York: John Wiley & Sons, 1984, 675 p.
- ARCHAK, S.; GAIWARD, A.B.; GUATAM, D.; RAO, E.V.V.B.; SWAMY, K.R.M. Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. **Genome**, v. 46, p. 362-369. 2003a.
- ARCHAK, S.; GAIWARD, A.B.; GUATAM, D.; RAO, E.V.V.B.; SWAMY, K.R.M.; KARIHALOO, J.L. DNA fingerprinting of Indian cashew (*Anacardium occidentale* L.) varieties using RAPD e ISSR techniques. **Euphytica** v. 230, p. 397-404. 2003b.
- ASSEFA, K.; MERKER, A.; TEFERA, H. Inter simple sequence repeat (ISSR) analysis of genetic diversity in tef [*Eragrostis tef* (Zucc.) Trotter]. **Hereditas**, n. 139, p. 174-183. 2003.
- ASSUMPÇÃO, J.; NASCIMENTO, M.T. Fitofisionomia de uma restinga no extremo norte do litoral fluminense: um novo elemento no mosaico? In: SIMPÓSIO DE ECOSSISTEMAS BRASILEIROS, 4., 1998, Águas de Lindóia. **Anais...v. 3** São Paulo: Anciesp, 1998. p.158-164.
- AZEVEDO, D.M.P.; CRISÓSTOMO, J.R.; ALMEIDA, F.C.G.; ROSSETTI, A.G. Estimates of genetic correlations and correlated responses to selection in cashew (*Anacardium occidentale* L.). **Genetics and Molecular Biology**, v. 21, n. 3, p. 399-402. 1998.
- BAJRACHARYA, J.; STEELE, K.A.; JARVIS, D.I.; STHAPIT, BR.; WITCOMBE, J.R. Rice landrace diversity in Nepal: variability of agro-morphological traits and SSR markers in landraces from a high-altitude site. **Field Crop Research**, n. 95, p. 327-335. 2005.
- BARBOSA-NETO, J.F.; BERED, F. 1998. Marcadores moleculares e diversidade genética no melhoramento de plantas. In: MILACH, S.C.K. (Ed.) **Marcadores Moleculares em Plantas**. Porto Alegre: UFRGS. p. 29-40.

BARROS, L.M. **Caracterização, morfológica e isoenzimática do cajueiro (*Anacardium occidentale* L.) tipos comum e anão precoce, por meio de técnicas multivariadas.** 1991. 256p. Tese (Doutorado em Genética e Melhoramento de Plantas)-Escola Superior de Agricultura Luiz de Queiroz/Universidade de São Paulo, Piracicaba,SP, 1991.

BARROS, L.M. Botânica, origem e distribuição geográfica. In: ARAÚJO, J.P.P; SILVA, V.V. **Cajucultura: modernas técnicas de produção.** Fortaleza: EMBRAPA-CNPAT. 1995. p. 53-69.

BARROS, L.M. Introdução. In: BARROS, L.M. (ed.) **Caju - Produção: aspectos técnicos.** Fortaleza: Embrapa Agroindústria Tropical. Brasília: Embrapa Informação Tecnológica, série frutas do Brasil, n. 30, 148p. 2002.

BARROS, L.M. & CRISÓSTOMO, J.R. Melhoramento do cajueiro. In: ARAÚJO, J.P.P; SILVA, V.V. **Cajucultura: modernas técnicas de produção.** Fortaleza: EMBRAPA-CNPAT. 1995. p. 73-93.

BARROS, L.M.; CAVALCANTI, J.J.V.; PAIVA, J.R.; CRISÓSTOMO, J.R.; CORRÊA, M.P.F.; LIMA, A.C. Seleção de clones de cajueiro-anão para o plantio comercial no Estado do Ceará. **Pesquisa Agropecuária Brasileira**, v. 35, n.11, p.2197-2204. 2000.

BARROS, L.M.; PAIVA, J.R.; CAVALCANTI, J.J.V.; ARAÚJO, J.P.P. Cajueiro. In: BRUCKNER, C.H. (ed.) **Melhoramento de fruteiras tropicais.** Viçosa, MG: UFV, 2002. p. 160-176.

