

**UNIVERSIDADE FEDERAL DO PARANÁ**

**RICARDO ALEXANDRE VALGAS**

**ANÁLISE MULTIVARIADA APLICADA NO MAPEAMENTO DA DIVERGÊNCIA  
GENÉTICA DE SUBPOPULAÇÕES DE *Araucaria angustifolia* POR  
MARCADORES ISOENZIMÁTICOS**

**CURITIBA**

**2008**

**RICARDO ALEXANDRE VALGAS**

**ANÁLISE MULTIVARIADA APLICADA NO MAPEAMENTO DA DIVERGÊNCIA  
GENÉTICA DE SUBPOPULAÇÕES DE *Araucaria angustifolia* POR  
MARCADORES ISOENZIMÁTICOS**

Dissertação apresentada ao curso de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração em Programação Matemática, Departamentos de Construção Civil e de Matemática, setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Anselmo Chaves Neto  
Co-orientador: Prof. Dr. Osmir José Lavoranti

**CURITIBA**

**2008**

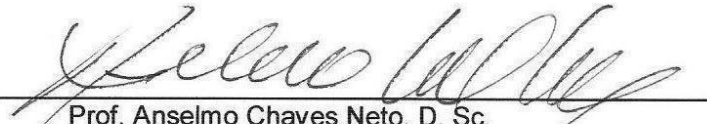
## TERMO DE APROVAÇÃO


RICARDO ALEXANDRE VALGAS

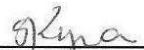
### ANÁLISE MULTIVARIADA APLICADA NO MAPEAMENTO DA DIVERGÊNCIA GENÉTICA DE SUBPOPULAÇÕES DE *Araucaria angustifolia* POR MARCADORES ISOENZIMÁTICOS


Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no curso de Pós-Graduação em Métodos Numéricos em Engenharia – Área de Concentração em Programação Matemática, Setores de Tecnologia e de Ciências Exatas da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador:

  
Prof. Anselmo Chaves Neto, D. Sc.  
Departamento de Estatística, UFPR

  
Prof.<sup>a</sup> Graciela Inês Bolzón de Muñiz, Dr.  
Departamento de Engenharia Florestal, UFPR

  
Sachiko Araki Lira, D. Sc.  
IPARDES

  
Valderês Aparecida de Sousa, Ph.D.  
EMBRAPA Florestas

Curitiba, 29 de abril de 2008.

A Deus, meus pais Jorge e Carmen, e minha irmã Sílvia.

À minha namorada Daniele.

Com muito carinho, dedico.

## **AGRADECIMENTOS**

A Deus, por ter me proporcionado esta oportunidade e ouvir minhas preces.

À minha família, que sempre esteve ao meu lado durante minha vida e me deu forças para continuar e vencer.

À Daniele Damian de Almeida, uma pessoa maravilhosa, pela sua paciência, carinho e incentivo durante esta fase.

Ao professor Dr. Anselmo Chaves Neto, que se propôs a orientar este trabalho e se tornou um grande amigo e companheiro de boas conversas e grandes risadas.

Ao professor Dr. Osmir José Lavoranti, pesquisador da Embrapa Florestas, por contribuído de forma significativa para a realização deste trabalho.

À Dra. Valderês Aparecida de Sousa, pesquisadora da Embrapa Florestas, por tornar este trabalho viável e ter colaborado com informações importantes.

À Embrapa Florestas, por tornar este trabalho possível.

Aos meus amigos do mestrado, aos professores e funcionários do Departamento de Estatística da UFPR, por sempre estarem presentes no meu dia a dia, contribuindo, de alguma maneira, para a realização deste trabalho.

Em especial à Maristela Bandil, secretária do CESEC e Henrique Túlio, secretário do Departamento de Estatística da UFPR, por atenção e ajuda prestadas.

## RESUMO

Dado o crescente número de espécies da Floresta Ombrófila que se encontram em risco de extinção, é muito difícil implementar programas de conservação para todas elas. Portanto, é urgente a adoção de medidas para a conservação dos recursos genéticos remanescentes das araucárias, dado seu grande valor econômico e social. Este trabalho teve por objetivo mensurar a divergência genética de 15 subpopulações de *Araucaria angustifolia*, por marcadores isoenzimáticos, considerando três análises diferentes. Foram amostradas 70 árvores da população de Campos do Jordão - SP, sendo 35 da Subpopulação Floresta Intocada e 35 da Floresta Explorada; 50 árvores da população Irati - PR, sendo 10 de cada subpopulação, a saber: Floresta Intocada, Capão 1 e 2, Floresta Explorada e Floresta Plantada; 60 árvores da população de Caçador - SC, sendo 10 de cada subpopulação, a saber: Floresta Intocada, Capão 1, 2, 3 e 4 e Floresta Explorada; 70 árvores da população de Camanducaia - MG e 70 árvores da população de Barbacena - MG. Para cada árvore foram avaliados seis *locus*: *GOT-B*, *GOT-C*, *PGMA-A*, *MDH-B*, *SKDH-B* e *6PGDH-B* e observados os alelos dos seguintes materiais: endosperma e embriões (homozigotos e heterozigotos). Foram utilizados os coeficientes de similaridades de Jaccard, Sneath e Sokal, Southwood, Gower, Holliday e Koeff, e medidas de dissimilaridade como distância Euclidiana e de Manhattan, além dos métodos de ligação de Ward, do Vizinho Mais Próximo e o Método da Ligação Média para obter os dendogramas. Aplicou-se a MANOVA e o teste  $T^2$  de Hotelling para a comparação entre os vetores de médias nos agrupamentos formados. Observou-se a menor divergência genética entre as amostras das populações de Irati e Caçador, seguido de Campos do Jordão e Camanducaia. Em Irati as subpopulações mais semelhantes foram Capão 1 e Capão 2, seguido de Floresta Intocada e Floresta Explorada. Em Caçador existem evidências que as subpopulações de Capão 1 e Capão 3 são muito semelhantes. Porém, há pouca variabilidade genética entre as árvores dessa população. A primeira situação de análise, que comparou vetores amostrais, foi considerada a mais adequada às análises estatísticas. A segunda situação de análise, que comparou vetores de médias, também foi satisfatória, e a terceira situação de análise, que comparou as árvores uma a uma, formou apenas alguns agrupamentos. As árvores de decisão classificaram corretamente alguns agrupamentos. As regiões mais críticas para conservação observadas foram Barbacena, Floresta Plantada em Irati e os Capões 2 e 4 em Caçador. Dessa forma, em média, pode-se reduzir para 62% as áreas de conservação das populações de *Araucaria angustifolia* nestes locais.

Palavras-chave: *Araucaria angustifolia*. Mapeamento genético. Análise de Agrupamento.

## ABSTRACT

Given the increasing number of species of the Ombrofila Mixed Forest threatened of extinction, it is very difficult to implement programs of conservation for all species. Therefore, the adoption of measures for the conservation of the remaining genetic resources of araucaria is urgent, given its great economic and social value. This work has as objective to measure the genetic divergence of 15 subpopulations of *Araucaria angustifolia*, by isoenzyme markers, considering three different analysis. For this 70 trees from the population of Campos do Jordão – SP were taken, being 35 from the subpopulation Natural Forest and 35 from Exploited Forest; 50 trees from Irati – PR population, being 10 of each subpopulation, to know: Natural Forest, Forest Island 1 and 2, and Plantation; 60 trees of Caçador – SC population, being 10 from each subpopulation, to know: Natural Forest, Forest Island 1, 2, 3 and 4 and Exploited Florest; 70 trees from Camanducaia – MG population and 70 trees from Barbacena – MG population. For each tree six *locus* were evaluated: *GOT-B*, *GOT-C*, *PGMA-A*, *MDH-B*, *SKDH-B*, *6PGDH-B* and allelos observed for endosperms and embryos (homozigotos and heterozigotos). Were used Coefficients of similarities of Jaccard, Sneath and Sokal, Southwood, Gower, Holliday and Koleff, and also measures of dissimilarities as Euclidean and Manhattan distances, beyond the methods of Ward linkage, the Next Neighborhood and the Average Linkage method to perform the dendograms. MANOVA was applied and T<sup>2</sup> Hotteling test for the comparison between the vectors of averages in the performed groupings. It was observed little genetic divergence between the samples of Irati and Caçador populations, followed by Campos do Jordão and Camanducaia. In Irati Forest Island 1 and Forest Island 2 subpopulations are the most similar, followed by Natural Forest Exploited Forest. In Caçador there are evidence that Forest Island 1 and Forest Island 3 subpopulations are very similar. However, a little genetic variability was observed between the trees of this population. The first situation, that compares sample vectors, was considered the most adjusted to proceed the statistical analysis. The second situation, that compares vectors of averages, also was satisfactory, and the third situation, that compare trees one to one, formed only some groupings. The decision trees had classified some groupings correctly. The most critical regions observed for conservation had been Barbacena, Plantation in Irati and Forest Island 2 and Forest Island 4 in Caçador. Therefore, in average, the conservation area can be reduced to 62% of the *Araucaria angustifolia* in these sites.

Key words: *Araucaria angustifolia*. Genetic mapping. Analysis of grouping.

