

**ANÁLISE DA ESTRUTURA GENÉTICA DE
Eugenia dysenterica DC UTILIZANDO
MARCADORES RAPD E SSR**

MARIA IMACULADA ZUCCHI

Tese apresentada à Escola Superior
de Agricultura "Luiz de Queiroz",
Universidade de São Paulo, para
obtenção do título de Doutor em
Agronomia, Área de Concentração:
Genética e Melhoramento de Plantas.

P I R A C I C A B A

Estado de São Paulo - Brasil

Novembro - 2002

**ANÁLISE DA ESTRUTURA GENÉTICA DE
Eugenia dysenterica DC UTILIZANDO
MARCADORES RAPD E SSR**

MARIA IMACULADA ZUCCHI

Bacharel em Ciências Biológicas

Orientador: Prof. Dr. **ROLAND VENCOVSKY**

Tese apresentada à Escola Superior
de Agricultura "Luiz de Queiroz",
Universidade de São Paulo, para
obtenção do título de Doutor em
Agronomia, Área de Concentração:
Genética e Melhoramento de Plantas.

P I R A C I C A B A

Estado de São Paulo - Brasil

Novembro - 2002

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Zucchi, Maria Imaculada

Análise da estrutura genética de *Eugenia dysenterica* DC utilizando marcadores RAPD e SSR / Maria Imaculada Zucchi. - - Piracicaba, 2002.
130 p.

Tese (doutorado) - Escola Superior de Agricultura Luiz de Queiroz, 2002.
Bibliografia.

1. Fruta tropical 2. Cagaita 3. Genética população 4. Marcador genético 5. Progenie I. Título

CDD 634.6

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

Dedico ao Professor Vencovsky com todo meu coração pela amizade
e exemplo de vida profissional,
Ao Baldin pelo apoio e carinho durante todos estes anos, e
A Deus pela força em todos os momentos desta caminhada, pois sem sua luz este
trabalho não teria sido realizado.

AGRADECIMENTOS

Agradeço a todos que de alguma forma contribuíram para a realização deste trabalho. Em especial:

Ao professor Dr. Roland Vencovsky pela orientação, amizade e confiança.

Ao Departamento de Genética da ESALQ/USP e a todos os seus professores pelos conhecimentos transmitidos. Aos funcionários em especial: Silvana, Berdan, Léia, Fernandinho e Macedônio.

À Universidade Federal de Goiás, Instituto de Ciências Biológicas, em especial ao professor Alexandre Siqueira Guedes Coelho, pelo apoio no trabalho, pelos ensinamentos e pela amizade.

À Universidade Federal de Goiás, Escola de Agronomia, em especial aos professores: Lázaro José Chaves e Ronaldo Naves Veloso, pelas coletas do material do presente estudo.

À Embrapa Arroz e Feijão, CNPAF, em especial para os pesquisadores Rosana Brondani e Cláudio Brondani pelo apoio nos experimentos com marcadores microssátelites. Aos novos e velhos amigos, Priscila, Tatiane, Tereza, Mara e Ana Cláudia.

Aos colegas do Curso de pós graduação em Genética e Melhoramento de Plantas da ESALQ/USP: Elizabete Takahashi, Mariza, Carlos Alberto, Lucianinha, Maria Cristina, Luciana Carlini, Caroline, Fernandinha, Leonardex, Maria Aldete, Alessandra, Aurélio e Andréa Milteman.

Ao laboratório de Genética Vegetal/IB-UFG e ao Laboratório de Biologia Molecular de Plantas do Setor de Melhoramento Vegetal/EA-UFG pela estrutura disponibilizada e pela amizade durante este período, em especial para Lizz, Nara, Mara, Mansuêmia, Ananda, Nozimeire, Ludmila, Sílvia, Susete, Glória, Gladys, Jaison, Flávio, Franklin, Cláudio, Reginaldo, Deys, Mônica, Alexandre, Naira e Fábio.

À CAPES, pela concessão da bolsa de estudo e à FUNAPE e CNPq pelo apoio financeiro para este trabalho.

A Maria Olávia Vencovsky pela amizade e carinho durante este período.

Aos meus pais: Ivani e Mathilde, minhas irmãs: Silvana, Denize, Maria do Rosário e Maristela e à toda família, pelo constante apoio e carinho.

Ao Baldin, pela paciência durante a realização deste trabalho, sempre me animando e me apoiando nas horas de desânimo.

SUMÁRIO

	Página
LISTA DE FIGURAS.....	xi
LISTA DE TABELAS.....	xi
RESUMO.....	xiv
SUMMARY.....	xvi
1 INTRODUÇÃO	1
1.2 Objetivos Gerais.....	3
2 REVISÃO DE LITERATURA.....	5
2.1 Importância da Espécie.....	5
2.2 Estrutura Genética de Populações de Plantas.....	7
2.2.1 Análise da Estrutura Genética com Marcadores Codominantes	7
2.2.1 Análise da Estrutura Genética com Marcadores Dominantes	11
2.3 Fluxo Gênico em Plantas.....	13
2.3.1 Modelos de Fluxo Gênico.....	15
2.3.2 Métodos de Estimação do Fluxo Gênico.....	16
2.4 Marcadores Moleculares no Estudo de Populações.....	19
2.5 Marcadores RAPD.....	23
2.6 Marcadores SSR.....	24
3 ESTRUTURA GENÉTICA E FLUXO GÊNICO DE <i>Eugenia dysenterica</i> DC, UTILIZANDO MARCADORES RAPD.....	28
Resumo.....	28
Summary.....	29
3.1 Introdução.....	29
3.2 Material e Métodos.....	31
3.2.1 Material Vegetal.....	31
3.2.2 Extração do DNA Genômico.....	34
3.2.3 Quantificação do DNA.....	35

3.2.4 Condições de Amplificação.....	35
3.2.5 Seleção de <i>Primers</i>	35
3.2.6 Análise Estatística dos Dados.....	36
3.3 Resultados.....	37
3.3.1 Seleção de <i>Primers</i>	37
3.3.2 Estrutura Genética e Fluxo Gênico.....	40
3.4 Discussão.....	42
3.5 Conclusões.....	45
4 ESTRUTURA GENÉTICA E FLUXO GÊNICO DE <i>Eugenia dysenterica</i> DC, UTILIZANDO MARCADORES SSR.....	46
Resumo.....	46
Summary.....	47
4.1 Introdução.....	47
4.2 Material e Métodos.....	49
4.2.1 Coleta do Material.....	49
4.2.2 Transferibilidade dos Locos de SSR e Condições de Amplificação.....	50
4.2.3 Metodologia de Análise Estatística dos Dados	51
4.3 Resultados.....	52
4.3.1 Transferibilidade de <i>Primers</i> SSR de <i>Eucalyptus</i> ssp. para <i>E.</i> <i>dysenterica</i>	52
4.3.2 Variação Genética.....	55
4.3.3 Estrutura Genética	60
4.4 Discussão.....	62
4.5 Conclusões.....	66
5 ESTRUTURA GENÉTICA DE <i>Eugenia dysenterica</i> DC, UTILIZANDO DADOS DE MARCADORES RAPD EM DUAS GERAÇÕES.....	67
Resumo.....	67
Summary.....	68
5.1 Introdução.....	69
5.2 Material e Métodos.....	71
5.2.1 Material Vegetal.....	71
5.2.2 Condições de Amplificação.....	72
5.2.3 Metodologia de Análise Estatística dos Dados	73

5.2.3.1 Estimativa do Sistema Reprodutivo Utilizando Dados de Progênies.....	73
5.2.3.2 Estimativa do Sistema Reprodutivo Utilizando Dados de Plantas Matrizes.....	74
5.2.3.3 Caracterização da Variabilidade e Estrutura Genética.....	74
5.2.3.4 Caracterização da Variabilidade e Estrutura Genética com Marcadores Dominantes Utilizando Dados de Progênies.....	76
5.3 Resultados.....	77
5.3.1 Seleção de <i>Primers</i>	77
5.3.2 Sistema Reprodutivo Estimado com Progênies.....	79
5.3.3 Estrutura Genética e Fluxo Gênico (Dados de uma Geração).....	80
5.3.4 Caracterização, Estrutura Genética e Fluxo Gênico (Dados de uma Geração).....	81
5.4 Discussão.....	86
5.5 Conclusões.....	90
6 ESTRUTURA GENÉTICA POPULACIONAL E PARÂMETROS AFINS AVALIADOS POR MARCADORES RAPD, SSR E ISOENZIMÁTICOS.....	91
Resumo.....	91
Summary.....	92
6.1 Introdução.....	94
6.2 Material e Métodos.....	97
6.2.1 Material Vegetal.....	97
6.2.2 Análise Estatística dos Dados.....	97
6.3 Resultados.....	100
6.3.1 Comparação da Taxa de Cruzamento.....	100
6.3.2 Caracterização, Estrutura Genética e Fluxo Gênico.....	101
6.3.3 Correlação entre Distâncias Genéticas e Geográficas.....	105
6.3.4 Correlação entre Matrizes de Distâncias Genéticas.....	107
6.3.4 Correlação entre Padrões Genéticos Edáficos e Morfológicos.....	107
6.4 Discussão.....	109
6.5 Conclusões.....	112
7 CONCLUSÕES GERAIS	114
REFERÊNCIAS BIBLIOGRÁFICAS.....	116

LISTA DE FIGURAS

	Página
1 Modelos de fluxo gênico. a) Modelo continente-ilha, b) Modelo de ilhas, c) Modelo de “alpondras” e d) Modelo de isolamento por distância.....	15
2 Municípios e localização das áreas de coleta de <i>E. dysenterica</i> DC. na região Sudeste do Estado de Goiás.....	32
3 Perfil de um gel de RAPD utilizando o primer OPA-11 em 74 plantas de cagaiteira; à esquerda está o padrão de peso molecular (ladder 100 pb).....	39
4 Padrão de similaridade genética obtido entre as 10 populações de cagaiteira, definido pelo critério de agrupamento UPGMA, com base na médias dos índices de Jaccard.....	41
5 Representação da correlação (r) entre a matriz de distâncias geográficas e a matriz de distâncias genéticas obtida por marcador RAPD. Significância de r dada pela probabilidade p	42
6 Histograma das frequências alélicas dos 7 locos de SSR, estimados para 116 indivíduos de <i>Eugênia dysenterica</i> . O eixo Y indica a frequência alélica e o eixo X indica o número do alelo.....	54
7 Perfil de um gel de SSR utilizando os primers EMBRA-210 e EMBRA-14 em 44 plantas de cagaiteira; à esquerda está o padrão de peso molecular um ladder 10 pb.....	56
8 Número de alelos exclusivos obtidos em 73 alelos de marcadores SSR, em 10 populações de <i>Eugênia dysenterica</i>	58

9 Padrão de divergência genética entre 10 populações de cagaiteira, definido pelo agrupamento UPGMA, com base nas identidade genética obtida a partir das distâncias genéticas de Nei (1978). Correlação cofenética igual a 0.97...	61
10 Representação da correlação entre a matriz de distâncias geográficas e a matriz de distâncias genéticas (r) obtida por marcador SSR. Significância de r dada pela probabilidade p	62
11 Perfil de um gel de RAPD utilizando o primer OPA-13 em 74 plantas de cagaiteira; à esquerda está o padrão de peso molecular de um ladder 100 pb.....	78
12 Padrão de estruturação genética entre 10 populações de cagaiteira, definido pelo agrupamento UPGMA, com base no índice de similaridade de Jaccard e distâncias genéticas de Nei (1978).....	85
13 Representação da correlação (r) entre a matriz de distâncias geográficas e matriz de distâncias genéticas, obtida por marcador RAPD com dados de duas gerações. p : probabilidade associada a significância de r	86
14 Padrão de estruturação genética de 10 populações de cagaiteira, definido pelo agrupamento UPGMA, com base distâncias genéticas de Nei (1972, 1978) e índice de similaridade de Jaccard. M: dados de plantas matrizes (adultos); M/P: dados de plantas matrizes genotipadas através das progênes e P: dados de progênes.....	104
15 Representação correlação (r) entre distâncias geográficas e distâncias genéticas obtidas por diferentes tipos de marcadores moleculares. Significância de r dada pela probabilidade p . M: plantas-mãe; p ; progênes...	106

LISTA DE TABELAS

	Página
1 Esquema de Análise de Variância (ANAVA).....	9
2 Esquema de Análise de Variância Molecular (AMOVA).....	12
3 Estimativas de F_{ST} e fluxo gênico em populações de espécies tropicais.....	19
4 Localidades do Estado de Goiás e número de árvores amostradas de cagaiteira, e suas respectivas informações geográficas.....	33
5 Sequência dos <i>primers</i> selecionados para <i>Eugenia dysenterica</i> e padrão de polimorfismo.....	37
6 Resultado da análise de variância molecular (AMOVA) e estimativa de fluxo gênico (Nm) calculado pelo método de Wright (1951).....	40
7 Localidades do Estado de Goiás e número de árvores amostradas de cagaiteira com suas respectivas informações geográficas.....	50
8 Sequências dos pares de primers desenvolvidos para <i>Eucalyptos</i> que amplificaram locos microssatélites em <i>Eugênia dysenterica</i> , com as respectivas amplitudes alélicas e temperatura de anelamento (T_a).....	53
9 Estimativas de parâmetros genéticos de diversidade em 10 populações de <i>Eugenia dysenterica</i> , sendo N: número de indivíduos amostrados; L: número de locos microssatélites; n_A : número médio de alelos; A: número total de alelos; H_o : heterozigosidade observada; H_e : heterozigozidade esperada sob equilíbrio de Hardy-Weinberg; f: índice de fixação, t_a : taxa de cruzamento aparente.....	57
10 Probabilidade do Teste Exato de Fisher para aderência ao equilíbrio de Hardy-Weinberg.....	58
11 Alelos exclusivos e freqüência com que ocorrem em cada população.....	59

12 Estimativa das estatísticas F de Wright, de R_{ST} e do número de migrantes por geração (Nm) em 10 populações naturais de <i>Eugenia dysenterica</i> . Intervalo de confiança (IC) de 95% de probabilidade.....	60
13 Localidades do Estado de Goiás, número de árvores amostradas de cagaiteira, e suas respectivas informações geográficas.....	72
14 Esquema de Análise de Variância Molecular (AMOVA).....	75
15 Sequência dos <i>primers</i> selecionados para <i>Eugenia dysenterica</i> e padrão de polimorfismo.....	77
16 Estimativa de parâmetros do sistema reprodutivo, sendo t_m : taxa de cruzamento multiloco; t_s : taxa de cruzamento baseado na média dos locos individuais. Entre parênteses: erro-padrão estimado através de 1.000 <i>bootstraps</i>	80
17 Resultado da análise de variância molecular (AMOVA) e estimativa do fluxo gênico (Nm).....	81
18 Estimativas relativas à diversidade genética em 10 populações de <i>Eugenia dysenterica</i> , sendo N: número de indivíduos amostrados; L: número de locos RAPD; A: número médio de alelos; H_b : heterozigosidade observada; H_e : heterozigosidade esperada; f: índice de fixação, t_a : taxa de cruzamento aparente.....	82
19 Probabilidade do Teste Exato de Fisher para aderência ao equilíbrio de Hardy-Weinberg.....	83
20 Estatísticas F de Wright, R_{ST} eo número de migrantes por geração (Nm) para 10 populações naturais de <i>Eugenia dysenterica</i> respectivas intervalos de confiança da 95% de probabilidade.....	84
21 Estimativas da taxa de cruzamento para 10 populações naturais de <i>Eugenia dysenterica</i> baseadas em diferentes marcadores.....	101
22 Estimativas de parâmetros genéticos em 10 populações de <i>Eugenia dysenterica</i> obtidas com diferente tipos de marcadores, sendo H_o : heterozigosidade observada; H_e : heterozigozidade esperada e f: índice de fixação.....	102
23 Estimativa das estatísticas F de Wright, R_{ST} e do número de migrantes por geração (Nm) para 10 populações naturais de <i>Eugenia dysenterica</i>	103

24	Correlação linear simples entre as distâncias genéticas, obtidas com os três marcadores. Os números entre parênteses indicam a significância de cada correlação.....	107
25	Correlação linear simples entre as distâncias genéticas, obtidas com os três diferentes marcadores e distâncias geográficas, edáficas, fenotípicas, a partir de dados de árvores, dos frutos e das progênies. Os números entre parênteses indicam a significância de cada correlação.....	108
26	Correlações parciais entre as distâncias genéticas e geográficas as, fixando a matriz de dados de árvores, de solos e das progênies.....	109

**ANÁLISE DA ESTRUTURA GENÉTICA DE *Eugenia dysenterica* DC
UTILIZANDO MARCADORES RAPD E SSR**

Autora: MARIA IMACULADA ZUCCHI

Orientador: Prof. Dr. ROLAND VENCOVSKY

RESUMO

Os marcadores moleculares têm sido frequentemente utilizados em estudos sobre a diversidade e a estrutura genética populacional. A introdução dos marcadores moleculares revolucionou a genética de populações na década de 50, com a técnica de isoenzimas e recentemente tem conseguido enormes avanços com a aplicação de tecnologias baseadas em DNA. Os marcadores moleculares baseados em DNA permitiram uma ampla cobertura genômica e tornaram-se poderosas ferramentas para estudos de genética de populações. Marcadores codominantes têm sido muito utilizados na genética de populações por serem mais informativos que os marcadores dominantes. O objetivo principal deste trabalho foi utilizar marcadores dominantes com o intuito de contornar o problema da dominância. Para isso, foram utilizadas neste trabalho progênies para genotipagens com RAPD visando inferir o genótipo das plantas matrizes. Foi encontrando um valor de variação entre populações de plantas jovens, igual a $\hat{f}_{ST}=0,328$. E estimativa similar a esta, foi calculada com os dados de freqüências alélicas das plantas matrizes, obtendo-se um valor de $\hat{F}_{ST}=0,318$. O objetivo secundário deste trabalho foi gerar várias estimativas de parâmetros populacionais com a finalidade de comparar os diferentes tipos de marcadores moleculares. Foram comparadas as estimativas relacionadas ao sistema reprodutivo (\hat{f}), heterozigidade esperada

(\hat{H}_e) e observada (\hat{H}_o), estatísticas F de Wright, o parâmetro \hat{f}_{ST} e dendrogramas obtidos com os diferentes tipos de marcadores (SSR, RAPD e isoenzimas) em diversas abordagens. A taxa de cruzamento (\hat{f}) obtida com os diferentes tipos de marcadores foram congruentes, sendo que a menor foi encontrada com isoenzimas utilizando progênies ($\hat{f}_m=0,835$), e a maior obtida com marcadores SSR ($\hat{f}_a=1,07$) utilizando indivíduos adultos. Isto levou a concluir que a espécie é alógama ou com tendência para alogamia. Com relação à estrutura genética e o fluxo gênico as estimativas não foram inteiramente concordantes. A menor estimativa de divergência foi obtida com marcadores isoenzimáticos ($\hat{q}_p=0,154$, $Nm=1,370$) e a maior com RAPD ($\hat{f}_{ST}=0,328$, $Nm=0,512$). Porém é importante ressaltar, que estas estimativas não são diretamente comparáveis, pois são obtidas de maneiras diferentes e com dados de progênies ou adultos. Quanto à estruturação da variabilidade visualizada através de agrupamentos, observou-se que os dendrogramas apresentam um padrão concordante sendo que a maior sensibilidade (maior amplitude de distâncias) foi obtida com os dados de SSR. Além disso, foram feitas correlações entre as matrizes de distâncias genéticas e geográficas e entre as matrizes de distâncias genéticas (correlações entre os diferentes marcadores). Estas correlações foram altas indicando que todos os marcadores amostram bem o genoma. Este estudo permitiu comparar dados de duas gerações e uma metodologia diferente para análise de dados com marcadores dominantes, sem impor restrições nos modelos ($f=1$ ou $f=0$). Pôde-se concluir que houve congruência entre as estimativas dos parâmetros populacionais obtidas com os diferentes tipos de marcadores, embora tenha ocorrido algumas discrepâncias como foi o caso da estimativa de fluxo gênico com marcadores isoenzimáticos. Apesar de os marcadores possuírem diferentes naturezas todos foram igualmente informativos para este estudo populacional, inclusive os dominantes quando foram usados dados de duas gerações.

**ANALYSIS OF THE GENETIC STRUCTURE OF *Eugenia dysenterica* DC USING
MOLECULAR MARKERS (RAPDs and SSRs)**

Author: MARIA IMACULADA ZUCCHI

Adviser: Prof. Dr. ROLAND VENCOVSKY

SUMMARY

Molecular markers have frequently been used in studies on diversity and population genetic structure. The introduction of these markers revolutionized the genetics of populations in the decade of 50 with the isoenzymes technique and recently enormous progresses have been achieved with the application of technologies based on DNA. Molecular markers based on DNA allowed a wide coverage of the genome and therefore became a powerful tool for these studies. Codominant markers, being more informative than the dominant ones are now being widely used. The main objective of this work was to use dominant (RAPD) markers overcoming the problem of dominance through progeny testing. Offspring were genotyped in order to infer the maternal genotype. Using the AMOVA procedure the variation value found among populations of seedlings was equal to $\hat{f}_{ST} = 0.328$. The corresponding F statistic, based on allelic frequencies of seed parents, was $\hat{F}_{ST} = 0.318$. The second objective of this work was to obtain different estimates of population parameters with the purpose of comparing the different types of molecular markers. In addition to Wright's F statistics, estimates related to the reproductive system (\hat{t}), expected (\hat{H}_e) and observed heterozygosities (\hat{H}_o) and dendrograms obtained with markers SSR, RAPD and isoenzymes were compared. The outcrossing rate (\hat{t}) obtained agreed reasonably well, and the

smallest value was found with isozymes using progenies ($\hat{t}_m=0.835$), and the largest one, obtained with SSR markers ($\hat{t}_a=1.07$) using adult individuals. The results led to a conclusion that the species is allogamic or with tendency to allogamy. Estimates related with the genetic structure and gene flow, however, were not entirely congruent. The smallest divergence estimate was obtained with isozymes ($\hat{q}_p=0.154$, $Nm=1.370$) and the largest one with RAPD ($\hat{f}_{ST}=0.328$, $Nm=0.512$). It is important to point out, that these estimates are not directly comparable, due to the difference in the underlying model and the basic material used (progenies or adults). About the structuring of the variability visualized through groupings, dendrograms presented a congruent pattern and the largest sensibility (larger range of distances) was obtained with SSR data. In addition, correlations were calculated among the matrices of genetic and geographical distances and among the matrices of genetic distances obtained from the different markers. High correlations were verified among these matrices, indicating that all markers sampled well the genome under study. This study allowed to compare data of two generations and a different methodology for data analysis with dominant markers, without imposing restrictions on the model ($f=1$ or $f=0$). It could be concluded that a reasonable consistency was verified among the population parameters obtained with different types of markers, although some discrepancies existed as in the case of estimates of gene flow with isozymes markers. In spite of their different origins, the markers used here were equally informative for this population study, including the dominant one, when data of two generations were used.

1 INTRODUÇÃO

A espécie *E. dysenterica*, vulgarmente conhecida como cagaiteira, uma frutífera nativa da região de Cerrados, foi escolhida como modelo para este estudo de análise da estrutura genética populacional utilizando diferentes abordagens com marcadores moleculares dominantes e codominantes.

Grande parte das frutíferas nativas apresenta potencial de utilização em sistemas tradicionais de produção agrícola (Almeida et al., 1997) sendo este um motivo para a escolha da espécie. Outro motivo é que esta espécie permite a obtenção de um grande número de indivíduos por matriz, o que é necessário para o estudo utilizando marcador dominante, em duas gerações.

As frutíferas nativas do Cerrado são espécies de diversos gêneros e famílias que produzem frutos de interesse tanto para a alimentação "in natura" quanto para a industrialização. Há um mercado potencial e crescente para as frutíferas nativas, porém pouco explorado pelos agricultores. Todo o aproveitamento desses frutos tem sido feito de forma extrativista.

Várias espécies frutíferas nativas estão sendo estudadas atualmente e a maioria delas encontra-se em estado silvestre, sem qualquer grau de domesticação. Uma vez domesticadas e cultivadas em lavouras comerciais, evitar-se-ia o extrativismo predatório, ao mesmo tempo em que estas se conservariam na natureza. Dentre as espécies que apresentam potencial de utilização em sistemas tradicionais de produção agrícola, a cagaiteira (*E. dysenterica*), merece destaque pelo seu potencial econômico.

Um amplo conhecimento das características demográficas e variabilidade da espécie é fundamental para a conservação e manejo sustentável. Dentre as características mais importantes para a conservação, destacam-se: o conhecimento do fluxo gênico, do sistema reprodutivo e da diversidade entre populações.

A avaliação da variabilidade genética sempre foi de interesse para os geneticistas, que desenvolveram vários métodos para detectá-la e analisá-la. A

introdução das técnicas de eletroforese de isoenzimas revolucionou a genética de populações na década de 50 e recentemente tem conseguido enormes avanços com a aplicação de tecnologias baseadas em polimorfismo da molécula de DNA. Os marcadores moleculares baseados em DNA (RAPD, RFLP, SSR, SCAR, AFLP e outros) permitiram uma ampla cobertura genômica, quando comparados com as isoenzimas.

Porém a realidade dos laboratórios que estudam genética de populações até recentemente estavam limitados aos marcadores isoenzimáticos pelo fato de serem marcadores codominantes e de fácil elaboração da técnica. Os marcadores codominantes (RFLP, SSR e isoenzimáticos) têm vantagens sobre os dominantes (RAPD, AFLP, VNTR, etc) pelo fato que os primeiros conseguem distinguir o genótipo heterozigoto do homozigoto, enquanto que os demais apresentam limitações, na medida em que não discriminam o heterozigoto. Desta forma os marcadores codominantes são mais informativos, enquanto que com os dominantes, só é possível o cálculo das frequências alélicas, sob condições especiais: equilíbrio em autogamia, quando o valor de $f=1,0$ ou equilíbrio em alogamia (ou equilíbrio de Hardy-Weinberg), quando $f=0,0$.

Atualmente, existe uma tendência de utilização de marcadores RAPD, apesar de sua dominância (não distinção do heterozigoto) e segundo Lynch & Milligan (1994), dentre os diversos tipos de marcadores atualmente disponíveis, os marcadores RAPD destacam-se pelas vantagens que apresentam em termos de simplicidade (como exemplo: a visualização das marcas RAPD pode ser realizada sem o uso de sondas radioativas), pelo fato de serem aplicados a um grande número de espécies e por permitirem a análise de um grande número de locos. Porém, o uso destes tipos de marcadores possibilita ainda, uma amostragem aleatória mais ampla do genoma do que a proporcionada por outros métodos convencionais. O uso dos marcadores dominantes, no entanto, tem sido muito limitado pela falta de informação genotípica do marcador (Ferreira & Grattapaglia, 1996). Embora alguns modelos estatísticos tenham sido desenvolvidos no sentido de contornar estas limitações, permitindo a estimação de parâmetros populacionais com a utilização de marcadores dominantes (Excoffier et al., 1992 e Lynch & Milligan, 1994), o uso destes modelos ainda é limitado.

Os marcadores microssatélites são marcadores moleculares considerados ideais para a estimação de parâmetros genéticos de populações e para a

compreensão de padrões de fluxo gênico e parentesco. Estes marcadores são abundantes e uniformemente distribuídos pelo genoma de plantas, além de serem codominantes, multialélicos e com alta heterozigosidade. Ressalta-se que sua utilização em espécies nativas ainda é complicada, pela dificuldade de desenvolver *primers* para estas espécies e devido ao elevado custo e tempo.

O presente estudo teve como objetivo principal avaliar a utilização de marcadores RAPD em duas gerações (matrizes e progênies) com o objetivo de inferir o genótipo materno para o posterior cálculo das frequências alélicas sem admitir restrições ($f=0$ ou $f=1$), ou seja, para espécies de reprodução mista como é o caso de algumas espécies nativas. Além disso, buscou-se avaliar a transferibilidade, bem como a caracterização de marcadores microssatélites em cagaiteira, para estudos sobre variabilidade genética, estrutura populacional, fluxo gênico e sistema reprodutivo. Este trabalho permitiu ainda comparar os dois tipos de marcadores moleculares (SSR e RAPD) no estudo da estrutura de populações fornecendo informações para a domesticação, conservação e pré-melhoramento da cagaiteira.

1.2 Objetivos Gerais

- ✓ Selecionar *primers* de marcadores SSR (baseados em polimorfismos de microssatélites) para *Eugenia dysenterica* a partir de *primers* desenvolvidos para *Eucalyptus*.
- ✓ Comparar as metodologias de caracterização genética populacional por marcadores dominantes (RAPD) (utilizando progênies) com dados de marcadores codominantes (SSR).
- ✓ Desenvolver metodologia para o estudo da estrutura genética de populações naturais utilizando marcadores dominantes (RAPD), analisando dados de duas gerações, visando dar subsídios às pesquisas básicas dessa natureza desenvolvidas em laboratórios que estudam diversidade.

- ✓ Comparar as estimativas obtidas de marcadores RAPD e SSR com as obtidas de marcadores isoenzimáticos existentes para as mesmas populações utilizadas neste estudo.

- ✓ Estabelecer correlações entre caracteres fenotípicos (das árvores e progênies) e de solo, obtidos das mesmas plantas e regiões estudadas neste trabalho, e os dados obtidos com marcadores RAPD e SSR.

2 REVISÃO DE LITERATURA

2.1 Importância da Espécie

Segundo Cruz (1979), o nome vulgar cagaiteira tem sido utilizado para designar várias espécies da família Myrtaceae cujos frutos produzem disenteria quando ingeridos em excesso ou fermentados pelo sol. A árvore estudada era designada como *Stenocalyx dysentericus* Berg e apresenta como sinonímia *Eugenia dysenterica* DC e *Myrthus dysentericus* M.

A cagaiteira (*Eugenia dysenterica* DC) é uma árvore mediana (quatro a dez m de altura) de tronco e ramos tortuosos, com a casca profundamente sulcada, folhas glabras, simples e opostas, um tanto alongadas, de pecíolos curtos e brilhantes. As flores são hermafroditas, de cor branca, grandes, nascendo isoladamente. Apresenta o fruto arredondado, amarelo-esverdeado, com cerca de dois cm de comprimento, geralmente com uma só semente envolta em polpas aciduladas e com a casca do fruto membranácea (Corrêa, 1984).

A espécie *E. dysenterica* é pertencente à família Myrtaceae. Esta família é composta por 14 gêneros, representados por 211 espécies de ocorrência natural no bioma cerrado, sendo uma das 10 famílias mais representadas do bioma que juntas contribuem com mais de 51% de sua riqueza florística. É uma espécie típica da região de Cerrados brasileiros, ocorrendo nos cerradões, cerrado *stricto sensu* e campos sujos do estado de Goiás, Tocantins, Bahia, Minas Gerais, São Paulo e Distrito Federal (Donadio et al., 1992).

Quanto à distribuição espacial, a cagaiteira é considerada rara com ocorrência em agregados (Chaves & Naves, 1998). Em levantamento realizado em 50 áreas de 1,0 ha de cerrado pouco antropizado, Naves (1999) encontrou a espécie ocorrendo em apenas 10 áreas.

De acordo com Proença & Gibbs (1994) a cagaiteira apresenta estratégia de florescimento tipo “big-bang”, na qual ocorre uma sincronização do florescimento em um período de tempo muito curto. Segundo Heringer & Ferreira (1974), a maior frequência da floração desta espécie está no mês de agosto, sendo as flores de cor branca e em grande número. Como principais polinizadores foram identificadas as abelhas, incluindo espécies do gênero *Bombus* (*B. atratus* e *B. morio*) e outras espécies como *Apis mellifera*, que assumem um importante papel na polinização da espécie.

Segundo Ferreira & Cunha (1980) a cagaiteira, assim como outras espécies da família Myrtaceae, possui dois tipos principais de dispersão, o primatocórico e antropocórico. O primatocório refere-se a um grupo de plantas que são procuradas por macacos sagüis e no antropocórico a disseminação é feita pelo homem, onde as sementes são indiretamente disseminadas através de implementos agrícolas. Segundo Rizzini (1971) a cagaita, de uma maneira geral, é dispersa por animais e portanto denominada de zoocórica.

Dentre as diversas frutíferas nativas dos cerrados que apresentam potencial de utilização em sistemas de produção agrícola, a cagaiteira (*E. dysenterica* DC) tem se destacado pela importância sócio-econômica devido ao seu grande potencial para o processamento de vários subprodutos, o que poderia contribuir para a geração de emprego e renda para as comunidades regionais.

A cagaiteira além de ser uma planta ornamental e melífera, presta-se à extração de cortiça e ainda encontra utilização em pequenas construções civis ou na fabricação de carvão, fornecendo lenha de boa qualidade, sendo sua casca utilizada em curtumes. Sua folha tem propriedades antidiarréicas, existindo relatos do seu uso para o tratamento de diabetes e icterícia; e seus frutos têm qualidade laxativa (Heringer & Ferreira, 1974).

A produção dos frutos é alta, chegando a 2000 frutos por árvore (Almeida et al., 1987). Desta forma, apresenta elevado potencial para a comercialização dos frutos e de seus derivados na forma de sorvetes, licores e sucos.

2.2 Estrutura Genética de Populações de Plantas

2.2.1 Análise da Estrutura Genética com Marcadores Codominantes

A estrutura genética refere-se à distribuição heterogênea (não aleatória) dos alelos e genótipos no espaço e no tempo resultante da ação de forças evolutivas tais como: mutação, migração, seleção e deriva genética que atuam dentro do contexto de cada espécie e população (Hamrick, 1982).

O número de alelos por loco, a heterozigozidade esperada e observada têm sido os parâmetros genéticos mais utilizados em vários estudos para quantificar a variabilidade genética em populações de plantas (Pinto, 2001).

A estimativa da frequência de um alelo particular em uma população, chamada frequência gênica ou alélica, é considerada fundamental nos estudos evolutivos, pois a mudança genética de uma população pode ser avaliada pela mudança nas suas frequências gênicas (Nei, 1978). A frequência de um determinado alelo em um grupo de indivíduos diplóides é igual a duas vezes o número de indivíduos homozigóticos para o alelo, mais o número de indivíduos heterozigóticos, dividido por duas vezes o número de indivíduo na amostra, em virtude de cada indivíduo carregar dois alelos em um loco.

O número de alelos observados por loco aumenta em função do tamanho da amostra. Deste modo, amostras grandes possuem maior chance de detectar alelos raros.

O conhecimento da frequência dos heterozigotos apresenta importância na medida que cada heterozigoto carrega diferentes alelos, os quais demonstram a existência de variação genética na população (Weir, 1996).

De acordo com o princípio de Hardy-Weinberg em uma população de tamanho infinito, praticando cruzamentos ao acaso, as frequências gênicas e genotípicas permanecem constantes de geração para geração na ausência de migração, seleção e deriva. Este princípio permite o cálculo teórico da frequência de um determinado genótipo independente do número de alelos existente.

A caracterização da estrutura genética entre populações, através de marcadores codominantes pode ser abordada de três maneiras diferentes: estatísticas F de Wright (Wright, 1965), análise da diversidade gênica em

populações subdivididas (Nei, 1977) e os coeficientes de coancestralidade de Cockerham (Cockerham, 1969; Vencovsky, 1992 e Weir, 1996).

As três abordagens apresentam bases genéticas similares, porém, são complementares em relação ao significado biológico das estimativas obtidas (Reis, 1996). Desta forma as estatísticas F permitem a caracterização da distribuição da variabilidade genética entre as populações (F_{ST}), assim como dos níveis médios de endogamia ao nível populacional (F_{IS}) e total (F_{IT}). Os coeficientes de coancestralidade (θ) possibilitam a avaliação da divergência em diferentes níveis de hierarquia, além de possibilitar a obtenção de estimativas de endogamia, a partir de uma base não viesada. E a análise da divergência gênica em populações subdivididas permite a comparação dos níveis de heterozigosidade entre e dentro das populações, bem como a obtenção de uma estimativa de divergência, a partir de uma base diferente daquela que fundamenta as estimativas de F_{ST} e θ .

As estatísticas F, podem ser apresentadas de várias maneiras, conforme Wright (1951, 1965), Cockerham (1969) ou sob a ótica da diversidade de acordo com Nei (1977).