BAXEVANIS, A.D.; TRIANTAPHYLLIDIS, G.V.; KAPPAS, I.; TRIANTAPHYLLIDIS, A.; TRIANTAPHYLLIDIS, C.D.; ABATZOPOULOS, T.J. Evolutionary assessment of *Artemia tibetiana* (Crustacea, Anostraca) based on morphometry and 16S rRNA RFPL analysis. **JZS**, n. 43, v. 3, p. 189-198. 2005.

BEHARAV, A.; NEVO, E. Predictive validity of discriminant analysis for genetic data. **Genetica**, v. 119, p. 259-267. 2003.

BEKELE, F.; BEKELE, I. A sampling of the phenetic diversity of cacao in the International Cacao Gene Bank of Trinidad. **Crop Science**, n. 36, p. 57-64. 1996.

BLAIR, M.W.; PANAUD, O.; McCOUCH, S.R. Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). **Theor Appl Genet**, v. 98, p. 780-792. 1999.

BORNET, B.; BRANCHARD, M. Nonanchored inter simple sequence repeat (ISSR) markers: reproducible and specific tools for genome fingerprinting. **Plant Molecular Biology Reporter**, v. 19, p. 209-215. 2001.

BUENO, D.M. **Estudo da floração, frutificação, embriogênese final zigótica e anatomia do pericarpo do cajueiro anão precoce (*Anacardium occidentale* L.)**. 1996. 95p. Tese (Doutorado em Fitotecnia). Universidade Federal de Viçosa, Viçosa, MG, 1997.

BURSTIN, J.; CHARCOSSET, A. Relationship between phenotypic and marker distances: theoretical and experimental investigations. **Heredity**, v. 79, p. 477-483. 1997.

CARDOSO, J.E.; CAVALCANTI, J.J.V.; CAVALCANTE, M. J.B.; ARAGÃO, M.L.; FELIPE, E.M. Genetic resistance of dwarf cashew (*Anacardium occidentale* L.) to anthracnose, black mold, and angular leaf spot. **Crop Protection**, v. 18, n. 1, p. 23-27. 1999.

CARLINI-GARCIA, L. A.; VENCOVSKY, R.; COELHO, A.S.G. Métodos bootstrap aplicados em níveis de reamostragem na estimação de parâmetros genéticos populacionais. **Scientia Agricola**, v. 58, n. 4, p. 785-793, 2001.

CAVALCANTI, J.J.V. **Genetic mapping and QTL identification in cashew (*Anacardium occidentale* L.)**. 2004. Tese (Doutorado em Biologia Molecular). University of Reading, UR, Inglaterra. 2004.

CAVALCANTI, J.J.V.; PAIVA, J.R.; BARROS, L.M.; CRISÓSTOMO, J.R. & CORRÊA, M.P.F. Repetibilidade de Caracteres de produção e porte da planta em clones de cajueiro-anão precoce. **Pesquisa Agropecuária Brasileira**, v. 35 n. 4, p. 773-777. 2000a.

CAVALCANTI, J.J.V.; PAIVA, J.R.; PINTO, C.B.P.; CRISÓSTOMO, J.R. & FERREIRA, D.F. Análise dialélica para avaliação de híbridos interpopulacionais de cajueiro. **Pesquisa Agropecuária Brasileira**, v.35, n.8, p.1567-1575. 2000b.

CAVERS, S.; NAVARRO, C.; LOWE, A.J. A combination of molecular markers identifies evolutionary significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. **Conservation Genetics**, v. 4, p. 571-580. 2003.

CHARTERS, Y.M.; WILKINSON, M.J. The use of self-pollinated progenies as 'in groups' for the genetic characterization of cocoa germplasm. **Theor. Appl. Genet.** v. 100, p. 160-166. 2000.

COCKERHAM, C.C. Analysis of gene frequencies. **Genetics**, n.74, p. 679-700. 1973.

COLE-RODGERS, P.; SMITH, D.W.; BOSLAND, P.W. A novel statistical approach to analyze genetic resource evaluations using *Capsicum* as an example. **Crop Science**, v. 37, p. 1000-1002. 1997.