## LISTA DE ILUSTRAÇÕES

FIGURA 2.1 – ÁRVORE DA <i>Araucaria angustifolia</i> , SÍMBOLO DO ESTADO DO PARANÁ .....	19
FIGURA 2.2 – CONE DA <i>Araucaria angustifolia</i> , CONHECIDA POPULARMENTE COMO PINHA .....	21
FIGURA 2.3 – ETAPAS DO PROCESSO DE ELETROFORESE EM GEL DE AMIDO .....	26
FIGURA 3.1 – MAPA DO BRASIL APRESENTANDO AS 5 REGIÕES DE ESTUDO .....	52
QUADRO 3.1 – DISTÂNCIAS (EM KM) ENTRE AS REGIÕES DE ESTUDO .....	52
QUADRO 3.2 – DESCRIÇÃO DO CONJUNTO DE DADOS.....	54
QUADRO 3.3 – COEFICIENTES DE SIMILARIDADE UTILIZADOS NO TRABALHO .....	57
FIGURA 4.1 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO.....	60
FIGURA 4.2 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO .....	61
FIGURA 4.3 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO.....	61
FIGURA 4.4 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO .....	62
FIGURA 4.5 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO .....	63
FIGURA 4.6 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE KOLEFF E 3 MÉTODOS DE AGRUPAMENTO.....	64
FIGURA 4.7 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO .....	65
FIGURA 4.8 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO .....	66
FIGURA 4.9 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO .....	67
FIGURA 4.10 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO .....	67
FIGURA 4.11 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO .....	68
FIGURA 4.12 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO .....	70
FIGURA 4.13 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO .....	70



FIGURA 4.14 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO .....	71
FIGURA 4.15 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO .....	71
FIGURA 4.16 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO .....	72
FIGURA 4.17 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO A DISTÂNCIA EUCLIDIANA E 3 MÉTODOS DE AGRUPAMENTO .....	73
FIGURA 4.18 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO A DISTÂNCIA DE MANHATTAN E 3 MÉTODOS DE AGRUPAMENTO.....	74
FIGURA 4.19 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO A DISTÂNCIA EUCLIDIANA E 3 MÉTODOS DE AGRUPAMENTO .....	76
FIGURA 4.20 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO A DISTÂNCIA DE MANHATTAN E 3 MÉTODOS DE AGRUPAMENTO .....	76
FIGURA 4.21 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO A DISTÂNCIA EUCLIDIANA E 3 MÉTODOS DE AGRUPAMENTO .....	78
FIGURA 4.22 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO A DISTÂNCIA DE MANHATTAN E 3 MÉTODOS DE AGRUPAMENTO .....	78
FIGURA 4.23 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO .....	81
FIGURA 4.24 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH .....	82
FIGURA 4.25 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOUTHWOOD .....	82
FIGURA 4.26 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE .....	83
FIGURA 4.27 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE HOLLIDAY .....	84
FIGURA 4.28 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE KOLEFF .....	84
FIGURA 4.29 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO .....	87
FIGURA 4.30 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH .....	87
FIGURA 4.31 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOUTHWOOD .....	88
FIGURA 4.32 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE .....	88
FIGURA 4.33 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE HOLLIDAY.....	89
FIGURA 4.34 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE KOLEFF .....	89

FIGURA 4.35 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO .....	92
FIGURA 4.36 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH .....	93
FIGURA 4.37 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOUTHWOOD .....	93
FIGURA 4.38 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE .....	94
FIGURA 4.39 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE HOLLIDAY.....	94
FIGURA 4.40 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE KOLEFF .....	95
FIGURA 4.41 – ÁRVORE DE CLASSIFICAÇÃO DAS POPULAÇÕES .....	96
FIGURA 4.42 – ÁRVORE DE CLASSIFICAÇÃO DAS SUBPOPULAÇÕES DE IRATI .....	97
FIGURA 4.43 – ÁRVORE DE CLASSIFICAÇÃO DAS SUBPOPULAÇÕES DE CAÇADOR .....	98

## LISTA DE TABELAS

TABELA 2.1 – OCUPAÇÃO TERRITORIAL DA FLORESTA COM ARAUCÁRIA NO SUL DO BRASIL - INÍCIO DO SÉCULO XX .....	18
TABELA 2.2 – UTILIZAÇÃO DA VARIÁVEL BINÁRIA PARA REPRESENTAR A PRESENÇA (1) OU A AUSÊNCIA (0) DE CARACTERÍSTICAS (VARIÁVEIS) .....	42
TABELA 2.3 – CÁLCULO DOS COEFICIENTES DE SIMILARIDADE COM BASE NAS OBSERVAÇÕES .....	42
TABELA 2.4 – CÁLCULO DOS COEFICIENTES DE SIMILARIDADE COM BASE NAS VARIÁVEIS BINÁRIAS .....	44
TABELA 3.1 – CONJUNTO DE DADOS ANALISADO (5 PRIMEIRAS OBSERVAÇÕES) .....	56
TABELA 3.2 – CONJUNTO DE DADOS CODIFICADO (5 PRIMEIRAS OBSERVAÇÕES) .....	57
TABELA 4.1 – NUMERAÇÃO DAS ÁRVORES DAS AMOSTRAS DAS POPULAÇÕES .....	79
TABELA 4.2 – NUMERAÇÃO DAS ÁRVORES DAS SUBPOPULAÇÕES DE IRATI	85
TABELA 4.3 – NUMERAÇÃO DAS ÁRVORES DAS SUBPOPULAÇÕES DE CAÇADOR .....	90

## LISTA DE SIGLAS

AFLP	– <i>Amplified Fragment Length Polymorphism</i>
APP	– Área de Proteção Permanente
ISSR	– <i>Inter Simple Sequence Repeats</i>
MANOVA	– Análise de Variância Multivariada
PCR	– <i>Polimerase Chain Reaction</i>
RAPD	– <i>Random Amplified Polymorphic DNA</i>
RFLP	– <i>Restriction Fragment Length Polymorphism</i>
UMVU	– <i>Uniform Minimum Variance Unbiased</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>15</b>
1.1 PROBLEMA DE ESTUDO .....	15
1.2 OBJETIVOS DO TRABALHO .....	16
1.2.1 Objetivo Geral .....	16
1.2.2 Objetivos Específicos .....	16
1.3 JUSTIFICATIVA .....	16
<b>2 REVISÃO DE LITERATURA</b> .....	<b>17</b>
2.1 INTRODUÇÃO .....	17
2.1.1 A espécie <i>Araucaria angustifolia</i> (Bert.) O. Ktze .....	17
2.1.1.1 Características morfológicas .....	19
2.1.1.2 Importância social e ecológica .....	21
2.1.2 Diversidade genética em espécies arbóreas.....	23
2.1.2.1 Diversidade de espécies .....	23
2.1.2.2 Diversidade genética.....	24
2.1.3 Marcadores genéticos no estudo de populações naturais.....	24
2.1.3.1 Isoenzimas .....	26
2.1.3.2 Eletroforese em gel de amido.....	27
2.1.4 Marcadores baseados em reação de polimerase em cadeia (PCR) .....	27
2.1.4.1 Marcadores RFLP (Restriction Fragment Length Polymorphism) .....	28
2.1.4.2 Marcadores AFLP (Amplified Fragment Length Polymorphism) .....	28
2.1.4.3 Marcadores RAPD (Random Amplified Polymorphic DNA).....	29
2.1.4.4 Microssatélites.....	29
2.1.4.5 Análise comparativa .....	30
2.2 ANÁLISE MULTIVARIADA .....	31
2.2.1 Estatísticas descritivas .....	32
2.2.2 Técnicas de agrupamento .....	35
2.2.2.1 Introdução .....	35
2.2.2.2 Definição e objetivos da análise de agrupamento .....	36
a) Medidas de Similaridade e Dissimilaridade.....	40
b) Agrupamento Hierárquico .....	44

c) Agrupamento não hierárquico .....	49
2.2.3 Árvore de Decisão .....	49
<b>3 MATERIAL E MÉTODO.....</b>	<b>51</b>
3.1 MATERIAL.....	51
3.1.1 Amostras das populações e subpopulações .....	51
3.1.2 Eletroforese em gel .....	55
3.2 MÉTODO.....	55
<b>4 RESULTADOS E DISCUSSÃO .....</b>	<b>59</b>
4.1 PRIMEIRA SITUAÇÃO DE ANÁLISE.....	59
4.2 SEGUNDA SITUAÇÃO DE ANÁLISE .....	73
4.3 TERCEIRA SITUAÇÃO DE ANÁLISE.....	79
4.4 ÁRVORES DE CLASSIFICAÇÃO .....	95
<b>5 CONSIDERAÇÕES FINAIS E CONCLUSÃO.....</b>	<b>99</b>
<b>REFERÊNCIAS.....</b>	<b>101</b>
<b>APÊNDICES .....</b>	<b>108</b>

# 1 INTRODUÇÃO

## 1.1 PROBLEMA DE ESTUDO

Muitas comunidades biológicas que foram se desenvolvendo ao longo de milhares de anos, agora estão sendo devastadas pela ação antrópica. Com isso, muitas espécies estão sob o risco de extinção nesta e nas próximas décadas. A principal causa desse problema, a destruição de *habitats*, é o crescimento da população humana, que causa impactos ambientais negativos pela exploração indiscriminada, transformação de florestas em pastagens e terras agricultáveis.

Nos *habitats* naturais preservados como reservas naturais ou parques nacionais, são necessários cuidados especiais para prevenir a extinção das espécies remanescentes, pois a conservação “in situ” permite a continuidade da evolução natural das espécies.