De acordo com Wright (1965):

$$1 - \hat{F}_{IT} = (1 - \hat{F}_{IS})(1 - \hat{F}_{ST})$$

\hat{F}_{IT} = índice de fixação ou coeficiente de endogamia para o conjunto das populações (devido ao sistema reprodutivo e subdivisão)

\hat{F}_{IS} = índice de fixação ou coeficiente de endogamia intrapopulacional (devido ao sistema reprodutivo)

\hat{F}_{ST} = índice de fixação ou coeficiente de endogamia entre populações (devido à subdivisão)

Sob modelo aleatório, pode-se considerar que as populações amostradas representam a espécie tendo história evolutiva comum. Mesmo que as populações tenham se diferenciado com o decorrer do tempo, a análise é construída sob a hipótese de que há uma única população de referência. Na ausência de forças evolutivas que alterem as frequências gênicas, como diferentes pressões de seleção em diferentes populações, espera-se que todas as populações tenham as mesmas frequências alélicas. A análise da diferenciação em modelos aleatórios é

subordinada ao fato de que amostragem genética faz com que diferentes alelos na população sejam dependentes ou relacionados. Mesmo que indivíduos ou alelos possam ser amostrados aleatoriamente, quando se obtêm as esperanças, deve-se saber que são dependentes de um ancestral comum.

O interesse deve ser voltado ao quanto as populações se diferenciam dentro de uma espécie, a qual se diferenciou no decorrer do tempo. A ação das forças evolutivas, ou da amostragem genética, resultarão na diferenciação intraespecífica, quantificada com as estatísticas F de Wright ou medidas análogas de Cockerhan. Estas medidas também quantificam o grau de parentesco entre pares de alelos. As três quantidades básicas, na situação onde indivíduos diplóides são amostrados em várias populações, de acordo com Cockerhan (1969). Considerando os níveis de hierarquia (s = populações, n =indivíduos e $g=2$ em organismos diplóides), pode-se verificar no esquema de análise de variância na Tabela 1:

Tabela 1. Esquema de Análise de Variância (ANAVA).

FONTE DE VARIAÇÃO	GL	SQ	QM	E(QM)
Entre populações	$s-1$	SQ_p	QM_p	$\mathbf{s}_G^2 + 2\mathbf{s}_I^2 + 2n\mathbf{s}_p^2$
Entre indivíduos dentro de populações	$n(s-1)$	SQ_I	QM_I	$\mathbf{s}_G^2 + 2\mathbf{s}_I^2$
Genes dentro de indivíduos	$ns(2-1)$	SQ_G	QM_G	\mathbf{s}_G^2
TOTAL	$2ns-1$			

$$\mathbf{s}_T^2 = \mathbf{s}_G^2 + \mathbf{s}_I^2 + \mathbf{s}_p^2$$

Os componentes de variância podem ser definidos como:

$$\mathbf{s}_p^2 = p(1-p)\mathbf{q} \text{ para populações}$$

$$\mathbf{s}_I^2 = p(1-p)(F - \mathbf{q}) \text{ para indivíduos dentro de populações}$$

$$\mathbf{s}_G^2 = p(1-p)(1-F) \text{ para alelos dentro de indivíduos}$$

Seguindo a analogia, podemos reescrever as equações acima para os parâmetros das Estatísticas F.

$$F_{ST} = \frac{\mathbf{s}_p^2}{\mathbf{s}_T^2}$$

$$F_{IT} = \frac{\mathbf{s}_I^2 + \mathbf{s}_p^2}{\mathbf{s}_T^2}$$

$$F_{IS} = \frac{\mathbf{s}_I^2}{\mathbf{s}_I^2 + \mathbf{s}_G^2}$$

F_{IT} de Wright = F de Cockerham: é a correlação entre alelos dentro de indivíduos em todas as populações. É o coeficiente de endogamia que se refere aos indivíduos em relação ao conjunto de populações, reunindo a informação dos índices de fixação F_{ST} e F_{IS} .

F_{ST} de Wright = θ de Cockerham: é a correlação dos alelos dentro de indivíduos na mesma população. É o coeficiente de coancestralidade.

F_{IS} de Wright = f de Cockerham: correlação dos alelos dentro de indivíduos dentro das populações.

De acordo com Nei (1977):

$$\hat{F}_{IT} = 1 - \frac{\hat{H}_{ot}}{\hat{H}_{et}}$$

$$\hat{F}_{IS} = 1 - \frac{\hat{H}_{oi}}{\hat{H}_{ei}}$$

$$\hat{F}_{ST} = 1 - \frac{\hat{H}_{ei}}{\hat{H}_{et}}$$

onde:

$$\hat{H}_{ot} = \frac{\sum \sum \hat{x}_{il}}{ls}$$

$$\hat{H}_{ei} = 1 - \frac{\sum \sum \hat{x}_{ilk}^2}{ls}$$

$$\hat{H}_{et} = 1 - \frac{\sum \left(\frac{\sum \hat{x}_{ilk}}{s} \right)^2}{l}$$

sendo:

\hat{H}_{et} = heterozigosidade esperada total

\hat{H}_{oi} = heterozigosidade observada total

\hat{H}_{ei} = heterozigosidade esperada média total

\hat{x}_{il} = frequência de heterozigotos do loco I na população i

\hat{x}_{ilk} = frequência do alelo k do loco I na população i

s = número de populações

l = número de locos

2.2.2 Análise da Estrutura Genética com Marcadores Dominantes

Para analisar marcadores dominantes, Excoffier et al. (1992) pela introdução da estatística ϕ , proporcionaram uma nova alternativa para esse tipo de marcadores, que foi utilizada pela primeira vez por Huff et al. (1993).

Informações sobre a divergência de DNA de dados provenientes de haplótipos foram incorporadas na forma de análise de variância, derivada da matriz de distâncias quadradas entre todos os pares de haplótipos. Esta análise de variância, denominada AMOVA, produz estimativas dos componentes de variância das análogas estatísticas F, que os autores chamaram de estatísticas ϕ , que refletem a correlação da diversidade dos haplótipos em diferentes níveis de subdivisão hierárquica. O método acomoda diversos tipos de matrizes de entrada fornecidas por diversos tipos de marcadores moleculares, e diferentes tipos de pressuposições evolutivas, sem modificar a estrutura básica da análise. A significância dos componentes de variância e das estatísticas ϕ , é testada através do uso de permutações. Os autores, ao aplicarem esta análise a dados de haplótipos de DNA mitocondrial humano, verificaram que as subdivisões de populações são melhor resolvidas se alguma medida de diferenças moleculares entre haplótipos for introduzida na análise. Segundo os autores, a AMOVA é facilmente aplicável em diferentes situações e constitui uma estrutura coerente e flexível para análise de dados moleculares.

A base da análise apresentada por Excoffier et al. (1992) é que as somas de quadrados convencionais (SQ) podem ser escritas na forma de somas de quadrados de diferenças entre pares de observações. Desta forma, eles construíram uma análise de variância molecular hierárquica, partindo diretamente da matriz das distâncias quadradas de todos os pares de haplótipos. Os autores estudaram dez populações humanas, para as quais amplos conjuntos de dados são disponíveis na literatura. Estas populações representam cinco grupos regionais, com duas populações cada. O polimorfismo do material amostrado foi analisado através de cinco enzimas de restrição.

A análise de variância hierárquica de mtDNA de humanos mostrou significativa subdivisão entre populações, mas com grande variação dentro de populações (acima de 74%). Os resultados sugeriram que estudos dentro de cada

região são necessários para determinar se o fato de a diversidade observada entre regiões ser maior que a observada entre populações dentro de regiões é um artefato da escolha arbitrária das populações, uma consequência do isolamento pela distância, ou zonas de fronteira limite de trocas genéticas entre regiões (Excoffier et al., 1992).

Huff et al. (1993) utilizaram a AMOVA pela primeira vez em plantas. Eles estudaram a variação genética entre e dentro de populações naturais de *Buchloe dactyloides*. Baseando-se em dados provenientes de quatro populações provenientes do México e duas do Texas (EUA). Buso et al. (1998) estudaram a variabilidade genética de populações de arroz da América do Sul utilizando marcadores isoenzimáticos e RAPD.

Exemplificando, podemos considerar os níveis de hierarquia (P= número total de populações, N=número total de dados genotípicos ou número total de genes para dado haplótipo), pode-se verificar na Tabela 2:

Tabela 2. Esquema de Análise de Variância Molecular (AMOVA).

FONTE DE VARIAÇÃO	GL	SQ	QM	E(QM)
Entre populações	P-1	SQ _a	QM _a	$\mathbf{s}_b^2 + 2\mathbf{s}_a^2$
Dentro de populações	N-P	SQ _b	QM _b	\mathbf{s}_b^2
TOTAL	N-1			

$$\mathbf{s}^2 = \mathbf{s}_b^2 + \mathbf{s}_a^2$$

Os componentes de variâncias de cada nível de hierarquia são extraídos das esperanças dos quadrados médios. Segundo Cockerham (1969), temos:

$$\mathbf{f}_{ST} = \frac{\mathbf{s}_a^2}{\mathbf{s}^2}$$

Onde: \mathbf{f}_{ST} = a proporção da variabilidade molecular de haplótipos entre populações

2.3 Fluxo Gênico em Plantas

A perda da variabilidade genética é o tópico central da conservação genética (Avice, 1994). Áreas fragmentadas onde existem pequenas populações principalmente de plantas alógamas estão propensas à endogamia e deriva genética resultante da subdivisão. A endogamia pode atuar desmascarando determinados alelos deletérios recessivos, diminuindo num curto espaço de tempo o valor adaptativo da população, nas quais ocorre redução da heterozigosidade e pode resultar em perda da diversidade alélica. Biologistas e conservacionistas acrescentam que este declínio na variação genética pode inibir no futuro a adaptação do organismo às mudanças ambientais e conseqüentemente limitar seu potencial evolucionário, podendo levar essas populações a um possível risco de extinção.

A terminologia aplicada para o fluxo gênico em modelos populacionais é potencialmente confusa por causa dos muitos significados dos termos migração e dispersão, que não correspondem aos seus significados quando aplicados ao fluxo. Em geral, migração pode ser definida como qualquer movimento, incluindo movimentos cíclicos que podem regularmente retornar o organismo ao seu local de origem. Em contraste, o termo dispersão é mais precisamente restrito para movimentos que aumentam a distância entre os organismos, gametas ou propágulos da fonte dispersora (Neigel, 1997).

Neste contexto, fluxo gênico é definido como o movimento de genes em populações e, portanto inclui todos os movimentos de gametas, propágulos e indivíduos que efetivamente trocam genes na distribuição espacial (Neigel, 1997).

Conforme Slatkin (1981, 1985) fluxo gênico (ou fluxo alélico) é um termo coletivo que inclui todos os mecanismos que resultam no movimento de alelos de uma população para outra.

Ainda, Levin & Kerster (1974) definem fluxo gênico potencial como a deposição do pólen e sementes a partir de uma população fonte em função da distância. O fluxo gênico efetivo, refere-se à incidência de fertilização (no caso de pólen) e ao estabelecimento de indivíduos reprodutivos (no caso de sementes), em função da distância da população fonte e fluxo gênico no tempo. É devido à

ocorrência de dormência nas sementes de algumas espécies, gerando uma sobreposição de gerações sucessivas, podendo funcionar como mecanismo de fluxo gênico no tempo.

O fluxo gênico tem sido amplamente discutido em relação à sua magnitude e influência na estrutura genética das populações. A importância do fluxo gênico principalmente em populações naturais está na homogeneização das frequências alélicas entre as populações pequenas, deste modo, mesmo que separadas geograficamente elas comportam-se como uma grande população panmítica. Com frequências alélicas que antes do fluxo eram diferentes, depois do fluxo elas se tornam homogêneas entre si. A sua maior importância está na manutenção da diversidade genética e do polimorfismo. Em plantas a transferência de genes pode ocorrer tanto pelo movimento de organismos individuais (sementes, rizomas, estolões), como, pelo movimento de gametas (pólen).

Segundo Futuyma (1993), o grau no qual uma população pode ser delimitada de outras depende do nível de fluxo gênico entre elas. A taxa de fluxo m_j da população j para a população i é a proporção de indivíduos que se reproduzem na população i que migram para a população j naquela geração. Se m_j é muito alta, próxima de 0,5, as duas populações são, na realidade, uma população panmítica cujo tamanho populacional é a soma das duas. Desta forma, a taxa de fluxo gênico influencia o tamanho efetivo da população.

A importância do fluxo gênico está justamente em contrapor os efeitos da deriva genética, permitindo a homogeneização das frequências alélicas. O fluxo pode ser quantificado através de medidas diretas e indiretas. Os métodos diretos são via: corantes, marcadores morfológicos e análise da paternidade. Os métodos indiretos são via: F_{ST} , alelos privados, autocorrelação espacial e coalescência. Recentemente, abordagens mais precisas têm sido utilizadas neste sentido como: marcadores organelares (cpDNA e mtDNA), análise da paternidade utilizando microssátelites, aliados a estudos ecológicos.

2.3.1 Modelos de Fluxo Gênico

Existem diversos modelos de fluxo gênico:

- Modelo continente-ilha, no qual existe um movimento unidirecional de uma população grande continental para uma população menor isolada (Figura 1a);
- Modelo de ilha, no qual a migração ocorre ao acaso entre grupos de pequenas populações (Figura 1b);
- Modelo de alpondras (ou stepping-stone), no qual a população recebe migrantes somente de populações vizinhas (Figura 1c);
- Modelo de isolamento por distância, no qual o fluxo ocorre localmente entre os vizinhos, em uma população de distribuição contínua (figura 1d).

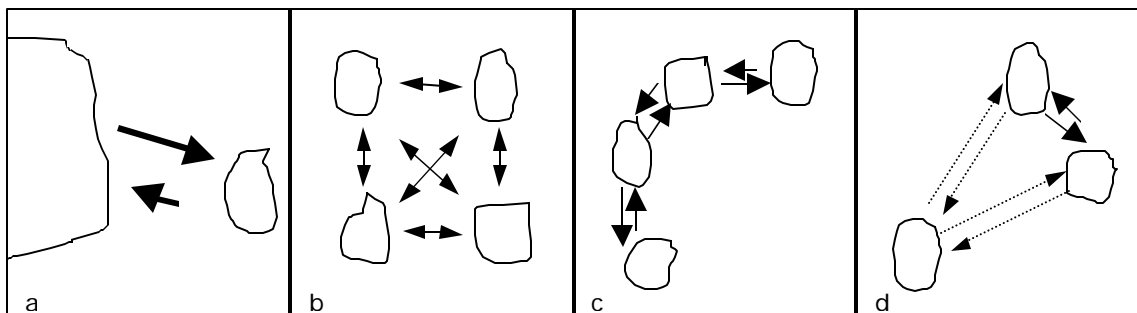


Figura 1 - Modelos de fluxo gênico. a) Modelo continente-ilha, b) Modelo de ilhas, c) Modelo de "alpondras" e d) Modelo de isolamento por distância.

Atualmente novas abordagens têm sido usadas para estudar o fluxo gênico, como o modelo de metapopulação e o de paisagens que têm sido muito utilizados na literatura (Sork et al., 1998 e Sork et al., 1999).

O modelo de ilhas infinitas de Sewall Wright, é o modelo convencional de genética de populações de troca de genes entre populações. Neste modelo, o número efetivo de migrantes, N_m , é estimado através da estatística F para um conjunto n de populações. O modelo assume equilíbrio entre migração e deriva genética entre todas as populações. Há igual troca de genes entre populações. Este modelo assume que todas as populações são iguais fontes de migrantes e produz

estimativas que não refletem a variação contemporânea na troca de genes entre populações ou mudanças atuais no processo dispersivo.

O modelo de metapopulação pode ser definido como um conjunto de populações (ou subpopulações locais) dinâmicas, interagindo através do fluxo gênico. Assume um conjunto de populações conectadas por fluxo gênico, porém controlada por extinções e novas colonizações.

Em populações de plantas são muito raros estudos comprovando que populações estudadas são realmente uma metapopulação, pois medir a extinção e a recolonização em espécies mesmo anuais é uma tarefa muito árdua. Husband & Barret (1995) usaram a abordagem de metapopulação para determinar se manchas dinâmicas podem explicar a persistência da espécie aquática *Eichhornia paniculata* na paisagem sazonal árida do nordeste do Brasil. Recenseamentos anuais de 167 populações, durante 7 anos consecutivos e avaliação da sobrevivência dessas populações através de habitats aquáticos, revelou que a extinção anual em populações de *E. panniculata* foi de 34% devido à seca, distúrbios (interrupção) e severa inundação com persistência de 66% das populações. Este resultado sugere indubitável reflexo da estocasticidade ambiental que é experimentada por muitas plantas encontradas em habitats efêmeros, onde mudanças catastróficas no local podem causar a extinção, independentemente da característica demográfica das populações locais.

2.3.2 Método de Estimação do Fluxo Gênico

O fluxo gênico pode ser quantificado através de medidas diretas e indiretas. As medidas diretas referem-se ao fluxo gênico contemporâneo, enquanto que as indiretas são baseadas na estrutura de populações e referem-se ao fluxo gênico histórico, ou passado. Os métodos diretos são inferidos através da utilização de corantes, marcadores morfológicos e análise da paternidade. Há relativamente poucas medidas diretas de fluxo gênico em populações naturais de plantas. Uma complicação que acontece com plantas é que os genes podem se mover em dois caminhos (semente e pólen) e para uma acurada medida do fluxo gênico ambos devem ser descritos e quantificados.

O movimento de agentes de dispersão são presumidos como representantes da dispersão do pólen ou sementes. Há um aumento da evidência de que o movimento do polinizador sub-representa a atual movimentação gênica devido ao alto carregamento de pólen (Levin & Kerster, 1974).

A dispersão por sementes tem grande potencial para documentar o fluxo gênico. Porém, não fornece informações sobre o sistema reprodutivo das espécies. Espécies com alta taxa de auto-fecundação (autogamia) com movimento a longa distância de liberação de pólen, podem ter menos fluxo gênico que espécies predominantemente alógamas com liberação de pólen local (Levin & Kerster, 1974).

Finalmente, movimentos de agentes dispersores não fornecem nenhuma informação a respeito do sucesso do grão-de-pólen na fertilização ou sobre a sobrevivência e estabelecimento das sementes.

Outro método direto é a análise da paternidade, que consiste na utilização de locos gênicos para identificar o mais provável pai de um conjunto de candidatos. Uma vez que o pai do indivíduo é identificado, o padrão do movimento do pólen pode ser determinado na população. No entanto, a limitação geral dos métodos diretos é que somente podem ser aplicados a populações pequenas.

Nos anos 80, modelos estatísticos de análise de paternidade foram desenvolvidos para realizar medidas de movimentação de genes, usualmente empregando genótipos multilocos para inferir os pais de progênes de sementes de mães conhecidas.

Este método de estimação do fluxo gênico é baseado na análise de paternidade que fornece com detalhes a descrição da estrutura individual. A análise de paternidade usa um loco genético para identificar o mais provável pai de um conjunto de pais candidatos (Hamrick & Schnabel, 1985). A resolução deste procedimento é dependente do número de locos polimórficos, número de alelos por loco, frequência alélica e número de pais potenciais. Uma vez o pai de um indivíduo ter sido identificado, o padrão do movimento do pólen dentro da área estudada pode ser determinado.

Melo (2000) estudou a paternidade de *Eucalyptus* de sementes oriundas do pomar AR6 da empresa ARACRUZ, e verificou que com 14 locos (101 alelos) a paternidade de 3 lotes de sementes foi determinada com precisão, em 75%, 62%, e 41,6% dos indivíduos, respectivamente. Também verificou uma mistura de

sementes nestes lotes onde o alelo da mãe não aparecia em alguns indivíduos da progênie.

Como já mencionado os métodos indiretos são via: F_{ST} , alelos privados e auto-correlação espacial. Estes métodos são baseados em inferências indiretas do fluxo gênico. O método do F_{ST} , é baseado na análise da estrutura genética de populações, sendo rotineiramente usado para estimar o número de migrantes por geração (Nm) para um conjunto de populações ou sub-populações. O fluxo calculado através do parâmetro F_{ST} é chamado de fluxo gênico aparente, pois este método admite a estrutura genética populacional sob o modelo de ilhas de Wright (1951) e assume equilíbrio entre migração e deriva genética. O F_{ST} é uma função efetiva do número de migrantes por geração, Nm , onde N é o tamanho populacional e m é a proporção de migrantes por geração. Desta forma tem-se que:

$$Nm = \frac{1}{4} \left(\frac{1}{F_{ST}} - 1 \right), \text{ Wright, 1951}$$

onde:

m = taxa de migração;

N = tamanho populacional;

Nm = número de migrantes/geração;

Em espécies tropicais, os valores obtidos para Nm têm se mostrado, em geral superiores a 1,0 (Reis, 1996; Ciampi, 1999; Moraes, 1997; Gaiotto, 2001). Segundo Wright (1951), quando $Nm \geq 1,0$, ou seja, quando um ou mais indivíduos migram por geração, os efeitos da migração são suficientes para contrapor os efeitos da deriva e, portanto, o número de migrantes por geração impede a divergência entre populações.

Esta estimativa de fluxo tem sido utilizada no conservacionismo de uma maneira bem simplista. Como se pode observar na tabela 3 a seguir o fluxo em várias espécies: *Eichhornia panicullata*, *Euterpe edulis*, *Copaifera langsdorffii*, *Cryptocarya moscata* e *M. aquifolia* é muito variável, com valores de 0, 31 a 10,07. Contudo um cuidado especial deve ser tomado na interpretação desta estimativa, porque como dito anteriormente o fluxo não é uma medida contemporânea, quando é estimado a partir do F_{ST} . Na espécie *E. panicullata* o fluxo é muito pequeno, igual

0,31 migrantes por geração e portanto atualmente estas populações estão provavelmente passando por sério risco de extinção. Já para o palmito (*E. edulis*) estudado por Reis (1996) a situação é menos dramática, pois foi estimado um fluxo de 10,07 migrantes por geração. Esta estimativa refere-se a um fluxo histórico, pois as populações estudadas por Reis encontram-se a grandes distâncias, estando algumas em São Paulo e outras em Santa Catarina, sendo praticamente impossível a existência de fluxo gênico atual entre elas.

Tabela 3. Estimativas de F_{ST} e fluxo gênico em populações de espécies tropicais.

Espécie	F_{ST}	Nm	Referências
<i>E. panicullata</i>	0,450	0,31	Husband & Barret, 1995
<i>E. edulis</i>	0,023	10,07	Reis, 1996
<i>C. langsdorfi</i>	0,050	5,00	Ciampi, 1999
<i>C. moscata</i>	0,140	0,90	Moraes, 1997
<i>M. aquifolia</i>	0,147	1,45	Perecin, 2001

Uma aplicação do F_{ST} é o cálculo da razão fluxo pólen/sementes. A avaliação de marcadores moleculares biparentais (nucleares) e maternos (organelares), tais como heranças genômicas permitem distinguir a contribuição histórica estimada de sementes e pólenes no fluxo gênico (Ennos, 1994), pela estimativa da razão do fluxo de pólen para fluxo de semente, através de Nm por herança materna e biparental (Ouborg et al., 1999). O método, descrito por Slatkin (1985) é baseado na frequência média de alelos privados, ou seja, alelos exclusivos a uma ou poucas populações. Admite-se que o log de Nm é linearmente relacionado com o log da frequência média dos alelos privados [p(1)].

Outro método indireto baseia-se na autocorrelação espacial, e tem sido utilizado a fim de detectar o padrão espacial da variabilidade genética e inferir sobre os processos microevolutivos na diferenciação de populações.

2.4 Marcadores Moleculares no Estudo de Populações

A conservação de recursos genéticos dos biomas tropicais é um dos tópicos mais prementes na atualidade. Com as elevadas taxas de destruição das

comunidades naturais, massiva extinção de espécies é iminente. Especificamente, no Brasil, um dos ecossistemas que vêm sofrendo maior pressão antrópica é o Cerrado. Essa pressão traduz-se principalmente pela expansão da fronteira agrícola, onde em extensas áreas contínuas, a vegetação natural vem sendo substituída por monoculturas agrícolas, reflorestamentos monoespecíficos e pastagens, sem deixar amostras das formações vegetais, que possam funcionar como banco genético e refúgio da fauna e flora. A conservação e o manejo da biodiversidade, mesmo em áreas protegidas, constituem-se em desafios complexos que requerem conhecimento básico sobre a distribuição e abundância de espécies, suas interações mutualísticas, sua biologia reprodutiva e a estrutura genética de suas populações (Mittermeier et al., 1992).

O sucesso de qualquer programa de pré-melhoramento ou de conservação é dependente do conhecimento da quantidade de variação presente na espécie de interesse. Caracteres morfológicos e agrônômicos, usados para a medição da diversidade genética em determinada população de indivíduos, possuem uma grande dificuldade na identificação de grupos taxonômicos discretos. Esta dificuldade deve-se ao fato de que a grande maioria dos caracteres vegetais são influenciados por fatores ambientais, exibindo variação contínua e um alto grau de plasticidade fenotípica. Para tentar solucionar estes problemas, técnicas moleculares têm sido utilizadas para monitorar a variabilidade genética (Parker et al., 1998).

Os recentes avanços na biologia molecular abriram novas perspectivas para a pesquisa em conservação de espécies e para os estudos de biologia populacional como um todo. Pela primeira vez a variação encontrada em plantas ou animais pode ser analisada ao nível do DNA. Diferenças na seqüência gênica podem ser diretamente observadas e descritas com um grau de precisão impraticável há pouco tempo atrás. Os marcadores moleculares são uma destas novas técnicas. Estes marcadores são utilizados no estudo da extensão e distribuição da variação entre espécies como também para investigar questões taxonômicas e evolutivas (Ferreira & Grattapaglia, 1996 e Haig, 1998).

Uma vez que os recursos genéticos referentes às espécies frutíferas nativas estão seriamente ameaçados pela progressiva destruição de seu habitat, torna-se necessária a utilização de uma técnica eficiente para se obterem informações sobre a quantidade e a distribuição da variação genética dentro e entre populações destas

espécies. O uso de marcadores moleculares torna-se portanto, uma opção viável para atingir estes objetivos, além de gerar informações importantes para um eficiente plano de conservação e manejo.

A introdução da técnica de eletroforese de isoenzimas no início da década de 60, além de iniciar a era dos marcadores moleculares, ampliou o número de marcadores que poderiam ser utilizados. A técnica baseia-se em mutações no código genético, as quais alteram a carga elétrica de algumas proteínas com função enzimática. Após eletroforese das amostras de proteína, o polimorfismo é detectado através da visualização do produto enzimático por métodos histoquímicos. Os marcadores isoenzimáticos podem ser obtidos de uma maneira relativamente rápida, barata e são disponíveis para praticamente todas as espécies de plantas. A utilização de isoenzimas permitiu a detecção de polimorfismo entre plantas que não mostravam diferenças morfológicas. Assim, foi possível uma melhor avaliação da variação genética (Torggler et al., 1995). Os marcadores isoenzimáticos, já na década de 60, foram utilizados para importantes estudos como o de Lewontin & Hubby (1966), de estrutura genética populacional, o que permitiu desde então uma grande disseminação desta técnica. Atualmente, este tipo de marcador tem sido ainda, muito utilizado no estudo de genética de populações (Sebben, 2001; Telles, 2000 e Moraes 1997).

O poder de detecção da variabilidade existente diretamente ao nível do DNA, só foi alcançado com o desenvolvimento de técnicas em biologia molecular. A primeira delas baseia-se na ação de enzimas de restrição, as quais reconhecem uma seqüência de DNA, e clivam a molécula nestes sítios específicos. Os efeitos de mutações pontuais, inserções, deleções e rearranjos nos sítios de clivagem, são detectados através de *Southern blots* como polimorfismos no tamanho dos fragmentos de DNA após digestão com uma enzima de restrição. Este marcador chamado de RFLP (Restriction Fragment Length Polymorphism) surgiu na década de 70 em um experimento destinado à detecção de mutação de DNA de vírus (Grodzicker et al., 1974). Ele foi usado mais tarde por Botstein et al. (1980) para análise genômica. Estes marcadores RFLP tornaram-se ferramenta útil e importante para várias áreas da biologia. Porém, em estudos de genéticas de populações este marcador não foi muito difundido e utilizado devido à grande morosidade de técnica a ser aplicada a um grande número de indivíduos.

Os marcadores minisátélites surgiram também na década de 80. A primeira região hipervariável foi isolada ao acaso a partir de uma biblioteca genômica descrita por Wyman & White, 1980. Os minisátélites tem sido utilizados no melhoramento de plantas para a identificação de variedades e clones, na análise de diversidade genética e na determinação da paternidade (Dallas, 1988; Nybom & Hall, 1991 e Broun et al., 1992).

O surgimento da técnica de PCR na década de 80 desenvolvida por Mullis & Faloona (1987), permitindo a síntese enzimática de milhões de cópias de um segmento específico de DNA, provocou uma verdadeira revolução nas técnicas de biologia molecular, facilitando muito o trabalho de laboratório. AFLPs, RAPDs, SCARs, são exemplos das muitas técnicas baseadas em PCR que abriram novas e inúmeras possibilidades de utilização do polimorfismo encontrado na molécula de DNA. Dentre estas técnicas, destaca-se o surgimento dos marcadores microssatélites.

Dentre os diversos tipos de marcadores atualmente disponíveis, os marcadores RAPD destacam-se pelas vantagens que apresentam em termos de simplicidade (ex. a visualização das marcas RAPD pode ser realizada sem o uso de sondas radioativas), pelo fato de serem aplicáveis a um grande número de espécies, e por permitirem a análise de um grande número de locos. Segundo Lynch & Milligan (1994), o uso deste tipo de marcador possibilita, ainda, uma amostragem aleatória mais ampla do genoma do que aquelas proporcionadas por outras classes de marcadores. No entanto, o uso dos marcadores dominantes tem sido muito limitado pela falta de informação genotípica do marcador, não sendo possível identificar o genótipo heterozigótico (Ferreira & Grattapaglia, 1996). Embora alguns modelos estatísticos tenham sido desenvolvidos no sentido de contornar este problema, o uso destes ainda é limitado, pois as estimativas de parâmetros populacionais com marcadores dominantes são sempre tendenciosas (Excoffier et al., 1992 e Lynch & Milligan, 1994).

Marcadores microssatélites possuem todas as características desejáveis para serem utilizados em estudos de genética de populações (Powell et al., 1996). Locos de microssatélites são constituídos por seqüências curtas de DNA repetitivo 1 a 6 pb repetidas várias vezes de maneira idêntica e adjacente (repetição em tandem). As seqüências de DNA que flanqueiam microssatélites são conservadas, o que permite a seleção de um par de pequenos fragmentos iniciadores da fita réplica, denominados

primers, de 20 a 30 pares de bases e sua amplificação via PCR. O polimorfismo é baseado nas diferenças de comprimento das sequências amplificadas, pois o número de repetições em cada microssatélite é altamente variável (Litt & Luty, 1989; Weber & May, 1989 e Tautz, 1989). Atualmente, estes marcadores têm sido muito utilizados em estudos de genética de populações de plantas (Taylor et al., 2002; Danquah et al., 2002; Lopes et al., 2002; Ramakrishnan et al., 2002; Lefort et al., 2002; Yamamoto et al., 2002; Souza, 2002 e Isagi & Suhandono, 1997).

Contudo, todas as facilidades deste marcador dependem da disponibilidade das seqüências de DNA que flanqueiam o microssatélites: os *primers*. O uso de marcadores microssatélites em espécies nativas é limitado devido às escassas informações sobre sequências de DNA de espécies tropicais em geral e do alto custo e trabalhos envolvidos na construção de bibliotecas genômicas, clonagens e sequenciamentos necessários para o isolamento de microssatélites e o desenvolvimento dos *primers*.

2.5 Marcadores RAPD

Atualmente, muitos métodos estão disponíveis para analisar a diversidade molecular em populações de plantas. Essas metodologias diferenciam-se pela técnica utilizada para revelar a variabilidade ao nível de DNA, e assim variam quanto à habilidade em detectar diferenças entre indivíduos, custo, facilidade de uso, consistência e reprodutibilidade (Milach, 1998).

Marcadores RAPD (Random amplified polymorphic DNA) (Williams et al., 1990) têm sido muito empregado no estudo de populações naturais nos últimos anos. Eles são tecnicamente mais simples que outros tipos de marcadores (por exemplo, RFLP ou SSR), facilitando o estudo de um grande número de locos, e fornecem uma amostragem aleatória maior do DNA.

Williams et al. (1990) foram os primeiros a descrever os marcadores RAPD, os quais são gerados pela amplificação de segmentos anônimos de DNA com um simples e curto oligonucleotídeo (decâmeros) de seqüência arbitrária. Welsh & McClelland (1990), na mesma época, publicaram um trabalho utilizando o mesmo princípio do RAPD e denominaram a técnica de AP-PCR (Arbitrarily Prime - Polymerase Chain Reaction). Porém esta técnica, utilizando oligonucleotídeos

arbitrários, ficou mais conhecida por RAPD, que é a denominação mais comumente encontrada na literatura. Inúmeras são as vantagens oferecidas pelos marcadores RAPD, como exemplo: a técnica não utiliza radioatividade, não requer o conhecimento prévio da seqüência alvo a ser amplificada, necessita de pequenas quantidades de DNA, além de ser aplicável a qualquer espécie. Aliados a estas vantagens estão o baixo custo, facilidade e rapidez no emprego da técnica. Embora haja poucos estudos comparando diretamente os diferentes tipos de marcadores, a maioria das comparações sugere que os marcadores RAPD e isoenzimas revelam padrão similar de diversidade genética, mas que os RAPDs tendem a fornecer marcas relativamente específicas de populações, raças ou espécies (Peakall et al., 1998). Estas características podem ser importantes para estudos de populações visando a conservação genética.

Entretanto RAPD também tem limitações significantes quando comparados com marcadores codominantes. A dominância limita as inferências feitas com RAPDs, especialmente em estudos com organismos diplóides, e podem fornecer algum viés em alguns parâmetros genéticos quando comparados com marcadores multi-alelicos e codominantes (Lynch & Milligan, 1994). Estudos com coníferas (Isabel et al., 1995) têm sugerido que freqüências alélicas inferidas com RAPDs em tecidos diplóides podem resultar em superestimativas de diferenciação entre populações quando comparadas com marcadores isoenzimáticos.

Marcadores RAPD possuem caráter dominante, limitando seu uso na estimação de parâmetros genéticos, além de apresentar problemas de reprodutibilidade (Karp et al., 1996).

2.6 Marcadores SSR

Segundo Charlesworth et al. (1995) grande parte do genoma de eucariotos é constituída por seqüências de DNA repetitivo. O DNA repetitivo é composto por DNA satélite (DNA altamente repetitivo), as seqüências de DNA microssatélites e minissatélites (DNA moderadamente repetitivo) e os elementos transponíveis (DNA moderadamente repetitivo, móvel e de seqüências dispersas).

Microsatélites ou SSR podem ser definidos como seqüências repetidas em tandem composta de número variáveis de motivo com 1 a 6 pb, encontrados no

genoma de vertebrados, insetos e plantas. Pelo menos 30.000 locos estão presentes no genoma humano (Charlesworth, 1995).

As “ilhas” de microssatélites são regiões instáveis do genoma, que estão sob alterações mutacionais, geralmente do tipo adição ou deleção de repetições em taxas mais elevadas que as observadas nas sequências de DNA não repetitivo, sendo que esta taxa pode chegar até em 1 a cada 1000 pb. Esta instabilidade gera um grande número de locos polimórficos o que é interessante no estudo de populações e mapeamento de genes.

A variabilidade do comprimento destas regiões repetidas pode ser analisada por PCR usando pares de *primers* que flanqueiam estas regiões. Devido ao grande número de motivos repetidos, os SSR são considerados marcadores altamente polimórficos e multi-alelicos. Entre as classes de marcadores existentes, os SSR são os que se aproximam mais do marcador ideal para os estudos de genética de populações (Rafalski et al., 1996).

Grande parte do conhecimento a respeito dos microssatélites foi obtido a partir do reino animal, principalmente em mamíferos. Em humanos, foi mostrado que a repetição (AC)_n, presente em aproximadamente 5×10^4 locos por genoma haplóide, é um dos motivos mais comuns (Hamada et al., 1984 e Moore et al., 1991), enquanto todas as repetições de três ou quatro nucleotídeos foram estimados em aproximadamente 4×10^5 locos (Edwards et al., 1991). A maioria dos microssatélites ocorre em regiões não codantes, e presumivelmente comportam-se como marcadores neutros.

Em plantas a presença da repetição (AC)_n e (AG)_n foi descrita pela primeira vez por Condit & Hubbell (1991). No genoma de plantas superiores a repetição (AC)_n é geralmente menos freqüente do que em mamíferos, sendo o motivo (AT)_n o mais encontrado, seguido por (A)_n e (AG)_n. Repetições de trinucleotídeos e tetranucleotídeos também estavam presentes nos genomas de plantas, sendo os motivos mais freqüentes (AAT)_n, (AAC)_n, (AGC)_n, (AAG)_n, (AATT)_n, (AAAT)_n (Akkaya et al., 1995).