COLOMBO, C.; SECOND, G.; VALLE, T.L.; CHARRIER, A. Genetic diversity characterization of cassava cultivars (*Manihot esculenta* Crantz). I) RAPD markers. **Genetics and Molecular Biology**, n. 21, v. 1, p. 105-113. 1998.

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

CRISÓSTOMO, J.R.; GADELHA, J.W.R.; ARAÚJO, J.P.P. & BARROS, L.M. **Conseqüências do plantio de sementes de plantas enxertadas ou de plantas de pé franco de cajueiro**. Fortaleza: Embrapa-CNPc – informativo, 3. 3p.1992

CRISÓSTOMO, J.R.; CAVALCANTI, J.J.V.; BARROS, L.M. et al. Melhoramento do cajueiro-anão-precoce: avaliação da qualidade do pedúnculo e a heterose de seus híbridos. **Rev. Bras. Frutic.** v. 24, n. 2 p. 477-480. 2002.

CROXFORD, A.E.; ROBSON, M.; WILKINSON, M.J. Characterization and PCR multiplexing of polymorphic microsatellite loci in cashew (*Anacardium occidentale* L.) and their cross-species utilization. **Molecular Ecology Notes**, v. 6, p. 249-251. 2006.

CRUZ, C.D. **Programa GENES**: Versão Windows, aplicativo computacional em genética e estatística. Viçosa-MG, UFV, 2001. 648p.

CRUZ, C.D. **Programa GENES**: Análise multivariada e simulação. Viçosa-MG, UFV, 2006a.175p.

CRUZ, C.D. **Programa GENES**: Estatística experimental e biometria. Viçosa-MG, UFV, 2006b. 285p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético** – volume 2. Viçosa-MG: UFV, 2003, 585p.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético** – volume 1. 2ª ed. Viçosa-MG: UFV, 2004, 480p.

CRUZ, P.J.; CARVALHO, F.I.F.; OLIVEIRA, A.C. et al. Genetic dissimilarity among wheat genotypes for lodging-associated traits. **Crop Breeding and Applied Biotechnology**, v. 4, p. 427-433. 2004.

DESHPANDE, A.U.; APTE, G.S.; BAHULIKAR, R.A.; LAGU, M.D.; KULKARNI, B.G.; SURESH, H.S.; SINGH, N.P. RAO, M.K.V.; GUPTA, V.S.; PANT, A. RANJEKAR, P.K. Genetic diversity across natural population of three montane plant species from the Western Ghats, India revealed by intersimple sequence repeats. **Molecular Ecology**, v. 10 p. 2397-2408. 2001.

DHANARAJ, A.L.; RAO .E.V.V.B.; SWAMY K.R.M.; BHAT, M.G.; PRASAD, D.T.; SONDUR, S.N. Using RAPDs to assess the diversity in Indian cashew (*Anacardium occidentale* L.) germplasm. **Journal of Horticultural Science and Biotechnology**, v. 77, n.1, p. 41-47. 2002.

DILLMANN, C.; BAR-HEN, A.; GUÉRIN, D.; CHARCOSSET, A.; MURIGNEUX, A. Comparison of RFLP and morphological distances between maize *Zea mays* L. inbred lines. Consequences for germplasm protection purposes. **Theor. Appl. Genet.** v. 95, p. 92-102. 1997.

DURÁN, L.A.; BLAIR, M.W.; GIRALDO, M.C.; MACCHIAVELLI, R.; PROPHETE, E.; NIN, J.C.; BEAVER, J.S. Morphological and molecular characterization of common bean landraces and cultivars from the Caribbean. **Crop Science**, n. 45, p. 1320-1328. 2005.

EIADTHONG, W.; YONEMORI, K.; SUGIURA, A.; UTSUNOMIYA, N.; SUBHADRABANDHU, S. Identification of mango cultivars of Thailand and evaluation of their genetic variation using the amplified fragments by simple sequence repeat-(SSR-) anchored primers. **Scientia Horticulturae**, n. 82, p. 57-66. 1999.