Diante deste contexto surgiu a biologia da conservação, que investiga os impactos negativos das atividades humanas sobre os ecossistemas e desenvolvem práticas que previnem a extinção de espécies, reintegrando-as quando possível a ecossistemas funcionais. No Brasil, as florestas naturais têm sido foco da biologia da conservação devido a sua riqueza de espécies e complexas estruturas de organização em comunidades e ameaças constantes aos ecossistemas.

A necessidade de se conservar os remanescentes florestais é urgente. Para que esse procedimento seja corretamente conduzido são necessários conhecimentos biológicos e ecológicos relacionados às populações a serem conservadas. Parâmetros como a quantidade e distribuição da variabilidade genética, biologia reprodutiva das espécies, fluxo gênico, regeneração, são, dentre outras, informações essenciais nesse processo (TRINDADE, 2001, p.1).

## 1.2 OBJETIVOS DO TRABALHO

### 1.2.1 Objetivo Geral

Este trabalho tem como objetivo fundamental fornecer uma parcela de contribuição ao conhecimento da estrutura genética em populações da Floresta Ombrófila e gerar informações técnico-científicas que possibilitarão o delineamento de estratégias de conservação e aproveitamento racional para o manejo extrativista sustentado.

### 1.2.2 Objetivos Específicos

Mapear a variabilidade genética em 15 subpopulações de *Araucaria angustifolia* oriundas de 5 populações naturais, através de marcadores isoenzimáticos, utilizando técnicas estatísticas multivariadas. Busca-se analisar qual a metodologia estatística é a mais adequada para determinar padrões de variabilidade genética, visando identificar centros de diversidade e subpopulações mais críticas para conservação.

## 1.3 JUSTIFICATIVA

Dado o crescente número de espécies da Floresta Ombrófila que se encontram em risco de extinção, não é possível implementar programas de conservação para todas elas. Portanto, é urgente a adoção de medidas para a conservação dos recursos genéticos remanescentes, e conseqüentemente, estudos abrangentes sobre populações de maior importância econômica, social e ecológica podem trazer contribuições importantes neste tema.



## 2 REVISÃO DE LITERATURA

### 2.1 INTRODUÇÃO

Os recursos naturais existentes em nosso planeta são elementos que podem ser utilizados pelo homem a fim de se promover o desenvolvimento das populações. No Brasil encontramos uma grande reserva de recursos naturais, sendo considerado o país de maior biodiversidade do planeta graças às variações climáticas e geomorfológicas aqui encontradas, variando desde a Mata Atlântica, cerrados, savanas, caatingas, até a Floresta Amazônica.

As florestas tropicais ocupam, aproximadamente, 20% da área terrestre (cerca de 17 milhões de km<sup>2</sup>) e abrigam mais de um terço das espécies do planeta. Durante séculos, as florestas tropicais brasileiras foram consideradas fontes inesgotáveis de matéria-prima para os mais diversos fins, tais como: madeira, frutos, fibras, entre outros produtos. Entretanto, esses recursos foram explorados de forma predatória e muitos deles encontram-se ameaçados de extinção.

Um exemplo dessa exploração está na Floresta Ombrófila e, principalmente, nas Florestas com Araucária (Floresta Ombrófila Mista), ecossistema encontrado no Brasil em altitudes elevadas, freqüente nos estados do Paraná, Santa Catarina e Rio Grande do Sul. Estas regiões de clima temperado possuem as estações do ano bem definidas e recebem chuva o ano todo.

#### 2.1.1 A espécie *Araucaria angustifolia* (Bert.) O. Ktze

O primeiro naturalista a descrever cientificamente a araucária foi o europeu Antonio Bertolini, no ano de 1820. Isso foi feito a partir de uma amostra plantada na cidade do Rio de Janeiro (ANSELMINI, 2005, p. 2). Um ano depois, foi determinada sua classificação científica: divisão Gymnospermae, classe Coniferopsida, ordem

Coniferae, família Araucariaceae, gênero *Araucaria* e espécie *angustifolia* (JOLY<sup>1</sup>, 1983, *apud* ANSELMINI, 2005, p. 2).

O surgimento da família Araucariaceae deu-se exclusivamente no Hemisfério Sul e evoluiu na Era Mesozóica durante os períodos Jurássico e Cretáceo, há cerca de 250 milhões de anos. Das 30 espécies do gênero *Araucaria* que acredita-se existir, apenas duas ocorrem na América do Sul: a *Araucaria araucana*, no sul do Chile e Argentina, e a *Araucaria angustifolia*, no sul e sudeste do Brasil, além da Argentina e Paraguai (ANSELMINI, 2005, p. 3).

No Brasil, segundo Mantovani *et al.* (2004), a espécie *Araucaria angustifolia* corresponde ao principal componente das Matas de Araucária, estendendo-se desde o oeste de Minas Gerais até o Rio Grande do Sul, principalmente no Planalto Sul Brasileiro em altitudes elevadas - de 600 a 1.200 m.

Até o início do século XX, a área original da Floresta com Araucária dominava as paisagens do Sul do Brasil, ocupando aproximadamente 200.000 km<sup>2</sup> de extensão. Conforme a Tabela 2.1 pode-se verificar o percentual ocupado por esta mata em cada um dos estados sul brasileiros nesse período.

TABELA 2.1 – OCUPAÇÃO TERRITORIAL DA FLORESTA COM ARAUCÁRIA NO SUL DO BRASIL - INÍCIO DO SÉCULO XX

ESTADO	OCUPAÇÃO TERRITORIAL (%)
Paraná	40
Santa Catarina	31
Rio Grande do Sul	25

FONTE: CARVALHO (2003)

Existe ainda, em sua composição florística, a Imbuia (*Ocotea porosa*), a Sassafrás (*Ocotea odorifera*), a Canela-preta (*Ocotea catharinensis*), o cedro (*Cedrella fissilis*) e a erva-mate (*Ilex paraguariensis*), entre outras espécies com grande valor econômico, que foram e ainda são utilizadas em diversos ramos industriais.

De acordo com Souza (2006, p. 6), no estado do Paraná a espécie *Araucaria angustifolia* é muito conhecida como pinheiro-do-paraná, entre outras denominações que também são empregadas: pinho, curi, curiúva, pinheiro-brasileiro, pinheiro-

<sup>1</sup> JOLY, A. B. **Botânica: introdução à taxonomia vegetal**. 6 ed. São Paulo: Editora Nacional, 1983, p. 777.

elegante, pinheiro-macaco, sendo a única espécie do gênero de ocorrência natural no Brasil e que se tornou o símbolo do estado do Paraná.

Tal espécie forma um ecossistema com uma enorme diversidade de espécies vegetais, devido à posição da copa de suas árvores estarem num extrato acima das demais árvores (ANSELMINI, 2005, p. 5).

#### 2.1.1.1 Características morfológicas

Vegetação tipicamente perene, as árvores da *Araucaria angustifolia* são de grande porte e atingem de 20 a 50 m de altura com 50 a 120 cm de DAP (diâmetro a altura do peito) quando adulta. Sua casca é grossa (aproximadamente 10 cm) e resinosa, de cor marrom-arroxeadada, a qual protege um tronco cilíndrico, reto e com raras ramificações. As árvores são dióicas, raramente apresentando plantas monóicas, com os cones femininos (conhecidos popularmente como pinha) localizados no ápice de um ramo protegido por folhas. Quando maduros, os cones possuem de 10 a 20 cm de diâmetro, com cerca de 150 sementes cada. Os estróbilos masculinos são cilíndricos, alongados e com escamas coriáceas; quando maduros, possuem de 10 a 15 cm de comprimento por 2 a 4 cm de diâmetro (INSTITUTO de pesquisas ... , 2007).



FIGURA 2.1 – ÁRVORE DA *Araucaria angustifolia*, SÍMBOLO DO ESTADO DO PARANÁ  
FONTE: (CAMPOS do Jordão ... , 2007)

As árvores jovens apresentam a copa em forma cônica, enquanto que as mais velhas formam um candelabro terminal podendo ultrapassar os 200 anos de idade e produzindo, em média, 40 pinhas por ano ao longo de toda sua vida. A idade da árvore pode ser calculada de acordo com anéis de crescimento que vão sendo formados no tronco. Durante os meses de agosto a setembro ocorre o desenvolvimento das estruturas reprodutivas, e a polinização se dá através do vento durante os meses de outubro e novembro. Apesar das flores ocorrerem anualmente, as sementes demoram aproximadamente 3,5 anos para se formarem.

As acículas possuem de 3 a 6 cm de comprimento, sendo coriáceas, sésseis, simples, de coloração verde-escuro. As sementes, chamadas de pinhão, são bastante carnosas e possuem de 3 a 8 cm de comprimento, servindo de alimento para muitos animais que vivem nas Florestas com Araucária como, por exemplo, gralhas (*Cyanocorax sp.*), bugios (*Alouatta fusca*), pacas (*Agouti paca*), serelepes (*Sciurus ingrami*), entre outros. Cabe aqui ressaltar que as Florestas com Araucária abrigam uma grande diversidade de aves, cerca de 250 espécies, que representam 15% do total brasileiro. Esses animais desempenham um papel fundamental na reprodução das árvores pois contribuem para a dispersão das sementes, principalmente a cutia *D. azarae*. Este roedor merece destaque, pois seu hábito de enterrar as sementes, para depois comê-las, torna-o um dos maiores disseminadores da araucária (MAIS interior, 2007).