Quanto à distribuição dos microssatélites pelo genoma, há aproximadamente um sítio de (AC)_n a cada 360-450kb em arroz, em comparação com um sítio de (AC)_n a cada 40-80kb em humanos (McCouch et al., 1997). Baseado na pesquisa em bibliotecas genômicas de outras plantas, a freqüência de (AC)_n está disponível para

trigo em um sítio a cada 700kb, no milho em um sítio a cada 100-1000kb (Condit & Hubbel, 1991) e em *Arabidopsis* em um sítio a cada 430 kb (Bell & Ecker, 1994). Em contraste com mamíferos, a frequência observada de $(GA)_n$ em plantas é consistentemente maior do que $(AC)_n$ (Condit & Hubbel, 1991). Estima-se que ocorra um sítio de $(GA)_n$ a cada 225-330kb em arroz (Panaud et al., 1995 e Wu & Tanksley, 1993), um sítio a cada 440kb em trigo (Röder et al. 1995), um em cada 168-710 em milho (Condit & Hubbel, 1991) e um em cada 244 kb em *Arabidopsis* (Bell & Ecker, 1994). Comparando-se com as estimativas de ocorrência do sítio $(GA)_n$ em humanos a cada 75kb aproximadamente, pode-se concluir que estes motivos são de 2 a 10 vezes mais freqüentes em humanos do que em espécies de plantas, porém, amplamente distribuídos no genoma vegetal (McCouch et al., 1997).

Como outras classes de DNA repetitivo, os microssatélites possuem altas taxas de mutação, numa faixa entre 10^{-3} e 10^{-4} por loco por gameta por geração. Esta instabilidade surge através de um mecanismo específico de mutação chamado deslizamento (*slippage*) da DNA polimerase (Tautz & Schlötterer, 1994). Outro motivo pelo qual a taxa de mutação é alta nos microssatélites é devido ao fenômeno do crossing-over desigual, causado pelo pareamento errôneo destas seqüências durante o quiasma. Em uma grande variedade de organismos, como a levedura, *Drosophila* e humanos, foi demonstrado que as taxas de mutação dos microssatélites estavam positivamente correlacionadas com o número de repetições (Jin et al., 1996; Wierdl et al., 1997; Schlötterer et al., 1998 e Brown et al., 1996).

Microssatélites são marcadores codominantes gerando dados similares àqueles gerados por isoenzimas, porém com um número de alelos e uma heterozigidade muito maior. Devido ao alto grau de polimorfismo, os microssatélites tornaram-se os marcadores ideais para o mapeamento genético e estudos populacionais. Microssatélites foram empregados com sucesso na discriminação entre acessos e cultivares de bancos de germoplasma, detectando duplicações, mistura de sementes, deriva, e cruzamentos não controlados (Melo, 2000; Souza, 2002 e Olufowote et al., 1997). Também foram empregados na determinação do grau de parentesco entre indivíduos (Yang et al., 1994 e Melo, 2000), no esclarecimento da estrutura genética ou da divisão da variação entre indivíduos, populações e espécies (White et al., 1999; Dayanandan et al., 1997; Dayanandan et al., 1999 e Collevatti et al., 2001), além de

possibilitar a construção de mapas genéticos (Bell & Ecker, 1994; Akkaya et al., 1995; Cregan et al., 1999 e Wu & Tanksley, 1993).

3 ESTRUTURA GENÉTICA E FLUXO GÊNICO DE *Eugenia dysenterica* DC, UTILIZANDO MARCADORES RAPD

Resumo

Estudos de conservação de espécies nativas requerem um pleno entendimento da diversidade genética entre as populações, bem como da ocorrência de fluxo gênico. Nas regiões de cerrado é comum a ocorrência natural de algumas frutíferas nativas que possuem um grande potencial econômico. Dentre estas espécies frutíferas nativas, tem destaque a cagaiteira (*Eugenia dysenterica*), que é uma árvore pertencente à família Myrtaceae, com origem e dispersão natural na região dos cerrados. Informações sobre a variabilidade genética das populações existentes no Estado são escassas. Este trabalho tem por objetivo o estudo da variabilidade genética em 10 populações de cagaiteira da região sudeste do Estado de Goiás. Para tanto, foram identificados 54 locos marcadores RAPD, para a caracterização da variabilidade genética. Esta foi avaliada através da AMOVA, onde se verificou que 27,03% da variabilidade genética está entre populações e 72,97% dentro de populações e o valor de Φ_{ST} portanto igual a 0,2703. Foi calculada uma matriz de índices médios de similaridade de Jaccard para cada população e a diversidade foi estruturada por uma análise de agrupamento seguindo o critério UPGMA. Verificou-se a existência de dois grupos distintos, em conformidade com sua localização espacial na região sudeste do Estado de Goiás. Com a finalidade de analisar padrões de variação espacial, foi calculada a estimativa do coeficiente de correlação de Pearson (r) entre a matriz de distâncias genéticas (1 – índice de similaridade de Jaccard) e de distâncias geográficas, sendo encontrada assim, forte correlação positiva $r=0,770$ ($p<0,001$). Estes resultados sugerem que estas populações estão se diferenciando por um processo estocástico, havendo fluxo dependente da distribuição geográfica e que, provavelmente, essa diferenciação esteja ocorrendo de acordo com o modelo de isolamento por distância.

Summary

In Cerrado areas several native species occur with fruits having great economical potential. Among these native fruit tree species, the cagaiteira (*Eugenia dysenterica*), is prominent. This species belongs to the Myrtaceae family, with origin and natural dispersion in the Cerrado area. Informations about the genetic variability of the existent populations are scarce. Studies on the conservation of native species requires a full understanding the genetic diversity among the populations, as well as of the occurrence of gene flow. The objective of this work was to study the genetic variability in 10 populations of cagaiteira in the southeast area of the State of Goias. The genetic variability of these populations was evaluated using 54 RAPD marker loci. Using the AMOVA procedure it was verified that 27.03% of the genetic variability occurred among populations and consequently that 72.97% is within populations. The matrix of Jaccard's indices similarities was calculated and the diversity was structured grouping the populations through the UPGMA criterion. The existence of two different groups was verified, in conformity with their spacial location in the southeast area of the State of Goias. With the purpose of analyzing patterns of space variation, Pearson's correlation coefficient (r) was estimated among the matrix of genetic distances (1 - index of similarity of Jaccard) and of geographical distances, and a strong positive correlation ($r=0.770$, $p < 0.001$) was found. These results suggest that these populations are differentiating through a stocastic process, with gene flow being dependent on the geographical distribution and that this differentiation is probably occurring according the a model of isolation by distance.

3.1 Introdução

A cagaiteira é uma fruteira nativa da região de cerrado com possibilidade de utilização na agricultura, pois tem importância sócio-econômica devido ao grande potencial no processamento de seus subprodutos. Além de ser uma espécie ornamental e melífera sua madeira é usada para construções civis. Suas folhas são utilizadas para tratamento de moléstias e seus frutos são conhecidos por terem propriedades laxativas (Heringer & Ferreira, 1974).

As frutíferas nativas do Cerrado são espécies de diversos gêneros e famílias que produzem frutos de interesse tanto para a alimentação regional quanto para a industrialização. Há um mercado potencial e crescente para as frutíferas nativas, porém pouco explorado pelos agricultores. Todo o aproveitamento desses frutos tem sido feito de forma extrativista e predatória.

Para o estudo de populações naturais de espécies nativas torna-se necessário o conhecimento da diversidade entre populações, fluxo gênico e outros parâmetros genéticos.

Os marcadores moleculares têm sido freqüentemente utilizados em estudos sobre diversidade e estrutura genética populacional. Na literatura diversos trabalhos são encontrados utilizando esta técnica em estudos de conservacionismo como os de Chalmers et al. (1992), Chase (1996), Palacios & Gonçalves-Candelas (1997), Maguire & Sedgley (1997), Bielawski & Pumo (1997), Lou et al. (1998), Skabo et al. (1998), Wolf et al. (1998), Ayres & Ryan (1999), Roa et al. (2000), Hagen et al. (2001), Taylor et al. (2002), Danquah et al. (2002), Lopes et al. (2002), Ramakrishnan et al. (2002), Lefort et al. (2002), Yamamoto et al. (2002), entre muitos outros. Dentre os diversos tipos de marcadores atualmente disponíveis, os marcadores RAPD destacam-se pelas vantagens que apresentam em termos de simplicidade (por exemplo, a visualização das marcas RAPD pode ser realizada sem o uso de sondas radioativas), pelo fato de serem aplicáveis a um grande número de espécies, e por permitirem a análise de um grande número de locos. Segundo Lynch & Milligan (1994), o uso deste tipo de marcador possibilita, ainda, uma amostragem aleatória mais ampla do genoma do que aquelas proporcionadas por outras classes de marcadores.

A principal desvantagem do uso deste marcador é devido à dominância, que impede distinguir o homozigoto dominante do heterozigoto, não sendo possível obter informações das freqüências alélicas, a não ser em condições especiais, ou seja, admitindo equilíbrio em alogamia ou (de Hardy-Weinberg) quando o índice de fixação $f=0$ ou em equilíbrio em autogamia, quando $f=1$.

Huff et al. (1993) estudaram a variação genética entre e dentro de populações pela primeira vez utilizando populações naturais de *Buchlœe dactyloides* através da metodologia de análise de variância molecular descrita por Excoffier et al. (1992) para estudar haplótipos de mitocôndria em humanos. Embora não tenha sido desenvolvida para estudos de dados gerados por marcadores RAPDs, esta

análise tem sido muito aplicada atualmente em estudos de populações de plantas (Fisher et al., 2001; Sun & Wong, 2001 e Hagen et al., 2001).

O presente trabalho avaliou a utilização de marcadores RAPD para caracterização genética de populações de *E. dysenterica* DC destacando a sua variabilidade genética, a estrutura populacional e o fluxo gênico. O objetivo principal da pesquisa foi a geração de informações para a domesticação e melhoramento da espécie, além de sua conservação.

3.2 Material e Métodos

3.2.1 Material vegetal

O material para o presente estudo é originado de 10 áreas da região sudeste do Estado de Goiás, destas foram coletadas 10 populações com cerca de 12 plantas em cada população. Os municípios amostrados e as coordenadas geográficas das áreas de coleta estão relacionados na Tabela 4. As coordenadas (latitude e longitude) foram estabelecidas utilizando-se o Sistema de Posicionamento Global (GPS) e a altitude foi medida com altímetro em precisão de 10 m. Estas regiões foram selecionadas a partir de viagens de prospecção que ocorreram durante os meses de agosto e setembro de 1996. As populações de 1 a 8 foram coletadas nesta mesma época (outubro de 1996) e as plantas das populações 9 e 10 foram coletadas em novembro de 1999.

O material de coleta constituiu-se de folhas jovens acondicionadas em gelo e posterior estocagem em "freezer" com temperatura de -20°C , até o momento de extração de DNA para análise.

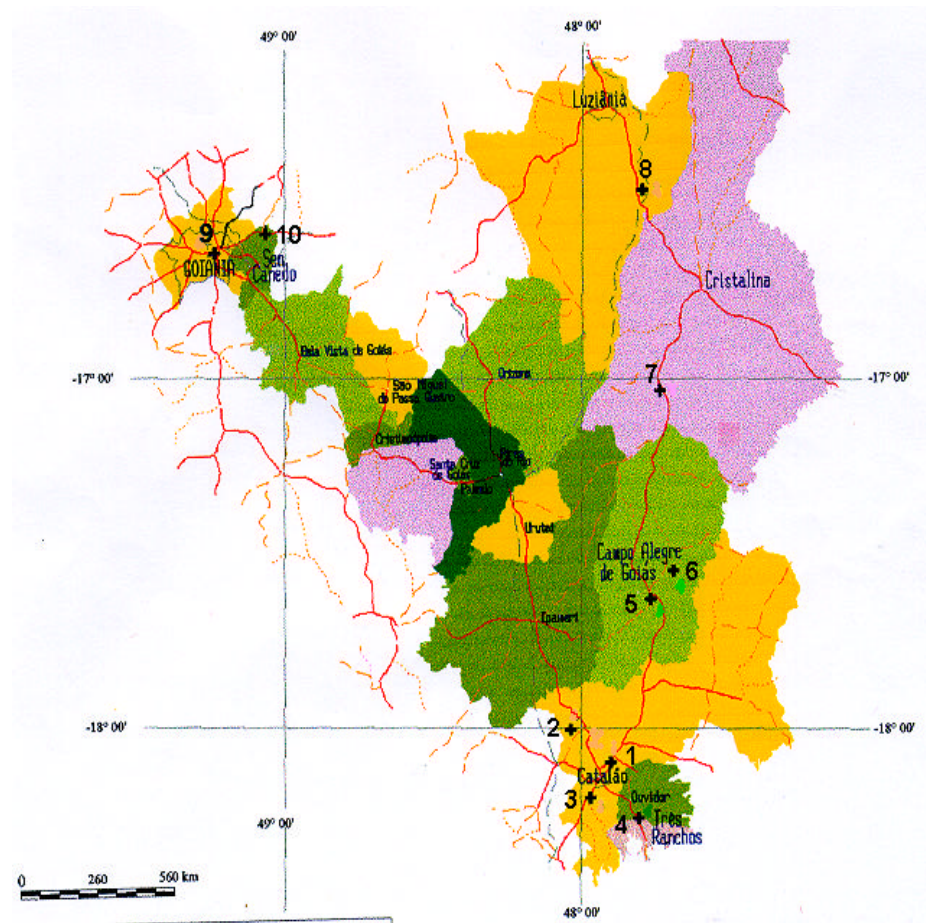


Figura 2 - Municípios e localização das áreas de coleta de *E. dysenterica* DC. na região Sudeste do Estado de Goiás.

Tabela 4. Localidades do Estado de Goiás e número de árvores amostradas de cagaiteira, e suas respectivas informações geográficas.

Áreas	Localidade Município	Número de árvores amostradas	Altitude (m)	Latitude	Longitude
1	Catalão	12	880	18°07' 35"	47°54' 20"
2	Catalão	12	860	18°02' 0,3"	48°02' 31"
3	Catalão	12	800	18°13' 39"	47°58' 12"
4	Três Ranchos	12	820	18°17' 15"	47°48' 41"
5	Campo Alegre de Goiás	12	930	17°39' 11"	47°46' 37"
6	Campo Alegre de Goiás	12	780	17°34' 24"	47°42' 12"
7	Cristalina	6	890	17°10' 47"	47°31' 07"
8	Luziânia	12	900	16°28' 48"	47°48' 40"
9	Goiânia	12	740	16°40' 30"	49°14' 42"
10	Senador Canedo	12	840	16°37' 13"	49°04' 29"

Foram consideradas como áreas propícias para as coletas aquelas que possuíam no mínimo 20 plantas adultas de cagaiteira em um raio de 1 km e que, de cada área, fosse possível coletar dados de no mínimo 10 matrizes, as quais pudessem produzir frutos para análises posteriores (Telles, 2000). A menor distância entre as populações coletadas foi de 13,12 km entre as populações 1 e 3, e a maior de 234,72 km entre as populações 4 e 9.

Foram amostradas 12 matrizes em cada área de coleta (exceto na população 7), representando a população. Este estudo baseou-se em 114 matrizes provenientes das coletas realizadas em 10 populações naturais distribuídas na região mencionada (Figura 2).

3.2.2 Extração do DNA Genômico

Para extração de DNA utilizou-se o método do CTAB descrito por Murray & Thompson (1980) e Rogers & Bendich (1985), com modificações. Foram utilizados, aproximadamente, 250 mg de tecido fresco de cada planta, que foram macerados em nitrogênio líquido. Após a maceração, o material foi transferido para tubos Eppendorf de 2 mL, contendo 1 mL do tampão de extração (100 mM Tris-HCl pH 8,0; 1,4 M NaCl; 20 mM EDTA; 2% (p/v) CTAB, 1% (p/v) PVP 40.000) ao qual foi adicionado 1% β -mercaptoetanol, previamente aquecido a 65°C. O material foi homogeneizado com uma espátula e mantido a 65°C por 30 minutos, em banho-maria, com agitações periódicas. Após a retirada dos tubos do banho-maria, acrescentou-se uma mistura de 700 μ L de clorofórmio-octanol (24:1), homogeneizando-se por inversões dos tubos.

As amostras foram centrifugadas por 15 minutos a 14.000 rpm em microcentrifuga Eppendorf, recuperando 700 μ L do sobrenadante, ou fase aquosa, os quais foram transferidos para novos tubos de 2 mL. Para a precipitação do DNA, acrescentaram-se 520 μ L de isopropanol, invertendo-se cuidadosamente os tubos, que foram mantidos por 30 minutos à temperatura ambiente ou colocados sob refrigeração a 4°C para favorecer a precipitação do DNA. O precipitado foi recuperado através de centrifugação a 10.000 rpm durante 10 min. O sobrenadante foi descartado e, em seguida, adicionou-se 1 mL de solução de lavagem (etanol 76%, acetato de NH_4 10 mM), por 30 minutos. A solução de lavagem foi descartada e os pellets foram secos a temperatura ambiente. Acrescentaram-se 100 μ L de solução TE (10 mM Tris-HCl pH 8,0; 1mM EDTA pH 8,0) aos tubos Eppendorf, os quais permaneceram à temperatura de aproximadamente 4°C por cerca de 16 h para a ressuspensão do DNA. O DNA foi novamente precipitado acrescentando-se 50 μ L de acetato de NH_4 7,5 M e 375 μ L de etanol absoluto. O conteúdo do tubo foi misturado por inversões e o DNA foi recuperado por centrifugação por 20 min a 10.000 rpm. O sobrenadante foi descartado e a última lavagem foi feita adicionando-se 500 μ L de etanol 70% por 20 min. O etanol foi descartado e ao DNA foram acrescentados cerca de 100 μ L de TE.

3.2.3 Quantificação do DNA

A quantificação foi realizada em géis de agarose 0,8% (p/v) submetidos à eletroforese. Aliquotas de cada DNA foram aplicadas nos poços do gel ao lado de uma série de concentrações conhecidas de DNA do fago λ (20 a 400 ng). A concentração das amostras foi estimada por comparação visual da intensidade de fluorescência das bandas do DNA do fago λ . Os géis foram visualizados após a coloração com 10 μ L de brometo de etídeo (10 mg/mL) diluído em 100 mL de tampão TEB 1X. Posteriormente o DNA foi diluído (5 ng/ μ L) para as reações de RAPD.

3.2.4 Condições de Amplificação

Após a quantificação do DNA, as reações de amplificação foram feitas em termociclador modelo PTC 100 da MJ, num volume de 25 μ L contendo: 10 mM Tris-HCl pH 8,3; 50 mM de KCl; 2,0 mM MgCl₂; 0,2 mM dNTPs; 0,25 μ M de *primer* (Operon Technologies), 5,0 ng de DNA molde, 1 unidades de Taq polimerase e H₂O q.s.p.. As reações foram submetidas a 48 ciclos de amplificação após desnaturação inicial a 94°C por 5 min. Cada ciclo consistiu de 30 seg a 92°C, 1 min e 30 seg a 37°C e 1min e 30 seg a 72°C. Ao final de 48 ciclos, foi realizada uma extensão final de 5 min a 72 °C.

Os produtos de amplificação foram submetidos à eletroforese (3V/cm) em géis de agarose 1,4% (p/v), utilizando o tampão de corrida TEB 1 X. O ladder 100bp (Gibco) foi usado como marcador de peso molecular. Os géis foram corados com brometo de etídeo e fotografados sob luz UV (EDA-KODAK).

3.2.5 Seleção de *Primers*

Foram testados 80 *primers* dos respectivos Kit A, C, G, H da Operon Technologies. Para verificação do perfil de amplificação de cada *primer* foram utilizadas quatro plantas matrizes. Trindade et al. (2001) realizaram um ensaio preliminar para determinação do número de locos a serem utilizados em *E. dysenterica*, amostrando 25 plantas com 36 locos RAPD (7 *primers*). Uma matriz binária 1 (presença) e 0 (ausência da banda) foi gerada e o procedimento de

randomização por bootstrap foi utilizado para estimar o coeficiente de variação associado ao número de bandas utilizados, através do Simple Matching com 5000 permutações utilizando-se o software Dboot (Coelho, 2000). Verificou-se que para um CV de 10% apenas 33 locos eram necessários para análise da estrutura genética.

3.2.6 Análise Estatística dos Dados

A partir da leitura dos géis gerou-se uma matriz binária em que os indivíduos foram genotipados quanto a presença (1) e ausência (0) de bandas. Com esta matriz calculou-se a porcentagem de polimorfismo obtido com cada primer utilizado através da fórmula:

$$P = \frac{nbp}{nbt}$$

Onde:

P = porcentagem de polimorfismo (ou taxa de polimorfismo);

nbp = número de bandas polimórficas;

nbt = número de bandas total.

Com esta matriz de dados calculou-se o coeficiente de similaridade de Jaccard entre todos os indivíduos conforme a expressão abaixo:

$$S_{ij} = \frac{a}{a+b+c}$$

Onde:

a = número de casos em que ocorre a presença da banda em os ambos indivíduos, simultaneamente;

b = número de casos em que ocorre a presença da banda somente no indivíduo i ;

c = número de casos em que ocorre a presença da banda somente no indivíduo j .

O dendrograma foi feito a partir dos índices médios de similaridade para cada população. Este foi o motivo de não ser possível calcular a estabilidade dos

agrupamentos devido a ausência de um programa computacional que permitisse através do procedimento de *bootstrap* analisar as distâncias médias entre populações.

A análise da variância de dados moleculares (AMOVA) foi realizada através da decomposição total nas suas componentes entre e dentro de populações utilizando-se as distâncias ao quadrado, conforme descrito por Excoffier et al. (1992) com o auxílio do programa Arlequin (Schneider et al., 2000). As distâncias genéticas foram obtidas conforme Nei & Li (1979):

$$D = \{d_{xy}^2\} = 100 \left[1 - \frac{2n_{xy}}{n_x + n_y} \right]$$

Onde, n_x e n_y são números de marcadores observados em indivíduos x e y, respectivamente, e $2n_{xy}$ é o número de marcas existentes em ambos os indivíduos.

A significância destas estimativas foi obtidas pelo método de randomização utilizando 1000 bootstrap.

Com a finalidade de analisar os padrões de variação espacial, foi obtida estimativa do coeficiente de correlação de Pearson (r) entre matrizes de distâncias genéticas calculadas a partir do índice de similaridade de Jaccard e as distâncias geográficas entre as populações. A significância desta correlação matricial foi testada pelo estatística Z de Mantel, utilizando 9999 permutações aleatórias, com emprego do programa NTSYS (Rohlf, 1989).

3.3 Resultados

3.3.1 Seleção de *Primers*

A partir da seleção de *primers* foram identificados os que geraram um maior número de bandas polimórficas, sendo eles: OPA-01, OPA-02, OPA-03, OPA-11, OPA-13, OPA-19, OPH-04 e OPH-08. Conforme a Tabela 5, pode-se verificar a seqüência dos *primers* selecionados e o padrão de polimorfismo de cada um deles. Na Figura 3, pode se observar o perfil do primer OPA-11. O primer que produziu um maior número de locos e também de locos polimórficos foi o OPA-11 e o que gerou um menor número de locos polimórficos foi o OPA-13.

Tabela 5. Sequência dos *primers* selecionados para *Eugenia dysenterica* e padrão de polimorfismo.

Primer	Seqüência 5'-3'	Número de locos obtidos	Número de locos polimórficos	Porcentagem de polimorfismo
OPA_01	CAGGCCCTTC	09	05	55,5
OPA-02	TGCCGAGCTG	06	06	100,0
OPA-03	AGTCAGCCAC	09	05	55,5
OPA-11	CAATCGCCGT	11	11	100,0
OPA-13	CAGCACCCAC	09	03	33,3
OPA-19	CAAACGTCGG	09	05	55,5
OPH-04	GGAAGTCGCC	10	09	90,0
OPH-08	GAAACACCCC	11	10	90,9
TOTAL		74	54	73,0

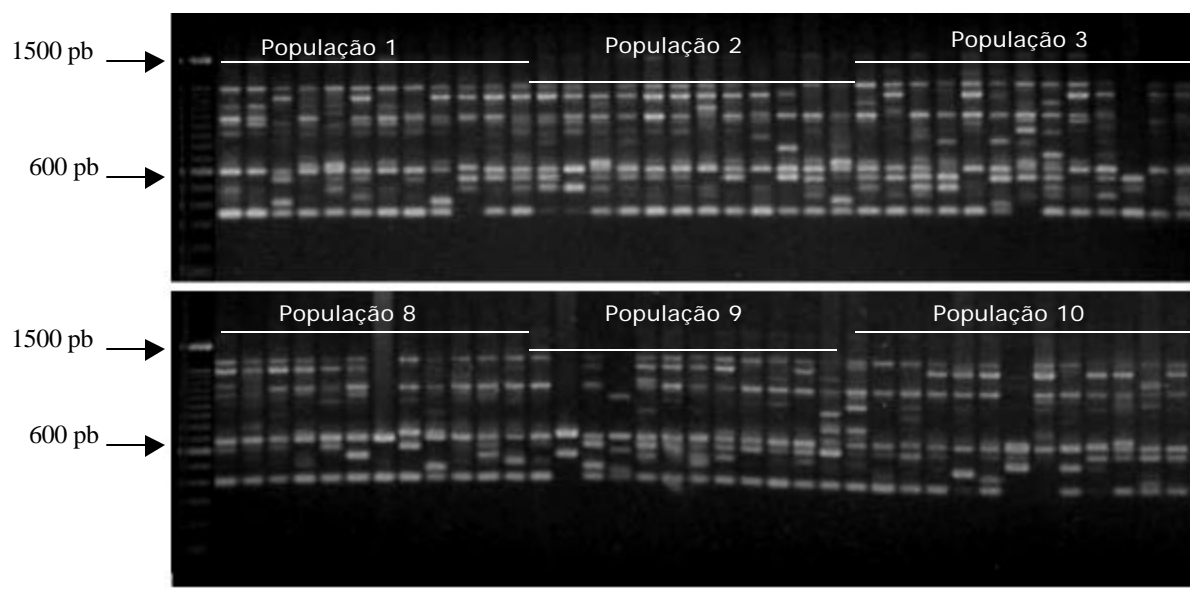


Figura 3 - Perfil de um gel de RAPD utilizando o primer OPA-11 em 74 plantas de cagaiteira; à esquerda está o padrão de peso molecular (ladder 100 pb).

3.3.2 Estrutura Genética e Fluxo Gênico

Através da AMOVA, com base nos dados de 54 locos polimórficos verificou-se que 27,03% da variabilidade genética está entre populações e 72,97% dentro de populações e o valor de ϕ_{ST} foi, pois, igual a 0,2703, conforme a Tabela 6.

Tabela 6. Resultado da análise de variância molecular (AMOVA) e estimativa de fluxo gênico (Nm) calculado pelo método de Wright (1951).

Fonte de variação	GL	SQ	Porcentagem total da variação	P	Estatística f'	Nm
Entre Populações	09	191,107	27,03%	(<0,001)	$\phi_{ST}=0,2703$	0,675
Dentro de populações	103	423,548	72,97%	(<0,001)	$1 - \phi_{ST}=0,7297$	
TOTAL	112	614,655				

As distâncias genéticas médias obtidas a partir do índice de similaridade de Nei & Li (1979) variaram de 0,455 a 0,653. O respectivo dendrograma está apresentado na Figura 4. A correlação cofenética do agrupamento UPGMA desta matriz foi elevada (0,915). A análise de agrupamento revela que populações 9 e 10 são as mais semelhantes formando um grupo geneticamente distinto dos demais. Estas localizam-se no extremo oeste da região amostrada. As populações de 1 a 8 formam outro grupo, as quais estão situadas na região leste da área estudada. É importante salientar que as populações 9 e 10 estão geograficamente separadas das demais pela bacia do Rio Corumbá, não se observando a ocorrência frequente da espécie nas áreas de menor altitude correspondentes à depressão formada pelo rio e seus afluentes. As demais populações (1 a 8) situam-se ao longo do divisor de águas das bacias dos rios Corumbá e São Marcos. Embora atualmente haja descontinuidade entre estas populações, esta pode ter sido acentuada pela ação antrópica recente por ser esta uma região de agricultura de larga escala. O padrão de agrupamento do dendrograma mostra, para estas populações, um padrão clinal de variação (Figura 4). Esta estruturação das populações está congruente com os

dados obtidos por Telles (2000) utilizando isoenzimas, onde foram avaliadas as mesmas populações deste estudo.

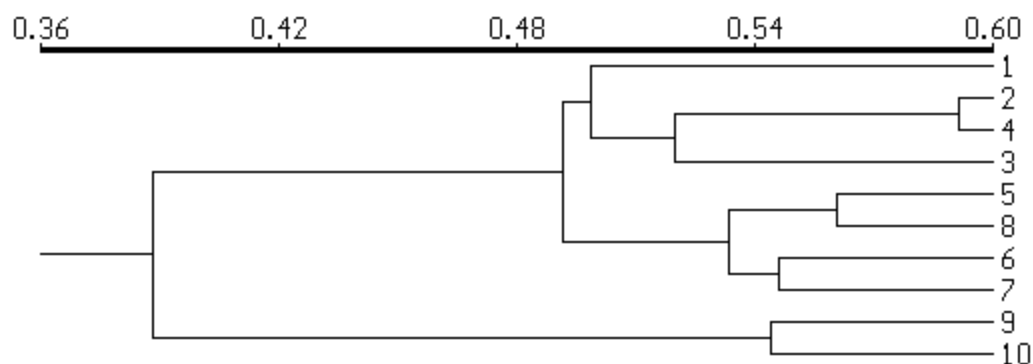


Figura 4 - Padrão de similaridade genética obtido entre as 10 populações de cagaiteira, definido pelo critério de agrupamento UPGMA, com base na médias dos índices de Jaccard.

O alto valor encontrado na correlação entre matriz de distâncias geográficas e genéticas ($r=0,770$) sugere um padrão espacial da variabilidade genética existente entre as populações e que estas estão estruturadas no espaço, conforme pode ser visualizado na Figura 5. É provável que esta estrutura tenha se originado por um modelo de diferenciação estocástica, existindo fluxo a pequenas distâncias e deriva nas populações.

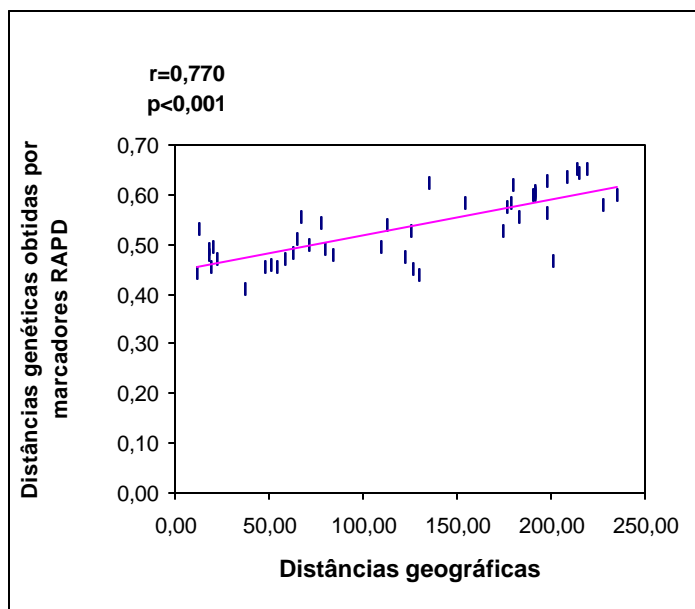


Figura 5 - Representação da correlação (r) entre a matriz de distâncias geográficas e a matriz de distâncias genéticas obtida por marcador RAPD. Significância de r dada pela probabilidade p .

3.4 Discussão

O conhecimento da distribuição da variabilidade genética entre e dentro de populações naturais de *E. dysenterica* é essencial para a adoção de estratégias eficientes para conservação de germoplasma em condições *ex situ* e *in situ*. Os marcadores RAPD atualmente têm sido muito utilizados para estudos de variabilidade genética em populações naturais.

Araújo (2001), estudando o pequi (*Caryocar brasiliense*) uma arbórea de ampla distribuição no cerrado, analisou 23 populações distribuídas na região sudeste e sudoeste do estado de Goiás. Com 46 locos marcadores RAPD foi encontrado um valor de $\hat{f}_{ST}=0,260$ e $\hat{G}_{ST}=0,304$. Estes parâmetros estimados com RAPD foram elevados quando comparados com os dados obtidos por Collevati et al. 2001, que analisando a mesma espécie, obteve $\hat{q}_p=0,05$. Os dados obtidos no presente estudo em cagaiteira, com $\hat{f}_{ST}=0,2703$, também detectaram diversidade

entre populações maior que a obtida por Telles (2000), que obteve $\hat{q}_p = 0,156$, usando isoenzimas.

Trindade (2001) utilizando 35 marcadores RAPD em três populações de cagaiteira da região nordeste do Estado de Goiás, verificou um valor de $\hat{f}_{ST} = 0,086$, sendo este, um valor bem menor do que o encontrado no presente estudo. Isso pode ser devido à proximidade entre as três populações estudadas. Estas podem estar em um processo de diferenciação populacional muito mais lento que o encontrado na região sudeste do Estado.

Mariot (2000) analisou a estrutura genética de populações naturais de *Piper cernnum* e observou que a diferenciação genética entre quatro populações da Mata Atlântica foi elevada ($\hat{F}_{ST} = 0,29$), com forte estruturação espacial. O autor atribui a diferenciação encontrada ao efeito fundador das populações, visto ser esta uma espécie pioneira que coloniza clareiras dentro da floresta.

Estudando pimenta longa em áreas antropizadas, no estado do Acre, Wadt (2001) avaliou a estrutura genética de 13 populações naturais de *P. hispidinervum* utilizando 44 locos de marcadores RAPD. Verificou que a variabilidade entre população foi alta $\theta_p = 0,25$. Duas regiões, com dois grupos distintos representaram o alto e o baixo Acre, dando um valor de $\hat{f}_{ST} = 0,2061$.

A estimativa $\hat{f}_{ST} = 0,2703$, baseada nos dados binários gerados por marcadores RAPD, foi mais alta que as encontradas com isoenzimas $F_{ST} = 0,156$, conforme Telles (2000), já mencionado. Esta comparação deve ser feita com ressalvas, uma vez que os dados obtidos com marcadores isoenzimáticos ou com outro qualquer tipo de marcador, não são diretamente comparáveis, devido à própria natureza do marcador. Isso pode ser devido ao fato das enzimas serem marcadores que provavelmente estão sofrendo seleção, e que de maneira geral estão relacionados a caracteres adaptativos, já com os marcadores RAPD cerca de 95% das bandas são oriundas de regiões não codantes do genoma e fazem parte do DNA repetitivo, sendo portanto neutros. Outro ponto a ser salientado é a amostragem genômica. Os marcadores RAPD, devido à facilidade da técnica, permitem a obtenção de um número muito maior de locos do que os obtidos com marcadores isoenzimáticos. Outro aspecto importante a ressaltar, é a dominância, pois os RAPD sendo marcadores dominantes, segundo Lynch & Milligan (1994),

podem causar um viés nas estimativas obtidas; isto já não acontece com os marcadores codominantes (como é o caso das isoenzimas).

Buso et al. (1998) estudando a estrutura genética de populações de arroz selvagem, utilizaram marcadores isoenzimáticos e RAPD para estimar o nível de diversidade de quatro populações de arroz selvagem da América do Sul (*Oryza glumepatula*), coletadas na floresta amazônica e em rios do Oeste do Brasil. O padrão de diversidade genética entre e dentro de populações foi calculado para ambos os tipos de marcadores. Com quatro locos de isoenzimas os autores encontraram um $\hat{F}_{ST}=0,31$ para e com 156 locos de marcadores RAPD um $\hat{F}_{ST}=0,64$.

O fluxo gênico encontrado neste estudo foi igual a 0,675. Este valor pode ser considerado como intermediário. Segundo Slatkin (1985), a deriva genética resultará em diferenciação populacional se o valor do fluxo for menor do que um migrante por geração. Isso não deverá ocorrer se este valor for maior que 1. Ainda, de acordo com Valva & Coelho (1998), quando o fluxo gênico for restrito, as populações tenderão a ter um menor tamanho efetivo e mais endogamia e, como resultado, uma maior probabilidade de se diferenciarem ainda mais. Por outro lado, uma alta taxa de fluxo gênico homogeneiza as diferenças genéticas entre populações, mesmo em presença de uma seleção intensa.