ENGELS, J.M.M. A systematic description of cacao clones II. The discriminative value of qualitative characteristics and the practical compatibility of the discriminative value of quantitative and qualitative descriptors. **Euphytica**, n. 32, p. 387-396. 1983a.

ENGELS, J.M.M. A systematic description of cacao clones III. Relationships between characteristics and some consequences for the cacao breeding. **Euphytica**, n. 32, p. 719-733. 1983b.

EXCOFFIER, L.; SMOUSE, P.E.; QUATTRO, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. **Genetics** v. 131, p.179-191. 1992.

FANG, D.Q.; ROOSE, M.L.; KRUEGER, R.R.; FEDERICI, C.T. Fingerprinting trifoliolate orange germplasm accessions with isozymes, RFPLs, and inter-simple sequence repeats markers. **Theor. Appl. Genet.**, v. 95, p. 211-219. 1997.

FAO (Food and Agriculture Organization of the United Nations). **FAOSTAT**. Disponível em: <http://faostat.fao.org/site/567>. Acesso em 16 de mar. 2007.

FEDERIZZI, L.C. Estrutura de um programa de melhoramento de plantas e possíveis aplicações de marcadores moleculares: visão do melhorista. In: Milach, S.C.K. (ed.) **Marcadores Moleculares em Plantas**. Porto Alegre: UFRGS. p. 03-16. 1998.

FERRÃO, J.E.M. **O cajueiro (*Anacardium occidentale* L.)**. Lisboa: Instituto de Investigação Científica Tropical, 1995. 299p.

FERREIRA, M.B. Frutos comestíveis do DF(II): gabiobas, araçás, amoreira e cajus. **Cerrado**, v.05, p. 25-29. 1973.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3ª ed. Brasília: EMBRAPA – CERNAGEN, 1998. 220p.

FERREIRA, R.P.; CRUZ, C.D.; SEDIYAMA, C.S. FAGUERIA, N.K. Identificação de cultivares de arroz tolerantes à toxidez de alumínio por técnica multivariada. **Pesq. Agropec. Bras.** v. 30, n. 6, p. 789-795. 1995.

FONSECA, A.F.A.; SEDIYAMA, T.; CRUZ, C.D. et al. Discriminant analysis for the classification and clustering of robusta coffee genotypes. **Crop Breeding and Applied Biotechnology**, v. 4 p. 285-289, 2004.

FRANCO, J. ; CROSSA, J.; RIBAUT, J.M.; BETRAN, J.; WARBURTON, M.L.; KHAIRALLAH, M. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. **Theor. Appl. Genet.**, v. 103, p. 944-952. 2001.

FREITAS, B.M.; PAXTON, R.J. A comparison of two pollinators: the introduced honey bee *Apis mellifera* and an indigenous bee *Centris tarsata* on cashew *Anacardium occidentale* in its native range of NE Brazil. **Journal of Applied Ecology**, v. 35 p. 109-121. 1998.

GE, X.J.; SUN, M. Reproductive biology and genetic diversity of a cryptoviviparous mangrove *Aegiceras corniculatum* (Myrsinaceae) using allozyme and intersimple sequence repeat (ISSR) analysis. **Molecular Ecology**, v. 8, p. 2061-2069. 1999.

GE, X.J.; YU, Y.; YUAN; Y.M.; HUANG, H.W.; YAN, CHENG. Genetic diversity and geographic differentiation in endangered *Ammopiptanthus* (Leguminosae) populations in desert regions of northwest China as revealed by ISSR analysis. **Annals of Botany**, v. 95, p. 843-851. 2005.

GODWIN, I.A.; AITKEN, E.A.B.; SMITH, L.W. Application of inter simple sequence repeat (ISSR) markers to plant genetics. **Electrophoresis**, v. 18, p. 1524-1528. 1997.

GOLDBLATT, P. (Ed.). **Index to plant chromosome numbers 1975 -1978**. Missouri Botanical Garden, 1981.

GOLDBLATT, P. (Ed.). **Index to plant chromosome numbers 1979 -1981**. Missouri Botanical Garden, 1984.