A madeira dessas árvores apresenta coloração branco-amarelada, bastante uniforme com densidade moderada. Por ser altamente permeável pode ser facilmente tratada para se obter um resultado de qualidade, apesar de apresentar tendência à distorção. Mainieri e Chimelo (1989, p. 418) indica algumas aplicações: caixotaria, movelaria, laminados, tábuas para forro, ripas, caibros, lápis, carpintaria, palitos de fósforos, formas para concreto, marcenaria, compensados, pranchas, postes e mastros de navios.

Os nós da madeira apresentam alto poder calorífico e, portanto, tornam-se um bom combustível que já foi largamente usado nas caldeiras de locomotivas e de embarcações (INSTITUTO de pesquisas ... , 2007).

### 2.1.1.2 Importância social e ecológica

De acordo com Souza (2006, p. 15) a araucária é altamente valorizada devido ao consumo de suas sementes e a utilização da madeira para a confecção de artefatos em marcenaria. O pinhão é rico em amido e pode ser utilizado tanto na alimentação de animais quanto na alimentação humana, integrando por um longo período a alimentação de alguns grupos indígenas, dos primeiros colonos e dos caboclos que habitavam o interior dos estados do Sul.



FIGURA 2.2 – CONE DA *Araucaria angustifolia*, CONHECIDA POPULARMENTE COMO PINHA  
FONTE: (CAMPOS do Jordão ... , 2007)

De acordo com Caldeira (2003), dentre as árvores presentes no sul do Brasil, a *Araucaria angustifolia* foi a primeira espécie vegetal a proporcionar um grande retorno financeiro àqueles que deram início a sua exploração, tanto no estado do Rio Grande do Sul quanto nos outros estados da mesma região. Durante décadas, esta espécie desempenhou um papel muito importante na economia brasileira fornecendo matéria-prima para a indústria obtendo, nos anos 70, grande destaque na economia de exportação. A madeira serrada e laminada apresenta alta qualidade para ser utilizada na fabricação de móveis e de papel. Carvalho (1994, p. 640) descreve que: “A resina proveniente da casca serve para fabricação de vernizes, terebintina, acetona, ácido pirolenhoso e outros produtos químicos.”

Anselmini (2005, p. 1) relata que a intensa exploração da floresta com araucária foi responsável pelo primeiro ciclo econômico no sul do Brasil, resultando em um enorme desmatamento nesta região. No Paraná, a criação da Estrada da

Graciosa, em 1873, serviu para facilitar o transporte de toras e tábuas da cidade de Curitiba até Antonina, assim como a Ferrovia Paranaguá-Curitiba em 1885 (KOCH e CORRÊA, 2002, p. 148).

Alguns anos depois, com o início da Primeira Guerra Mundial, a madeira do pinheiro foi destinada ao mercado interno, principalmente para São Paulo e Rio de Janeiro. A partir de 1934 a exploração foi intensificada de tal forma que aproximadamente 18 bilhões de m<sup>3</sup> de madeira foram exportados, quase toda ela extraída das Matas de Araucária, reduzindo a apenas 8% da sua distribuição original em 1978 (FUPEF<sup>2</sup>, 1978, *apud* SOUZA, 2006, p. 18).

Koch e Corrêa (2002, p. 148) relatam uma grande vantagem comercial no fornecimento de madeira. Enquanto um hectare da Floresta Amazônica fornece aproximadamente 200 m<sup>3</sup> de madeira, a Floresta com Araucária pode chegar a cerca de 516 m<sup>3</sup> de madeira por hectare, comprovando praticamente a ausência de concorrentes.

Toda a exploração ocorrida durante este ciclo resultou em uma devastação na biodiversidade das matas nativas de araucária. Em 2002 restavam cerca de 10% das florestas manejadas nos três estados do sul brasileiro e cerca de 1% das reservas naturais. Nesta época, o crescimento populacional do país gerou a necessidade da obtenção de terras para o plantio e para a criação de gado de corte.

No Paraná a situação é considerada crítica pois as florestas primárias ou intocadas praticamente não existem mais. Restam menos de 0,8% (cerca de 66.109 hectares) de florestas em estágio avançado de sucessão ocasionado pela agricultura, agropecuária, reflorestamentos com espécies exóticas, expansão desordenada das cidades, represas e manejo inadequado, representam os últimos remanescentes da biodiversidade da Floresta com Araucária no estado (SOUZA, 2006, p. 17-21).

Laboreau e Matos Filho (1948) citam a exploração da Matas de Araucária em três fases: a primeira fase explorou as árvores de maior valor comercial; a segunda explorou as árvores remanescentes da primeira fase e a terceira proporcionou a queimada da mata para a criação de áreas de pastoreio e agricultura.

Além da morte por causas naturais (1% ao ano), as araucárias ainda continuam sendo cortadas pelo homem. O reflorestamento utilizando o pinus (*Pinus*

---

<sup>2</sup> FUPEF (Fundação de Pesquisas Florestais do Paraná). **Inventário florestal do pinheiro no sul do Brasil**. Relatório Final. Curitiba, Brasil, 1978.

*elliottii*), espécie originária dos Estados Unidos, aumenta a ameaça ao que resta de araucárias no Paraná. A monocultura de pinus está sendo implantada pelas indústrias em áreas cada vez maiores a fim de triplicar a produção de papel, móveis e celulose nos próximos 10 anos (BIODIVERSITY reporting ... , 2007).

Neste ano de 2008, ainda restam araucárias confinadas em "museus", ou seja, em áreas de preservação aos cuidados dos estados e do governo federal, totalizando uma área inferior a 300.000 hectares, área equivalente a uma das grandes fazendas do norte do país. Contudo, essas áreas de preservação não são suficientes para evitar a extinção das araucárias (PORTAL São Francisco, 2007).

## 2.1.2 Diversidade genética em espécies arbóreas

### 2.1.2.1 Diversidade de espécies

A forma mais simples de caracterizar a diversidade de espécies é elaborar uma listagem daquelas que existem em uma determinada região de interesse; na prática, entretanto, esta não é uma tarefa muito fácil. Filósofos e naturalistas antigos como Aristóteles e Plínio (historiador romano) listaram os organismos vivos que eram reconhecidos em suas épocas e esboçaram esquemas para classificá-los.

Pesquisadores que estudavam a História Natural (Biologia, Geologia, entre outras áreas) demonstravam que o interesse em classificar organismos vivos estava crescendo. Um dos motivos que impulsionou esta atividade foi a descoberta de formas de vida desconhecidas e estranhas, as quais desafiavam os esquemas de classificação. Também o surgimento do microscópio no século XVII revelou um universo de organismos vivos que não podiam ser vistos a olho nu. Anos depois, no final do século XVIII, o naturalista sueco Carl Von Linné propôs um sistema de classificação de organismos que serve de base para a classificação até hoje (REVISTA eletrônica ... , 2007).

Geralmente, para se expressar a diversidade entre espécies, é necessário apresentar medidas de diversidade que podem ser classificadas em três grupos: diversidade taxonômica, diversidade filogenética e diversidade genética.

### 2.1.2.2 Diversidade genética

A diversidade genética (ou variabilidade genética) mensura as diferenças existentes tanto entre indivíduos de uma mesma espécie quanto entre populações naturais, as quais podem estar separadas entre si pela fragmentação dos *habitats* naturais.

De acordo com Francisco (2005, p. 2) a diversidade genética é criada pelo processo de mutação, a qual gera diferentes alelos. Entende-se por alelo como sendo cada uma das várias formas alternativas do mesmo gene ocupando um *locus* (posição) em um cromossomo.

Cada vez mais, métodos empregando marcadores genéticos (bioquímicos e moleculares) são utilizados para se examinar as diferenças na constituição do DNA, RNA ou certas proteínas entre indivíduos ou populações. Aprofundar esse conhecimento é fundamental para evitar que a perda de diversidade genética aumente a probabilidade de que uma espécie venha a se extinguir.

Durante muitos anos, observou-se a exploração descontrolada para fins comerciais, além de desmatamentos sem nenhum gerenciamento sustentável, que resultaram numa mudança significativa na aparência das Matas de Araucária. As populações naturais foram, ao longo do tempo, substituídas por áreas destinadas à pecuária e à agricultura. Uma consequência disso é a perda de variabilidade genética das araucárias. Um estudo realizado pela Universidade Federal de Santa Catarina (UFSC) para comparar áreas de florestas intactas com áreas já exploradas mostrou uma perda de mais de 50% na variabilidade genética destas (BIODIVERSITY reporting ... , 2007).

### 2.1.3 Marcadores genéticos no estudo de populações naturais

No período anterior aos anos 60, a variabilidade genética era expressa exclusivamente através das características quantitativas ou morfológicas. Conte (2004, p. 25) esclarece que marcadores morfológicos apresentam certas características durante uma determinada fase da vida da planta, sendo identificados geralmente na fase adulta. Nos casos de espécies com um ciclo de vida longo, sua



identificação torna-se muito difícil. Ainda, o estudo de populações naturais torna-se limitado, uma vez que um número muito pequeno desse tipo de marcador pode ser obtido. Entretanto, a fácil utilização e identificação visual tornaram-no extensamente utilizado na análise genética.