Apesar de ser uma espécie com tendência para a alogamia, conforme Telles (2000) o fluxo gênico entre as populações é pequeno e considerado restrito, provavelmente, em consequência da antropização do cerrado. O fluxo sendo restrito a deriva provocou maior diversidade entre estas populações. Um fato interessante a discutir é o agrupamento formado pelas populações 9 e 10, que pode ter sido causado deriva genética e pela ausência de fluxo gênico, pois a população 9 localiza-se em região urbana e a população 10 apesar de ser uma população natural, acha-se completamente isolada das demais.

O alto valor obtido da correlação entre as matrizes de distâncias geográficas e genéticas sugere que um padrão espacial da variabilidade genética existe entre as populações e que estas estão, portanto, estruturadas no espaço.

Os resultados sugerem que as populações estudadas estão se diferenciando por um processo estocástico, havendo fluxo dependente da distribuição geográfica. Sugerem ainda que elas estão sofrendo os efeitos da deriva genética devido a

antropização e à ausência de fluxo, causada pelo isolamento por distância. O alto valor de ϕ_{ST} encontrado sugere grande divergência entre as populações e restrito fluxo gênico. Estas estimativas devem ser consideradas para a conservação e o pré-melhoramento desta espécie.

3.5 Conclusões

Os resultados apresentados permitem concluir que:

- a) A estimativa $\hat{f}_{ST}=0,2703$, baseada nos dados binários gerados por marcadores RAPD, foi mais alta que a encontrada com isoenzimas $\hat{q}_p=0,156$, obtida por Telles (2000).
- b) O fluxo gênico encontrado neste estudo foi igual a 0,675, ou seja, menor que 1, este resultado é discordante com os dados obtidos com marcadores isoenzimáticos.
- c) O alto valor obtido da correlação entre as matrizes de distâncias geográficas e genéticas sugere que um padrão da variabilidade genética entre as populações estão estruturadas no espaço.

4 ESTRUTURA GENÉTICA E FLUXO GÊNICO EM *Eugenia dysenterica* DC, UTILIZANDO MARCADORES SSR

Resumo

A cagaiteria é uma espécie frutífera nativa da região dos cerrados pertencente à família Myrtaceae que apresenta potencial de utilização em sistemas de produção agrícola e tem se destacado pelo potencial sócio-econômico para o processamento de vários subprodutos. As frutíferas nativas do Cerrado são espécies de diversos gêneros e famílias que produzem frutos de interesse tanto para a alimentação regional quanto para a industrialização. Há um mercado potencial e crescente para as frutíferas nativas, porém pouco explorado pelos agricultores. Todo o aproveitamento desses frutos tem sido feito de forma extrativista e predatória. Este estudo teve como finalidade a caracterização de marcadores microssatélites em cagaiteira para estudos sobre variabilidade genética, estrutura populacional, fluxo gênico e sistema reprodutivo. O objetivo principal foi permitir a geração de informações para a domesticação e melhoramento da espécie, além de sua conservação. Uma bateria de 356 pares de *primers* desenvolvido para *Eucalyptus* spp. foram testados em cagaita. Somente 10 pares de *primers* foram transferidos de uma espécie para outra. Usando gel de poliacrilamida, foi encontrada uma média de 10,4 alelos por loco, em uma amostra de 116 indivíduos de 10 populações naturais de cagaiteira. Os 7 locos polimórficos permitiram estimar parâmetros genéticos, como a heterozigosidade esperada média ($H_e=0,442$), a diversidade entre populações $R_{ST}=0,268$ e o fluxo gênico $Nm=0,680$. Conclui-se que, é possível a transferibilidade de *primers* de microssatélites de diferentes gêneros, como, no caso de *Eucalyptus* para *E. dysenterica*. Sugere-se ainda que, a espécie estudada é predominantemente alógama e que os marcadores microssatélites são mais sensíveis que RAPD e isoenzimas para cálculo de estimativas populacionais.

Summary

The Cagaita (*Eugenia dysenterica*) is a widespread plant found in the Brazilian cerrado. Its fruit is used for popular consumption and for industrial purposes. This study opens a new perspective for the generation of population genetic data and parameter estimates for devising sound collection and conservation procedures for *Eugenia dysenterica*. A battery of 356 primer pairs developed for *Eucalyptus* spp. was tested on cagaita. Only 10 primer pairs were found to be transferable between the two species. Using a polyacrilamide gel, an average of 10.4 alleles per locus was detected, in a sample of 116 individuals from 10 natural populations of cagaita. Seven polymorphic loci allowed estimation of genetic parameters, including expected average heterozygosity $H_e=0.442$, diversity among population, $R_{ST} = 0.268$ and gene flow, $Nm = 0.680$. Results indicated a potential of SSR locus transferability developed for *Eucalyptus* to other species of different genera, such as in the case of the "cagaita tree". The high genetic diversity among populations detected with SSR markers indicated that these markers are highly sensitive to detect population structure. Estimated Nm values and the existence of private alleles indicated reduced gene flow and consequently possible damages of the metapopulation structure due to human activities.

4.1 Introdução

A cagaiteira é uma espécie frutífera nativa da região dos cerrados pertencente à família Myrtaceae que apresenta potencial de utilização em sistemas de produção agrícola e tem se destacado pelo potencial sócio-econômico para o processamento de vários subprodutos. A espécie pode contribuir para a geração de empregos e renda para as comunidades regionais. Além de ser uma planta ornamental e melífera, presta-se à extração de cortiça e ainda encontra utilização em pequenas construções civis ou na fabricação de carvão, fornecendo lenha de boa qualidade, sendo sua casca utilizada em curtumes. Sua folha tem propriedades anti-diarréicas, existindo relatos do seu uso para o tratamento de diabetes e icterícia e seus frutos têm qualidade laxativa.

Há um mercado potencial e crescente para essas frutíferas, porém pouco explorado pelos agricultores. Todo o aproveitamento desses frutos tem sido feito de forma extrativista e predatória.

A fragmentação da vegetação do Cerrado no Brasil é causada pela expansão das fronteiras agrícolas, que têm afetado a dinâmica de populações de muitas espécies, inclusive as populações de *E. dysenterica*. A alteração destas áreas pode reduzir a variabilidade genética através do efeito fundador, ou efeito de gargalo. A deriva genética e o restrito fluxo gênico, provocam um acréscimo na endogamia promovendo um aumento da divergência genética entre populações. A endogamia pode levar à fixação de alelos deletérios, colocando em risco de extinção determinadas populações presentes neste habitat (Gilpin & Soulé, 1986 e Young et al., 1996).

Para o estudo de populações naturais de espécies nativas torna-se necessário o conhecimento da diversidade entre populações, fluxo gênico e outros parâmetros genéticos.

O surgimento dos marcadores moleculares possibilitou a ampliação dos conhecimentos sobre populações de espécies nativas. Marcadores microssatélites (SSR) têm sido utilizados nos estudos de população naturais (Collevatti et al., 1999; Daynadan et al., 1997; Daynadan et al. 1999 e Gaiotto, 2001), pois são altamente polimórficos quando comparados com outras classes de marcadores. O marcador SSR tem sido muito utilizado como ferramenta para responder a diversas questões sobre a genética de populações, fluxo gênico e análise de paternidade (Wright & Bentzen, 1994).

Os microssatélites apresentam grandes vantagens devido à codominância por serem altamente polimórficos e multialélicos. Microssatélites ou SSRs (Short Sequence Repeats) são definidos como sequências repetidas compostas de número variável de motivos com 1 a 6 pb, encontrados no genoma de vertebrados, insetos e plantas. Pelo menos 30.000 locos estão presentes no genoma humano (Charlesworth, 1995).

Embora o sistema de análise de microssatélites tenha sido originalmente desenvolvido para humanos, essa tecnologia recentemente tem se tornado uma poderosa ferramenta para pesquisas em plantas (Dow & Ashley, 1996; Daynadan et al., 1997; Write & Powell, 1997; Kirst et al., 1997; Steinkellner et al., 1997; Cipriani et al., 1999; Collevatti et al., 2001; Roa et al., 2000 e Zucchi et al., 2002b).

Entretanto, avanços na utilização de microssatélites têm sido dificultados devido ao custo elevado e ao tempo para o desenvolvimento de *primers* específicos para cada loco de espécies nativas. A chance do sucesso na transferibilidade (amplificação heteróloga) de seqüências de DNA por PCR é inversamente relacionada com a distância evolucionária entre as duas espécies. Muitos estudos têm mostrado a possibilidade de usar pares de *primers* desenhados para uma espécie pertencente ao mesmo gênero (Cipriani et al., 1999; Isagi & Suhandono, 1997), ou até mesmo entre diferentes gêneros (Roa et al., 2000; White & Powell, 1997 e Zucchi et al., 2002a).

O presente trabalho utilizou a transferibilidade para *E. dysenterica* de pares de *primers* desenvolvidos em *Eucalyptus* spp., visando a caracterização de marcadores microssatélites em cagaiteira para estudos da variabilidade genética, estrutura populacional, fluxo gênico e sistema reprodutivo. O objetivo principal desta pesquisa é permitir a geração de informações para a domesticação e melhoramento da espécie, além de sua conservação.

4.2 Material e Métodos

4.2.1 Coleta do Material

O material de estudo foi coletado em dez locais da região sudeste do Estado de Goiás, perfazendo dez populações, representadas por 116 matrizes. Para a genotipagem das plantas coletou-se material vegetal (folhas) em cada uma das 116 matrizes de cagaiteira. Na tabela 7, encontram-se as localidades onde foram coletadas as populações. Estas regiões foram selecionadas através de viagens de prospecção que ocorreram durante os meses de agosto e setembro de 1996, em que a finalidade foi a delimitação ou demarcação das áreas de coleta das populações da espécie estudada. As populações de 1 a 8 foram coletadas nesta mesma época (final de 1996) e as matrizes das populações 9 e 10 foram coletadas em novembro de 1999.

Tabela 7. Localidades do Estado de Goiás e número de árvores amostradas de cagaiteira com suas respectivas informações geográficas.

Áreas	Localidade Município	Número de árvores amostradas	Altitude (m)	Latitude	Longitude
1	Catalão	12	880	18°07' 35"	47°54' 20"
2	Catalão	12	860	18°02' 0,3"	48°02' 31"
3	Catalão	12	800	18°13' 39"	47°58' 12"
4	Três Ranchos	12	820	18°17' 15"	47°48' 41"
5	Campo Alegre de Goiás	12	930	17°39' 11"	47°46' 37"
6	Campo Alegre de Goiás	12	780	17°34' 24"	47°42' 12"
7	Cristalina	5	890	17°10' 47"	47°31' 07"
8	Luziânia	11	900	16°28' 48"	47°48' 40"
9	Goiânia	16	740	16°40' 30"	49°14' 42"
10	Senador Canedo	12	840	16°37' 13"	49°04' 29"

4.2.2 Transferibilidade dos Locos de SSR e Condições de Amplificação

Neste trabalho avaliou-se a transferibilidade de pares de *primers*, desenvolvidos para *Eucalyptus* em *Eugenia dysenterica*, pertencente à mesma família Mirtaceae, porém de gêneros diferentes.

Foram utilizados 356 pares de *primers* desenvolvidos para *Eucalyptus* spp. (*Eucalyptus grandis* X *Eucalyptus urophylla*) por Brondani et al. (1998). Esses pares de *primers* foram testados para amplificação em *Eugenia dysenterica*.

Para a reação de amplificação por PCR foram utilizados 15 ng de DNA genômico, em volume de 25 µL contendo 50mM KCl, 20mM Tris-HCl pH 8.8; 1,5mM MgCl₂; 0,10mM dNTPs; 0,2 µM de cada primer (forward + reverse) e 1U de Taq polimerase.

O protocolo de PCR consistiu em uma prévia denaturação inicial a 96°C por 3 min, seguida por 30 ciclos de 94°C por 30 seg, 60°C por 1 min, 72°C por 1 min, e o último passo para extensão a 72°C por 7 min. Os fragmentos amplificados foram separados em gel de poliacrilamida 4%, em corrida com 1X TBE a 2000 V por 2h, e corados com nitrato de prata. Utilizou-se como marcador de peso molecular, o ladder 10 bp.

4.2.3 Metodologia de Análise Estatística dos Dados

A partir da leitura dos dados nos géis foram obtidas as frequências gênicas (alélicas) e genotípicas em cada loco. Estas frequências foram submetidas a um teste de aderência (teste exato de Fisher) às proporções de equilíbrio de Hardy-Weinberg, conforme definido por Weir (1996), utilizando o programa TFGA (Miller, 1997). O teste exato de Fisher foi feito por meio do método convencional de Monte Carlo utilizando 10 *batches* com 1000 permutação por *batch*.

A diversidade genética e as estatísticas F foram estimadas sob modelo aleatório de acordo com Weir (1996), em que as populações amostradas são consideradas como representativas da espécie e com uma história evolutiva comum. As frequências alélicas, o número de alelos por loco (A), a heterozigotidade observada (H_0) e esperada (H_e) e as estatísticas F de Wright (F_{IS} , F_{ST} e F_{IT}), foram estimadas utilizando o programa GDA (Lewis & Zaykin, 2000).

Em locos de microssatélites, o processo mutacional não está de acordo com o que se admite no modelo de alelos infinitos com baixas taxas de mutação. Por essa razão, foi utilizada, além da estatística de F_{ST} , uma análoga a ela, denominada de R_{ST} (Slatkin, 1995), desenvolvida especificamente para dados de microssatélites. Parâmetros como o R_{ST} e o fluxo gênico (Nm) foram estimados com o auxílio do programa $R_{ST}Cal$ (Goodman, 1997). A estruturação da variabilidade foi visualizada através de dendrogramas construídos pela matriz de distâncias genéticas de Nei e pelo critério de agrupamento UPGMA, e para isto utilizou-se o programa NTSYS (Rohlf, 1989). A estabilidade dos agrupamentos também foi testada através pelo procedimento de reamostragem por 10000 bootstrap.

Com a finalidade de analisar os padrões de variação espacial, foi obtida estimativa do coeficiente de correlação de Pearson (r) entre matrizes de distâncias genéticas de Nei e de distâncias geográficas entre as populações. A significância desta correlação matricial foi testada pelo estatística Z de Mantel, utilizando 9999 permutações aleatórias.

4.3 Resultados

4.3.1 Transferibilidade de *Primers* SSR de *Eucalyptus* ssp. Para *E. dysenterica*

Os 356 *primers* testados foram classificados conforme a qualidade obtida pela PCR, a saber: 2,8% (dez pares de *primers*) amplificaram claros produtos de SSR; 30,0% tiveram amplificação inespecífica de banda e 67,2% não amplificaram nenhuma banda.

Pela seleção foram identificados 10 pares de *primers* que foram utilizados para o estudo da estrutura genética populacional. Inicialmente, todos os *primers* foram submetidos ao programa básico com 56°C para o anelamento do primer. Aqueles que não amplificaram satisfatoriamente foram submetidos a ciclos de amplificação com temperaturas de anelamentos mais baixas. Os *primers* EMBRA 73 e EMBRA 134 produziram amplificação satisfatória a 52°C.

Na tabela 8 são apresentadas as condições de amplificação utilizadas para os 10 locos de SSR avaliados e suas respectivas amplitudes alélicas. A maior amplitude alélica observada foi de 157 pares de bases.

As frequências alélicas estão representadas graficamente na Figura 6.

Tabela 8. Sequências dos pares de *primers* desenvolvidos para *Eucalyptos* que amplificaram locos microssatélites em *Eugênia dysenterica*, com as respectivas amplitudes alélicas e temperatura de anelamento (T_a).

Locos de SSR	Sequências dos pares de primer	Amplitude alélica (pb)	A *	H_e	H_o	T_a (°C)	Número de acesso no Genbank
EMBRA 14	F-5'gCC TCA AAC CAA TTC AAA T3' R-5'CAT gAT TCT CCC ACT CCT C3'	95-170	22,0	0,837	0,514	56	G74881
EMBRA 210	F-5'CgT gTg gTT Atg TgA ACT3' R-5'CCT AAC AAT gCA TAA gCT C3'	88-245	14,0	0,831	0,706	56	G74883
EMBRA 63	F-5'-CAT CTg gAg ATC gAg gAA-3' R-5'-gAg AgA Agg ATC ATg CCA-3'	165-175	4,0	0,497	0,681	56	G74884
EMBRA 122	F-5'TTg CTC CAT CTT TCT TgC3' R-5'AAA ACg ATT AgA ggg TCA Tg3'	290-340	12,0	0,675	0,406	56	G74876
EMBRA 73	F-5'Cgg TCg TTg TCg gAA TCT C3' R-5'AgT Tgg gTA ACg CCA ggT TT3'	153-158	3,0	0,198	0,207	52	G74880
EMBRA 172	F-5'AAAgCgAACggTCACACC3' R-5'gTgCTTCTCCAggTTCTgATC3'	110-132	5,0	0,669	0,496	56	G74877
EMBRA 72	F-5'CTggTCAACgTCCgAAAg3' R-5'ATgCTgCAGAgggCATAA3'	100-190	13,0	0,405	0,200	56	G74879
EMBRA 134	F-5'CTC TgA ggA gTT ggC AgT AgC3' R-5'CAC gTT TAA ATg CgC AAg Tg3'	monomórfico	1,0	-	-	52	G74885
EMBRA 179	F-5'gTCggCTCACAgCATGAA3' R-5'gCCTCCAgTAgTTAACAGACG3'	monomórfico	1,0	-	-	56	G74878
EMBRA 243	F-5'gCgTACggATCAAgaACA3' R-5'gAAAggAACgCCAATAA3'	monomórfico	1,0	-	-	56	G74882

A*: número de alelos, H_e e H_o : heterozigosidades esperada e observada, respectivamente.

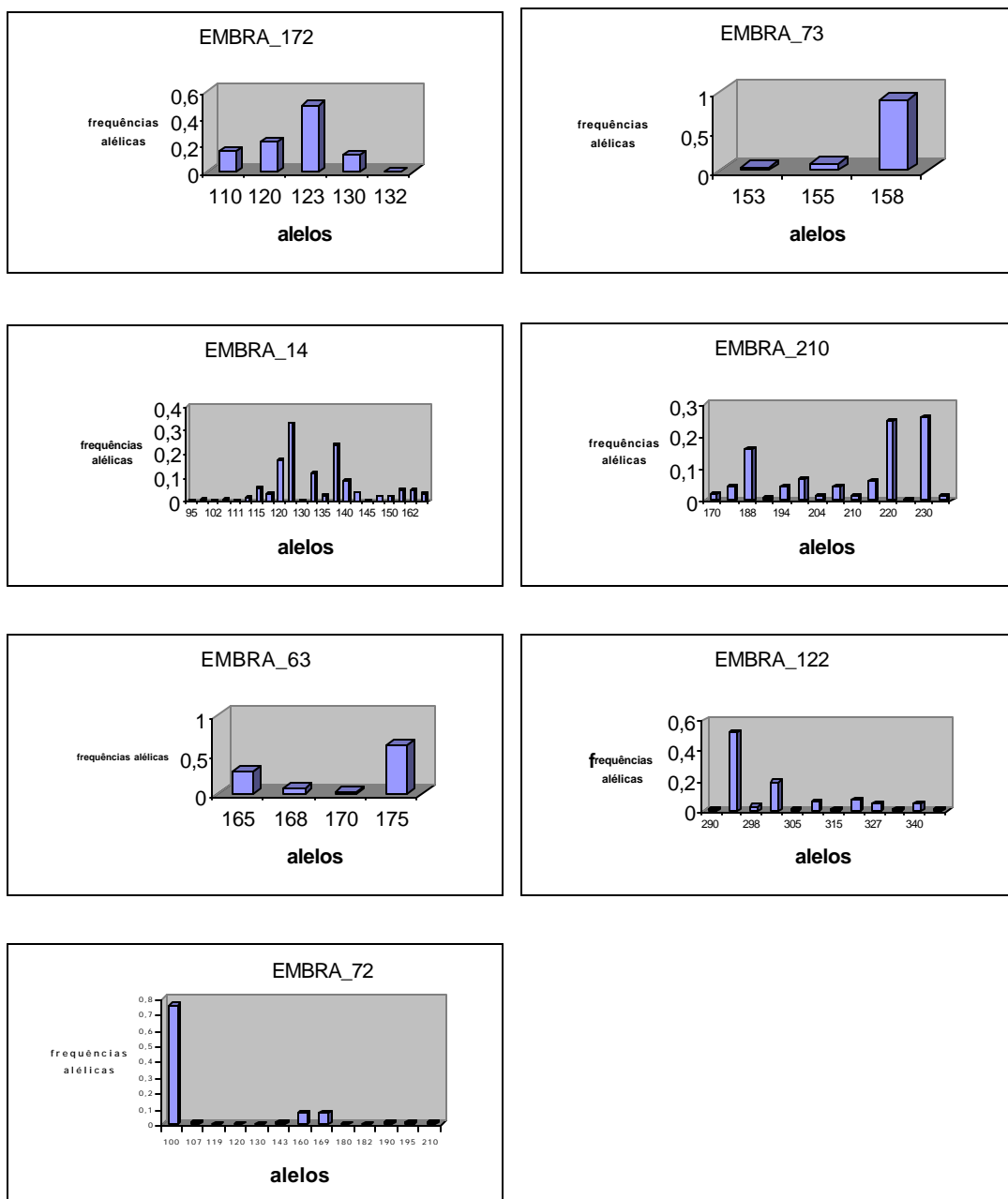


Figura 6 - Histograma das frequências alélicas dos 7 locos de SSR, estimados para 116 indivíduos de *Eugênia dysenterica*. O eixo Y indica a frequência alélica e o eixo X indica o número do alelo.

4.3.2 Variação Genética

O número médio de alelos por loco polimórfico foi de 10,4, variando de três alelos (loco EMBRA 73) ao máximo de 22 alelos (loco EMBRA 14), como pode-se observar na Tabela 8. Nesta tabela pode-se verificar a maior e a menor heterozigosidade observada e esperada nos locos. O loco EMBRA 210 apresentou a maior heterozigosidade observada (0,706) e o EMBRA 72 a menor (0,200). A maior heterozigosidade esperada foi verificada no loco EMBRA 210 (0,831) e a menor (0,198) no loco EMBRA_73. A Figura 7 apresenta a variabilidade alélica existente nos locos EMBRA-210 e EMBRA-14.

Conforme pode ser visto na Tabela 9 a heterozigosidade observada variou entre populações desde 0,253 (população 1) a 0,599 (população 10), sendo a média igual a 0,458. Já a menor heterozigosidade esperada foi de 0,276 para a população 1, sendo que a população 9 apresentou o maior valor desta estimativa (0,670).

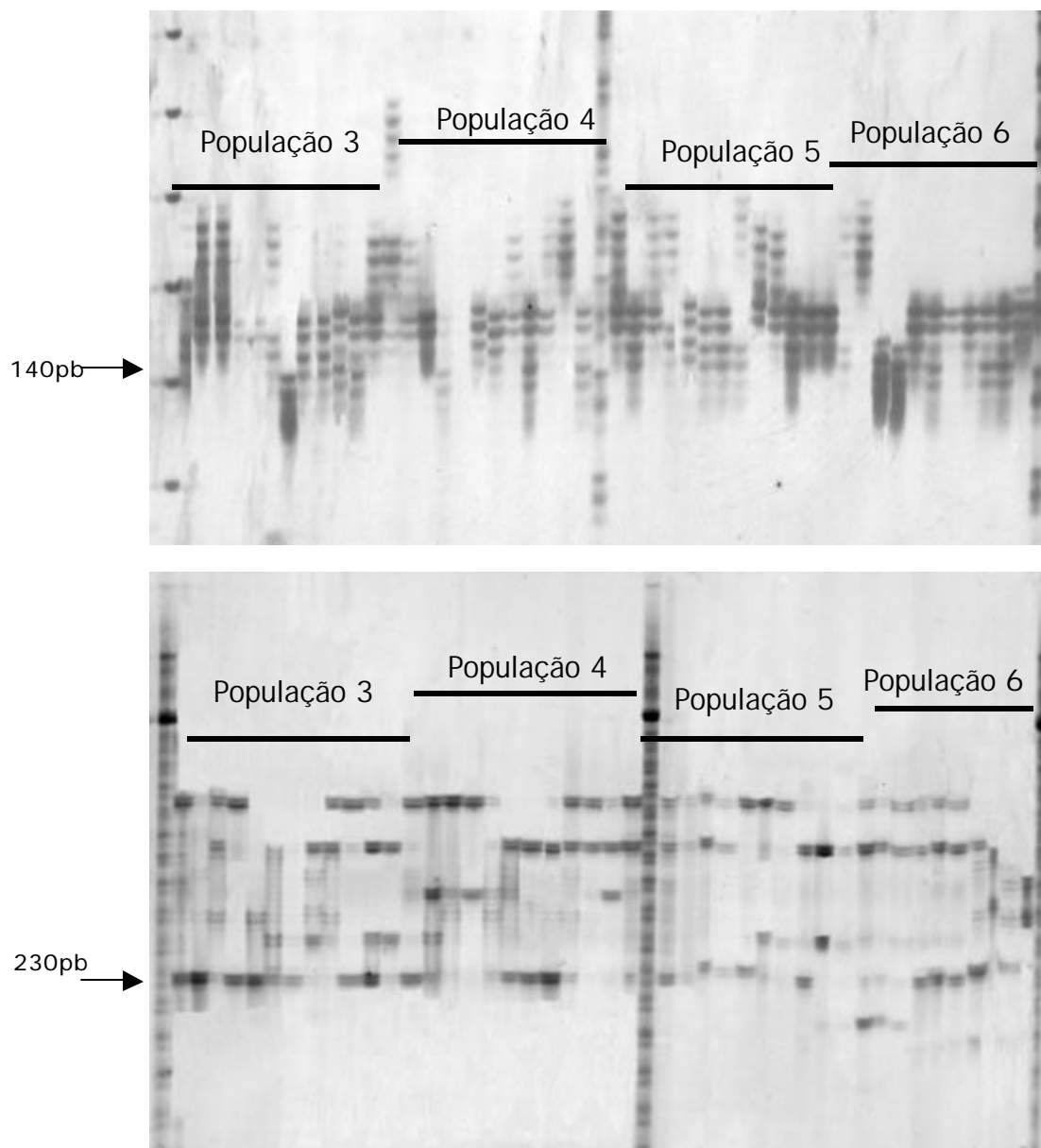


Figura 7 - Perfil de um gel de SSR utilizando os *primers* EMBRA-210 e EMBRA-14 em 44 plantas de cagaiteira; à esquerda está o padrão de peso molecular um ladder 10 pb.

Tabela 9. Estimativas de parâmetros genéticos de diversidade em 10 populações de *Eugenia dysenterica*, sendo N: número de indivíduos amostrados; L: número de locos microsatélites; n_A : número médio de alelos; A: número total de alelos; H_o : heterozigosidade observada; H_e : heterozigosidade esperada sob equilíbrio de Hardy-Weinberg; f: índice de fixação, t_a : taxa de cruzamento aparente.

Pop.	N	L	n_A	A	H_o	H_e	f	t_a
1	12	7	2,7	19	0,253	0,276	0,088	0,838
2	12	7	2,3	16	0,449	0,403	-0,123	1,280
3	12	7	2,7	19	0,366	0,394	0,076	0,859
4	12	7	2,6	18	0,453	0,408	-0,117	1,265
5	12	7	2,6	18	0,497	0,404	-0,245	1,649
6	12	7	2,4	17	0,411	0,374	-0,106	1,237
7	5	7	2,2	15	0,450	0,389	-0,183	1,448
8	11	7	3,0	21	0,563	0,438	-0,316	1,924
9	16	7	6,0	42	0,541	0,670	0,197	0,671
10	12	7	4,8	34	0,599	0,667	0,105	0,810
MÉDIA	11,6	7	3,1	21,9	0,458	0,442	-0,037	1,077

O índice de fixação intrapopulacional ($\hat{f} = \hat{F}_{IS}$) estimado em cada população através das heterozigosidades (H_o e H_e), estão apresentados na tabela 09. O valor médio foi de $\hat{f} = -0,037$, com variação de $-0,316$ a $0,197$. Pelo processo da análise de variância obteve-se $\hat{f} = -0,017$ (IC 95% $-0,276$ a $0,134$; Tabela 12). Estes resultados, juntamente com os testes exatos de Fisher, indicam que as populações são preferencialmente alógamas.

Através do através do Teste Exato de Fisher (Tabela 10) verificou que algumas populações não se encontram em equilíbrio de Hardy-Weinberg, para a maioria dos locos estudados. A população 10 acusou significância no teste exato de Fisher em 6

dos 7 locos, portanto, em 7 locos estudados 6 deles não estão no equilíbrio de Hardy-Weinberg.

Tabela 10. Probabilidade do Teste Exato de Fisher para aderência ao equilíbrio de Hardy-Weinberg.

loco	Pop1	Pop2	Pop3	Pop 4	Pop 5	Pop 6	Pop 7	Pop 8	Pop 9	Pop 10
EMBRA 14	0,012	0,084	0,182	0.779	0.657	0,034	0.113	0.408	0,001	0,022
EMBRA 210	0.578	0,001	0,000	0.164	0.436	0.080	0.647	0.637	0.739	0,001
EMBRA 63	1.000	0.136	0.214	0,027	0,003	0,001	0.121	0,008	1.000	1.000
EMBRA 122	0,427	1.000	0.406	0.139	1.000	1.000	0.128	0.501	0,000	0,011
EMBRA 73	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0.414	0,004
EMBRA 172	1.000	1.000	1.000	0.395	0,002	0.842	0.331	0.097	0,000	0,007
EMBRA 72	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,022

Verifica-se na Figura 8, um grande número de alelos exclusivos em determinadas populações, que foram 15 na população 9, 11 na população 10, 3 alelos na população 8, 2 alelos na população 1 e 1 alelo exclusivo na população 6. Embora as populações compartilhem a maior parte dos 73 alelos, há um número expressivo de alelos que caracterizam determinadas populações (como no caso das populações 9 e 10). As respectivas frequências destes alelos podem ser observadas na tabela 11.

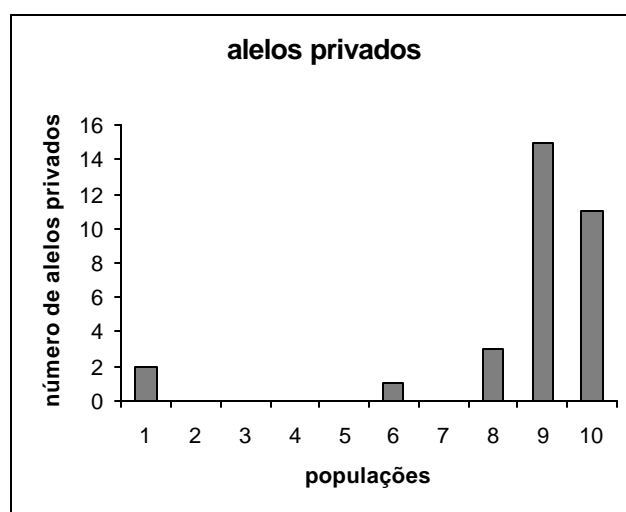


Figura 8 - Número de alelos exclusivos obtidos em 73 alelos de marcadores SSR, em 10 populações de *Eugênia dysenterica*.

Tabela 11. Alelos exclusivos e frequência com que ocorrem em cada população.

Locos	alelos	frequência	População
EMBRA 14	95	0.0454	Pop 10
EMBRA 14	111	0.0454	Pop 10
EMBRA 14	148	0.1818	Pop 10
EMBRA 14	155	0.0454	Pop 10
EMBRA 14	142	0.3181	Pop 10
EMBRA 14	102	0.0312	Pop 9
EMBRA 14	113	0.0937	Pop 9
EMBRA 14	100	0.0625	Pop 9
EMBRA 14	110	0.0625	Pop 9
EMBRA 14	162	0.0833	Pop 8
EMBRA 14	118	0.5000	Pop 8
EMBRA 14	145	0.0416	Pop 1
EMBRA 14	130	0.0416	Pop 1
EMBRA 210	210	0.1250	Pop 10
EMBRA 210	225	0.0333	Pop 9
EMBRA 210	190	0.0666	Pop 9
EMBRA 210	170	0.1818	Pop 8
EMBRA 210	204	0.1500	Pop 6
EMBRA 122	315	0.0556	Pop 10
EMBRA 122	290	0.0556	Pop 10
EMBRA 122	305	0.0333	Pop 9
EMBRA 122	343	0.0333	Pop 9
EMBRA 122	340	0.3000	Pop 9
EMBRA 172	132	0.0333	Pop 9
EMBRA 72	190	0.0909	Pop 10
EMBRA 72	195	0.1363	Pop 10
EMBRA 72	210	0.0454	Pop 10
EMBRA 72	130	0.0312	Pop 9
EMBRA 72	120	0.0312	Pop 9
EMBRA 72	143	0.0625	Pop 9
EMBRA 72	107	0.0625	Pop 9
EMBRA 72	199	0.0312	Pop 9

4.3.3 Estrutura Genética

O índice de fixação da espécie (\hat{F}_{IS}), estimado com sete locos de SSR (73 alelos) não diferiu de zero. Sendo este uma medida da endogamia localizada devido ao sistema reprodutivo, pode-se sugerir que a espécie é panmítica.

O valor de R_{ST} estimado foi de 0,269 e o de F_{ST} igual a 0,25. Estas estimativas foram congruentes e significativamente diferentes de zero, como pode se observar na Tabela 12. O parâmetro Nm estimado com base na estimativa de R_{ST} , resultou no valor de 0,680 indivíduos, indicando uma taxa de migrantes entre as populações de grandeza intermediária. Este valor pode ser considerado como indício que as populações encontram-se em um processo de diferenciação e com possível comprometimento de sua estrutura metapopulacional.

Tabela 12. Estimativa das estatísticas F de Wright, de R_{ST} e do número de migrantes por geração (Nm) em 10 populações naturais de *Eugenia dysenterica*. Intervalo de confiança (IC) de 95% de probabilidade.

	F_{IS}	F_{IT}	F_{ST}	R_{ST}	Nm
Estimativa	-0,017	0,238	0,250	0,269	0,680
Limite Superior (IC 95%)	0,1343	0,3881	0,3485	0,3588	0,7663
Limite Inferior (IC 95%)	-0,2765	0,0250	0,1941	0,2450	0,4424

As distâncias genéticas de Nei calculadas entre as populações variaram de 0,046 a 0,497. O respectivo dendrograma está apresentado na Figura 9. A correlação cofenética do agrupamento UPGMA desta matriz foi elevada (0,943). Esta análise revela que as populações 9 e 10 são as mais semelhantes entre si mas formando um grupo geneticamente diferenciado. Estas localizam-se no extremo oeste da região amostrada. As populações de 1 a 8 formam outro grupo e estão situadas na região leste da área estudada. É importante salientar que as populações 9 e 10 estão separadas das demais populações geograficamente, pela bacia do Rio Corumbá, formando, assim, dois grupos distintos. Esta estruturação foi visualizada através do dendrograma, onde fica evidente como as populações 9 e 10 diferenciam-se geneticamente das demais. Esta estruturação das populações está congruente com os dados obtidos por Telles (2000) utilizando isoenzimas e Zucchi et al. (2001) que usou RAPD. Nestes trabalhos foram avaliadas as mesmas populações deste estudo.

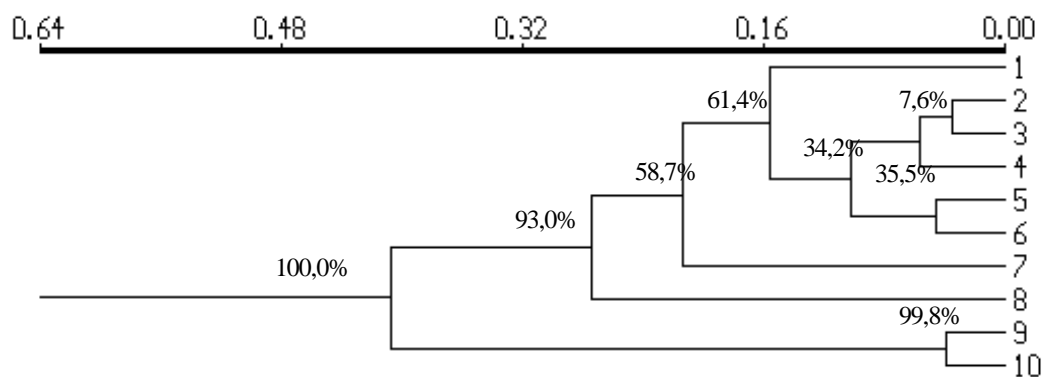


Figura 9 - Padrão de divergência genética entre 10 populações de cagaiteira, definido pelo agrupamento UPGMA, com base nas identidade genética obtida a partir das distâncias genéticas de Nei (1978). Correlação cofenética igual a 0.943.