GUPTA, M.; CHYI, Y.S.; ROMERO-SEVERSON, J.; OWEN, J.L. Amplification of DNA evolutionarily diverse genomes using single primers of simple sequence repeats. **Theor. Appl. Genet.**, v. 89, p. 998-1006. 1994.

HAIR JR., J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Análise multivariada de dados**. Tradução de Adonai Schlup Sant'Anna e Anselmo Chaves Neto. 5ª ed. Porto Alegre: Bookman, 2005. 593p.

HAND, D.J. A comparison of two methods of discriminant analysis applied to binary data. **Biometrics**, v. 39, p. 683-694. 1983.

HOLANDA-NETO, J.P.; FREITAS, B.M.; BUENO, D.M.; DE ARAUJO, Z.B. 2002. Low seed/nut productivity in cashew (*Anacardium occidentale*): Effects of self-incompatibility and honey bee (*Apis mellifera*) foraging behavior. **Journal of Horticultural Science & Biotechnology**, v.77 n.2, p. 226-231. 2002.

HOLSINGER, K.E.; LEWIS, P.O.; DEY, D.K. A Bayesian approach to inferring population structure from dominant markers. **Molecular Ecology**, v. 11, p. 1157-1164, 2002.

HUANG, J.C.; SUN, M. Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. **Theor. Appl. Genet.**, v. 100, p. 1050-1060. 2000.

HUFF, D.R.; PEAKALL, R.; SMOUSE, P.E. RAPD variation within and among natural populations of outcrossing buffalograss (*Buchloe dactyloides* (Nutt.) Engelm.). **Theor. Appl. Genet.** n. 86, p. 927-934. 1993.

IBPGR. **Cashew descriptors**. International Board for Plant Genetic Resources. Rome, 1986. Disponível em: <http://www.ipgri.cgiar.org/publications/HTMLPublications/348>. Acesso em 4 out. 2005.

JOHNSON, D.V. **O caju do Nordeste do Brasil: um estudo geográfico**. Trad. José Alexandre Robatto Orrico. Fortaleza: ETENE, 1974. 169p.

JOSHI, S.P.; GUPTA, V.S.; AGGARWAL, R.K.; RANJEKAR, P.K.; BRAR, D.S. Genetic diversity and phylogenetic relationship as revealed by inter-simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. **Theor. Appl. Genet.** v. 100, p. 1311-1320. 2000.

KAMADA, T. **Avaliação da diversidade genética de populações de fáfia (*Pfaffia glomerata* (Spreng.) Pedersen) por RAPD, caracteres morfológicos e teor de beta-ecdisona**. 2006. 106p. Tese (Doutorado em Genética e Melhoramento). Universidade Federal de Viçosa, Viçosa, MG, 2006.

KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS[®] software**. Cary, NC: SAS Institute Inc., 2000. 558p.

KRAUSS, S.L. Accurate gene Diversity estimates from amplified fragment length polymorphism (AFLP) markers. **Molecular Ecology**, v. 9, p.1241-1245. 2000.

KRUSKAL, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika**, v. 29, p. 1-27. 1964.

KUMARAN, P.M.; NAYAR, N.M.; MURTHY, K.N.; VIMALA, B. A study of variation in flowering characters of cashew germplasm. **Acta Horticulturae**, n. 108 p. 20-24. 1985.

LACERDA, D.R.; ACEDO, M.D.P.; LEMOS-FILHO, J.P.; LOVATO, M.B. Genetic diversity and structure of natural populations of *Plathymenia reticulata* (Mimosoideae), a tropical tree from the Brazilian Cerrado. **Molecular Ecology**, v. 10, p. 1143-1152. 2001.

LEÃO, A.C.M.; RODRIGUES, W.A.; CHAVES, L.J.; NAVES, R.V. Análise citogenética da espécie do caju-do-campo *Anacardium othonianum*. In: CONGRESSO NACIONAL DE GENÉTICA, 48., 2001, Águas de Lindóia-SP. **Anais...** Ribeirão Preto: SBG, 2001. 1 CD-ROM.

LEFEBVRE, V.; GOFFINET, B.; CHAUVET, J.C.; CAROMEL, B.; SIGNORET, P.; BRAND, R.; PALLOIX, A. Evaluation of genetic distances between pepper inbred lines for cultivar protection purposes: comparison of AFLP, RAPD and phenotypic data. **Theor. Appl. Genetic**. v. 102, p.741-750. 2001.