Os marcadores genéticos estão divididos em dois grupos, de acordo com seu tipo de expressão: marcadores codominantes e dominantes. Marcadores microssatélites, isoenzimas e RFLP (*Restriction Fragment Length Polymorphism*) são codominantes, enquanto que os marcadores RAPD (*Random Amplified Polymorphic DNA*), ISSR (*Inter Simple Sequence Repeats*), e AFLP (*Amplified Fragment Length Polymorphism*) segregam como dominantes, podendo ser tratados ao nível de genótipo formando dados de ausência e presença (0; 1, respectivamente). De acordo com Aguiar (2004, p. 22), para o cálculo das distâncias genéticas, estes marcadores utilizam matrizes de similaridade para representar os indivíduos em estudo.

Tanto na conservação como no melhoramento de recursos genéticos, é muito importante o conhecimento da estrutura genética das espécies naturais, bem como de seus sistemas reprodutivos, a fim de se determinar as áreas mais importantes para conservação.

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. Além disso, a utilização de isoenzimas permitiu a identificação de polimorfismo entre plantas que não apresentavam diferenças morfológicas (CONTE, 2004, p. 26).

Todo esse avanço tecnológico criou novas perspectivas para o acompanhamento da conservação de espécies e para o estudo populacional como um todo.

### 2.1.3.1 Isoenzimas

As isoenzimas surgiram na década de 60 e, a partir de então, têm sido extremamente importantes para as investigações sobre variação intra-específica, genética de populações, evolução e mapeamento genético, já realizados em centenas de espécies.

Após a introdução da eletroforese em gel de amido, as isoenzimas têm sido muito utilizadas em pesquisas na área biológica. Em plantas, elas têm sido utilizadas principalmente em genética de populações, evolução e caracterização de germoplasma.

A atividade de uma enzima específica pode ser detectada através da eletroforese: quanto maior a carga elétrica de uma proteína, mais rápida a sua migração no gel. A consequência é o aparecimento de uma ou mais bandas no gel que fornecem as distintas formas de uma mesma enzima, as isoenzimas, sendo identificadas em diferentes regiões do gel caso apresentem diferentes mobilidades eletroforéticas. Com esta técnica, o estudo da variabilidade genética de populações é baseado na variação observada nas isoenzimas.



(1) Gel de amido colocado em estufa para revelação dos diferentes sistemas enzimáticos. (2) Bandas reveladas para a enzima GOT.

FIGURA 2.3 – ETAPAS DO PROCESSO DE ELETROFORESE EM GEL DE AMIDO

### 2.1.3.2 Eletroforese em gel de amido

O procedimento de eletroforese consiste na separação de moléculas de um tecido vegetal e da migração destas proteínas num gel de amido submetido a uma corrente elétrica contínua. O sentido e a velocidade de migração são determinados pelo tamanho e carga das moléculas.

### 2.1.4 Marcadores baseados em reação de polimerase em cadeia (PCR)

Em meados da década de 80, Kary Mullis causou uma revolução na biologia quando apresentou a tecnologia do PCR, levando o pesquisador a ganhar o prêmio Nobel de medicina no início da década de 90.

A sensibilidade, a rapidez e a versatilidade da técnica de PCR têm apresentado um grande impacto sobre a biologia molecular, expandindo diversas áreas que antigamente eram inatingíveis, devido à carência de um método que apresentasse as vantagens do PCR. “O desenvolvimento e aprimoramento das técnicas de marcadores moleculares baseados em PCR tem permitido a obtenção de mapas genéticos saturados, mais rápido e economicamente mais viável.” (FERREIRA, 2003, p. 45).

É possível adicionar, excluir ou alterar as informações dos *primers* feitos no PCR. Isso o torna ideal para proferir mutações específicas em posições determinadas em um fragmento de DNA, facilitando o estudo das interações entre proteínas e DNA (FERREIRA, 2003, p. 16).

Três etapas dividem o método: a primeira quando ocorre a desnaturação da fita dupla de DNA através da elevação da temperatura (cerca de 94°C). A segunda, chamada anelamento, reduz drasticamente a temperatura (entre 35 e 60°C) para que ocorra a hibridação DNA – DNA de cada *primer*. Na terceira etapa, chamada extensão, adicionam-se nucleotídeos empregando como molde uma seqüência alvo, de forma que é gerada uma cópia desta seqüência. Repetem-se estas etapas, várias vezes, para que uma grande quantidade de DNA da seqüência de interesse seja produzida.

Muitos marcadores derivados desta técnica foram surgindo, como RAPD (*Random Amplified Polymorphic DNA*), RFLP (*Restriction Fragment Length*

*Polymorphism*), AFLP (*Amplified Fragment Length Polymorphism*) e microssatélites, os quais possibilitaram a detecção de polimorfismo genético diretamente no DNA.

O método PCR apresenta uma desvantagem em relação à rapidez na renaturação após a desnaturação dos produtos amplificados. A magnitude desta desvantagem está relacionada com a diferença de tamanho entre o *primer* e o produto do PCR. Por este motivo, diversos métodos têm sido propostos baseados no princípio de que somente uma simples fita de DNA ou RNA pode ser avaliada por *primer* na reação.

#### 2.1.4.1 Marcadores RFLP (*Restriction Fragment Length Polymorphism*)

O primeiro marcador codominante desenvolvido utilizando a técnica de reação em cadeia de polimerase PCR ficou conhecido como RFLP (*Restriction Fragment Length Polymorphism*), durante a década de 70. Este tipo de marcador baseia-se na fragmentação da molécula de DNA pela ação de enzimas de restrição. Os fragmentos originados são separados de acordo com seu tamanho no processo de eletroforese em gel de agarose, sendo posteriormente hibridizados em uma membrana de celulose (WADT, 1997, p. 14).

Os RFLPs têm proporcionado a construção de mapas genéticos de ligação que possibilitam a realização de análises genéticas e moleculares com várias aplicações no melhoramento de plantas; porém, o elevado custo e o tempo necessário na geração destes marcadores restringem drasticamente seu uso, principalmente em países como o Brasil.

#### 2.1.4.2 Marcadores AFLP (*Amplified Fragment Length Polymorphism*)

Este tipo de marcador tem como característica combinar os princípios dos marcadores RAPD e RFLP. As vantagens dos AFLP correspondem à formação de grande número de polimorfismo por reação e não existe a necessidade do conhecimento prévio a respeito da seqüência de DNA para a construção dos *primers*

utilizados. Porém, as desvantagens são o alto custo, a necessidade de reagentes e as várias etapas de obtenção do marcador, as quais compreendem: a extração do DNA, a digestão com endonucleases de restrição, a ligação de adaptadores, a amplificação seletiva via PCR, a eletroforese em gel de alta resolução e a visualização dos fragmentos (FALEIRO, 2007, p. 23).

#### 2.1.4.3 Marcadores RAPD (*Random Amplified Polymorphic DNA*)

Marcadores RAPD são os mais freqüentemente utilizados por serem de fácil implementação, bastante polimórficos e de rápida execução, possibilitando resultados de grande validade prática. Segundo Wadt (2001, p. 20), a técnica de RAPD utiliza-se de uma reação de PCR (*Polymerase Chain Reaction*), a qual se baseia no anelamento e extensão de um par de iniciadores (*primers*) que são pequenas moléculas de DNA de fita simples, as quais delimitam a seqüência-alvo da molécula de DNA. O fragmento de DNA produzido pode ser visualizado a olho nu, por meio de corante específico (brometo de etídio), como uma banda em gel de agarose.

Com o uso de marcadores moleculares, pode-se gerar uma grande quantidade de informações que, associada a características fenotípicas, permitem o agrupamento de genótipos e o planejamento de cruzamentos de forma rápida e muito eficiente.

Portanto, a técnica de RAPD é eficiente nos estudos de divergência genética em que os indivíduos analisados são geneticamente próximos, ou seja, pertencem a uma mesma espécie.

#### 2.1.4.4 Microssatélites

Microssatélites, ou STR ou SSRP (*Simple Sequence Repeat Polymorphisms*) ou STMS (*Sequence Tagged Microsatellite Sites*) são seqüências repetidas de um, dois, três ou quatro nucleotídeos e que estão espalhadas pelo genoma de um indivíduo. São altamente polimórficos em animais, microorganismos e plantas.

O mapeamento genético, para fins de proteção e de germoplasma visando à conservação de espécies, está sendo feito com o uso dos marcadores microssatélites. Comparado ao RFLPs, os microssatélites proporcionam 3 a 4 vezes mais polimorfismo ou informação; porém, há a necessidade de primeiro amplificar uma região, posteriormente seqüenciá-la e depois sintetizar os iniciadores específicos para cada *locus*. Uma vez feito isto, o *locus* marcador pode ser utilizado indefinidamente naquela espécie.

No entanto, conforme Faleiro (2007, p. 22):

As principais desvantagens são o alto custo requerido no desenvolvimento de *primers* específicos, quando eles não estão disponíveis para a espécie a ser estudada e o fato de bandas inespecíficas ou géis de baixa resolução poderem dificultar a avaliação dos polimorfismos.

Os microssatélites já estão sendo utilizados em larga escala para as principais espécies de importância agrícola. Em breve, eles se tornarão os marcadores mais comuns nos laboratórios.

#### 2.1.4.5 Análise comparativa

Em programas de melhoramento genético, a caracterização da diversidade genética por meio de marcadores moleculares, seja de coleções de trabalho ou de populações naturais, é um método rápido e eficiente para selecionar os parentais que constituirão as populações de melhoramento.