A correlação matricial entre as distâncias genéticas de Nei (obtida por marcadores SSR) e as distâncias geográfica foi elevada e positiva ($r=0,872$) e

significativa a 1% de probabilidade, como pode-se observar na Figura 10. Este resultado sugere um padrão espacial da variabilidade genética entre as populações estando elas estruturadas no espaço. Resultado similar, porém com correlação menor ($r=0,725$), foi encontrado por Telles (2000) com dados de marcadores isoenzimáticos.

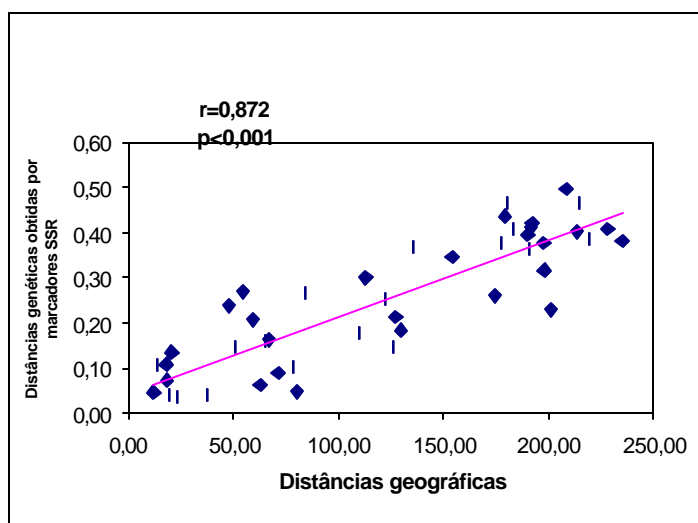


Figura 10 - Representação da correlação entre a matriz de distâncias geográficas e a matriz de distâncias genéticas (r) obtida por marcador SSR. Significância de r dada pela probabilidade p .

4.4 Discussão

O conhecimento da distribuição da variabilidade genética entre e dentro de populações naturais de *E. dysenterica* é essencial para a adoção de estratégias eficientes para conservação de seu germoplasma em condições *ex situ* e *in situ*. Os marcadores SSR constituem uma poderosa ferramenta para este tipo de estudo.

Entretanto, o progresso do uso de marcadores baseados em microsátélites tem sido dificultado devido ao alto custo e tempo despendidos no desenvolvimento de *primers* para cada espécie. A possibilidade de sucesso de amplificação heteróloga para qualquer sequência de DNA por PCR é inversamente relacionada com a distância evolucionária entre duas espécies. Muitos estudos têm mostrado, entretanto, a possibilidade da utilização de pares de *primers* desenvolvidos para

uma espécie em outras espécies do mesmo gênero (Cipriani et al., 1999), ou entre diferentes gêneros (Roa et al., 2000). O primeiro estudo demonstrou uma transferibilidade de 59% em espécies de *Prunus* e o segundo que a transferibilidade foi possível em cerca de 80% de espécies selvagens de diferentes gêneros de mandioca.

A transferibilidade de microssátelites entre espécies relacionadas é uma consequência da homologia da seqüência de DNA das regiões que flanqueiam os microssátelites. Outros estudos com árvores tropicais já têm demonstrado uma alta razão de transferibilidade de locos de SSR entre espécies arbóreas taxonomicamente relacionadas, como ocorre dentro das Leguminosae (Dayanadan et al., 1997), Meliaceae (White & Powell, 1997) e entre espécies de *Eucalyptus* (Brondani et al., 1998). A absoluta transferibilidade (100% de 10 locos) foi verificada em microssátelites desenvolvidos com *Caryocar brasiliense* para cinco outras espécies do mesmo gênero (*C. coriaceum*, *C. edule*, *C. glabrum*, *C. pallidum* e *C. villosum*) indicando um alto nível de homologia do genoma e permitindo, dessa forma, estudos comparativos da estrutura genética em todas estas espécies (Colevati et al., 1999)

Steinkellner et al. (1997) estudaram a amplificação heteróloga de locos de SSRs na família *Fagaceae* utilizando três diferentes gêneros e seis diferentes espécies de carvalho. Estes autores utilizaram 17 pares de *primers* desenvolvidos para o carvalho (*Quercus petraea*) e verificaram que 66% destes *primers* amplificaram locos de SSR com sucesso e 74% destes revelaram polimorfismo. Foi observado que o sucesso de amplificação declinava com o aumento da divergência genética entre as espécies.

Roa et al. (2000) estudaram a transferibilidade em mandioca (*Manihot esculenta*) para seis diferentes espécies (todas selvagens) do gênero *Manihot*. Em oito locos amplificados, apenas dois (ou dois pares de *primers*) não amplificaram para duas espécies selvagens mais distantes. Verificaram que muitos *primers* de microssátelites podem ser utilizados para amplificação heteróloga entre gêneros diferentes. Apesar de a transferibilidade diminuir com o aumento da distância genética, pesquisas com espécies afins, como a de Roa et al. (2000), demonstraram a possibilidade da utilização de *primers* de SSR para amplificação heteróloga em diferentes espécies e gêneros (Byrne et al., 1996; Isagi & Suhandono, 1997; Smulder et al. 1997 e Steinkellner et al. 1997).

Dayanandan et al. (1997) utilizaram pares de *primers* desenvolvidos para uma árvore tropical, *Pithecellobium elegans*, visando detectar locos de SSR que amplificassem em outras espécies da mesma família (Leguminosae). Utilizaram 13 espécies da família Leguminosa, sendo 12 da sub-família Mimosidae e uma da família Pappilionoidae. Os seis pares de *primers* desenvolvidos para *P. elegans* tiveram sucesso na amplificação em espécies do mesmo gênero e de gêneros diferentes.

Em relação à variação genética observou-se no presente trabalho um nível relativamente alto de multialelismo em todos os sete locos polimórficos. O número médio de alelos por loco foi de 10,4 e a heterozigotidade esperada média foi de 0,442, sendo maior que a encontrada por Telles (2000) em um estudo similar com as mesmas populações de cagaiteira, porém utilizando marcadores isoenzimáticos.

O valor médio encontrado de F_{IS} pode ser considerado nulo nas populações de *E. dysenterica*. Este valor sugere que a espécie aproxima-se da alogamia, e indica que as frequências alélicas destas populações encontram-se nas proporções de equilíbrio de Hardy-Weinberg, para esta geração. Este resultado contrasta com o valor de $F_{IS}=0,243$ obtido por Telles (2000) com isoenzimas e que sugeria desvios de panmixia. Este conflito pode ser devido à utilização de dados de gerações diferentes, ou até mesmo pela própria diferença da natureza do marcador genético utilizado. As enzimas são marcadores que provavelmente sofrem seleção, pois estão relacionados a caracteres adaptativos, já os microssatélites são regiões não codantes do genoma, fazem parte do DNA repetitivo e portanto são neutros.

A taxa de fecundação cruzada aparente, estimada foi considerada muito alta, tendo um valor médio de 1,08. Esta estimativa é maior que a descrita por Telles (2000), que encontrou uma média de 0,83, utilizando marcadores isoenzimáticos e dados de progênies.

Govindajaru (1989) distinguiu três níveis de fluxo gênico: alto $N_m > 1$, intermediário ($0,25 < N_m < 0,99$) e baixo $N_m < 0,25$. De acordo com esta definição o fluxo gênico encontrado neste estudo foi intermediária (0,68). Sendo a estimativa de fluxo gênico calculada através do parâmetro R_{ST} o fluxo estimado neste trabalho não deve ser considerado contemporâneo, mas baseado na história genética destas populações. Dessa forma, deve-se atentar para este fato nos estudos de conservação genética dessa espécie, bem como em outros estudos que utilizam esta estimativa.

Quanto à medida da diversidade entre populações, notou-se que as estimativas de R_{ST} e F_{ST} foram muito semelhantes. Acredita-se que mais informações de outras populações naturais, sejam necessárias para verificar as tendências da diferença entre R_{ST} e F_{ST} . Estes valores foram mais elevados que os encontrados por Telles (2000) com isoenzimas o que sugere serem os microssátélites mais sensíveis para mensurar a diferenciação entre populações, em comparação com as isoenzimas. Alíás, isso também ficou patente pela detecção de alelos privados ou exclusivos neste trabalho, o que não ocorreu com os dados obtidos com marcadores isoenzimáticos.

Apesar de ser uma espécie com grau provavelmente alto de alogamia o fluxo gênico estimado entre as populações foi pequeno e deve ser considerado restrito, provavelmente como consequência da antropização do cerrado. O fluxo é restrito e a deriva vem atuando nestas populações. Um fato interessante a discutir é a alta frequência de alelos privados nas populações 9 e 10, o que pode ter sido causado pela deriva genética e ausência de fluxo gênico, pois a população 9 localiza-se em região urbana e a população 10, apesar de ser uma população natural, está completamente isolada das demais.

Uma outra hipótese a ser considerada em relação à alta taxa de alelos privados na população 9 é que esta pode ter uma constituição genética alterada por introduções feita pelo próprio homem. Em relação à população 10 o resultado pode ser devido ao isolamento da mesma e à ausência de fluxo gênico, o que faz com que esta não compartilhe estes alelos com as demais populações. Slatkin, (1985) descreve uma metodologia para avaliação do fluxo gênico a partir de alelos raros (ou privados). A distribuição das frequências destes alelos raros (alelos que aparecem em uma única população) é utilizada para estimar a média do número de migrantes permutados entre populações locais. O logaritmo de Nm é aproximadamente relacionado de forma linear com o logaritmo da média de frequência de alelos privados (Slatkin, 1985 e Slatkin & Barton, 1986). O método de Slatkin (1985), considera a relação:

$Ln[p(1)] = a \ln(Nm) + b$, onde $p(1)$ é a frequência média dos alelos encontrados em apenas uma população, e a e b são constantes que dependem do número de indivíduos amostrados por população. Segundo os autores, a deriva genética resultará em diferenciação populacional se $Nm \leq 1$, mas não resultará, se $Nm \geq 1$ (Slatkin, 1985).

O alto valor encontrado da correlação entre as matrizes de distâncias geográficas e genéticas sugere um padrão espacial da variabilidade genética existente entre as populações estando elas estruturadas no espaço. É provável ainda que esta estrutura tenha se originado por um modelo de diferenciação estocástica, existindo fluxo a pequenas distâncias (isolamento pela distância) e que decresce com o aumento da distância.

4.5 Conclusões

Os resultados indicam um potencial para a transferibilidade de locos de SSR desenvolvidos em *Eucalyptus* para outras espécies de gêneros diferentes, como no caso da cagaita. Sugerem ainda que a espécie tem uma tendência à alogamia. A elevada diversidade genética entre populações detectada com marcadores SSR, indica ser este marcador mais sensível para detectar a estrutura populacional. Com relação ao fluxo gênico aparente e a diversidade, os resultados do presente estudo não foram totalmente concordantes com os obtidos com marcadores isoenzimáticos. Sugerem também, um comprometimento da estrutura metapopulacional, indicando que as populações estudadas sofreram deriva genética devido à antropização e têm fluxo restrito, devido ao isolamento pela distância. Estes resultados devem ser considerados para a conservação e o pré-melhoramento desta espécie.

5 ESTRUTURA GENÉTICA DE *Eugenia dysenterica* DC, UTILIZANDO DADOS DE MARCADORES RAPD EM DUAS GERAÇÕES

Resumo

Os marcadores moleculares têm sido freqüentemente utilizados em estudos sobre a diversidade e a estrutura genética populacional. Dentre os diversos tipos de marcadores atualmente disponíveis, os marcadores RAPD destacam-se pelas vantagens que apresentam em termos de simplicidade, pelo fato de poderem ser aplicados a um grande número de espécies e por permitirem a análise de um grande número de locos. Este marcador, no entanto, apresenta dominância o que constitui sua principal desvantagem no estudo de populações. Com o intuito de contornar este problema, sem considerar condições especiais como o equilíbrio de Hardy-Weinberg ($f=0$) ou equilíbrio com autogamia ($f=1$) foram utilizadas neste trabalho progênies para genotipagens com RAPD visando inferir o genótipo das plantas matrizes e posterior cálculo das frequências alélicas. Foram utilizadas 114 famílias (matrizes) com 7 filhas cada uma, num total de 798 plantas. Estes indivíduos foram analisados com 40 locos de marcadores RAPD, o que permitiu inferir os genótipos maternos. Desse modo foi possível estimar as estatísticas F de Wright, o fluxo gênico e a taxa de cruzamento em 10 populações da espécie nativa do Cerrado, *Eugenia dysenterica*. A estimativa da taxa de cruzamento utilizando essas progênies foi de 0,937, o que sugere que a espécie é preferencialmente alógama. A variabilidade entre populações foi calculada através da AMOVA com os dados binários de progênies sendo a variabilidade entre populações igual 32,82% ($\hat{f}_{ST}=0,3282$). Estimativa similar a esta foi calculada com os dados de frequências alélicas das plantas matrizes, obtendo-se um valor de $\hat{F}_{ST}=0,318$. Este estudo permitiu comparar dados de duas gerações e uma metodologia diferente para análise de dados com marcadores dominantes, sem impor restrições nos modelos.

Summary

Molecular markers have frequently been used in studies on the diversity and population genetic structure. Among the several types of markers now available, the RAPD markers stand out for the advantages they present in terms of simplicity, its applicability to a great number of species and the possibility of doing the analysis with a great number of loci. This type of marker, however is dominant, and this constitutes its main disadvantage in population studies. Aiming at overcoming this problem, without restrictions such as Hardy-Weinberg equilibrium ($f=0$) or complete self-fertilization ($f=1$), RAPD genotyping with progeny tests was used, in order to infer the genotype of the seed parents and subsequent calculation of allelic frequencies. 114 maternal families were used each with seven offspring, totalizing 798 plants. These individuals were analyzed with 40 RAPD markers loci, which allowed inferring the maternal genotypes. Wright's F statistics, gene flow and the outcrossing rate of 10 populations of the native species of the Savannah, *Eugenia dysenterica* were then calculated. The estimate of the outcrossing rate was 0.937, suggesting that the species is predominantly allogamic. The variability among populations was also calculated through AMOVA with the binary data of the progenies, being the variability among populations equal to 32.82% ($\hat{f}_{ST}=0.3282$). The corresponding value based on allelic frequencies of the seed parents was $\hat{F}_{ST}=0.318$. Results indicated that using RAPD markers with the genotyping of offspring of maternal progenies and of the corresponding seed parents (data of two generations) allowed estimating parameters generally used with codominant markers. The procedure used imposes no restrictions about the genetic constitution of the population.

5.1 Introdução

Os marcadores moleculares possuem diversas aplicações práticas em estudos de estrutura genética populacional e elas concentram-se em quantificar a variabilidade genética, descrever como ela se distribui nas populações e como pode ser manipulada (Robinson, 1998). Os marcadores moleculares têm sido freqüentemente utilizados em pesquisa sobre a diversidade genética e em estudos da estrutura genética populacional. Na literatura podem-se encontrar muitos exemplos em estudos de conservacionismo como os de Chase (1996), Palacios & Gonçalves-Candelas (1997), Ayres & Ryan (1999), Maguire & Sedgley (1997); Wolf et al (1998), Danquah et al. (2002), Lopes et al. (2002), Lefort et al. (2002), Ramakrishnan et al. (2002), Lowe et al (2002) entre muitos outros.

A estrutura genética populacional simplificada refere-se à distribuição da variabilidade entre e dentro de populações. A formação desta estrutura resulta de vários fatores como: o sistema de acasalamento, níveis de endogamia, fluxo gênico, bem como a deriva genética entre e dentro das populações.

Para que se possa entender a estrutura de uma população, é necessário estimar parâmetros populacionais tais como: frequências genotípicas e alélicas, índices de fixação, taxa de cruzamento natural, tamanho efetivo populacional, grau de heterozigosidade média, entre outros. Atualmente estes parâmetros têm sido estimados com a utilização de marcadores moleculares codominantes.

A tentativa de utilizar marcadores dominantes deve-se às vantagens de técnicas como RAPD (Random Amplified Polymorphic DNA), as quais são de grande simplicidade. De modo contrário, os marcadores codominantes apresentam desvantagens tais como: limitação do número de locos, no caso dos marcadores isoenzimáticos, amostrando pouco o genoma; a necessidade de se dispor de um banco prévio de sondas radioativamente marcadas, no caso do RFLP, além da morosidade destas técnicas. No caso dos marcadores baseados no polimorfismo de microssatélites, sabe-se que eles precisam ser desenvolvidos para as pesquisas com espécies nativas, embora já se disponha dos *primers* necessários para análises nas espécies economicamente mais importantes.

Diversos modelos estatísticos têm sido desenvolvidos para estimar parâmetros genéticos populacionais com a utilização de marcadores dominantes

(Excoffier et al., 1992 e Lynch & Milligan, 1994), porém o uso destes modelos ainda é limitado pelos pressupostos que apresentam.

A metodologia desenvolvida por Excoffier et al. (1992) é uma alternativa para o estudo da estrutura genética de populações de uma única espécie, através de uma metodologia conhecida como análise molecular de variância (AMOVA). Informações ao nível de DNA, obtidas de haplótipos, foram incorporadas no formato da Análise de Variância, tradicionalmente utilizadas em estatística experimental, derivada da matriz de distâncias quadradas entre todos os pares de haplótipos. Huff et al. (1993) aplicaram pela primeira vez a metodologia de AMOVA na plantas de *Buchlœe dactyloides* (Nutt.). Muitos estudos têm utilizado a metodologia de AMOVA para a análise de dados binários, gerados por marcadores dominantes em estudos da estrutura genética populacional em diversas espécies (Maguire & Sedgley, 1997; Lou et al., 1998; Wadt, 2001, Trindade, 2001 e Araújo, 2001).

Segundo Lynch & Milligan (1994), a análise dessa estrutura com dados de marcadores dominantes é complicada devido à falta de informação genotípica completa, causada pela dominância. Eles demonstraram que desta maneira aumenta-se a variância amostral associada a um único loco e se induz a uma tendenciosidade na estimação dos parâmetros. Os autores sugerem alguns passos para tornar este viés desprezível. Verificaram que é preciso analisar duas vezes mais indivíduos do que os analisados com marcadores codominantes e também que os alelos marcadores da maioria dos locos devem estar em frequências relativamente baixas.

Essa metodologia, sem dúvida, tem sido aplicada a um número restrito de estudos. Uma das pressuposições, nesse caso, é que a população de estudo esteja obedecendo às proporções do equilíbrio de Hardy-Weinberg, o que, no entanto, requer estudos prévios da espécie, geralmente realizados com marcadores codominantes (Aagaard et al., 1998; Ayres & Ryan, 1999; Bielawski & Pumo, 1997; Smith & Pham, 1996 e Wolf et al., 1998). Portanto, para espécies de sistema misto de reprodução, essa pressuposição é irreal e a metodologia proposta não pode ser utilizada.

Estudos prévios em populações de *Eugenia dysenterica* feitos por Telles (2000), demonstraram que a estimativa da taxa de fecundação cruzada foi de 0,835, em termos médios, sugerindo que a espécie possui sistema de reprodução misto com tendência a se comportar como preferencialmente alógama.

Existem diferentes alternativas para a análise da estrutura genética populacional utilizando marcadores codominantes, porém para dados obtidos a partir de marcadores dominantes as possibilidades são ainda restritas, sendo as estimativas dos parâmetros populacionais sempre tendenciosas.

Neste contexto, fez-se necessário um estudo para o desenvolvimento de uma nova metodologia aplicada à estimação de parâmetros populacionais a partir de dados provenientes de marcadores dominantes (RAPD). Espera-se que esta pesquisa contribua para o planejamento de estratégias de conservação genética. Conforme já enfatizado, ela permitirá estimar frequências alélicas e parâmetros afins, contornando as dificuldades inerentes à não distinção entre homo e heterozigotos. Convém reforçar que atualmente dados de marcadores dominantes são em geral analisados somente pela Análise da Variância Molecular – AMOVA (Excoffier et al., 1992) ou, alternativamente, com a pressuposição de que as populações estão em equilíbrio de Hardy-Weinberg (Lynch & Milligan, 1994). Percebe-se que tais procedimentos são insuficientes ou muito restritos em sua aplicabilidade. Espera-se que esta metodologia (utilização de marcadores RAPD com duas gerações) sirva de suporte para outras pesquisas com enfoque na conservação genética. Conforme já enfatizado, será possível estimar as frequências alélicas e parâmetros afins, contornando as dificuldades inerentes à não distinção entre homo e heterozigotos.

5.2 Material e Métodos

5.2.1 Material Vegetal

O material experimental deste estudo foi coletado em 10 áreas da região sudeste do Estado de Goiás, onde foram coletadas 10 populações nas quais foram amostrados 12 matrizes em cada população bem como sementes destes frutos. Os municípios amostrados e as respectivas coordenadas geográficas estão relacionados na Tabela 13. Estas regiões foram selecionadas através de viagens de prospecção que ocorreram durante os meses de agosto e setembro de 1996, em que a finalidade foi a delimitação ou demarcação das áreas de coleta das populações da espécie estudada. As populações de 1 a 8 coletadas nesta mesma época (final de 1996) e as matrizes e progênies das populações 9 e 10 foram coletadas em

novembro de 1999. A localização dos pontos foi feita através de coordenadas Geográficas, a partir de GPS (Sistema de Geoposicionamento Global) e a altitude foi obtida através de altímetro com precisão de 10 m.

Tabela 13. Localidades do Estado de Goiás, número de árvores amostradas de cagaiteira, e suas respectivas informações geográficas.

Áreas	Localidade Município	Número de árvores amostradas	Altitude (m)	Latitude	Longitude
1	Catalão	12	880	18°07' 35"	47°54' 20"
2	Catalão	12	860	18°02' 0,3"	48°02' 31"
3	Catalão	12	800	18°13' 39"	47°58' 12"
4	Três Ranchos	12	820	18°17' 15"	47°48' 41"
5	Campo Alegre de Goiás	12	930	17°39' 11"	47°46' 37"
6	Campo Alegre de Goiás	12	780	17°34' 24"	47°42' 12"
7	Cristalina	6	890	17°10' 47"	47°31' 07"
8	Luziânia	12	900	16°28' 48"	47°48' 40"
9	Goiânia	12	740	16°40' 30"	49°14' 42"
10	Senador Canedo	12	840	16°37' 13"	49°04' 29"

Foram consideradas como áreas propícias para as coletas aquelas que possuíam no mínimo 20 plantas adultas de cagaiteira em um raio de 1 Km e que, de cada área fosse possível coletar dados de no mínimo 10 matrizes, as quais produzissem frutos para análises posteriores.

As 12 matrizes de cada área, com exceção da localidade Cristalina, representam a população de cada área. Este estudo envolveu 114 matrizes provenientes das coletas realizadas representando 10 populações naturais distribuídas da região investigada. De cada uma das 114 matrizes, foram amostradas 7 plântulas num total de 798 indivíduos para a genotipagem.

5.2.2 Condições de Amplificação

Foram obtidas amostras foliares para a extração de DNA visando análise de RAPD, conforme o protocolo descrito por Murray & Thompson (1980) e Rogers & Bendich (1985), com modificações. Após a quantificação do DNA, as reações de amplificação foram feitas em termociclador modelo PTC-100 da MJ, de acordo com as condições descritas a seguir. As reações foram feitas em volume de 25 µL contendo: 10 mM Tris-HCl pH 8,3; 50 mM de KCl; 2,0 mM MgCl₂; 0,2 mM dNTPs;

0,25 μM de *primer* (Operon Technologies), 5,0 ng de DNA molde, 1 unidades de Taq polimerase e H_2O q.s.p.. As reações foram submetidas a 48 ciclos de amplificação após desnaturação inicial a 94°C por 5 min. Cada ciclo consistiu de 30 seg a 92°C, 1 min e 30 seg a 37°C e 1min e 30 seg a 72°C. Ao final de 48 ciclos, foi realizada uma extensão final de 5 min a 72 °C.

Os produtos de amplificação foram submetidos à eletroforese (3V/cm) em géis de agarose 1,4% (p/v), utilizando o tampão de corrida TEB 1 X. Um ladder de 100bp (Gibco) foi usado como marcador de peso molecular. Os géis foram corados com brometo de etídeo e fotografados sob luz UV (EDA-KODAK)

5.2.3 Metodologia de Análise Estatística dos Dados

5.2.3.1 Estimativa do Sistema Reprodutivo Utilizando Dados de Progênes

O estimativa da taxa de cruzamento da espécie, foi feita a partir de uma matriz binária gerada após da leitura dos géis de RAPD, onde se utilizou o valor 1 para a presença do marcador e 2 para sua ausência. A taxa de fecundação cruzada dentro de famílias foi estimada sob o modelo de sistema de cruzamento (*mixed mating model*), no qual se assume que estejam ocorrendo simultaneamente eventos de fecundação cruzada e autofecundação (Shaw & Allard, 1982), através de um algoritmo que utiliza o método Newton-Raphson e o da maximização e expectativa (Expectation – Maximization) implementados pelo software MLDT (Multilocus Estimation of Outcrossing with Dominant Markers; Ritland, 1990). Esse programa fornece o valor de t e F via método Newton-Raphson e as frequências gênicas via EM. O método de Newton-Raphson é usado para estimar t e F uma vez que é mais rápido e fornece estimativas válidas. Em contrapartida o método EM é mais lento, é porém usado para obter estimativas das frequências gênicas do pólen (Ritland, 1990).

A taxa de alogamia foi estimada com base em múltiplos locos (t_m multilocos) e também para cada loco (t_s ; locos simples). O método multilocos permite, além do cálculo da taxa de alogamia, estimar os prováveis genótipos maternos e as frequências gênicas (Ritland & Jain, 1981), a partir dos marcadores tipo RAPD na população. O parâmetro t estimado é a proporção de óvulos fertilizados por

cruzamentos aleatórios e, s é a proporção de óvulos auto-fertilizados, sendo, $t = 1 - s$. As variâncias dos estimadores das taxas de fecundação cruzada foram obtidas através do método de reamostragem por randomizações de 1000 *bootstrap*.

5.2.3.2 Estimativa do Sistema Reprodutivo Utilizando Dados de Plantas Matrizes

A partir dos dados dos géis de progênies foram inferidas as frequências gênicas (alélicas) das plantas matrizes. Estas frequências foram submetidas a um teste de aderência (teste exato de Fisher) às proporções de equilíbrio de Hardy-Weinberg, conforme definido por Weir (1996).

Admitindo equilíbrio, a taxa de fecundação cruzada foi estimada conforme o procedimento descrito por Weir (1996) e Vencovsky (1994). Essa taxa foi obtida por:

$$\hat{t}_a = \frac{(1 - \hat{f})}{(1 + \hat{f})}$$

onde:

\hat{t}_a = taxa de cruzamento aparente

\hat{f} = índice de fixação

5.2.3.3 Caracterização da Variabilidade e Estrutura Genética

A partir da leitura dos géis gerou-se uma matriz binária contendo a genotipagem dos indivíduos quanto à presença (1) e ausência (0) de bandas. Com esta matriz calculou-se a porcentagem de polimorfismo obtida com cada primer utilizado.

A análise da variância de dados moleculares (AMOVA) foi realizada através da decomposição total nas suas componentes (como apresentado na Tabela 14) entre e dentro de populações utilizando-se as distâncias ao quadrado, conforme descrito por Excoffier et al. (1992) com o auxílio do programa Arlequin (Schneider et al., 2000).

Tabela 14. Esquema de Análise de Variância Molecular (AMOVA).

FONTE DE VARIACÃO	GL	SQ	QM	E(QM)
Populações	p-1	SQ _a	QM _a	$\mathbf{s}_c^2 + n'\mathbf{s}_b^2 + n''\mathbf{s}_a^2$
Famílias/Populações	F-P	SQ _b	QM _b	$\mathbf{s}_c^2 + n\mathbf{s}_b^2$
Indivíduos/Famílias	N-F	SQ _c	QM _c	\mathbf{s}_c^2
TOTAL	N-1			

Variância total $\mathbf{s}^2 = \mathbf{s}_c^2 + \mathbf{s}_b^2 + \mathbf{s}_a^2$

Através dos componentes de variância são definidos os seguintes parâmetros:

$$\mathbf{f}_{ST} = \frac{\mathbf{s}_a^2}{\mathbf{s}^2} \qquad \mathbf{f}_{FT} = \frac{\mathbf{s}_a^2 + \mathbf{s}_b^2}{\mathbf{s}^2} \qquad \mathbf{f}_{FS} = \frac{\mathbf{s}_b^2}{\mathbf{s}_b^2 + \mathbf{s}_c^2}$$

Onde:

$\hat{\mathbf{f}}_{ST}$: fração da diversidade total (da espécie como um todo) que é devida a populações.

$\hat{\mathbf{f}}_{FT}$: fração da variação total que é devida às famílias, ignorando populações, ou seja, entre todas as famílias em relação à espécie como um todo.

$\hat{\mathbf{f}}_{FS}$: fração da variação dentro de populações que é devida às famílias. Esta medida é considerada ao nível de populações individualizadamente.

Para medir a fração da diversidade que é devida a indivíduos dentro de populações foi estimado o valor de $\hat{\mathbf{f}}_{IF}$ que corresponde a:

$$\mathbf{f}_{IF} = 1 - \mathbf{f}_{FS} = \frac{\mathbf{s}_c^2}{\mathbf{s}_b^2 + \mathbf{s}_c^2}$$

As distâncias genéticas foram obtidas como $D = 1 - \{d_{xy}^2\}$ (Nei & Li, 1979) conforme descrito por Excoffier et al. (1992) utilizando o programa Arlequin (Schneider et al. 2000).

$$D = \{d_{xy}^2\} = 100 \left[1 - \frac{2n_{xy}}{n_x + n_y} \right]$$

Onde, n_x e n_y são números de marcadores observados em indivíduos x e y , respectivamente, e $2n_{xy}$ é o número de marcas existentes em ambos os indivíduos multiplicados por 100.

As significâncias destas estimativas foram obtidas pelo método de randomização utilizando 1000 bootstrap.

5.2.3.4 Caracterização e Estrutura Genética com Marcadores Dominantes e Dados de Progênes

Esta metodologia é a utilizada com marcadores codominantes. A partir da inferência do genótipo materno (como descreve o item 5.2.3.1) feita pelo software MLDT (Ritland, 1990), obtiveram-se as freqüências alélicas e as freqüências genóticas maternas submetidas a um teste de aderência (teste exato de Fisher) às proporções de equilíbrio de Hardy-Weinberg, conforme definido por Weir (1996), utilizando o programa TFGA (Miller, 1997). O teste exato de Fisher foi feito utilizando do método convencional de Monte Carlo com 10 batches e 1000 permutação por batch.

A diversidade genética e as estatísticas F foram estimadas nas plantas-mãe sob modelo aleatório de acordo com Weir (1996), em que as populações amostradas são consideradas como representativas da espécie e com uma história evolutiva comum. As freqüências alélicas, o número de alelos por loco (A), a heterozigotidade observada (H_0) e esperada (H_e) e as estatística F de Wright (F_{IS} , F_{ST} e F_{IT}), e o fluxo gênico foram estimados utilizando o programa GDA (Lewis & Zaykin, 2000).

A estruturação da variabilidade foi visualizada através de dendrogramas construídos com dados da matriz de distâncias genéticas de Nei e pelo critério de agrupamento UPGMA, nesta etapa empregou-se o programa NTSYS (Rohlf, 1989). A estabilidade dos agrupamentos também foi testada através de procedimento de reamostragem com 10000 *bootstraps*.

Com a finalidade de analisar os padrões de variação espacial, foi calculada estimativa do coeficiente de correlação de Pearson (r) entre as matrizes de distâncias genéticas de Nei e as distâncias geográficas tomando as populações duas

a duas. A significância desta correlação matricial foi testada pelo estatística Z de Mantel, utilizando 9999 permutações aleatórias através do programa NTSYS (Rohlf, 1989).

5.3 Resultados

5.3.1 Seleção de *Primers*

A partir da identificação de primers foram selecionados os que geraram um maior número de bandas polimórficas, sendo estes: OPA-01, OPA-02, OPA-03, OPA-11, OPA-13, OPA-19, OPH-04 e OPH-08. Conforme a Tabela 15, pode-se verificar a seqüência dos primers selecionados e o padrão de polimorfismo de cada um. Na Figura 11, pode-se observar o perfil do primer OPA-13. O primer que produziu um maior número de locos polimórficos foi o OPA-11 sendo o OPA-13, o que gerou um menor número.

Tabela 15. Sequência dos *primers* selecionados para *Eugênia dysenterica* e padrão de polimorfismo.

Primer	Seqüência 5'-3'	Número de locos obtidos	Número de locos polimórficos	Número de locos polimorficos analisados	Porcentagem de polimorfismo
OPA_01	CAGGCCCTTC	09	05	04	55,5
OPA-02	TGCCGAGCTG	06	06	03	100,0
OPA-03	AGTCAGCCAC	09	05	05	55,5
OPA-11	CAATCGCCGT	11	11	06	100,0
OPA-13	CAGCACCCAC	09	03	03	33,3
OPA-19	CAAACGTCGG	09	05	02	55,5
OPH-04	GGAAGTCGCC	10	09	09	90,0
OPH-08	GAAACACCCC	11	10	08	90,9
TOTAL		74	54	40	73,0

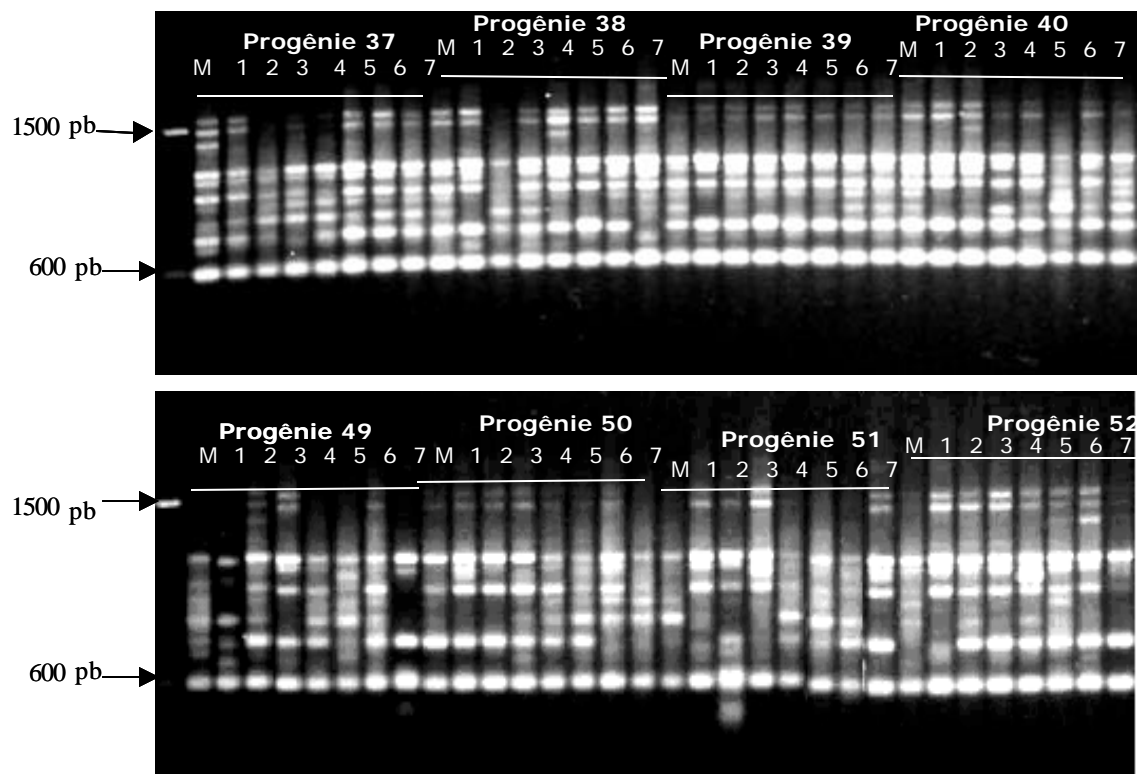


Figura 11 - Perfil de um gel de RAPD utilizando o primer OPA-13 em 74 plantas de cagaiteira; à esquerda está o padrão de peso molecular de um ladder 100 pb.

5.3.2 Sistema Reprodutivo Estimado com Progênies

A estimativa da taxa de fecundação cruzada foi realizada com base em 40 locos marcadores RAPD em 84 indivíduos em 114 famílias. Como é sabido, os marcadores dominantes não discriminam o genótipo heterozigótico (Aa), do homozigótico num determinado loco analisado (AA). Devido a esse problema os locos marcadores mais informativos para a determinação da taxa de cruzamento de uma espécie, são aqueles em que a banda aparece em baixa frequência. Por isso foram selecionados entre 54 locos 40 destes os quais apresentavam as menores frequências. Alguns marcadores apresentaram-se com bandas frequentes em algumas famílias e raras em outra, e estes foram considerados ideais para a determinação da taxa de cruzamento.