LYNCH, M.; MILLIGAN, B.G. Analysis of population genetic structure with RAPD markers. **Molecular Ecology**, v.3, p.91-99. 1994.

LONG, J. The allelic correlation structure of Gainj- and Kalan-speaking people. I. The estimation and interpretation of Wright's F-statistics. **Genetics**, v. 112, p. 629-647, 1986.

MANLY, B.F.J. Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. **Research in Population Ecology**, n. 28, p. 201-218. 1986.

MANLY, B.F.J. **Multivariate statistical methods: a primer**. 2^a ed. Londres: Chapman & Hall, 1998. 215p.

MARDIA, K.V.; KEMT, J.T.; BIBBY, J.M. **Multivariate analysis**. Londres: Academic Press, 1997. 518p.

MATTIONI, C.; CASASOLI, M. GONZALES, M.; IPINZA, R. Comparison of ISSR and RAPD markers to characterize three Chilean *Nothofagus* species. **Theor. Appl. Genetic**. v. 104, p.1064-1070. 2002.

MICHALAKIS, Y. EXCOFFIER L. A generic estimation of subdivision using distances between alleles with special reference for microsatellite loci. **Genetics**, v. 142, p. 1061-1064, 1996.

MITCHELL, J.D.; MORI, S.A. The cashew and its relatives (*Anacardium*: Anacardiaceae). **Memoirs Of The New York Botanical Garden**, v. 42, 1987, 76p.

MITCHELL, J.D.; YOUNG, D.A. Cladistic analysis. In: MITCHELL, J.D.; MORI, S.A. The cashew and its relatives (*Anacardium*: Anacardiaceae). **Memoirs Of The New York Botanical Garden**, v. 42, p. 24-57. 1987.

MNENEY, E.E.; MANTELL, S.H. & BENNETT, M. Use of random amplified polymorphic DNA (RAPD) markers to reveal genetic diversity within and between populations of cashew (*Anacardium occidentale* L.). **Journal of Horticultural Science and Biotechnology**, v.76, n.4, p. 375-383. 2001.

MOHAMMADI, S.A.; PRASANNA, B.M. Analysis of genetic diversity in crop plants: salient statistical tools and considerations. **Crop Science**, v. 432, p. 1235-1248. 2003.

MOORE, D.H. Evaluation of five discrimination procedures for binary variables. **J. Am. Statist. Assoc.** v. 68, p.3721-3323. 1973.

NIELSEN, L.R.; COWAN, R.S.; SIEGISMUND, H.R.; ADSERN, H.; PHILIPP, M.; FAY, M.F. Morphometric, AFLP and plastid microsatellite variation in populations of *Scalesia divisa* and *S. incisa* (Asteraceae) from the Galápagos Islands. **Botanical Journal of the Linnean Society**, v. 143, p. 243-254. 2003.

NYBOM, H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. **Molecular Ecology**, v.13, p.1143-1155. 2004.

PAIVA, J.R.; BARROS, L.M.; CRISÓSTOMO, J.R.; ARAÚJO, P.J.P.; ROSSETTI, A.G.; CAVALCANTI, J.J.V.; FELIPE, E.M. Depressão por endogamia em progênies de cajueiro anão precoce var. nanum. **Pesquisa Agropecuária Brasileira** v.33, n.4, p. 425-431. 1998.

PAIVA, J.R.; BARROS, L.M.; CRISÓSTOMO, J. R. CAVALCANTI, J.J.V. **Recursos genéticos de cajueiro: coleta conservação, caracterização e utilização**. Fortaleza: Embrapa Agroindústria Tropical, 2003. 43 p.

PAVITHRAN, K.; SHAFFI, M; INDIRA,E.P. Development, differentiation and evolution of sex-dimorphism in cashew. **Acta Horticulturae**, n. 108 p. 17-19. 1985.