Atualmente, os RFLPs e os RAPDs apresentam o maior potencial de uso nos diversos tipos de estudos e aplicações em plantas, cada um com suas particularidades. Esses dois tipos de marcadores genéticos apresentam algumas semelhanças entre si, mas os RAPDs são mais eficientes no mapeamento genético que os RFLPs, mesmo com baixo nível de informação.

A eficiência decorre de maior número de bandas por reação, menor trabalho e menor quantidade e qualidade do DNA para as análises. Na realidade, marcadores RAPDs têm sido utilizados para proporcionar uma maior saturação dos mapas de ligação previamente estabelecidos com isoenzimas e/ou RFLPs.

Futuramente, a utilização de microssatélites via PCR poderá se tornar tão eficiente quanto a utilização de RAPDs no mapeamento genético e caracterização de germoplasma.

## 2.2 ANÁLISE MULTIVARIADA

A Análise Multivariada compreende um conjunto de métodos estatísticos que, a partir de observações multivariadas, busca identificar a relação existente entre as variáveis envolvidas em um determinado fenômeno. Assim, ao invés de uma única variável aleatória para representar uma só característica de uma unidade experimental (ou observacional), tem-se um vetor aleatório para fazer a representação de várias características.

Um dos objetivos dessas técnicas é o estudo da estrutura de covariância do vetor aleatório.

Geralmente as variáveis estão correlacionadas e, quanto maior o número de variáveis em estudo, mais complexa torna-se a análise por métodos clássicos da estatística univariada (MINGOTI, 2005, p. 21). Não basta conhecer informações estatísticas isoladas, é necessário conhecer a totalidade das informações fornecidas pelo conjunto das variáveis.

Segundo Lima (2002), cada vez mais se observa a importância da utilização das técnicas multivariadas no setor empresarial (e em todos os ramos da ciência) devido a sua potencialidade para resolver problemas envolvendo, ao mesmo tempo, muitas características. Também, os programas (*softwares*) estatísticos desenvolvidos tornaram tais técnicas populares e de fácil emprego. Johnson e Wichern (1998, p. 3) comentam o enorme crescimento da aplicação da Análise Multivariada nos últimos anos nas diversas áreas do conhecimento, tais como: economia, ciências médicas, biologia, além da indústria, entre outras.

Chaves Neto (1997, p. 5) descreve que a Análise Multivariada é utilizada na solução de problemas relacionados com:

- Inferências sobre vetores médios;
- Análise da estrutura de covariância de um vetor aleatório;
- Técnicas de agrupamentos;
- Reconhecimento de padrões e classificação.

Especificamente podem-se citar as técnicas multivariadas mais utilizadas: Análise de Componentes Principais, Análise Fatorial, Análise de Correlação Canônica, Análise de Agrupamento (*Cluster Analysis*), Reconhecimento de Padrões e, neste contexto, Regressão Logística e Análise Discriminante.

Nos estudos contendo  $p \geq 1$  variáveis são tomadas, geralmente,  $n$  observações de cada variável. Nesse caso está sendo amostrado um vetor aleatório  $\underline{X}$  de dimensão  $p$ . Desta forma pode-se agrupar as observações em uma matriz de dados de ordem  $n$  por  $p$ , tal como:

$${}_n X_p = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.1)$$

Assim, as linhas dessa matriz são as  $n$  observações do vetor aleatório  $p$ -dimensional  $\underline{X}' = [X_1, X_2, \dots, X_p]$  e cada entrada da matriz ( $x_{ij}$ ) corresponde a observação  $i$  da variável  $j$ . Pode-se considerar que as linhas da matriz (2.1) representam as observações (objetos, pessoas, etc.) e as colunas representam as características (ou variáveis) envolvidas, como por exemplo, preço, altura, raça, etc.

### 2.2.1 Estatísticas descritivas

Um grande conjunto de dados contendo muitas variáveis não é facilmente representado graficamente; com isso, interpretações visuais não são triviais. Johnson e Wichern (1988, p. 8) relatam que a matriz de dados apresentada em (2.1)



pode ser representada por números sumários conhecidos como estatísticas descritivas os quais são calculados a partir de dados amostrais. Então, as estatísticas descritivas funcionam como estimadores de parâmetros populacionais. O vetor médio populacional  $\underline{\mu}' = E(\underline{X}') = [E(X_1) \ E(X_2) \ \dots \ E(X_p)]$  é estimado pelo vetor médio amostral  $\underline{\bar{X}}$  o qual representa o centro de gravidade dos pontos amostrais  $x_i, i = 1, \dots, n$ , definido pela expressão:

$$\underline{\bar{X}}' = [\bar{X}_1 \ \bar{X}_2 \ \dots \ \bar{X}_p] \quad (2.2)$$

em que  $\bar{X}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$  para  $j = 1, 2, \dots, p$ , que é um estimador não-viciado, consistente e eficiente, ou melhor, UMVU (*Uniform Minimum Variance Unbiased*) do parâmetro  $\underline{\mu}$ .

A matriz de covariância populacional  $\Sigma = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})']$  é escrita como:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{bmatrix} \quad (2.3)$$

e possui nas entradas da diagonal principal as variâncias das variáveis aleatórias que compõem o vetor e fora dela, as covariâncias entre elas. O parâmetro  $\Sigma$  é estimado pela matriz de covariância amostral  $S$ , dada por:

$$S = \frac{1}{n-1} \sum_{i=1}^n (\underline{X}_i - \underline{\bar{X}})(\underline{X}_i - \underline{\bar{X}})' = \begin{bmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{bmatrix} \quad (2.4)$$

tal que a variância amostral de  $X_j$  é dada por:

$$s_{jj} = s_j^2 = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1} \quad (2.5)$$

e a covariância entre as variáveis aleatórias  $X_j$  e  $X_k$  é dada por:

$$s_{jk} = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{n-1} \quad (2.6)$$

Como são utilizados dados amostrais para estimar o par de médias  $(\mu_j, \mu_k)$ , perde-se um grau de liberdade.

Padronizando as variáveis de estudo, a matriz de covariância (2.3) torna-se a matriz de correlação que, no caso populacional, é definida por  $\rho = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1}$  onde  $V^{1/2}$  é a matriz desvio-padrão definida pela matriz diagonal:

$$V^{1/2} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix} \quad (2.7)$$

Assim, tem-se a expressão da matriz de correlação:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \quad (2.8)$$

A matriz  $\rho$  é estimada pela matriz de correlação amostral:

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (2.9)$$

na qual:

$$r_{jj} = \frac{s_{jj}}{\sqrt{s_{jj}} \cdot \sqrt{s_{jj}}} = 1 \quad \text{e} \quad r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}} \cdot \sqrt{s_{kk}}} = \frac{s_{jk}}{s_j s_k} \quad (2.10)$$

em que:

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$

$$s_{jj} = s_j = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1}}$$

$$s_{kk} = s_k = \sqrt{\frac{\sum_{i=1}^n (X_{ik} - \bar{X}_k)^2}{n-1}}$$

## 2.2.2 Técnicas de agrupamento

### 2.2.2.1 Introdução

Uma das maiores habilidades dos seres humanos envolve a capacidade de agrupar e classificar objetos similares. Há muitos anos o homem foi capaz de observar que havia objetos com propriedades comuns, como por exemplo, cores, formatos, minerais, e assim por diante. A idéia de agrupar “coisas” semelhantes em categorias é claramente uma idéia primitiva de classificação, ou de forma mais ampla, é necessária para o desenvolvimento da linguagem a qual consiste em utilizar palavras para reconhecer e discutir os diferentes tipos de eventos, objetos e pessoas que se possam encontrar. Um rótulo é usado para descrever uma classe de “coisas” que possuem características notáveis em comum. Então, os animais foram

chamados de gatos, cães, cavalos, etc., substantivos que acabam coletando indivíduos para dentro de cada grupo.

A atividade conceitual humana básica de classificar é também importante para os demais ramos da ciência. Em Biologia, por exemplo, classificar organismos tem sido uma preocupação desde a primeira pesquisa biológica. Aristóteles construiu e elaborou um sistema para classificar as espécies do reino animal, o qual teve início dividindo os animais em dois grupos principais: aqueles com sangue vermelho (correspondente aos animais vertebrados), e aqueles sem sangue vermelho (correspondente aos animais invertebrados) (EVERITT, 1993, p. 1).

A classificação pode, de forma simples, organizar um grande conjunto de dados para que se possam obter informações relevantes de forma eficiente, considerando, por exemplo, a média de cada grupo formado como uma medida conveniente.

#### 2.2.2.2 Definição e objetivos da análise de agrupamento

De acordo com Braulio (2005, p. 29), as técnicas de agrupamento são técnicas exploratórias utilizadas para formar grupos homogêneos de elementos à partir das distâncias entre eles. Essas distâncias são calculadas com base nos vetores de características desses elementos. Assim, tem-se uma matriz de distâncias para todos os elementos.

O desenvolvimento dessas técnicas ocorreu na década de 30, mas sua real utilização deu-se com o surgimento dos computadores e programas estatísticos, pois os cálculos eram muito complexos e a manipulação de matrizes dispensava muito tempo e trabalho (BRANDALISE, 2003, p. 37).