Quando uma banda encontra-se presente em toda a progênie de uma matriz, conclui-se que provavelmente a matriz (mãe) é homozigótica dominante (AA), e desse modo toda a progênie recebeu seu alelo. Portanto, quando a banda se apresentava com alta frequência esta era descartada, por não ser tão informativa. Já uma banda (de um loco) que estando presente em aproximadamente metade das plantas filhas, deduz-se que o genótipo da planta matriz deve ser heterozigótico (Aa), e dessa forma a polinização cruzada não pode ser discriminada dos eventos de autofecundação devido à dominância do marcador. Já quando a banda aparece somente nas plantas filhas e não na matriz o genótipo desta matriz será provavelmente homozigótico recessivo (aa), não apresentando a banda no gel.

Na Tabela 16, pode-se verificar que a taxa t_m (taxa de cruzamento multilocos) média foi de 0,937, variando de 0,906 (população 7) a 0,977 (população 2). Já a taxa de cruzamento t_s (locos simples) média foi de 0,838, com amplitude de 0,799 (população 5) a 0,883 (população 8).

Tabela 16. Estimativa de parâmetros do sistema reprodutivo, sendo t_m : taxa de cruzamento multiloco; t_s : taxa de cruzamento baseado na média dos locos individuais. Entre parênteses: erro-padrão estimado através de 1.000 *bootstraps*.

População	t_m	t_s	t_m-t_s
Pop 1	0,968 (0,226)	0,850 (0,053)	0,118 (0,194)
Pop 2	0,977(0.444)	0,845(0.056)	0,133(0.398)
Pop 3	0,941(0.037)	0,858(0.033)	0,082(0.029)
Pop 4	0,945(0.040)	0,811(0.044)	0,135(0.019)
Pop 5	0,907(0.037)	0,799(0.043)	0,108(0.026)
Pop 6	0,911(0.044)	0,818(0.520)	0,094(0.027)
Pop 7	0,906(0.254)	0,846(0.530)	0,061(0.218)
Pop 8	0,936(0.082)	0,883(0.054)	0,053(0.041)
Pop 9	0,917(0.038)	0,840(0.360)	0,076(0.026)
Pop 10	0,965(0.110)	0,834(0.040)	0,131(0.100)
média	0,937	0,838	0,099

5.3.3 Estrutura Genética e Fluxo Gênico (Dados de uma Geração)

A seguir estão descritos os resultados relativos a uma só geração, e que incorporam o caráter dominante do marcador utilizado.

A variabilidade genética existente entre as 10 populações de cagaiteira pôde ser medida com base em 40 locos polimórficos. Verificou-se que 32,82% da variabilidade genética está entre as populações de modo que 67,18% concentrou-se dentro delas (Tabela 17). Nesta tabela os parâmetros: \hat{f}_{ST} refere-se à fração da variação total que é devida a populações considerando a espécie como um todo, \hat{f}_{FT} : fração da diversidade total que é devida às famílias em relação a espécie como um todo e \hat{f}_{FS} : fração da variação dentro de populações que é devido à famílias, ao nível de populações idividualmente.

Em relação a espécie como um todo foi verificado que 32,82% ($\hat{f}_{ST}=0,3283$) da variação total é devida às populações, portanto 67,18% da variação total esta dentro de populações ($1 - \hat{f}_{ST}=0,6718$). Pode-se observar também que 52,80% (\hat{f}_{FT}) da variação está entre todas as famílias, sem levar em conta a população a que pertence, esta medida incluye a variação de populações e famílias. Para

isolarmos só o efeito de famílias temos que subtrair do \hat{f}_{FT} o \hat{f}_{ST} (0,5280 – 0,3280), então temos que 20,00% da variação se encontra dentro de família dentro de populações.

Ao nível de populações individualmente, em termos médios 15,10% ($\hat{f}_{FS}=0,1510$) da diversidade ocorreu entre famílias e 84,90% é a fração da variabilidade que é devida a indivíduos dentro de populações. Ou seja, a nível de populações 15,10% da diversidade está entre famílias e 84,90% está dentro de famílias. Pode-se verificar que, a maior variabilidade se encontra dentro de populações e dentro de famílias.

Tabela 17. Resultado da análise de variância molecular (AMOVA) e estimativa do fluxo gênico (Nm):

FONTE DE VARIAÇÃO	Porcentagem total da variação	GL	SQ	p	Estatísticas ϕ	Nm
Entre populações	32,82	9	1381,962	<0.001	$\phi_{ST}=0,3282$	0,512
Famílias/Populações	15,10	104	909,869	<0.001	$\phi_{FS}=0,1510$	
Indivíduos/Famílias	52,8	684	1975,286	<0.001	$\phi_{FT}=0,5280$	
TOTAL		797	4267,117			

5.3.4 Caracterização, Estrutura Genética e Fluxo Gênico utilizando Marcadores Dominantes em Duas Gerações (Inferência Do Genótipo Materno)

Neste item constam os resultados referentes aos genótipos maternos inferidos da genotipagem das respectivas progênies.

O número médio de alelos por loco foi de dois, sendo constante este número devido à inferência do genótipo materno e por se tratar de marcador dominante, conforme apresentado na Tabela 18. Pode-se ainda observar nesta tabela, que a heterozigosidade observada variou de 0,114 (população 2) a 0,250 (população 7), com média igual a 0,182. Já a heterozigosidade esperada variou de 0,158 (população 3) a 0,236 (população 7), e média igual a 0,195. Nota-se que a

heterozigidade média observada é ligeiramente menor que a esperada, e que portanto existe um decréscimo de heterozigotos.

O índice f médio foi igual a 0,071 variando entre $-0,164$ a $0,490$ enquanto a taxa de cruzamento igual a $0,867$, em média. Esses resultados evidenciam a tendência da espécie à alogamia e serão melhor discutidos posteriormente com as estatísticas F.

Tabela 18. Estimativas relativas à diversidade genética em 10 populações de *Eugenia dysenterica*, sendo N: número de indivíduos amostrados; L: número de locos RAPD; A: número médio de alelos; H_o : heterozigosidade observada; H_e : heterozigosidade esperada; f : índice de fixação, t_a : taxa de cruzamento aparente.

População	N	L	A	H_o	H_e	f	t_a
1	12	40	2	0,160	0,194	0,181	0,695
2	12	40	2	0,114	0,220	0,490	0,345
3	12	40	2	0,154	0,158	0,028	0,945
4	12	40	2	0,173	0,199	0,138	0,758
5	12	40	2	0,206	0,204	-0,130	1,026
6	12	40	2	0,221	0,201	-0,105	1,234
7	6	40	2	0,250	0,236	-0,064	1,134
8	12	40	2	0,219	0,189	-0,164	1,392
9	12	40	2	0,175	0,163	-0,073	1,157
10	12	40	2	0,150	0,188	0,213	0,695
Média	11,4	40	2	0,182	0,195	0,071	0,867

A partir das freqüências alélicas, utilizando-se o Teste Exato de Fisher, verificou-se que várias populações não se encontram em equilíbrio de Hardy-Weinberg (Tabela 19).

Tabela 19. Probabilidade do Teste Exato de Fisher para aderência ao equilíbrio de Hardy-Weinberg.

Loco	Pop1	Pop2	Pop3	Pop 4	Pop 5	Pop 6	Pop 7	Pop 8	Pop 9	Pop 10
01	0.005	0.069	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
02	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
03	1.000	0.000	0.000	0.398	0.000	0.527	1.000	0.484	0.514	0.516
04	0.396	0.191	0.404	1.000	0.255	1.000	0.520	0.064	0.000	0.000
05	0.586	0.190	0.253	0.591	0.217	1.000	1.000	0.552	0.000	0.000
06	1.000	1.000	0.520	0.088	0.206	0.593	1.000	1.000	0.000	0.000
07	0.568	1.000	0.569	1.000	1.000	0.528	1.000	1.000	0.000	0.000
08	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
09	1.000	0.0447	1.000	0.407	1.000	0.488	0.393	0.000	1.000	1.000
10	1.000	0.000	1.000	0.000	0.000	1.000	1.000	0.549	0.000	0.000
11	1.000	0.195	0.401	1.000	0.014	0.095	1.000	1.000	0.000	0.000
12	0.208	0.281	1.000	0.589	0.132	1.000	0.000	0.000	0.000	0.000
13	1.000	0.126	1.000	1.000	1.000	0.218	1.000	1.000	1.000	1.000
14	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.595
15	0.000	0.000	0.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000
16	1.000	0.404	1.000	1.000	0.524	0.000	1.000	1.000	0.000	1.000
17	1.000	0.521	0.258	1.000	0.043	0.526	1.000	1.000	1.000	0.047
18	1.000	0.290	0.563	1.000	0.523	1.000	0.283	0.000	0.000	0.247
19	0.256	0.409	0.260	0.402	0.000	1.000	0.000	0.552	0.483	1.000
20	0.000	0.000	1.000	0.045	1.000	1.000	0.000	1.000	1.000	0.000
21	0.002	0.012	0.288	1.000	0.000	1.000	0.152	1.000	0.213	0.089
22	0.000	1.000	0.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
23	1.000	0.128	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	1.000	0.000	0.000	0.000	0.591	0.545	1.000	0.065	0.000	1.000
25	0.407	0.209	0.000	1.000	1.000	0.010	0.000	0.275	0.592	0.029
26	0.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000	1.000	0.000
27	1.000	0.000	0.129	1.000	1.000	0.000	1.000	1.000	1.000	1.000
28	0.000	0.012	0.000	1.000	0.000	0.000	1.000	0.000	0.538	0.026
29	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.595
30	1.000	0.000	0.000	1.000	0.000	0.000	1.000	1.000	0.000	0.000
31	0.194	0.006	1.000	0.043	0.248	0.216	1.000	1.000	1.000	0.000
32	0.000	0.000	1.000	1.000	1.000	0.000	0.000	1.000	0.000	0.000
33	0.406	0.043	0.000	0.000	0.000	0.000	1.000	1.000	1.000	0.134
34	0.000	0.045	0.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
35	0.255	0.043	1.000	1.000	0.592	1.000	1.000	1.000	0.544	0.592
36	1.000	0.046	0.000	1.000	0.193	1.000	1.000	0.000	0.000	0.000
37	1.000	1.000	1.000	0.000	0.000	0.046	0.083	0.000	0.127	0.403
38	1.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
39	0.256	0.031	1.000	0.000	0.288	0.000	1.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000

O índice de fixação da espécie ($\hat{F}_{IS} = \hat{f}$) estimado com os 40 locos de RAPD foi de 0,082, conforme a tabela 20. Esta estimativa é uma medida de endogamia localizada devida ao sistema reprodutivo e revela o que se notou anteriormente que a espécie é preferencialmente alógama. Pode-se observar que as estimativas de \hat{f} (tabela 18) e \hat{F}_{IS} (tabela 20) tem geneticamente o mesmo significado, porém diferem apenas no processo de estimação. Para a estimativa de F_{ST} o valor foi de 0,318. Estas estimativas são congruentes com as encontradas com marcadores

SSR, para as mesmas populações. O parâmetro N_m foi de 0,536 (Tabela 16), indicando uma baixa taxa de migrantes entre as populações. Este é um indício que as populações se encontram em um processo de diferenciação e que provavelmente, não podem ser tratadas como uma metapopulação.

Tabela 20. Estatísticas F de Wright, R_{ST} e o número de migrantes por geração (N_m) para 10 populações naturais de *Eugenia dysenterica* respectivas intervalos de confiança da 95% de probabilidade.

	F_{IS}	F_{IT}	F_{ST}	N_m
Todos os locos	0.082	0,374	0,318	0,536
Superior (IC 95%)	0,145	0,444	0,387	
Inferior (IC 95%)	0,017	0,306	0,246	

As distâncias genéticas de Nei entre as populações variaram de 0,005 a 0,229. A correlação cofenética do agrupamento UPGMA desta matriz foi elevado e igual a 0,977. Esta análise revela que as populações 9 e 10 são as mais semelhantes e formam um grupo geneticamente diferente estando localizadas no extremo oeste da região estudada e as populações 1 a 8 formam outro grupo e estando situadas na região leste da área estudada (Figura 12). Esta estruturação de populações está congruente com os dados obtidos por Zucchi et al. (2001) com marcadores RAPD e Telles (2000) com dados de isoenzimas onde foram estudadas as mesmas populações.

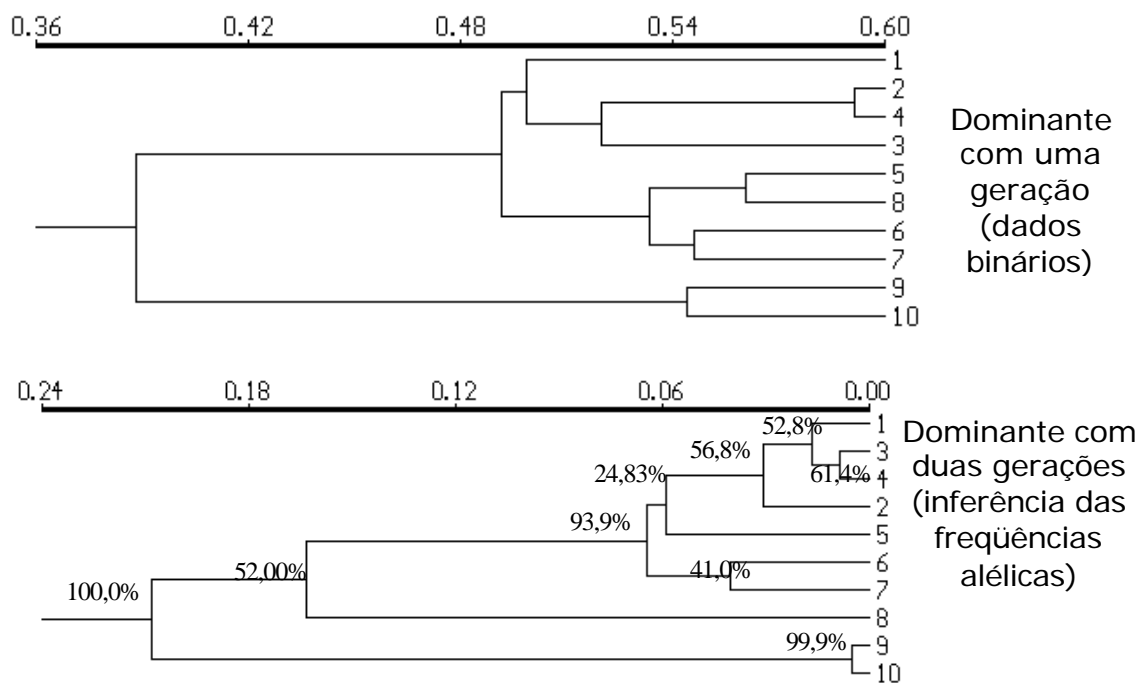


Figura 12 - Padrão de estruturação genética entre 10 populações de cagaiteira, definido pelo agrupamento UPGMA, com base no índice de similaridade de Jaccard e distâncias genéticas de Nei (1978).

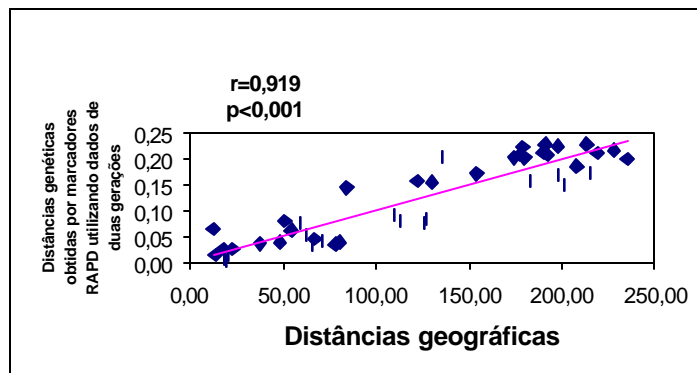


Figura 13 - Representação da correlação (r) entre a matriz de distâncias geográficas e matriz de distâncias genéticas, obtida por marcador RAPD com dados de duas gerações. p : probabilidade associada a significância de r .

Tal como ocorreu com os resultados de marcadores SSR (Zucchi et al., 2002b), os marcadores RAPD, com dados de duas gerações, também detectaram alta correlação entre distâncias genéticas e geográficas ($r=0,919$), como pode ser visualizado na Figura 13.

5.4 Discussão

O conhecimento da variabilidade entre e dentro de populações naturais de *E. dysenterica* é essencial para a adoção de estratégias eficientes para a conservação de germoplasma em condições *ex situ* e *in situ*. Atualmente diversos tipos de marcadores e diversas abordagens têm sido utilizados com esta finalidade. Porém, devido às facilidades encontradas na utilização de RAPD, este tem sido muito utilizado no estudo da variabilidade em populações naturais.

Este trabalho visou comparar estimativas de parâmetros populacionais obtidas de marcador dominante (em progênies de matrizes) com as obtidas dos genótipos das matrizes, inferidos através das progênies. Com a avaliação das progênies foi possível visar técnicas de marcadores codominantes, a partir dos genótipos maternos inferidos.

A taxa multilocos obtida com os dados de progênies foi de $t_m=0,937$. No processo de estimação deste parâmetro é possível inferir os prováveis genótipos maternos e estimar as frequências gênicas (Ritland & Jain, 1981) a partir dos

marcadores tipo RAPD na população. Esta estimativa pode ser comparada com a obtida por Telles (2000) com seis sistemas isoenzimáticos e utilizando as mesmas populações. Neste estudo a autora encontrou uma taxa de cruzamento de 0,835. Zucchi et al. (2002b) utilizando 73 alelos de marcadores SSR e analisando as mesmas matrizes obteve uma estimativa de $t_a = 1,08$. Estes dados mostram boa concordância entre a estimativa de t_m obtida com marcador RAPD e a obtida via SSR.

A estimativas de t_m obtidas com marcadores RAPD e SSR como dito anteriormente, mostram boa concordância. Porém, esta estimativa baseada e marcadores isoenzimáticos se mostram um pouco subestimada, isto pode ter sido ocorrido devido às possíveis hipóteses discutidas a seguir. A primeira hipótese é que os diferentes marcadores amostram de maneira diferente o genoma, as isoenzimas, são enzimas, portanto regiões codantes do genoma; enquanto que, os RAPD grande parte dos fragmentos é seqüência não codante ou DNA repetitivo e já os microsátélites são formados somente por seqüências repetitivas, sendo que a evolução dos SSR é muito mais rápida do que a dos outros marcadores. A segunda hipótese, refere-se à neutralidade do marcador, sabe-se que as isoenzimas sofrem seleção, pelo motivo de estarem relacionadas a caracteres adaptativos, enquanto que os marcadores baseados em DNA são neutros.

A técnica da AMOVA vem sendo freqüentemente utilizada. Estudando o pequi (*Caryocar brasiliense*), uma arbórea de ampla distribuição do cerrado, Araújo (2001) analisou 23 populações distribuídas na região sudeste e sudoeste do estado de Goiás. Com 46 locos marcadores RAPD foram encontrados os valores $\hat{F}_{ST} = 0,260$ e $\hat{G}_{ST} = 0,304$. Estes parâmetros estimados com RAPD foram elevados quando comparados com os dados obtidos por Collevati et al. (2000), analisando a mesma espécie, que obteve $\hat{p} = 0,05$.

Mariot (2000) analisou a estrutura genética de populações naturais de *Piper cernnum* e observou que a diferenciação genética entre quatro populações da Mata Atlântica foi elevada ($\hat{F}_{ST} = 0,29$) com forte estruturação espacial. O autor atribui a diferenciação encontrada ao efeito fundador das populações, visto tratar-se de uma espécie pioneira que coloniza clareiras dentro da floresta.

Estudando pimenta longa em áreas antropizadas, no estado do Acre, Wadt (2001) avaliou a estrutura genética de 13 populações naturais de *P. hispidinervum* utilizando 44 locos de marcadores RAPD. Verificou que a variabilidade dentro da população foi alta $\hat{\phi}=0,25$, Duas regiões, com dois grupos distintos, representaram o alto e o baixo Acre dando um valor de $\hat{f}_{ST}=0,2061$.

Buso et al. (1998), estudando a estrutura genética de populações de arroz selvagem utilizaram marcadores isoenzimáticos e RAPD para estimar o nível de diversidade de quatro populações de arroz selvagem da América do Sul (*Oryza glumepatula*), coletada na floresta amazônica e rios do oeste do Brasil. O padrão de diversidade genética entre e dentro de populações foi calculado para ambos os tipos de marcadores. Com os quatro locos de isoenzimas encontraram $\hat{F}_{ST}=0,31$ para e 156 locos de marcadores RAPD um $\hat{F}_{ST}=0,64$.

Com relação à estrutura genética estimada a partir da AMOVA (Análise da Variância Molecular) com base em dados de RAPD, verificou-se que a proporção da variabilidade genética dentro de populações foi maior que entre populações, ou seja, 67,18% versus 32,82%, respectivamente. Em estudo anterior com as matrizes destas progênies Zucchi et al. (2001) obtiveram valores bastante concordantes desta distribuição da variabilidade dentro e entre populações (72,97% versus 27,03% respectivamente). Apesar de serem dados obtidos em gerações e amostragem diferentes a concordância é bastante satisfatória.

Esta comparação deve ser feita com ressalvas, uma vez que os dados obtidos com marcadores isoenzimáticos ou com outro qualquer tipo de marcador, não são diretamente comparáveis, devido à própria natureza do marcador. Outro ponto a salientar é devido à amostragem genômica, pois os marcadores RAPD devido à facilidade da técnica possibilitam um maior número de locos do que os obtidos com marcadores isoenzimáticos. Outro fato é a dominância, já que o RAPD sendo um marcador dominante segundo Lynch & Milligan (1994), pode causar um viés nas estimativas obtidas da diversidade, o que não acontece com os marcadores codominantes.

O fluxo gênico encontrado neste estudo foi de $Nm=0,512$, um valor relativamente pequeno. Segundo Slatkin (1985), a deriva genética resultará em diferenciação populacional se $Nm \leq 1$, mas não resultará, se $Nm \geq 1$. Ainda, de acordo

com Valva & Coelho (1998), quando o fluxo gênico é restrito, as populações tenderão a ter um menor tamanho efetivo e mais endogamia e, como resultado, uma maior probabilidade de se diferenciarem ainda mais. Por outro lado, uma alta taxa de fluxo gênico homogeneiza as diferenças genéticas entre populações, mesmo em presença de uma seleção intensa.

As estatísticas F de Wright (1951) foram estimadas com base nos genótipos RAPD das matrizes, inferidos dos genótipos das progênes. Obteve-se $\hat{F}_{ST}=0,318$, sendo este mais elevado do que o encontrado por Telles (2000) com marcadores isoenzimáticos ($\hat{\phi}=0,156$). Zucchi et al. (2002b) utilizando marcadores SSR com as matrizes encontraram um valor de $\hat{R}_{ST}=0,268$. Estas diferenças podem ser atribuídas a várias causas como fora discutido anteriormente. De qualquer modo, novamente se verificou maior poder de detectar a divergência entre populações com os dados de RAPD, em comparação com os isoenzimáticos.

Apesar de *E. dysenterica* ser uma espécie tendendo à alogamia conforme verificado neste trabalho e também por Telles (2000), o fluxo gênico entre as populações é pequeno e pode ser considerado restrito, provavelmente, em consequência da antropização do cerrado. O fluxo sendo restrito, a deriva pode atuar nestas populações com mais intensidade. Conforme já verificado com os dados de microssatélites (Zucchi et al. 2002b) um agrupamento é formado pelas populações 9 e 10. Este pode ter sido causado pela deriva genética e pela ausência de fluxo gênico, pois a população 9 localiza-se em região urbana e a população 10, apesar de ser uma população natural, está completamente isolada das demais populações.

Como vantagens em se utilizarem progênes para inferir o genótipo materno, pode-se apontar: a possibilidade de se poder estimar um maior número de parâmetros relacionados à caracterização e estrutura genética de populações, como as estatísticas F, a heterozigosidade, entre outros, pois com esta metodologia é possível calcular frequências gênicas dos indivíduos mesmo usando marcador dominante. Permite, ainda, verificar o equilíbrio de Hardy-Weinberg. Além disso possibilita calcular a taxa de cruzamento utilizando progênes que é uma estimativa mais robusta que a taxa de cruzamento aparente. A grande desvantagem desta técnica, porém, é a dificuldade laboratorial, ou seja, o grande número de plantas para genotipar, visto que, a amostragem é multiplicada por no mínimo por 7.

Porém, sugere-se que esta metodologia seja aplicada com marcadores AFLP, o que facilitaria o trabalho laboratorial na genotipagem de um grande número de plantas, com um número razoável de locos polimórficos.

5.5 Conclusões

Conclui-se, portanto, que a metodologia proposta de análise de progênies para genotipagem materna e utilização de marcador dominante como codominante é válida para as estimativas populacionais, visto que estas estimativas se aproximaram das obtidas com marcadores SSR. Porém, é importante ressaltar, que a amostragem se torna muito maior e o trabalho laboratorial se torna laborioso.

6 ESTRUTURA GENÉTICA POPULACIONAL E PARÂMETROS AFINS AVALIADOS POR MARCADORES RAPD, SSR E ISOENZIMÁTICOS

Resumo

Um amplo conhecimento da variabilidade da espécie é fundamental para a conservação com base num manejo sustentável. Dentre as características mais importantes para a conservação, destacam-se: o conhecimento do fluxo gênico, do sistema reprodutivo e a diversidade e estrutura genética das populações. A avaliação da variabilidade genética sempre foi de interesse para os geneticistas, que desenvolveram vários métodos para analisá-la. A introdução dos marcadores moleculares revolucionou a genética de populações na década de 50, com a técnica de isoenzimas e recentemente tem conseguido enormes avanços com a aplicação de tecnologias baseadas em DNA. Os marcadores moleculares baseados em DNA (RAPD, RFLP, SSR, SCAR, AFLP, VNTR e outros) permitiram uma ampla cobertura genômica e tornaram-se poderosas ferramentas para estudos de genética de populações. Marcadores codominantes (como: isoenzimas e SSR) têm sido muito utilizados na genética de populações por serem mais informativos que os marcadores dominantes (RAPD, AFLP e VNTR). O objetivo deste trabalho foi gerar um grande número de estimativas de parâmetros populacionais com a finalidade de comparar os diferentes tipos de marcadores moleculares. Foram comparadas as estimativas relacionadas ao sistema reprodutivo (\hat{f}), heterozigidade esperada (\hat{H}_e) e heterozigidade observada (\hat{H}_o), índice de fixação (\hat{f}), estatísticas F de Wrigth, o parâmetro \hat{F}_{ST} , e dendrogramas obtidos com os diferentes tipos de marcadores (SSR, RAPD e isoenzimas) em diversas abordagens. A taxa de cruzamento (\hat{f}) obtida com os diferentes tipos de marcadores foram congruentes, sendo que a menor foi encontrada com isoenzimas utilizando progênies ($\hat{f}_m=0,835$),

e a maior obtida com marcadores SSR ($\hat{t}_a=1,07$) utilizando indivíduos adultos. Levando a concluir que a espécie provavelmente é alógama ou com tendência para alogamia. Com relação à estrutura genética e o fluxo gênico as estimativas não foram inteiramente concordantes sendo que a menor estimativa foi obtida com marcadores isoenzimáticos ($\hat{q}_p=0,154$ e a maior com RAPD ($\hat{f}_{ST}=0,328$). Porém é importante ressaltar, que estas estimativas não são diretamente comparáveis, pois são obtidas de maneiras diferentes e com dados de progênies ou adultos. Quanto à estruturação da variabilidade visualizada através de agrupamentos, observou-se que os dendrogramas apresentam um padrão concordante sendo que a maior sensibilidade (maior amplitude de distâncias) foi obtida com os dados de SSR. Além disso, foram feitas correlações entre as matrizes de distância genética e geográficas, entre as matrizes de distâncias genéticas (correlações entre os diferentes marcadores) e entre distâncias genéticas e padrões fenotípicos (dados de progênies, árvores, frutos e solo) em populações de *E. dysenterica*. As correlações foram elevadas e significativas a 5% de probabilidade. A maior correlação entre matrizes de distâncias genéticas e geográfica foi a obtida com marcadores RAPD (matrizes genotipadas através das progênies - M/P) ($r=0,919$) e a maior correlação entre as matrizes de distâncias genéticas foi a obtida entre SSR e RAPD (M/P) ($r=0,887$). Pode-se concluir que existe congruência entre os parâmetros populacionais obtidos com diferentes tipos de marcadores, embora existam algumas discrepâncias como é o caso da estimativa de fluxo gênico com marcadores isoenzimáticos. Foi verificado também alta correlação entre as matrizes de distância obtida com diferentes tipos de marcadores genéticos, indicando que todos eles amostram o genoma. Apesar dos marcadores possuírem diferentes naturezas todos são igualmente informativos para estudos populacionais.

Summary

A knowledge of the variability of a species is fundamental for developing genetic conservation strategies. Among the most important properties required for conservation are: a knowledge about the amount of gene flow, the reproductive system and the diversity and genetic structure of the populations. Several analytical procedures were developed for this purpose. The introduction of molecular markers

revolutionized the study of genetics of populations initially with isozymes. Recently enormous progresses have been achieved with the application of technologies based on DNA. Molecular markers based on DNA (RAPD, RFLP, SSR, SCAR, AFLP, VNTR and other) allowed a wide coverage of the genome and became a powerful tool for studies of the genetics of populations. Codominant markers (eg: isozymes and SSR) have been widely used as they are more informative than dominant markers (RAPD, AFLP and VNTR). The objective of this work was to obtain several population parameter estimates for comparing the different types of molecular markers. The estimates related to the reproductive system (\hat{f}), expected (\hat{H}_e) and observed heterozigotities (\hat{H}_o), Wrigth's F statistics were compared and dendrograms obtained with the different types of markers (SSR, RAPD and isoenzymes). The outcrossing rate (\hat{f}) obtained were reasonably congruent. The smallest value was found with isoenzymes using progenies (P) ($\hat{f}_m=0.835$), and the largest with SSR ($\hat{f}_a=1.07$) using adult individuals (M). This lead to the conclusion that the species is probably allogamic or with tendency to allogamic. Concerning the genetic structure and gene flow, estimates were not entirely congruent and the smallest estimate was obtained with isozymes markers ($\hat{q}_p=0.154$ and the largest with RAPD $\hat{f}_{ST}=0.328$). It is important to point out, that these estimates are not directly comparable, due to differences in the underlying model and the source of the data (progenies or adults). The structuring of the variability as visualized in the dendrograms showed a good pattern of congruence. The largest sensibility (larger width of distances) was obtained with the SSR data. Correlations between the genetic and geographical distance matrices, among the matrices of genetic distances (correlations between the different markers) and between genetic distances and phenotypic patterns (quantitative data of progenies, adult trees, fruits and soil) were also obtained. All correlations were high and significant. The largest correlation among the genetic and geographical distance matrices was obtained with RAPD markers (progeny data - M/P); ($r=0.919$). The largest r value was found between the genetic distance matrices using SSR and RAPD (M/P) ($r=0.887$). It could be concluded that the different procedures and markers used lead to estimated with reasonable congruence. Some discrepancies however were found estimates specially of gene flow obtained with isozymes markers. It was also verified that the high correlation

among the genetic distance matrices obtained with different types of genetic markers, indicated that all sampled well the genome under study. Despite the intrinsic difference among the markers used all were sufficiently and congruently informative for population studies, with some exception as pointed out.

6.1 Introdução

Os marcadores moleculares possuem diversas aplicações práticas em estudos de estrutura genética populacional e estas consistem em quantificar a variabilidade genética, descrever como ela se distribui nas populações e como esta pode ser manipulada (Robinson, 1998).

A introdução dos marcadores moleculares revolucionou a genética de populações na década de 50, com a técnica de isoenzimas e recentemente tem conseguido enormes avanços com a aplicação de tecnologias baseadas em DNA. Os marcadores moleculares baseados em DNA são: RAPD, RFLP, SSR, SCAR, AFLP, VNTR e outros.

O surgimento dos marcadores moleculares possibilitou a ampliação do conhecimento sobre populações de espécies nativas, sendo possível calcular parâmetros até então desconhecidos e que têm sido de fundamental importância para a conservação e o pré-melhoramento destas espécies.

Como se sabe, existem diversos tipos de marcadores, pertencentes a duas classes: os dominantes (RAPD, AFLP, VNTR, SCAR) e os codominantes (isoenzimas, microsatélites, RFLP). Os marcadores dominantes são menos informativos, pois não permitem distinguir o fenótipo heterozigótico do homozigótico o que dificulta sua utilização no cálculo de estimativas populacionais, que dependam de frequências alélicas.

Os primeiros marcadores utilizados foram as isoenzimas já na década de 60 em estudos de genética de populações (Lewontin & Hubby, 1966). Estes tiveram efeito na disseminação da técnica para esse tipo de estudos. Até hoje, a maioria dos laboratórios em que se estuda a estrutura genética de populações tem utilizado este marcador, apesar de a técnica ser mais laboriosa que a do RAPD. Mesmo assim, as isoenzimas ainda são bem vistas pelos geneticistas pelo fato de serem marcadores codominantes. Atualmente muitos estudos têm sido feitos como os de Coelho & Telles (1997), Moraes (1997), Sebben (2001) e Telles (2000).

O primeiro marcador de DNA foi o designado por RFLP (Restriction Fragment Length Polymorphism). Tendo surgido na década de 70 em um experimento destinado à detecção de mutações de DNA em vírus (Grodzicker et al., 1974), esta técnica foi utilizada mais tarde por Botstein et al. (1980) para análise genômica. Esse marcador tornou-se uma ferramenta útil e importante em várias áreas da biologia, porém foi pouco difundido no estudo de genética de populações devido a sua morosidade, e ao grande número de plantas requerido.

Os marcadores minissátélites ou VNTR (Variable Number of Tandem Repeats) surgiram na década de 80 sendo descritos pela primeira vez no genoma humano por Jeffreys et al. (1985). Estes marcadores têm sido utilizados no melhoramento de plantas para a identificação de variedades, cultivares e clones, na análise de diversidade genética e determinação da paternidade (Dallas, 1988; Nybom & Hall, 1991 e Broun et al., 1992).

Ainda na década de 80 surgiu a técnica da PCR (Polymerase Chain Reaction) desenvolvida por Mullis & Falona (1987). Essa tecnologia provocou uma revolução na biologia, pois possibilitou o surgimento de vários marcadores baseados nesta técnica. A facilidade, rapidez e sensibilidade da PCR tornou-a uma ferramenta poderosa para estudos genéticos envolvendo grande número de indivíduos de qualquer organismo.

O maior avanço na área de marcadores moleculares ocorreu em 1990, com a idéia de utilizar *primers* menores e de seqüências arbitrárias para dirigir a reação de amplificação, podendo ser usada sem a necessidade do conhecimento desta seqüência. Essa técnica foi desenvolvida independentemente por dois grupos nos EUA. Williams et al. (1990) patentearam a técnica como RAPD (Random Amplified Polymorphic DNA). Welsh & McClelland (1990), na mesma época, propuseram a mesma técnica com o nome de AP-PCR (Arbitrarily Primed - Polymerase Chain Reaction). Atualmente essa técnica tem sido amplamente difundida no estudo de populações naturais, como os estudos de Fisher et al. (2000), Sun & Wong (2001), Hangen et al. (2001), entre outros.

Diversos modelos estatísticos têm sido desenvolvidos para estimar parâmetros genéticos populacionais com a utilização de marcadores dominantes (Excoffier et al., 1992 e Lynch & Milligan, 1994), porém o uso destes modelos ainda é limitado pelos pressupostos que apresentam. A metodologia desenvolvida por Excoffier et al. (1992) é uma alternativa para o estudo da estrutura genética de

populações de uma única espécie, através de uma metodologia conhecida como análise molecular de variância (AMOVA).

A tentativa de utilizar marcadores dominantes deve-se às vantagens de técnicas como RAPD, as quais são de grande simplicidade. De modo contrário, os marcadores codominantes apresentam desvantagens tais como: a limitação do número de locos, no caso dos marcadores isoenzimáticos, amostrando pouco o genoma; a necessidade de se dispôr de um banco prévio de sondas radioativamente marcadas, no caso do RFLP, além da morosidade destas técnicas. No caso dos marcadores baseados no polimorfismo de microssatélites, sabe-se que eles precisam ser desenvolvidos para as pesquisas com espécies nativas, embora já se disponha dos *primers* necessários para análises nas espécies economicamente mais importantes.