PAULINO, F.D.G. 1992. **Polinização entomófila em cajueiro (*Anacardium occidentale* L.) no litoral de Pacajus –CE**. 1992. 70p. Dissertação (Mestrado em Entomologia)-Escola Superior de Agricultura Luiz de Queiroz/Universidade de São Paulo. Piracicaba, 1992.

PEJIC, I.; AJMONE-MARSAN, P.; MORGANTE, M.; KOZUMPLICK, V.; CASTIGLIONI, P.; TAMARINO, G.; MOTTO, M. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. **Theor. Appl. Genetic**, v. 97, p. 1248-1255. 1998.

PELSER, P.B.; HOUCHIN, R. Taxonomic studies on *Senecio aquaticus* (Asteraceae). A recommendation for the taxonomic status of *Aquaticus* and *Barbareifolius*. **Botanical Journal of the Linnean Society**, v. 145, p. 489-498. 2004.

PERRY, M.C.; BETTENCOURT, E. Sources of information on existing germplasm collections. In: GUARINO, L.; RAO, V.R.; REID, R. (Eds.). **Collecting plant genetic diversity: technical guidelines**. Cambridge: University Press, 1995. cap. 8, p.121-129.

PESSONI, L.A.; CRUZ, C.D. Repetibilidade de variáveis morfológicas utilizadas na caracterização da diversidade genética de populações de cajueiro (*Anacardium occidentale* L.). In: CONGRESSO BRASILEIRO DE MELHORAMENTO DE PLANTAS, 3., 2005, Gramado-RS. **Anais...** Passo Fundo-RS: SBMP, 2005. CD-ROM.

PICOLI, E.A.T.; ALFENAS, A.C.; LAIA, M.L.; MOURA, D.F.; DIAS, L.A.S.; FONSECA, S.S.; FERNANDES, D.; CRUZ, C.D. Divergência Genética entre árvores matrizes de *Eucalyptus urophylla* e *E. grandis* por meio de análise de RAPD. In: CONGRESSO BRASILEIRO DE MELHORAMENTO DE PLANTAS, 1., 2001, Goiânia. **Anais...** Goiânia: SBMP, 2001. CD-ROM.

PIMENTEL, C.R.M.; FILGUEIRAS, H.A.C.; ALVES, R.E. Mercado: Situação atual e perspectivas. In: Alves, R.E. (ed.): **Caju. Pós-colheita**. Fortaleza: Embrapa Agroindústria Tropical. Brasília: Embrapa Informação Tecnológica, série frutas do Brasil, n. 31, 36p. 2002.

PREVOST, A.; WILKINSON, M.J. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. **Theor. Appl. Genetic**. v. 98, p.107-112. 2001

REDDY, M.P.; SARLA, N.; SIDDIQ, E.A. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. **Euphytica**, v. 128, p. 9-17. 2002.

REIF., J.C.; MELCHINGER, A.E.; FRISCH, M. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. **Crop Science**, v. 45, p. 1-7. 2005.

RESENDE, V.F. **Análise genética da resistência foliar a antracnose foliar em milho**. 2004. 103p. Tese (Doutorado em Genética e Melhoramento de Plantas)-Escola Superior de Agricultura Luiz de Queiroz/Universidade de São Paulo. Piracicaba, 2004.

ROLDÁN-RUIZ, I.; van EEUWIJK, F.A.; GILLILAND, T.J.; DUBREUIL, P.; DILLMANN, C.; LALLEMAND, J.; De LOOSE, M.; BARIL, C.P. A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. **Theor. Appl. Genetic**. v. 103 p. 1138-1150. 2001.

RUAS, P.M.; RUAS, C.F.; RAMPIM, L.; CARVALHO, V.P.; RUAS, E.A.; SERA, T. Genetic relationship in *Coffea* species and parentage determination of interspecific hybrids using ISSR (Inter Simple Sequence Repeat) markers. **Genetics and Molecular Biology**, v. 26, n. 3, p. 319-327. 2003.

RUSSELL, J.R.; FULLER, J.D. MACAULAY, M.; HATZ, B.G.B.; JAHOR, A.; POWELL, W.; WAUGH, R. Direct comparisons of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. **Theor. Appl. Genetic.** v. 95, p. 714-722. 1997.