A união de objetos em grupos (ou *clusters*) com base na similaridade ou distância entre eles, considerando as variáveis a serem agrupadas, requer a aplicação de uma técnica estatística de classificação apropriada destinada a dividir em subconjuntos um conjunto de dados observados (SHARMA, 1996, p. 187).

Então, com base nessas premissas:

1 – Um *cluster* deve ser o mais homogêneo possível, considerando as características que compõem o vetor correspondente aos objetos.

2 – Cada *cluster* deve ser diferente dos demais, considerando as mesmas características.

Validar os agrupamentos encontrados significa garantir que realmente os grupos diferem entre si. Tal verificação pode ser realizada empregando-se desde testes estatísticos univariados para comparação de médias até testes multivariados, como, por exemplo, a MANOVA (Análise de Variância Multivariada) e o teste  $T^2$  de Hotelling.

Procura-se, através da MANOVA, verificar se há diferença estatisticamente significativa entre os vetores de médias amostrais de cada grupo. Já o teste  $T^2$  de Hotelling é utilizado nos casos onde existam múltiplas variáveis dependentes intervalares e dois grupos formados por variáveis independentes categóricas, busca-se verificar se as diferenças de médias entre eles diferem estatisticamente através da expressão:

$$T^2 = [(\bar{X}_1 - \bar{X}_2) - (\underline{\mu}_1 - \underline{\mu}_2)]' \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) S_p \right]^{-1} [(\bar{X}_1 - \bar{X}_2) - (\underline{\mu}_1 - \underline{\mu}_2)] \quad (2.11)$$

em que  $S_p$  é a estimativa da matriz de covariância conjunta. A estatística  $T^2$  possui distribuição  $F$  com  $p$  graus de liberdade no numerador e  $n_1 + n_2 - p - 1$  no denominador.

Na MANOVA testa-se, como hipótese nula ( $H_0$ ), que os vetores de médias amostrais não diferem entre si, ou seja,  $H_0 : \underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_k$ .

O Quadro 2.1 resume algumas informações:

FONTE DE VARIAÇÃO	MATRIZ SOMA DE QUADRADOS	GRAUS DE LIBERDADE
Tratamentos	$B = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})'$	k - 1
Resíduos	$W = \sum_{j=1}^k \sum_{i=1}^{n_k} (\bar{x}_{ij} - \bar{x}_j)(\bar{x}_{ij} - \bar{x}_j)'$	n - k
Total	$B + W$	n - 1

QUADRO 2.1 – MANOVA

FONTE: O autor (2008)

A Hipótese nula é testada usando-se a estatística de teste conhecida como Lâmbda de Wilks ( $\Lambda$ ), dada pela expressão:

$$\Lambda = \frac{|W|}{|B+W|} \quad (2.12)$$

em que:

B: é a matriz que resume a variação entre os efeitos dos tratamentos;

W: é a matriz que resume a variação dentro dos grupos.

No emprego da MANOVA é necessário que as amostras aleatórias sejam independentes e oriundas de populações Gaussianas, ou seja,  $\underline{X}_i \sim N_p(\underline{\mu}_i, \Sigma), i = 1, \dots, n$ .

O pressuposto da Gaussianidade para o vetor  $\underline{X}$  pode ser verificado por um teste estatístico. No caso da distribuição normal bivariada com  $p = 2$  variáveis têm-se:

$$P\left[(\underline{X} - \underline{\mu})' \Sigma^{-1} (\underline{X} - \underline{\mu}) \leq \chi_2^2(1-\alpha)\right] = 1 - \alpha \quad (2.13)$$

Fixando-se  $1 - \alpha$  em 90%, por exemplo, espera-se que 90% das observações amostrais sejam tais que:

$$(\underline{X} - \underline{\bar{X}})' S^{-1} (\underline{X} - \underline{\bar{X}}) \leq \chi_2^2(0,9) \quad (2.14)$$

onde  $\underline{\mu}$  é substituído pela estimativa  $\underline{\bar{X}}$  e  $\Sigma$  pela estimativa  $S$ . Caso contrário, a hipótese de normalidade é suspeita com esse nível de confiança.

O gráfico de silhueta também é um procedimento descritivo para avaliar a qualidade dos agrupamentos formados. Ele verifica se um determinado objeto está mais próximo de objetos do seu grupo ou de objetos dos grupos vizinhos. Para isso são consideradas duas medidas:

- a) distância média entre o objeto específico  $i$  e os objetos de seu próprio grupo;
- b) distância média entre o objeto específico  $i$  e os objetos do grupo mais próximo de  $i$ .

Seja o objeto  $i$  pertencente ao grupo  $A_{(i)}$  com  $n_{A(i)}$  elementos neste grupo. Então:

$$a(i) = \frac{\sum_{j \in A(i), j \neq i} d_{ij}}{n_{A(i)} - 1} \quad (2.15)$$

em que  $d_{ij}$  é a distância Euclidiana entre os objetos  $i$  e  $j$ .

Assim, para todos os demais grupos é calculada a distância média entre seus elementos e o objeto  $i$ . Seja  $B_{(i)}$  contendo  $n_{B(i)}$  o grupo com a menor distância média entre seus elementos e o objeto  $i$ . Portanto:

$$b(i) = \frac{\sum_{j \in B(i), j \neq i} d_{ij}}{n_{B(i)}} \quad (2.16)$$

O valor da silhueta no ponto  $i$  é definido por:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.17)$$

Esse valor reflete quão adequada foi a alocação do objeto  $i$  em seu grupo. O resultado  $s(i)$  oscila entre  $-1$  e  $1$ ; quanto mais próximo de  $1$  melhor a alocação do objeto, porém quanto mais próximo de  $-1$  pior a alocação, uma vez que o objeto está, em média, mais próximo dos elementos do grupo vizinho do que do seu próprio grupo.

Do ponto de vista geométrico, a análise de agrupamento é muito simples. Segundo Sharma (1996, p. 186), cada observação pode ser representada como um ponto em um espaço  $p$ -dimensional, em que  $p$  representa o número de variáveis ou características envolvidas na análise, de forma semelhante à Análise Fatorial.

Boscarioli (2006) esclarece:

Há na literatura vários algoritmos de agrupamento propostos. A escolha de um algoritmo depende do tipo de dados disponíveis, da aplicação e de objetivos particulares. Se a análise de agrupamentos é usada como uma ferramenta descritiva ou exploratória, é possível tentar vários algoritmos nos mesmos dados e averiguar o que os dados podem revelar.

## a) Medidas de Similaridade e Dissimilaridade

Conforme Johnson e Wichern (1998, p. 728), os objetos agrupados em *clusters* representam uma medida de “proximidade” através de algum tipo de distância, na maioria dos casos. Esquarcini (2005, p. 44) descreve que o agrupamento pode ser feito com base nos coeficientes de similaridade, que representam a maior semelhança entre os objetos, como o coeficiente de correlação, ou então com base nos coeficientes de dissimilaridade, os quais apresentam a menor semelhança entre os objetos. Assim, os coeficientes de similaridade indicam que quanto maior o valor, maior a semelhança entre os objetos.

Entretanto, qual o critério a ser utilizado para se decidir até que ponto dois objetos do conjunto de dados podem ser considerados como semelhantes ou não?

Para responder essa questão é necessário considerar medidas que descrevam a similaridade entre elementos amostrais de acordo com as características que neles foram medidas. (...) Assim, podemos calcular as distâncias entre os vetores de observações dos elementos amostrais e agrupar aqueles de menor distância. (MINGOTI, 2005, p.156).

### *Distâncias e Coeficientes de Similaridade para pares de objetos*

A distância Euclidiana é certamente a mais conhecida e mais usada para calcular a distância entre dois objetos no espaço  $p$ -dimensional. Considerando dois vetores das matrizes  $\underline{x}$  e  $\underline{y}$  de observações, a distância Euclidiana entre eles é definida por:

$$d(\underline{x}, \underline{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (2.18)$$

Segundo Everitt (1993, p. 47), a distância Euclidiana não preserva a ordem das distâncias entre os objetos quando se altera fortemente a escala das variáveis. Assim, considerando-se o quadrado da distância Euclidiana dado por:



$$d^2(\underline{x}, \underline{y}) = \sum_{i=1}^p (x_i - y_i)^2 \quad (2.19)$$

pode-se formular a distância de Mahalanobis trabalhando-se com valores padronizados na distância Euclidiana:

$$d(\underline{x}, \underline{y}) = \sqrt{\frac{(x_1 - y_1)^2}{s_1^2} + \dots + \frac{(x_p - y_p)^2}{s_p^2}} \quad (2.20)$$

Outras distâncias podem ser consideradas, tais como:

#### *Distância de Minkowski*

$$d(\underline{x}, \underline{y}) = \sqrt[n]{|x_1 - y_1|^n + |x_2 - y_2|^n + \dots + |x_p - y_p|^n} = \sqrt[n]{\sum_{i=1}^p |x_i - y_i|^n} \quad (2.21)$$

Quando  $n = 1$ , a equação (2.21) é conhecida por “métrica do quarteirão” ou “métrica de *city-block*” e para  $n = 2$  é a mesma que a distância Euclidiana, mas variações nos valores de  $n$  podem causar trocas nos pesos dados (GROBE, 2005, p. 53).