Os marcadores microssatélites ou SSR (Simple Sequence Repeats) foram descobertas por Litt & Luty (1989) e consistem na amplificação via PCR de seqüências de 1 a 6 pb repetidas em tandem. Estes locos são altamente polimórficos constituindo a classe mais polimórfica dos marcadores até então conhecidos. Devido a estas vantagens, os marcadores microssatélites (SSR) têm sido altamente difundidos nos estudos de população naturais (Collevatti et al., 1999; Daynadan et al., 1997; Danquah et al., 2002; Lopes et al., 2002; Ramakrishnan et al. 2002 e Lefort et al., 2002), pois são altamente polimórficos quando comparados com outras classes de marcadores. O marcador SSR tem sido muito utilizado como ferramenta para responder a diversas questões sobre a genética de populações, o fluxo gênico e a análise de paternidade (Wright & Bentzen, 1994). Entretanto, avanços na utilização de microssatélites têm sido dificultados devido ao custo elevado e ao tempo para o desenvolvimento de *primers* específicos para cada espécie nativa.

O objetivo deste trabalho foi gerar um grande número de estimativas de parâmetros populacionais com a finalidade de comparar os diferentes tipos de marcadores moleculares. Foram utilizados para comparação das estimativas utilizadas para caracterização e estrutura genética, e as relativas ao fluxo gênico e ao sistema reprodutivo em 10 populações de *E. dysenterica* geradas por diferentes marcadores moleculares (SSR, RAPD, isoenzimas), com diferentes abordagens. Além disso, foram comparadas as correlações entre as matrizes de distâncias genéticas e geográficas, entre as matrizes de distâncias genéticas (correlações

entre os diferentes marcadores) e entre distâncias genéticas e padrões fenotípicos (dados de progênies, árvores, frutos e solo).

6.2 Material e Métodos

6.2.1 Material Vegetal

O material de estudo foi coletado em dez locais da região sudeste do Estado de Goiás, totalizando dez populações, representadas por 114 matrizes. Para a genotipagem das plantas coletou-se material vegetal (folhas) bem como sementes (progênies maternas) em cada uma das 114 matrizes de cagaiteira. As localidades onde foram coletadas as populações, podem ser observadas na tabela 13 item 5.2.1.

6.2.2 Análise Estatística dos Dados

A comparação entre os tipos de marcadores foi feita através de estimativas de parâmetros genéticos, conforme segue:

a) Comparação entre o Sistema Reprodutivo obtido pelos Diversos Tipos de Marcadores

As estimativas da taxa de cruzamento, utilizadas para comparação com dados de duas gerações, foram obtidas conforme descrito no item 5.2.3.1. As referentes aos dados de uma só geração (matrizes) foram calculadas conforme o item 5.2.3.2.

b) Comparação da Estrutura Genética e Fluxo Gênico

Os parâmetros utilizados para comparação da estrutura genética e do fluxo gênico foram estimados como descrito nos itens 3.2.6, 4.3 e 5.2.5, constantes nos capítulos 3, 4, 5 respectivamente.

c) Correlação entre Distâncias Genéticas e Geográficas

Com o propósito de analisar os padrões de variação espacial em um contexto multivariado, foi obtida a estimativa do coeficiente de correlação de Pearson (r) entre as matrizes de distâncias genéticas de Nei e as respectivas distâncias geográficas entre as populações. O estimador do coeficiente de correlação simples, no caso entre duas matrizes X e Y , é dado por:

$$r = \frac{\sum_{i,j=1}^n x_{ij} y_{ij}}{\sqrt{\sum_{i,j=1}^n x_{ij}^2 \sum_{i,j=1}^n y_{ij}^2}}$$

onde $x_{ij} = X_{in} - \bar{X}$ e $y_{ij} = Y_{in} - \bar{Y}$, em que x_{ij} e y_{ij} representam distâncias genéticas e geográficas, respectivamente. A significância dessa correlação matricial não pode ser testada por testes estatísticos usuais, por apresentar problemas de independência entre elementos nas matrizes. Por isso utilizou-se a estatística Z de Mantel (1967), (Manly, 1991). O valor Z de Mantel 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 (no caso, como exemplo, as matrizes de distâncias geográficas e genéticas). A significância de Z pode ser obtida comparando-se esse valor observado com valores de uma distribuição nula, construída recalculando os valores de Z diversas vezes, aleatorizando, em cada uma delas, a ordem dos elementos de uma das matrizes (Manly, 1991). A estatística Z possui uma relação monotônica com o r de Pearson calculado entre matrizes (correlação matricial), de modo que ela pode ser utilizada para testar a significância de r (Manly, 1986 a, b). Neste trabalho, 9999 permutações aleatórias foram utilizadas para testar a significância das correlações matriciais.

d) Correlação entre Matrizes de Distâncias Genéticas

As correlações entre matrizes de distâncias genéticas foram obtidas conforme descrito item c. Neste caso porém, X_{ij} e Y_{ij} são distâncias genéticas calculadas com base em dois tipos de marcadores.

e) Correlação entre Padrões Genéticos, Fenotípicos e Edáficos

Os dados morfológicos referentes às características de árvores, frutos e progênes, assim como edáficos foram obtidos por Silva (1999) nas mesmas populações deste estudo e nas áreas onde foram amostradas. Os dados referentes a isoenzimas foram obtidos por Telles (2000) também considerando as mesmas populações deste trabalho. Assim sendo, as matrizes de distâncias genéticas foram calculadas para diferentes marcadores (RAPD, enzima e SSR) com dados de progênes e matrizes. A partir dos dados originais de Silva (1999) foi calculada a distância Euclidiana entre as 10 populações. Essas matrizes foram comparadas entre si e com as matrizes de distâncias genéticas e geográfica, através do teste de Mantel.

Os dados edáficos foram obtidos em duas profundidades (0-20 cm e 20-40 cm) nas dez localidades, sendo mensurados os seguintes componentes do solo: P, K, Ca, Mg, Al, H^+Al , MO, pH, CTC, Sat. Bases, Sat. Al, Zn, Cu, Fé, Mn, Argila e Areia (Silva, 1999).

Os dados fenotípicos coletados nas árvores foram: altura de planta, circunferência de colo, circunferência do(s) tronco(s) a 1,30 m, diâmetro de projeção da copa nos sentidos norte-sul e leste-oeste, área basal e área do tronco a 1,30 m do solo. Foram amostradas 10 a 12 árvores de cada população, exceto das populações 7 e 9, das quais a amostra conteve oito árvores (Silva, 1999).

As variáveis fenotípicas dos frutos foram obtidas como médias de 12 frutos por árvore, anotando-se os dados individualizados por fruto. Foram obtidas as seguintes medidas dos frutos: peso total, diâmetro transversal, cor da casca, número e peso de sementes/fruto, peso da polpa e casca (Silva, 1999).

As progênes foram representadas por, no máximo, 10 plantas descendentes das árvores da população original. As sementes foram plantadas no viveiro localizado na Escola de Agronomia da Universidade Federal de Goiás - UFG, Goiânia-GO, em delineamento experimental, para aleatorizar os efeitos ambientais. Em cada planta foram coletados os seguintes dados: velocidade de emergência,

porcentagem de emergência, altura da plântula, diâmetro do colo da planta. Os dados de altura e diâmetro do colo da planta foram tomados em duas etapas, sendo a primeira coleta com 30 dias após a emergência de cada plântula e a segunda realizada aos 250 dias após a semeadura (Silva, 1999).

As relações entre as matrizes foram analisadas pela correlação parcial, pois, os coeficientes de correlação simples podem produzir equívocos de interpretação a respeito da relação que há entre duas variáveis analisadas (Cruz & Regazzi, 1997), podendo não ser uma medida real de causa e efeito. Assim, um alto ou baixo coeficiente de correlação entre duas variáveis pode ser o resultado do efeito que exerce uma terceira (ou um grupo de matrizes) sobre aquelas duas.

O coeficiente de correlação parcial entre duas variáveis X e Y, controlando-se o efeito da variável Z, é dado por:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

6.3 Resultados

6.3.1 Comparação da Taxa de Cruzamento

Como se pode observar na Tabela 21, as taxas de cruzamento (\hat{f}) obtidas com os diferentes marcadores, foram relativamente congruentes, sendo que a menor foi obtida com os marcadores isoenzimáticos utilizando dados de progênies ($\hat{t}_m = 0,835$) e a maior com marcadores SSR ($\hat{t}_a = 1,08$). É importante, ressaltar que estas estimativas são calculadas de maneira diferente e devem ser interpretadas com ressalvas.

As estimativas mais robustas são as calculadas utilizando progênies denominada de taxa de cruzamento multilocos (\hat{t}_m) (descrita por Ritland & Jain, 1980) uma vez que a taxa aparente (\hat{t}_a) é obtida apenas com dados de uma geração, e é estimada a partir do índice de fixação (\hat{f}) intrapopulacional o que torna a estimativa dependente da pressuposição de equilíbrio de Wright. As estimativas mais confiáveis foram as obtidas com dados de progênies, ou seja, a

taxa de cruzamento multilocos (\hat{t}_m), $\hat{t}_m=0,835$ com isoenzimas e $\hat{t}_m=0,937$ obtidas com RAPD. Dessa forma, entende-se que a espécie é preferencialmente alógama ou tem tendência a alogamia.

Tabela 21. Estimativas da taxa de cruzamento para 10 populações naturais de *Eugenia dysenterica* baseadas em diferentes marcadores.

Tipo de marcador	Valor da estimativa		Limites do IC	
	\hat{t}_a	\hat{t}_m	Inferior	Superior
RAPD (M)	0,867		0,672	1,062
RAPD (P)		0,937	0,920	0,954
Microsatélites (M)	1,08		0,826	1,328
Isoenzimas (P) (Telles, 2000)		0,835	0,762	0,908

M: dados de plantas matrizes (adultas)

P: dados de progênies

IC: Intervalo de confiança a 95%

6.3.2 Caracterização, Estrutura Genética e Fluxo Gênico

Em relação à caracterização genética foi verificado que o marcador SSR fornece estimativas mais altas de heterozigosidade observada (\hat{H}_o), heterozigosidade esperada (\hat{H}_e), com os valores iguais a 0,458 e 0,442 respectivamente. Conforme a tabela 22, os marcadores que fornecem as estimativas mais baixas de (\hat{H}_e) e (\hat{H}_o), foram os RAPD com progênies sendo (0,195 e 0,182, respectivamente). Isto pode ser explicado como consequência da natureza do marcador. Os microssatélites produziram em média 10,4 alelos por locos, para as isoenzimas esta média foi de três e para RAPD, apenas dois alelos por loco. Já a maior estimativa do índice de fixação foi obtida com marcadores isoenzimáticos cujo valor foi de $\hat{f}=0,337$, estimativa esta não concordante com as obtidas com RAPD e SSR.

Tabela 22. Estimativas de parâmetros genéticos em 10 populações de *Eugenia dysenterica* obtidas com diferente tipos de marcadores, sendo H_o : heteroziguidade observada; H_e : heteroziguidade esperada e f : índice de fixação.

Pop.	SSR (M)			RAPD (M/P)			Isoenzimas (P)		
	\hat{H}_o	\hat{H}_e	\hat{f}	\hat{H}_o	\hat{H}_e	\hat{f}	\hat{H}_o	\hat{H}_e	\hat{f}
1	0,253	0,276	0,088	0,160	0,194	0,181	0,163	0,231	0,295
2	0,449	0,403	-0,123	0,114	0,220	0,490	0,160	0,268	0,403
3	0,366	0,394	0,076	0,154	0,158	0,028	0,175	0,223	0,217
4	0,453	0,408	-0,117	0,173	0,199	0,138	0,215	0,285	0,248
5	0,497	0,404	-0,245	0,206	0,204	-0,130	0,136	0,225	0,401
6	0,411	0,374	-0,106	0,221	0,201	-0,105	0,163	0,265	0,384
7	0,450	0,389	-0,183	0,250	0,236	-0,064	0,224	0,353	0,311
8	0,563	0,438	-0,316	0,219	0,189	-0,164	0,195	0,263	0,258
9	0,541	0,670	0,197	0,175	0,163	-0,073	0,214	0,377	0,435
10	0,599	0,667	0,105	0,150	0,188	0,213	0,213	0,334	0,365
Média	0,458	0,442	-0,037	0,182	0,195	0,071	0,188	0,282	0,337

M: dados de plantas matrizes (adultos)

M/P: dados de plantas matrizes genotipadas através das progênies

P: dados de progênies

Conforme a Tabela 23, pode-se observar que as estimativas relativas a diversidade entre populações (\hat{f}_{ST} , \hat{F}_{ST} e \hat{q}_p) foram mais altas quando obtidas a partir de marcador RAPD e SSR do que a obtida com isoenzimas ($\hat{q}_p=0,156$). A estimativa mais alta foi obtida com marcadores dominantes (RAPD), visto que $\hat{f}_{ST}=0,328$. Porém, estas estimativas foram calculadas de diferentes maneiras; desta forma esta comparação tem que ser feita com ressalvas.

As estimativas obtidas com marcadores microssátelites (\hat{R}_{ST} e \hat{F}_{ST}) foram congruentes sendo iguais a 0,268 e 0,250, respectivamente. Aquelas obtidas com marcadores RAPD (\hat{f}_{ST}) foram relativamente concordantes sendo o $\hat{f}_{ST}=0,2703$ calculado com as matrizes e $\hat{f}_{ST}=0,328$ quando calculado com as progênies. Esta variação pode ter sido causada pelo fato da amostragem ser 7 vezes maior no segundo caso.

Tabela 23. Estimativa das estatísticas F de Wright, R_{ST} e do número de migrantes por geração (Nm) para 10 populações naturais de *Eugenia dysenterica*

Tipo de marcador	Valor da estimativa	Limites de IC		p	Nm
		Inferior	Superior		
RAPD (M) \hat{f}_{ST}	0,2703	-	-	<0,001	0,575
RAPD (P) \hat{f}_{ST}	0,3280	-	-	<0,001	0,512
RAPD (M/P) \hat{F}_{ST}	0,3182	0,3873	0,2460	-	0,536
Microsatélites (M) \hat{F}_{ST}	0,250	0,3588	0,2450	-	0,750
Microsatélites (M) \hat{R}_{ST}	0,269	0,3485	0,1941	-	0,680
Isoenzimas (P) \hat{q}_p	0,154	0.182	0.110		1,370

M: dados de plantas matrizes (adultos)

M/P: dados de plantas matrizes genotipadas através das progênies

P: dados de progênies

IC: Intervalo de confiança a 95%

Verifica-se na Figura 14 o padrão da estruturação genética das 10 populações de cagaiteira, a partir dos dendrogramas construídos com as distâncias genéticas de Nei (1978) e com as similaridades genéticas de Jaccard. Pode-se verificar grande concordância entre eles, na medida em que os quatro dendrogramas formam 2 grupos distintos, um formado pela população 9 e 10, e outro pelas populações de 1 a 8. Para os dados de marcadores SSR, as distâncias genéticas de Nei (1978) variaram de 0,046 a 0,497. Com os dados de progênies com RAPD estas variaram de 0,005 a 0,229 e a para dados isoenzimáticos 0,006 a 0,243. Com RAPD, o índice de similaridade genética de Jaccard variou de 0,455 a 0,653. A sensibilidade desta técnica pode ser medida através da amplitude das distâncias e similaridades encontradas, e pode-se verificar que a maior amplitude obtida foi de 0,451, com marcadores SSR, enquanto que as demais amplitudes ficaram em torno de 0,200. As correlações cofenéticas do agrupamento UPGMA relativas a SSR, RAPD com progênies, RAPD e isoenzimas foram elevadas atingindo os valores 0,943; 0,977; 0,915 e 0,914, respectivamente.

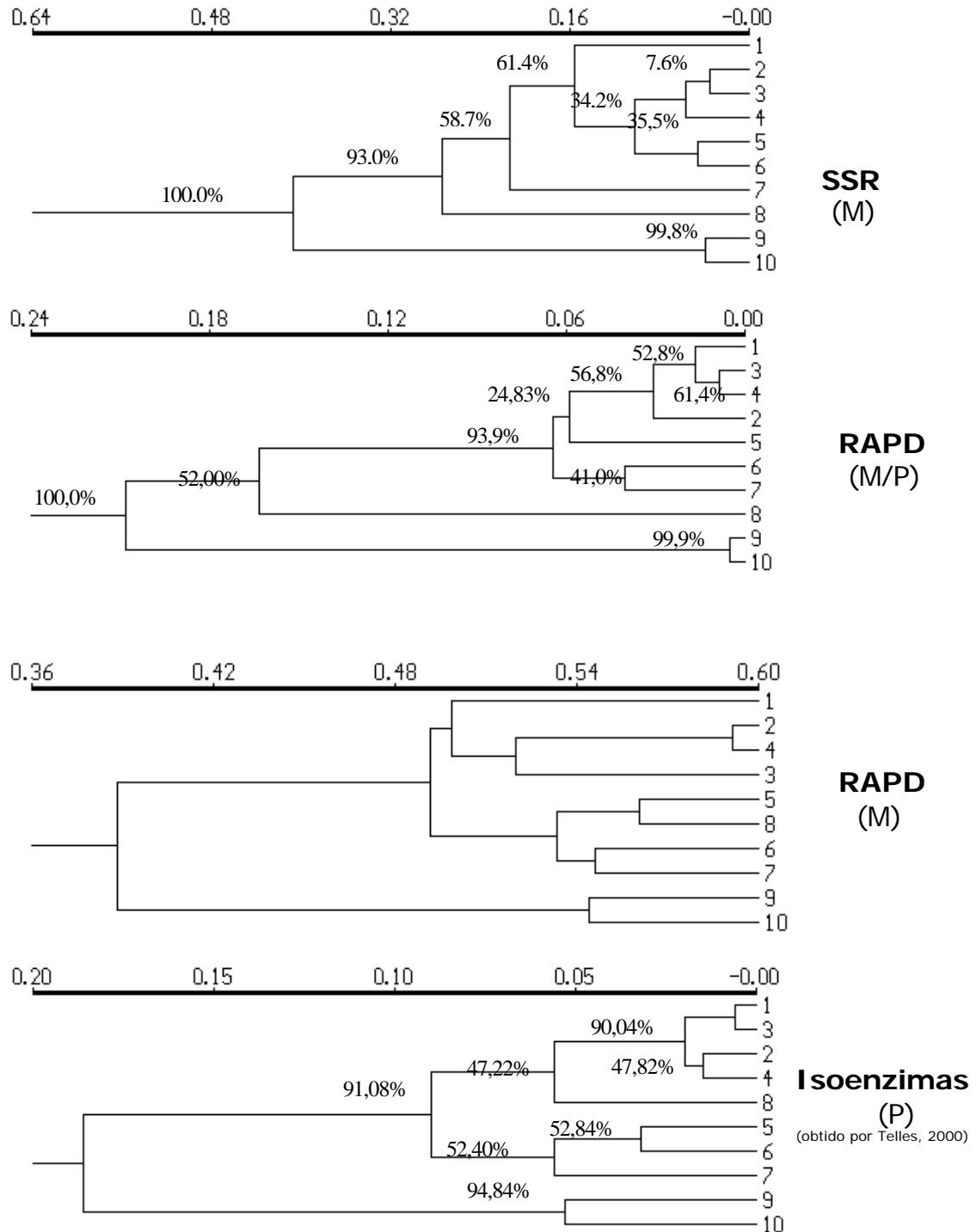


Figura 14 - Padrão de estruturação genética de 10 populações de cagaiteira, definido pelo agrupamento UPGMA, com base distâncias genéticas de Nei (1972, 1978) e índice de similaridade de Jaccard. M: dados de plantas matrizes (adultos); M/P: dados de plantas matrizes genotipadas através das progênies e P: dados de progênies

6.3.3 Correlação entre distâncias genéticas e geográficas

Por meio da análise de agrupamento UPGMA, como visto anteriormente no capítulo 5, foi detectado um padrão espacial implícito que pode ser confirmado pela análise da correlação entre as distâncias genéticas e as distâncias geográficas. As maiores correlações matriciais foram obtidas com os marcadores RAPD utilizando dados de plantas-mãe e progênies M/P e com marcadores SSR, sendo $r=0,919$ e $r=0,871$, respectivamente. Estas foram significativas a 1 % de probabilidade, pelo teste de Mantel (Figura 15).

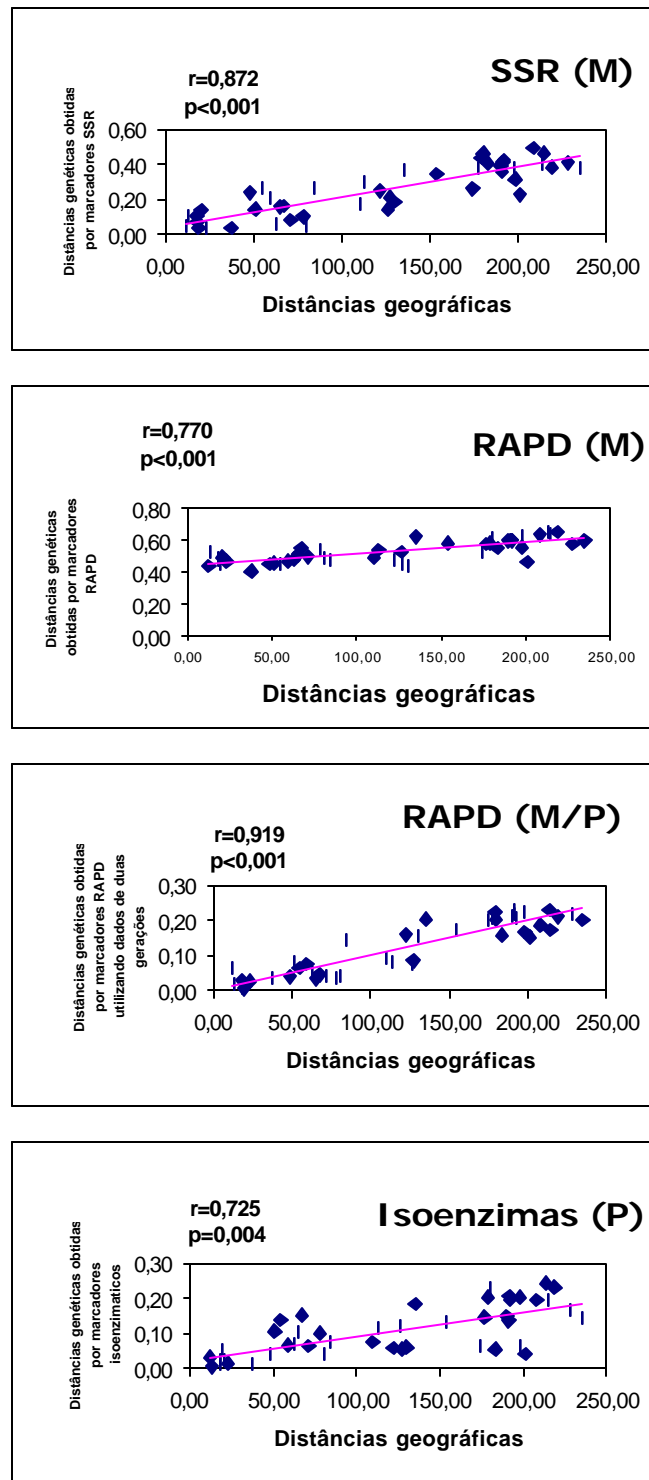


Figura 15 - Representação correlação (r) entre distâncias geográficas e distâncias genéticas obtidas por diferentes tipos de marcadores moleculares. Significância de r dada pela probabilidade p. M: plantas-mãe; p; progênies.

6.3.4 Correlação entre Matrizes de Distâncias Genéticas

A relativa congruência detectada entre as distâncias geográficas e as distâncias genéticas, calculadas com os diferentes marcadores (Figura 14), foi confirmada correlacionando as distâncias genéticas entre si, duas a duas (tabela 24). Todos os coeficientes r dados nesta tabela forma significativos a 1% de probabilidade (teste de Mantel). Sobressaem ligeiramente as correlações entre as distâncias obtidas com SSR e RAPD (M/P) bem como as de isoenzimas e RAPD (M).

Tabela 24. Correlação linear simples entre as distâncias genéticas, obtidas com os três marcadores. Os números entre parênteses indicam a significância de cada correlação.

	Isoenzima (P)	RAPD (M)	RAPD (M/P)	SSR (M)
Isoenzima (P)	-	-	-	-
RAPD (M)	0,847 (<0,001)	-	-	-
RAPD (M/P)	0,738 (0,005)	0,743 (0,007)	-	-
SSR (M)	0,803 (<0,001)	0,812 (0,002)	0,887 (<0,001)	-

M: dados de plantas matrizes (adultos)

M/P: dados de plantas matrizes genotipadas através das progênes

P: dados de progênes

6.3.5 Correlação entre Padrões Genéticos, Edáficos e Morfológicos

Ao correlacionar as distâncias genéticas com as caracteres edáficos (solo), os valores obtidos foram baixos, sendo que para as matrizes geradas por RAPD com M/P e SSR, obtiveram-se os valores de $r=0,508$ e $r=0,503$, respectivamente. Somente a correlação obtida com a matriz de distâncias genéticas (geradas por marcadores isoenzimáticos) não foi significativa a 5% de probabilidade, Tabela 25.

Para as distâncias baseadas em caracteres de frutos e as distâncias genéticas, todas as correlações lineares simples foram baixas e não significativas ao nível de 5%. As correlações entre as distâncias obtidas dos caracteres morfológicos de progênes e as distâncias genéticas revelaram-se fracas, sendo os maiores valores

aqueles obtidos pelas matrizes geradas por RAPD M/P e SSR, com correlações $r=0,425$ e $r=0,393$, respectivamente. Todas estas correlações foram significativas à 5%, Tabela 25.

Numa visão geral percebe-se conforme, já enfatizado no item anterior, que as maiores correlações foram obtidas com as distâncias geográficas, sendo em seguida, em ordem decrescente, as detectadas com os dados morfológicos das árvores, as de propriedades do solo, as características de progênies e por fim, as de caracteres dos frutos. Estas últimas foram nulas.

As correlações mencionadas, sendo simples, devem ser interpretadas com cautela, já que as variáveis podem estar correlacionadas entre si.

Tabela 25. Correlação linear simples entre as distâncias genéticas, obtidas com os três diferentes marcadores e distâncias geográficas, edáficas, fenotípicas, a partir de dados de árvores, dos frutos e das progênies. Os números entre parênteses indicam a significância de cada correlação.

	Isoenzima*(P)	RAPD(M)	RAPD(M/P)	SSR(M)
Geográfica	0,725 (0,004)	0,770 (<0,001)	0,919 (<0,001)	0,872 (<0,001)
Solo**	0,408 (0,102)	0,508 (0,017)	0,503 (0,006)	0,437 (0,042)
Árvores**	0,585 (0,018)	0,674 (0,007)	0,557 (0,024)	0,555 (0,019)
Frutos**	-0,140 (0,234)	0,139 (0,236)	0,084 (0,317)	0,002 (0,480)
Progênies**	0,346 (0,043)	0,316 (0,056)	0,425 (0,015)	0,393 (0,019)

*Dados isoenzimáticos obtidos por Telles (2000)

**Dados morfológicos e edáficos obtidos por Silva (1998)

M: dados de plantas matrizes (adultos)

M/P: dados de plantas matrizes genotipadas através das progênies

P: dados de progênies

Nota-se agora que a correlação parcial entre as distâncias genéticas e as geográficas permaneceram altas e significativas, mesmo quando mantida constante a matriz de dados das árvores. O mesmo acontece quando se fixou o efeito de solo e de progênies (Tabela 26) reforçando a idéia da grande influência da distribuição geográfica no padrão de diferenciação das frequências alélicas.

Tabela 26. Correlações parciais entre as distâncias genéticas e geográficas, fixando a matriz de dados de árvores, de solos e das progênies.

	Isoenzima* (P)	RAPD (M)	RAPD (M/P)	SSR (M)
$r_{\text{geo.gen}}$	0,725	0,770	0,919	0,872
$r_{\text{geo.gen_arv}}^{**}$	0,582	0,628	0,880	0,811
$r_{\text{geo.gen_sol}}^{**}$	0,656	0,681	0,889	0,841
$r_{\text{geo.gen_prg}}^{**}$	0,684	0,741	0,905	0,845

*Dados isoenzimáticos obtidos por Telles (2000)

**Dados morfológicos e edáficos obtidos por Silva (1998)

M: dados de plantas matrizes (adultos)

M/P: dados de plantas matrizes genotipadas através das progênies

P: dados de progênies

Os resultados do teste de Mantel múltiplo, reforçam os resultados das análises do padrão espacial realizadas a partir dos dados obtidos pelos diferentes marcadores, pois também indicam que o principal fator que está influenciando a divergência genética é a distribuição geográfica das populações. Se existisse um fator seletivo estruturado no espaço atuando nestes marcadores, a correlação parcial entre as matrizes de distâncias genética e geográfica deveria ter sido menor, ao se fixar as matrizes de caracteres de árvores, solos e progênies, já que o espaço geográfico não seria o papel predominante da estrutura genética.

6.4 Discussão

As estimativas da taxa de reprodução cruzada obtidas pelos diversos tipos de marcadores em diferentes gerações foram relativamente concordantes, sendo que a menor foi obtida com os marcadores isoenzimáticos utilizando dados de progênies ($\hat{t}_m = 0,835$) e a maior com marcadores SSR ($\hat{t}_a = 1,08$). É importante, ressaltar que estas estimativas são calculadas de maneira diferentes e devem ser comparadas com ressalvas. As estimativas menos dependentes de pressuposições foram as calculadas utilizando isoenzimas ($\hat{t}_m = 0,835$) e RAPD ($\hat{t}_m = 0,937$) obtidas com RAPD ambas baseadas em dados de progênies. Fica evidente a tendência para a prática da panmixia nesta espécie.

Em relação à caracterização genética verifica-se que o marcador SSR forneceu estimativas mais altas da heterozigosidade observada (\hat{H}_o), heterozigosidade esperada (\hat{H}_e), com os valores 0,458 e 0,442 respectivamente. A heterozigosidade observada foi ligeiramente maior que a esperada o que pode ser devido a uma seleção que aumenta a frequência de heterozigotos, durante o recrutamento. O recrutamento é um acontecimento comum, que ocorre em algumas espécies tropicais, conforme detectado com isoenzimas em palmitero (Conte, 2001).

Muitos estudos têm mostrado estimativas diferentes quando obtidas com marcadores RAPD e isoenzimáticos. Isto pode ser devido ao fato do viés causado nas estimativas obtidas com marcadores dominantes (Lynch & Milligan, 1994). Isabel et al. (1995), no entanto, encontraram idêntica heterozigosidade por marcadores RAPD e por isoenzimas dentro de populações de *Picea mariana*, usando frequências alélicas estimados por RAPD. Estas foram inferidas a partir de tecido de megagametófitos haplóides das sementes das árvores ($\hat{H}_o=0,35$ e $0,34$, respectivamente).

As estimativas relativas a diversidade genética molecular obtidas entre populações (\hat{f}_{ST} , \hat{F}_{ST} e \hat{q}_p), neste estudo, foram mais elevadas para os marcadores RAPD e SSR do que as obtidas com isoenzimas ($\hat{q}_p=0,156$). As obtidas com marcadores microssátelites (\hat{R}_{ST} e \hat{F}_{ST}) foram também congruentes 0,268 e 0,250, respectivamente. As calculadas com marcadores RAPD (\hat{f}_{ST}) foram relativamente concordantes entre si e o $\hat{f}_{ST}=0,2703$ e $\hat{f}_{ST}=0,328$ quando calculado com as plantas-mãe e as progênies, respectivamente.

Bucci et al (1997) estudaram 23 locos RAPD e 16 locos isoenzimáticos em sete populações de *Pinus leucodermis*, utilizando dados de megagametófitos (tecido haplóide), verificaram que as estimativas de \hat{F}_{ST} utilizando RAPD e isoenzimas foram 0,1792 e 0,0671, respectivamente, e portanto maiores com RAPD. O grau da sensibilidade genética utilizando RAPD, portanto, foi três vezes maior do que a obtida com isoenzimas.

Ayres & Ryan (1999) também utilizaram marcadores RAPD e isoenzimáticos para estudar populações de *Wyethia reticulata* e *W. bolanderi* (Asteradeae). Empregaram 19 locos RAPD e oito isoenzimáticos. Verificaram que a estimativa da diferenciação entre populações (\hat{F}_{ST}) obtida com marcadores RAPD foi duas vezes maior que a obtida com isoenzimas (0,36 e 0,18, respectivamente).

Ciampi (1999) estudou quatro populações de *Copaifera langsdorffi* (copaíba), em matas de Galeria do Cerrado com dados de marcadores AFLP, SSR e cpDNA. A análise comparativa das técnicas de marcadores dominantes AFLP e codominantes SSR para o genoma nuclear mostrou que ambas as técnicas são altamente eficientes para estimar as proporções da variabilidade genética entre e dentro de populações de *Copaifera langsdorffi*. Um resultado importante obtido foi a forte congruência nas estimativas da decomposição da variabilidade genética entre e dentro das populações e da significância destas estimativas, com os dois tipos de marcadores.

Buso et al. (1998) estudando a estrutura genética de populações de arroz selvagem utilizou marcadores isoenzimáticos e RAPD para estimar o nível de diversidade de quatro populações de arroz selvagem da América do Sul (*Oryza glumepatula*), coletadas na floresta amazônica e rios do Oeste do Brasil. O padrão de diversidade genética entre e dentro de populações foi calculado para ambos tipos de marcadores. Com quatro locos de isoenzimas foi encontrado um $\hat{F}_{ST}=0,31$ e 156 locos de marcadores RAPD um valor $\hat{F}_{ST}=0,64$ (calculado admitindo equilíbrio em autogamia).

Jenczewski et al. (1999) utilizou marcadores RAPD e isoenzimas para avaliar 20 populações de alfafa cultivadas naturais da Espanha e obtiveram $\hat{F}_{ST}=0,0717$ e $\hat{F}_{ST}=0,0088$, respectivamente. Estes foram estatisticamente diferentes entre si.

Outros estudos mostraram que marcadores RAPD podem ser eficientes ferramentas para análise da estrutura genética de populações diplóides e revelam padrões semelhantes aos marcadores isoenzimáticos (Isabel et al. 1995, Yeh et al. 1995, Aagard et al. 1998).

Aagard et al (1998) compararam níveis de diferenciação entre populações e raças de uma conífera (*Douglas-fir*) com 36 locos de RAPD e 20 locos de isoenzimas. Os autores obtiveram estimativas concordantes da diversidade entre

populações e raças com $\hat{G}_{ST}=0,05$ e $\hat{G}_{ST}=0,05$ e entre raças $\hat{G}_{ST}=0,25$ e $\hat{G}_{ST}=0,22$, respectivamente.

Neste estudo, a estruturação da variabilidade genética, visualizada pelo agrupamento feito pelo critério UPGMA, foi congruente nos dendrogramas obtidos pelos diferentes marcadores. A sensibilidade de uma técnica molecular pode ser medida através da amplitude das distâncias e similaridades encontradas. No presente estudo verificou-se que a maior amplitude observada foi de 0,451, obtida com marcadores SSR, enquanto que as demais amplitudes ficaram próximas de 0,200.

A maiores correlações matriciais, encontradas neste estudo, entre distâncias genéticas e geográficas foram obtidas com marcadores RAPD (M/P) e com marcadores SSR (M), tendo se obtido $r=0,919$ e $r=0,871$, respectivamente, ambas significativas a 1 % de probabilidade. Porém todas as correlações obtidas entre distâncias geográficas e genéticas foram altas e significativas a 1% de probabilidade.

Quanto à taxa de cruzamento, a espécie tem clara tendência à alogamia. Em relação à estimativa de taxa de cruzamento aparente (\hat{t}_a), não houve vantagem de utilizar os marcadores SSR em relação aos RAPD (M) pois os respectivos intervalos de confiança ambos incluíram o valor $\hat{t}_a=1,00$. Utilizando dados de progênies as estimativas multilocos foram altas mas menores que 1,00, pode-se portanto dizer que as técnicas RAPD (P) e isoenzimas (P) levaram a conclusões conflitantes.

Quanto à correlação entre distâncias genéticas e geográficas, a técnica com RAPD (M/P) parece ser a mais sensível, seguida da técnica SSR (M). Novamente pode-se dizer que, com relação a este fenômeno, não houve vantagem nítida do método que emprega microssátelites em comparação com de RAPD (M/P).