SAMAL, S.; ROUT, G.R.; LENKA, P.C. Analysis of genetic relationships between populations of cashew (*Anacardium occidentale* L.) by using morphological characterization and RAPD markers. **Plant Soil Environ.** v. 49, n. 4, p. 176-182. 2003.

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

SCHNEIDER, S.; ROESSLI, D.; EXCOFFIER, L. **Arlequin Ver. 2.000: A software for population genetics data analysis.** Genetics and Biometry Laboratory, University of Geneva, Switzerland. 2000.

SILVA-NETO, S.P.; MARUTA, I.; TAKAIWA, F.; OONO, K.; MATSUMOTO, K. Identification of cashew (*Anacardium occidentale* L.) seedlings with RAPD markers. In: International Symposium on Tropical Fruits. **Acta Horticulturae**, v. 370, p. 21-26. 1995.

SNEAT, P.H.; SOKAL, R.R. **Numerical taxonomy: the principles and practice of numerical classification.** São Francisco: W.H. Freeman, 1973. 573p.

SYED, N.H.; SURESHSUBDAR, S.; WILKINSON, M.J.; BHAI, B.S.; CAVALCANTI, J.J.V.; FLAVELL, A.J. Ty1-copia retrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.). **Theor. Appl. Genet.**, v. 110, p. 1195-1202. 2005.

THANKAMMA-PILLAI, P.K.; NAMBIER, M.C. Study of microsporogenesis in cashew. **Acta Horticulturae**, n. 108, p. 280-281. 1985.

van de WOUDE, M.; MAXTED, N.; FORD-LLOYD, B.V. A multivariate and cladística study of *Vicia* L. ser. *Vicia* (Fabaceae) based on analysis of morphological characters. **Plant. Syst. Evol.** V. 237, p. 19-39. 2003.

VIDIGAL, M.C.G.; VIDIGAL-FILHO, P.S.; AMARAL-JÚNIOR, A.T.; BRACCINI, A.de L. Divergência genética entre cultivares de mandioca por meio de estatística multivariada. **Bragantia**, v. 56, n. 2, p. 263-271. 1997.

VIJAYAN, K.; KAR, P.K.; TIKADER, A.; SRIVASTAVA, P.P.; AWASTHI, A.K.; THAGAVELU, K.; SARATCHANDRA, B. Molecular evaluation of genetic variability in wild populations of mulberry (*Morus serrata* Roxb.). **Plant Breeding**, v. 123, p. 568-572. 2004.

WADT, L.H.O.; KAGEYAMA, P.O. Estrutura genética e sistema de acasalamento de *Piper hispidinervum*. **Pesq. Agropec. Bras.** v. 39, n.2, p. 151-157. 2004.

WARD, J.H. Jr. Hierarchical grouping to optimize an objective function. **J. Am. Statist. Assoc.** v. 58, p. 236-244. 1963.

WEIDUSCHAT, A.A. **Elementos de ecologia e etnobotânica de *Anacardium occidentale* L. (Anacardiaceae) na área indígena Raposa Serra do Sol, Roraima-Brasil.** 1999. 112p. Dissertação (Mestrado em Ciências Florestais)-Instituto Nacional de Pesquisas da Amazônia/Fundação Universidade do Amazonas, Manaus, 1999.

WU, K.; JONES, R.; DANNAEBERGER, L. SCOLNIK, P.A. Detection of microsatellite polymorphisms without cloning. **Nucleic Acids Res.** v. 22, p. 3257-3258. 1994.

ZIETKIEWICZ, E.; RAFALSKI, A.; LABUDA, D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. **Genomics**, v. 20, p. 176-183. 1994.

ZHYVOTOVSKY, L.A. Estimating population structure in diploids with multilocus dominant DNA markers. **Molecular Ecology**, v.8, p.907-913. 1999.

ZUCCHI, M.I. **Análise da estrutura genética de *Eugenia dysenterica* DC utilizando marcadores RAPD e SSR.** 2002. 130p. Tese (Doutorado em Genética e Melhoramento de Plantas)-Escola Superior de Agricultura Luiz de Queiroz/Universidade de São Paulo. Piracicaba, 2002.