6.5 Conclusões

Os resultados apresentados permitem concluir que:

- a) Existe relativa congruência entre as estimativas dos parâmetros populacionais obtidos com diferentes tipos de marcadores, embora exista algumas discrepâncias como é o caso da estimativa de fluxo gênico com marcadores isoenzimáticos.
- b) Foi verificada alta correlação entre as matrizes de distância obtida com diferentes tipos de marcadores genéticos, indicando que todos eles amostram bem o genoma. Apesar dos marcadores possuírem diferentes naturezas todos são igualmente informativos para estudos populacionais

7 CONCLUSÕES GERAIS

- Existe potencial de transferibilidade de pares de *primers* de microsátelite para o mesmo gênero e outros gêneros, como no caso de *Eucalyptus* ssp. para *E. dysenterica*.
- As populações de cagaiteira, do sudeste do Estado de Goiás, apresentam uma alta variabilidade entre populações detectadas por marcadores dominantes e codominantes.
- A taxa de fecundação cruzada, obtida com dados de progênies, sugere um sistema de reprodução da espécie, com tendência a alogamia ($t_m=0,93$).
- Existe congruência entre as metodologias de caracterização genética populacional por marcadores dominantes (RAPD) através de progênies com dados de marcadores codominantes (SSR).
- A metodologia de análise da estrutura genética utilizando marcadores dominantes, no caso RAPD (analisando dados de duas gerações) é viável e congruente aos dados de marcadores codominantes (no caso SSR) e isoenzimáticos.
- A análise utilizando marcadores dominantes (em dados de duas gerações) é muito trabalhosa, pois são analisados de sete a dez vezes mais indivíduos. Porém, os resultados das estimativas populacionais são concordantes com dados de plantas matrizes utilizando marcadores codominantes.

- As estimativas de estrutura genética populacional obtidas com diferentes marcadores foram congruentes. Porém, os marcadores SSR são mais sensíveis para cálculo das estimativas populacionais.
- Correlações entre matrizes de distâncias genéticas obtidas por marcadores SSR, RAPD e isoenzimas foram fortes e positivas.
- Correlação entre matrizes de distâncias genéticas e geográficas obtidas com diferentes marcadores sugerem que a variabilidade destas populações estão estruturadas no espaço. E esta variabilidade provavelmente tenha sido gerada por deriva e fluxo a pequenas distâncias.

REFERÊNCIAS BIBLIOGRÁFICAS

- AAGAARD, J.E.; KRUTOVSKI, K.V.; STRAUSS, S.H. RAPD and allozymes exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir. **Heredity**, v.81, p.69-78, 1998.
- AKKAYA, M.S.; SHOEMAKER, R.C.; SPECHT, J.E.; BHAGWAT, A.A.; CREGAN, P.B. Integration of simple sequence repeat DNA markers into a soybean linkage map. **Crop Science**, v.35, n.5, p.1439-1445, 1995.
- ALMEIDA, S.P.; SILVA, J.A.; RIBEIRO, J.F. **Aproveitamento alimentar de espécies nativas dos cerrados**: araticul, barú, cagaita e jatobá. Planaltina: EMBRAPA, CPAC, 1987. 83p.
- ARAUJO, T.C.C. Estrutura da variabilidade genética e estrutura populacional de Caryocar brasiliense Camb. no Estado de Goiás utilizando marcadores RAPD. Goiânia, 2001. 87p. Dissertação (M.S.) – Instituto de Ciências Biológicas, Universidade Federal de Goiás.
- AVISE, J.C. Molecular markers. **Natural history and evolution**. New York: Chapman & Hall, 1994. 511p.
- AYRES, D.R.; RYAN, F.J. Genetic diversity and structure of the narrow endemic *Wyethia reticulata* and its congener *W. bolanderi* (Asteraceae) using RAPD and allozyme techniques. **American Journal of Botany**, v.86, p.344-353, 1999.
- BELL, C.J.; ECKER, J.R. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. **Genomics**, v.19, n.1, p.137-144, 1994.
- BIELAWSKI, J.P.; PUMO, D.E. Randomly amplified polymorphic DNA (RAPD) analysis of Atlantic coast striped bass. **American Journal of Botany**, v.78, p.32-40, 1997.
- BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphism. **American Journal Human Genetics**, v.32, p.314-331, 1980.

- BRONDANI, R.P.V.; BRONDANI, C.; TARCHINI, R.; GRATTAPAGLIA, D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics**, v.97, p.816-827, 1998.
- BROUN, P.; GANAL, M.W.; TANKSLEY, S.D. Telomeric arrays display high levels of heritable polymorphism among closely related plant varieties. **Proceedings of the National Academy of Sciences of the United States of America**, v.89, p.1354-1357, 1992.
- BROWN, S.M.; HOPKINS, M.S.; MITCHELL, S.E.; SENIOR, M.L.; WANG, T.Y.; DUNCAN, R.R.; GONZALEZCANDELAS, F.; KRESOVICH, S. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L) Moench]. **Theoretical and Applied Genetics**, v.93, n.1-2, p.190-198, 1996.
- BUCCI, G.; VENDRAMIN, G.G.; LELLI, L.; VICARÁIO, F. Assessing the genetic divergence of *Pinus leucodermis* Ant endangered populations: use of molecular markers for conservation purposes. **Theoretical and Applied Genetics**, n.95, p.1138-1146, 1997.
- BUSO, G.S.; RANGEL, P.H.; FERREIRA, M.E. Analysis of genetic variability of South American wild rice populations (*Oryza glumaepatula*) with isozymes and RAPD markers. **Molecular Ecology**, v.7, p.107-117, 1998.
- BYRNE, M.; MARQUEZ-GARCIA, M.I.; UREN, T.; SMITH, D.S.; MORAN, G.F. Conservation and genetic diversity of microsatellite loci, in the genus *Eucalipytus*. **Australian Journal of Botany**, v.44, p.331-341, 1996.
- CHALMERS, K.J.; WAUGH, R.; SPRENT, J.I.; SIMONS, A.J.; POWELL, W. Detection of genetic variation between and within population of *Gliricidia sepium* and *G. maculata* using RAPD markers. **Heredity**, v.69, p.465-472, 1992.
- CHARLESWORTH, B.; SIIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, v.371, p.215-220, 1995.
- CHASE, M.; KESSEL, R.; BAWA, K. Microsatellite markers for populations and conservation genetics of tropical trees. **American Journal of Botany**, v.83, p.51-57, 1996.
- CHAVES, L.J.; NAVES, R.V. **O cerrado do Brasil**: uma fonte potencial de recursos genéticos In: ENCONTRO DE GENÉTICA E MELHORAMENTO, 15., Piracicaba, 1998. **Anais**. Piracicaba: ESALQ, Depto Genética, 1998. p.74-86.

- CIAMPI, A. Desenvolvimento e utilização de marcadores microssatélites, AFLP e sequenciamento de cpDNA, no estudo da estrutura genética e parentesco em populações de Copaíba (*Copaifera langsdorffii*) em matas de galeria no cerrado. Botucatu, 1999. 204p. Tese (Doutorado) – Instituto de Biociências, Universidade Estadual Paulista.
- CIPRIANI, G.; LOT, G.; HUANG, W.G.; MARRAZZO, M.T.; PETERLUNGER, E.; TESTOLIN, R. AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. **Theoretical Applied Genetics**, v.99, p.65-72, 1999.
- COCKERHAM, C.C. Variance of gene frequencies. **Evolution**, v.23, p.72-84, 1969.
- COELHO, A.S.G. **Dboot**: avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores (software). Goiânia: UFG, Instituto de Ciências Biológicas, Laboratório de Genética Vegetal, 2000.
- COLLEVATTI, R.G.; BRONDANI, R.V.; GRATTAPAGLIA, D. Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. **Heredity**, v.83, n.6, p.748-756, 1999.
- COELHO, A.S.G.; TELLES, M.P.C. Caracterização genética de populações naturais de cagaita (*Eugenia dysinterica* DC.) pelo uso de marcadores isoenzimáticos. **Genetics and Molecular Biology**, v.20, n.3, p.324, 1997. Suplemento /Apresentado ao 43. Congresso Nacional de Genética, Águas de Lindóia, 1997 – Resumo/
- COLLEVATTI, R.G.; GRATTAPAGLIA, D.; HAY, J.D. Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. **Molecular Ecology**, v.10, n.2, p.349-356, 2001.
- CONDIT, R.; HUBBELL, S.P. Abundance and DNA-sequence of 2-base repeat regions in tropical tree genomes. **Genome**, v.34, n.1, p.66-71, 1991.
- CONTE, R. Genética do recrutamento de plantas de uma população natural do palmitero (*Euterpe edulis* Mart.). Florianópolis, 2001. 104p. Dissertação (MS) – Centro de Ciências Agrárias, Universidade Federal de Santa Catarina
- CORRÊA, M.P. **Dicionário das plantas úteis do Brasil**. Rio de Janeiro: Imprensa Nacional, 1984. p.1926-1978.

- CREGAN, P.B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER, R.C.; LARK, K.G.; KAHLER, A.L.; KAYA, N.; VANTOAI, T.T.; LOHNES, D.G.; CHUNG, J.; SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Science**, v.39, n.5, p.1464-1490, 1999.
- CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: Universidade Federal de Viçosa, 1997. 390p.
- CRUZ, G.L. **Dicionário das plantas úteis do Brasil**. Rio de Janeiro: Civilização Brasileira, 1979. 590p.
- DALLAS, J.F. Detection of DNA fingerprints of cultivated rice by hybridization with a human minisatellite DNA probe. **Proceedings of the National Academy of Sciences of the United States of America**, v.85, p.6831-6835, 1988.
- DANQUAH, E.Y.; HANLEY, S.J.; BROOKES, R.C.; ADAM, C.; KARP, A. Isolation and characterization of microsatellites in *Echinochloa* (L.) Beauv. spp. **Molecular Ecology Notes**, v.2, n.1, p.51-56, 2002.
- DAYANANDAN, S.; BAWA, K.S.; KESSELI, R. Conservation of microsatellites among tropical tree (Leguminosae). **American Journal of Botany**, v.84, n.12, p.1658-1663, 1997.
- DAYANANDAN, S.; DOLE, J.; BAWA, K.; KESSELI, R. Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). **Molecular Ecology**, v.8, n.10, p.1585-1592, 1999.
- DONADIO, L.C.; MARTINS, A.B.G.; VALENTE, J.P. **Fruticultura tropical**. Jaboticabal: FUNEP, 1992. 268p.
- DOW, B.D.; ASHLEY, M.V. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. **Molecular Ecology**, v.5, p.615-627, 1996.
- EDWARDS, A.; CIVITELLO, A.; HAMMOND, H.A.; CASKEY, C.T. DNA typing and genetic-mapping with trimeric and tetrameric tandem repeats. **American Journal of Human Genetics**, v.49, n.4, p.746-756, 1991.
- ENNOS, R.A. Estimating the relative rate of pollen and seed migration among plant populations. **Heredity**, v.72, p.250-259, 1994.
- EXCOFFIER, L.; SMOUSES, 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.479-491, 1992.

- FERREIRA, M.B.; CUNHA, L.H.S. Dispersão de plantas lenhosas de cerrado: germinação e desenvolvimento. **Informe Agropecuário**, v.16, n.173, p.12-17, 1980.
- FERREIRA, M.E.; GRATTAPLAGIA, D. **Introdução ao uso de marcadores moleculares na análise genética**. Brasília: EMBRAPA, CENARGEN, 1996. 220p.
- FISHER, M.; HUSI, R.; PRATI, D.; PEINTINGER, M.; KLEUNEN, M.V.; SCHMID, B. RAPD variation among and within small and large population of the rare clonal plant *Ranunculus reptans* (Ranunculaceae). **American Journal of Botany**, v.87, n.8, p.1128-1137, 2001.
- FUTUYMA, D.J. **Biologia evolutiva**. 2.ed. Ribeirão Preto: Sociedade Brasileira de Genética, 1993. 631p.
- GAIOTTO, F.A. Inferências sobre herança quantitativa e estrutura genética em populações naturais de *Euterpes edulis* Mart. utilizando marcadores microssatélites. Piracicaba, 2001. 122p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- GILPIN, M.E.; SOULÉ, M.E. Minimum viable populations: process of species extinction. In: SOULÉ, M.E. (Ed.). **Conservation biology, the science of scarcity and diversity**. Sunderland: Sinauer Associates, 1986. p.19-34.
- GOODMAN, S. J. R_{ST} Calc: a collection of computer program for calculating estimates of genetic differentiation from microsatellite and determining their significance. **Molecular Ecology**, v.6, p.881-885, 1997.
- GOVINDAJARU, R.D. Variation in gene flow levels among predominantly self-pollinated plants. **Journal of Evolutionary Biology**, v.2, p.173-181, 1989.
- GRODZICKER, T.; WILLIAMS, J.; SHARP, P.; SAMBROOK, J. Physical mapping of temperature-sensitive mutations of adenoviruses. **Cold Spring Harbor Symposia Quantitative Biology**, v.39, p.439-446, 1974.
- HAGEN, A.R.; GIESE, H.; BROCHMANN, C. Trans-atlantic dispersal and phylogeography of *Cerastium arcticum* (Caryophyllaceae) inferred from RAPD and SCAR markers. **American Journal of Botany**, v.88, p.103-112, 2001.
- HAIG, S.M. Molecular contributions to conservation. **Ecology**, v.79, n.2, p.413-425, 1998.

- HAMADA, H.; PETRINO, M.G.; KAKUNAGA, T.; SEIDMAN, M.; STOLLAR, B.D. Characterization of genomic poly (dT-dG) and poly (dC-dA) sequence: structure, organization, and conformation. **Molecular and Cellular Biology**, v.4, n.12, p.2610-2621, 1984.
- HAMRICK, J.L. Plant population genetics and evolution. **American Journal of Botany**, v.69, n.10, p.1685-1693, 1982.
- HAMRICK, J.L.; SCHNABEL, A. Understanding the genetic structure of plant populations: some old problems and a new approach. In: GREGORIUS, H.R. (Ed.). **Population genetics in forestry: lecture notes in biomathematics**. Berlin: Springer-Verlag, 1985. v.60, p.50-70.
- HERINGER, E.P.; FERREIRA, M.B. Informações preliminares acerca da floração precoce de vinte espécies arbóreas do Cerrado do Planalto Central. /Apresentado ao 25. Congresso Nacional de Botânica, Mossoró, 1974 – Resumos/
- HUFF, D.R.; PEAKALL, R.; SMOUSE, P.E. RAPD variation within and among natural populations of outcrossing buffalograss (*Buchoë dactuloides* (Nutt.) Elgen), **Theoretical and Applied Genetics**, v.86, p.927-934, 1993.
- HUSBAND, B.C.; BARRETT, S.C.H. Estimates of gene flow in *Eichhornia paniculata* (Pontederiaceae): effects of range substructure. **Heredity**, v.75, p.549-560, 1995.
- ISABEL, N.; BEAULIEU, J.; BOUSQUET, J. Complete congruence between gene diversity estimatives derived from genotypic data at enzyme and random amplified polymorphic DNA loci em black spruce. **Proceedings of the National Academy of Sciences of the United States of America**, v.92, p.6369-6373, 1995.
- ISAGI, Y.; SUHANDONO, S. PCR primers amplifying microsatellite loci of *Quercus mysinifolia* Blume and their conservation between oak species. **Molecular Ecology**, v.6, p.897-899, 1997.
- JENCZEWSKI, J.-M., P.; RONFORT, J. Evidence for gene flow between wild and cultivated *Medicago sativa* (leguminosae) based on allozyme markers and quantitative traits. **American Journal of Botany**, v.86, p.677-678, 1999.
- JIN, L.; MACAUBAS, C.; HALLMAYER, J.; KIMURA, A.; MIGNOT, E. Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. **Proceedings of the National Academy of Sciences of the United States of America**, v.93, n.26, p.15285-15288, 1996.

- KARP, A.; SEBERG, O.; BUIATTI, M. Molecular techniques in the assessment of botanical diversity. **Annual of Botany**, v.78, n.2, p.143-149, 1996.
- KIRST, M.; BRONDANI, R.P.V.; BRONDANI, C.; GRATTAPAGLIA, D. Screening of designed primers pair for recovery of microsatellite markers and their transferability among species of *Eucalyptus*. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF *EUCALYPTUS*, Salvador, 1997. **Anais**. Colombo: EMBRAPA, 1997. v.2, p.167-171.
- LEFORT F.; KYVELOS J; ZERVOU M.; EDWARDS K.J.; ROUBELAKIS-ANGELAKIS K. A. Characterization of new microsatellite loci from *Vitis vinifera* and their conservation in some *Vitis* species and hybrids. **Molecular Ecology Notes**, v.2, n.1, p.20-21, 2002.
- LEVIN, D.A.; KESTER, H.W. Gene flow in seed plants. **Evolutionary Biology**, v.7, p.139-220, 1974.
- LEWIS, P.; ZAYKIN, D. **Genetic data analysis**: computer program for the analysis of allelic data (software) 2000. version 1.0(d12). <http://alleyn.eeb.uconn.edu/gda/> (21 Nov. 2002)
- LEWONTIN, R.C.; HUBB, J.L. A molecular approach to the study of genetic heterozygosity in natural populations. II Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. **Genetics**, v.54, p.595-609, 1966.
- LITT, M.; LUTY, J.A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetics**, v.44, p.398-401, 1989.
- LOPES, M.S.; SEFC, K.M.; LAIMER, M; MACHADO, da C. A. Identification of microsatellite loci in apricot. **Molecular Ecology Notes**, v.2, n.1, p.24-26, 2002.
- LOU, K.F.; WEISS, M.J.; BRUCKNER, P.L.; MORRIL, W.L.; TALBERT, L.E.; MARTIN, J.M. RAPD variation within and among geographic population of wheat stem sawfly (*Cephus cinctus* Norton). **The Journal of Heredity**, v.89, n.4, p.329-335, 1998.
- LOWE A.J.; GOODALL-COPESTAKE, W.P.; CARON, H.; KREMER, A.; DECROOCCQ, S. A set of polymorphic microsatellites for *Vochysia ferruginea*, a promising tree for land reclamation in the Neotropics. **Molecular Ecology Notes**, v.2, n.3, p.209-210, 2002.

- LYNCH, M.; MILLIGAN, B. G. Analysis of population genetic structure with RAPD markers. **Molecular Ecology**, n.3, p.91-99, 1994.
- MAGUIRE, T.L.; SEDGLEY, M. Genetic diversity in *Banksia* and *Dryandra* (Proteaceae) with emphasis on *Banksia cuneata*, a rare and endangered species. **Heredity**, v.79, p.394-401, 1997.
- MANLY, B.F.J. **Randomization and Monte Carlo Methods in Biology**. London: Chapman & Hall. 1991. 281p.
- MANLY, B.F.J. **Multivariate statistical methods**: a primer. London: Chapman & Hall, 1986a. 159p.
- MANLY, B.F.J. Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. **Research Population Ecology**, v.28, p.201-208, 1986b.
- MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer Research**, v.27, p.209-220, 1967.
- MARIOT, A. Distribuição da diversidade genética e aspecto fa fenologia e dispersão de sementes de pariparoba (*Piper cernnum*). Florianópolis, 2000. 110p. Dissertação (M.S.) - Centro de Ciências agrárias, Universidade Federal de Santa Catarina.
- McCOUCH, S.R.; CHEN, X.L.; PANAUD, O.; TEMNYKH, S.; XU, Y.; CHO, Y.G.; HUANG, N.; ISHII, T.; BLAIR, M. Microsatellite marker development, mapping and applications in rice genetics and breeding. **Plant Molecular Biology**, v.35, n.1-2, p.89-99, 1997.
- MELO, V.J.R. Determinação de paternidade em pomares de sementes de *Eucalyptus* com marcadores microssatélites. Goiânia, 2000. 183p. Dissertação (M.S.) – Escola de Agronomia, Universidade Federal de Goiás.
- MILACH, S.C.K. **Marcadores moleculares em plantas**. Porto Alegre: UFRGS, 1998. 141p.
- MILLER, M. Tools for population genetic analyses (TFPGA) (software), version 1.3: a windows program for analyses of allozyme and molecular population genetic data, 1997. <http://herb.bio.nau.edu/~miller/tfpga.htm> (20 Nov. 2002)
- MITTERMEIER, R.A.; AYRES, J.M.; WERNER, T.; FONSECA, G.A.B. O país da megadiversidade. **Ciência Hoje**, v.14, n.81, p.20-27, 1992.

- MOORE, S.S.; SARGEANT, L.L.; KING, T.J.; MATTICK, J.S.; GEORGES, M.; HETZEL, D.J.S. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. **Genomics**, v.10, n.3, p.654-660, 1991.
- MORAES, P.L.R. Estrutura genética de populações de *Cryptocarya moschata* Nees & Martius Ex Nees (Lauraceae). Piracicaba, 1997. 177p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- MULLIS, K.; FALLONA, F. Specific synthesis of DNA *in vitro* via a polymerase catalysed chain reaction. **Methods Enzymology**, v.55, p.335-350, 1987.
- MURRAY, M.G.; THOMPSON, W.F. Rapid isolation of high molecular weight plant DNA. **Nucleic Acids Reserch**, v.8, p.1134-1137, 1980.
- NAVES, R.V. Espécies frutíferas nativas dos cerrados de Goiás: caracterização e influências do clima e dos solos. Goiânia, 1999. 206p. Tese (Doutorado) - Escola de Agronomia, Universidade Federal de Goiás.
- NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics**, v.89, n.3, p.583-590, 1978.
- NEI, M. F-statistics and analysis of gene diversity in subdivided population. **Annals of Human Genetics**, v.41, p.225-33, 1977.
- NEI, M. Genetics distance between populations. **American Naturalist**, v.106, p.283-292, 1972.
- NEI, M.; LI, W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences of the United States of America**, v.74, p.5267-5273, 1979.
- NEIGEL, J.E. A comparision of alternative strategies for estimating gene flow from genetic markers. **Annual Review Ecology Systematics**, v.28, p.105-128, 1997.
- NYBOM, H.; HALL, H.K. Minisatellite DNA fingerprints can distinguish *Rubus* na estimate their degree of relatedness. **Euphytica**, v.53, p.107-114, 1991.
- OLUFOWOTE, J.O.; XU, Y.B.; CHEN, X.L.; PARK, W.D.; BEACHELL, H.M.; DILDAY, R.H.; GOTO, M.; MCCOUCH, S.R. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L) using microsatellite and RFLP markers. **Genome**, v.40, n.3, p.370-378, 1997.

- OUBORG, N.J.; PIQUOT, Y.; GROENENDAEL, J.M.V. Population genetics, molecular markers and the study of dispersal in plants. **Journal of Ecology**, v.87, p.551-568, 1999.
- PALACIOS, C.; GONZÁLES-CANDELAS, F. Analysis of population genetic structure and variability using RAPD markers in the endemic and endangered *Limonium dufourii* (Plumbaginaceae). **Molecular Ecology**, n.6, p.1107-1121, 1997.
- PANAUD, O.; CHEN, X.L.; MCCOUCH, S.R. Frequency of microsatellite sequences in rice (*Oryza sativa* L). **Genome**, v.38, n.6, p.1170-1176, 1995.
- PARKER, P.G.; SNOW, A.A.; SCHUG, M.D.; BOOTON, G.C.; FUERST, P.A. What molecules can tell us about populations: choosing and using a molecular marker. **Ecology**, v.79, n.2, p.361-382, 1998.
- PEAKALL, R.; GILMORE, S.; KEYS, W.; MORGANTE, M.; RAFALSKI, A. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. **Molecular Biology and Evolution**, v.15, n.10, p.1275-1287, 1998.
- PERECIN, M.B. Diversidade genética em populações naturais de espinheira Santa, *Maytenus aquifolia* Mart e *M. ilicifolia* Mart ex. Reiss (Celastraceae). Piracicaba, 2001. 134p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- PINTO, L.R. Análise da estrutura genética das populações de milho (*Zea mays* L.) BR-105 e BR-106 e respectivos sintéticos IG-3 e IG-4 por meio de microssátelites, Piracicaba, 2001. 142p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- POWELL, W.; MACHRAY, G.C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends of Plant Science**, v.1, p.215-222, 1996.
- PROENÇA, C.E.B.; GIBBS, P.E. Reproductive biology of eight sympatric Myrtaceae from Central Brazil. **New Phytologist**, v.126, p.343-354, 1994.
- RAFALSKI, J.A.; MORGANTE, M.; POWELL, W.; VOGEL, J.M.; TINGEY, S.V. Generating and using DNA markers in plants. In: BIRREN, B.; LAI, E. (Ed.). **Analysis of non-mammalian genomes: a practical guide**. New York: Academic Press, 1996. p.75-134.
- RAMAKRISHMAN, A.P.; COLEMAN, C.E.; MEYER, S.E.; FAIRBANKS, D.J. Microsatellite markers for *Bromus tectorum* (cheatgrass). **Molecular Ecology Notes**, v.2, n.1, p.22-23, 2002.

- REIS, S.R. Distribuição e dinâmica da variabilidade genética em populações naturais de palmito (*Euterpe edulis* Martius). Piracicaba, 1996. 210p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- RITLAND, K. A serie FORTRAN computer programs for estimating plant mating systems. **Journal of Heredity**, v.81, p.235-237, 1990.
- RITLAND, K.; JAIN, S. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. **Heredity**, v.47, p.35-52, 1981.
- RIZZINI, C.T. Efeito tegumentar na germinação de *Eugenia dysenterica* DC. (Myrtaceae). **Revista Brasileira de Biologia**, v.30, p.381-402, 1971.
- ROA, A.C.; CHAVARRIAGA-AGUIRRE, P.; DUQUE, M.C.; MAYA, M.M.; BONIERBALE, M.W.; IGLESIAS, C.; THOME, J. Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. **American Journal of Botany**, v.87, n.11, p.1647-1655, 2000.
- ROBINSON, I.P. Aloenzimas na genética de populações de plantas. In: ACELINO, A.C. (Ed.). **Eletroforese de isoenzimas e proteínas afins: fundamentos e aplicações em plantas e microorganismos**. Viçosa: UFV, 1998. p.329-380.
- RÖDER, M.S.; PLASCHKE, J.; KONIG, S.U.; BORNER, A.; SORRELLS, M.E.; TANKSLEY, S.D.; GANAL, M.W. Abundance, variability and chromosomal location of microsatellites in wheat. **Molecular & General Genetics**, v.246, n.3, p.327-333, 1995.
- ROGERS, S.O.; BENDICH, A.J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plants tissues. **Plant Molecular Biology**, v.5, p.69-76, 1985.
- ROLF, F.J. **NTSYS-Pc: Numerical taxonomy and multivariate analysis system**. New York: Exeter Publisher, 1989. 210p.
- SCHLÖTTERER, C.; RITTER, R.; HARR, B.; BREM, G. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. **Molecular and Biology Evolution**, v.15, n.10, p.1269-1274, 1998.

- SCHLÖTTERER, C.; TAUTZ, D. Slippage synthesis of simple sequence DNA. **Nucleic Acids Research**, v.20, n.2, p.211-215, 1992.
- SCHNEIDER, S.; ROESSLI, D.; EXCOFFIER, L. **Arlequin ver. 2000**: a software for population data analysis. (software). Geneva: University of Geneva, Genetic and Biometry Laboratory, 2000. <http://anthropologie.unige.ch/arlequin/> (21 Nov. 2002)
- SEBEN, A.M. Estrutura genética de populações de jequitibá-rosa [*Cariniana legalis* (Mart.) O. Ktze por caracteres quantitativos e isoenzimas. Piracicaba, 2001, 210p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- SHAW, D.V.; ALLARD, R.W. Estimation of outcrossing rates in Douglas-fir using isozyme markers. **Theoretical and Applied Genetics**, v.62, p.113-120,1982.
- SILVA, R.S.M. Caracterização de sub-populações de cagaita (*Eugenia dysenterica* DC.) da região sudeste do estado de Goiás. Goiânia, 1999. 107p. Dissertação (M.S.) - Escola de Agronomia, Universidade Federal de Goiás.
- SKABO, S.; VAILLANCOURT, R.E.; POTTS, B.M. Fine-scale genetic structure forest revealed by RAPDs. **Australian Journal of Botany**, v.46, p.538-594, 1998.
- SLATKIN, M. Estimating level of gene flow in natural populations. **Genetics**, v.99, p.323-335, 1981.
- SLATKIN, M. Rare alleles as indicators of gene flow. **Evolution**, v.39, n.1. p.53-65, 1985.
- SLATKIN, M.; BARTON, N.H. A comparison of three indirect methods for estimating average levels of gene flow. **Evolution**, v.43, n.7, p.1349-1368, 1989.
- SLATKIN, M. A measure of population subdivision based on microsatellite allele frequencies. **Genetics**, v.130, p.457-462, 1995.
- SMITH, J.F.; PHAM, T.V. Genetic diversity of narrow endemic *Allium aaseae* (Alliaceae) **American Journal of Botany**, v.93, n.6, p.717-726, 1996.
- SMULDER, M.J.M.; BREDEMEIJER, G.; RUS-KORTEKASS, W.; ARENS, P.; VOSMAN, B. Used of short microsatellite from base sequence to generate polymorphism among *Lycopersicum esculentum* cultivars and accession of other *Lycopersicum* species. **Theoretical and Applied Genetics**, v.97, p.264-272, 1997.

- SORK, V.L.; CAMPBELL, D.; DYER, R.; FERNANDEZ, R.; NASON, J.; PETIT, R.; SMOUSE, P.; STEINBERG, E. In: WORKSHOP ON GENE FLOW IN FRAGMENTED, MANAGED AND CONTINUOUS POPULATIONS, Santa Barbara, 1998. <http://www.nceas-web/projects/2057/nceas-paper3/> (20 Nov. 2002)
- SORK, V.L.; NANSON, J.; CAMPBELL, D.R.; FERNANDEZ, J.F. Landscape approaches to historical and contemporary gene flow in plants. **Trends Ecology and Evolution**, v.13, n.5, p.219-224, 1999.
- SOUZA, S.A.C.D. Avaliação da variabilidade genética em *Musa* spp utilizando marcadores microssatélites. Piracicaba, 2002. 86p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- STEINKELLNER, H.; LEXER, C.; TUTRESCHEK, E.; GLOSSIL, J. Conservation of (GA)_n microsatellite loci between *Quercus* species, **Molecular Ecology**, n.6, p.1189-1194, 1997.
- SUN, M.; WONG, K.C. Genetic structure of three orchid species with contrasting breeding systems using RAPD and allozyme markers. **American Journal of Botany**, v.88, n.12, p.2180-2188, 2001.
- TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, v.17, n.16, p.6463-6471, 1989.
- TAUTZ, D.; SCHLÖTTERER, C. Simple sequences. **Current Opinion in Genetics & Development**, v.4, p.832-837, 1994.
- TAYLOR, A.C.; KRAAIJEVELD, K.; LINDENMAYER, D.B. Microsatellites for the greater glider, *Petauroides volans*. **Molecular Ecology Notes**, v.2, n.1, p.57-59, 2002.
- TELLES, M. Diversidade genética e estrutura genética populacional de cagaiteira (*Eugenia dysenteria* DC.) do sudeste de Goiás. Goiânia, 2000. 129p. Dissertação (M.S) - Escola de Agronomia, Universidade Federal de Goiás.
- TORGGLER, M.G.F.; CONTEL, E.P.B.; TORGGLER, S.P. **Isoenzimas**: variabilidade genética em plantas. São Paulo: Sociedade Brasileira de Genética, 1995. 175p. (Série Monografias)
- TRINDADE, M.G. Estrutura genética de populações naturais de cagaiteira (*Eugenia dysenterica* DC) do nordeste de Goiás. Goiânia, 2001. 99p. Dissertação (M. S.) - Escola de Agronomia, Universidade Federal de Goiás.
- TRINDADE, M.G.; CHAVES, L.J.; COELHO, A.S.G.; ZUCCHI, M.I Determinação do número de marcadores RAPD para estudos de estrutura genética em *Eugenia dysenterica* utilizando o método *bootstrap*/(compact disc). In: CONGRESSO

- BRASILEIRO DE MELHORAMENTO DE PLANTAS, 1., Goiânia, 2001. **Anais**. Goiânia: Sociedade Brasileira de Melhoramento de Plantas, 2001.
- VALVA, F. A.; COELHO, A.S.G. Estrutura genética de populações vegetais: estratégias adaptativas do cerrado. /Apresentado ao 44. Congresso Nacional de Genética, Goiânia, 1998 - Resumo/
- VENCOVSKY, R. Variance of an estimative of the outcrossing rate. **Brazilian of Journal of Genetics**, v.17, n.3, p.349-351, 1994.
- VENCOVSKY, R. Análise de variância de frequências alélicas. **Genetics and Molecular Biology**, v.15, p.53-60, 1992.
- WADT L.H.O. Estrutura genética de populações naturais de pimenta longa (*Piper hispidinervum* C. DC.). Piracicaba, 2001. 95p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- WEBER, J.L.; MAY, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain-reaction. **American Journal of Human Genetics**, v.44, n.3, p.388-396, 1989.
- WEIR, B.S. **Genetics data analysis II**: methods for discrete population genetic data. Sunderland: Sinauer Associates, 1996. 455p.
- WELSH, J.; McCLELLAND, M. Fingerprinting genomes using PCR with arbitrary primers. **Nucleic Acids Research**, v.18, p.7213-7218, 1990.
- WHITE, G.M.; BOSHIER, D.H.; POWELL, W. Genetic variation within a fragmented population of *Swietenia humilis* Zucc. **Molecular Ecology**, v.8, n.11, p.1899-1909, 1999.
- WIERDL, M.; DOMINSKI, A.M.; PETES, T.D. Microsatellite instability in yeast: dependence on the length of the microsatellite. **Genetics**, v.146, n.3, p.769-779, 1997.
- WILLIAMS, J.G.; KUBELIK, A.R.; LIVAK, K.J.; RAFALSKI, L.A.; TINGEY, S.V. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research**, v.18, p.6531-6535, 1990.
- WOLF, H.; BACKELJAU, T.; VERHAGEN, R. Congruence between allozyme and RAPD data in assessing macrogeographical genetic variation in the periwinkle *Littorina striata* (Mollusca, Gastropoda). **Heredity**, v.81, p.486-492, 1998.
- WRIGHT, J.M.; BENTZEN, P. Microsatellites: genetic markers of the future. **Review in Fish Biology and Fisheries**, v.4, p.384-388, 1994.

- WRIGHT, S. The genetical structure of populations. **Annals of Eugenics**, v.15, p.395-420, 1951.
- WRIGTH, S. The interpretation of population structure by F-statistic with special regard to systems of mating. **Evolution**, v.19, p.395-420, 1965.
- WRITE, G.; POWELL, W. Cross-species amplification of SSR loci in the Meliaceae family. **Molecular Ecology**, v.6, p.1195-1197, 1997.
- WU, K.S.; TANSKLEY, D. Abundance, polymorphism and genetic mapping of microsatellites in rice. **Molecular & General Genetics**, v.241, p.225-235, 1993.
- WYMAN, A.R.; WHITE, R. A highly polymorphic locus in human DNA. **Proceedings of the National Academy of Sciences of the United States of America**, v.77, p.6754-6758, 1980.
- YAMAMOTO, T.; KIMURA, T.; SHODA, M.; BAN, Y.; HAYASHI, T.; MATSUTA, N. Development of microstellite markers in the Japanese pear (*Pyrus pyrifolia* Nakai). **Molecular Ecology Notes**, v.2, n.1, p.14-16, 2002.
- YANG, G.P.; MAROOF, M.A.S.; XU, C.G.; ZHANG, Q.F.; BIYASHEV, R.M. Comparative-analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. **Molecular & General Genetics**, v.245, n.2, p.187-194, 1994.
- YOUNG, A.; BOYLE, T.; BROWN, T. The population genetic consequences of habitat fragmentation for plants. **Trends Ecology Evolutionary**, v.11, p.413-418, 1996.
- ZUCCHI, M.I.; BRONDANI, R.P.V.; PINHEIRO, J.B.; BRONDANI, C.; VENCOVSKY, R. Transferability of microsatellite markers from *Eucalyptus* spp. to *Eugenia dysenterica* (Myrtaceae family). **Molecular Ecology Notes**, v.3, n.4, Dec. 2002.
- ZUCCHIA, M.I.; PINHEIRO, J.B.; VENCOVSKY, R.; COELHO, A.S.G.; CHAVES, L.J. Estrutura genética e fluxo gênico em *Eugenia dysenterica* dc, utilizando marcadores microssátelites. **Genetics and Molecular Biology**, (compact disc) / Apresentado ao 48. Congresso Nacional de Genética, Águas de Lindóia, 2002 – Resumo/ <http://www.sbg.gov.br/> (20 Nov. 2002)
- ZUCCHIB, M.I.; PINHEIRO, J.B.; VENCOVSKY, R.; COELHO, A.S.G.; MOURA, N.F.; CHAVES, L.J. Análise da estrutura genética de populações de *Eugenia dysenterica* utilizando marcadores RAPD. **Genetics and Molecular Biology** (compact disc) / Apresentado ao 47. Congresso Nacional de Genética, Águas de Lindóia, 2001 – Resumo/