#### *Distância city-block (Manhattan)*

$$d(\underline{x}, \underline{y}) = \sum_{i=1}^p |x_i - y_i| \quad (2.22)$$

Quando os objetos não puderem ser representados através de medidas  $p$ -dimensionais significativas, pares de objetos são geralmente comparados com base na presença ou ausência de certas características. Objetos semelhantes apresentam mais características em comum do que objetos não semelhantes (JOHNSON e WICHERN, 1998, p. 547).

Matematicamente estas características são representadas através de uma variável binária, a qual assume o valor 1 se a característica está presente, ou 0 se a característica está ausente. Um exemplo pode ser visualizado na Tabela 2.2, considerando dois objetos (*i* e *j*) e quatro características (variáveis):

TABELA 2.2 – UTILIZAÇÃO DA VARIÁVEL BINÁRIA PARA REPRESENTAR A PRESENÇA (1) OU A AUSÊNCIA (0) DE CARACTERÍSTICAS (VARIÁVEIS)

Objetos	Variáveis				
	A	B	C	D	E
<i>i</i>	1	0	1	0	0
<i>j</i>	1	1	0	0	1

FONTE: O autor (2008)

De acordo com Chaves Neto (1997, p. 66), a partir da Tabela 2.2 é necessário observar a frequência de empates do tipo 1-1, 0-0 e também dos desempates do tipo 1-0 e 0-1, pois a presença de uma determinada característica pode ser mais forte do que sua ausência. Ainda, o quadrado da distância Euclidiana (equação (2.19)) fornece o número de desempates de um par de objetos e, quanto maior o seu valor, menor a similaridade.

A partir desse tratamento definem-se os coeficientes de similaridade apresentados em uma tabela de contingência:

TABELA 2.3 – CÁLCULO DOS COEFICIENTES DE SIMILARIDADE COM BASE NAS OBSERVAÇÕES

		Objeto <i>i</i>		Total
		1	0	
Objeto <i>j</i>	1	<i>a</i>	<i>b</i>	<i>a + b</i>
	0	<i>c</i>	<i>d</i>	<i>c + d</i>
Total		<i>a + c</i>	<i>b + d</i>	<i>p = a + b + c + d</i>

FONTE: O autor (2008)

Assim, alguns exemplos de coeficientes de similaridade utilizados são descritos no Quadro 2.2.

COEFICIENTES	EXPRESSÃO	PONDERAÇÃO
Co incidência Simples	$\frac{a+d}{p}$	Pesos iguais para 1-1 e 0-0
Sokal & Sneath	$\frac{2(a+d)}{2(a+d)+b+c}$	Pesos em dobro para 1-1 e 0-0
Roger & Tanimoto	$\frac{a+d}{a+d+2(b+c)}$	Pesos em dobro para 1-0 e 0-1
Jaccard	$\frac{a}{a+b+c}$	Desconsiderando o empate 0-0 no numerador e no denominador
Simpson	$\frac{a}{a+b}$	Razão das igualdades para as desigualdades, excluindo 0-0
Sorensen	$\frac{2a}{2a+b+c}$	Peso em dobro para 1-1
Ochiai	$\frac{a}{\sqrt{(a+b)(a+c)}}$	Ponderando o empate 1-1

QUADRO 2.2 – PRINCIPAIS COEFICIENTES DE SIMILARIDADE

FONTE: O autor (2008)

Até agora foram descritas as maneiras de calcular as distâncias e os coeficientes de similaridade. Porém, segundo Johnson e Wichern (1998, p. 735), é possível encontrar as similaridades a partir das distâncias. Através da expressão:

$$s_{ij} = \frac{1}{1+d_{ij}} \quad (2.23)$$

encontram-se os valores das similaridades ( $0 < s_{ij} < 1$ ) entre dois objetos  $i$  e  $j$  a partir da distância  $d_{ij}$  correspondente. As distâncias nem sempre são encontradas a partir de similaridades. Desde que a matriz de similaridade seja não-negativa definida e o valor máximo de similaridade seja 1, então tem-se a seguinte propriedade:

$$d_{ij} = \sqrt{2(1-s_{ij})} \quad (2.24)$$

### *Similaridades e Medidas de Associação para pares de variáveis*

Em algumas situações, é mais importante agrupar as variáveis do que os objetos. Quando as variáveis a serem agrupadas forem binárias, pode-se organizar os dados em uma tabela de contingência, conforme Tabela 2.4.

TABELA 2.4 – CÁLCULO DOS COEFICIENTES DE SIMILARIDADE COM BASE NAS VARIÁVEIS BINÁRIAS

	Variável $i$		Total	
	1	0		
Variável $j$	1	$a$	$b$	$a + b$
	0	$c$	$d$	$c + d$
Total	$a + c$	$b + d$	$n = a + b + c + d$	

FONTE: O autor (2008)

Estas medidas de associação (ou similaridade) são exatamente análogas às listadas na Tabela 2.3.

#### b) Agrupamento Hierárquico

As técnicas de agrupamento hierárquico utilizam tanto uma série de sucessivas junções quanto uma série de sucessivas divisões. Neste caso, os dados não são particionados em um número determinado de *clusters* de uma única vez.

A primeira dessas técnicas é chamada de Método Hierárquico Aglomerativo. Este utiliza uma matriz de distâncias para identificar o par de objetos mais semelhantes que serão agrupados e considerados um único objeto. A partir daí, define-se uma nova matriz de distâncias na qual será novamente identificado o par mais semelhante para formar um novo grupo, e assim sucessivamente até que todos os objetos estejam incluídos em um mesmo grupo (48<sup>a</sup> reunião da RBRAS, 2003, p. 20).

A segunda, chamada de Método Hierárquico Divisível, trabalha no sentido inverso.

(...) Inicialmente um único grupo de itens é dividido em dois subgrupos de modo que os objetos em um subgrupo estejam o mais “distante” possível dos objetos do outro. Esses subgrupos são mais uma vez divididos em grupos distintos; este processo continua até que haja tantos grupos quanto itens iniciais; de fato, até que cada objeto forme um grupo. (JOHNSON e WICHERN, 1998, p. 739).

As classificações hierárquicas podem ser representadas por um diagrama bidimensional chamado *dendograma* (ou dendrograma) o qual ilustra as fusões ou divisões realizadas em cada passo da análise (EVERITT, 1993, p. 55).

(...) Os objetos mais similares são agrupados inicialmente, e esses grupos fundem-se de acordo com suas similaridades. Conforme as similaridades decrescem, todos os subgrupos fundem-se em um único grupo. (MARQUES, 2006, p. 48).

O Método Hierárquico Aglomerativo é descrito em etapas por Johnson e Wichern (1998, p. 740) e consiste em um estudo sistemático, a partir da etapa 1.

Etapa 1) Inicia-se o estudo com  $g$  grupos, sendo que cada um é formado por um único objeto. Calcula-se a matriz simétrica de distâncias  $D = (d_{ij})$  de dimensão  $n \times n$ , na qual  $d_{ij}$  representa a distância do objeto  $i$  ao objeto  $j$ :

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad (2.25)$$

em que  $d_{11} = d_{22} = \dots = d_{nn} = 0$ .

Etapa 2) Na matriz simétrica de distâncias  $D$  encontra-se o par de grupos mais similares (menor distância) representados por  $d_{AB}$  para indicar que  $A$  e  $B$  são os grupos mais próximos.

Etapa 3) Surge, então, um novo grupo denominado  $AB$ . Agora, atualiza-se as entradas da matriz de distâncias eliminando as linhas e colunas referentes aos grupos  $A$  e  $B$ , adicionando uma linha e uma coluna contendo a distância do grupo  $AB$  aos demais grupos restantes.

Etapa 4) Os passos 2 e 3 devem ser repetidos  $g - 1$  vezes (até que todos os objetos sejam incluídos em um único grupo) observando-se as identidades dos grupos que vão sendo formados e os níveis em que os mesmos se fundem.

Existem vários algoritmos ou métodos de agrupamento hierárquico que diferem, principalmente, do modo como as distâncias entre dois *clusters* são analisadas. Entre estes métodos, têm-se os descritos a seguir.

#### *Método do centróide*

De acordo com Everitt (1993, p. 62), no método do centróide os grupos formados são representados pela média dos valores de cada variável, ou seja, pelo seu vetor médio. Então a distância entre os grupos é definida em termos da distância entre dois vetores médios.

Conforme este método, um *cluster* é formado pelos objetos mais próximos e substituído pelo centróide (média de seus elementos). Na seqüência, calcula-se a matriz de similaridade considerando o primeiro *cluster* formado e verifica-se qual dos dois objetos possuem a maior similaridade para então serem agrupados em um novo *cluster*. Novamente, este novo *cluster* será representado pelo centróide dos seus elementos (SHARMA, 1996, p. 188).

Uma desvantagem deste método é que se o tamanho de dois grupos a serem agrupados for muito diferente, então o centróide do novo grupo será muito parecido com o do grupo maior, podendo remanescer dentro desse grupo. A estratégia utilizada independe do tamanho do grupo, ou seja, assume-se que os grupos a serem unidos possuem o mesmo tamanho, fazendo com que a posição evidente do novo grupo esteja sempre entre os grupos que foram unidos (EVERITT, 1993, p. 65).

#### *Ligações simples (vizinho mais próximo)*

Neste método, a distância entre dois *clusters* é representada pela menor distância (ou maior similaridade) entre todas as combinações possíveis de dois

pares de objetos. Assim, forma-se o primeiro *cluster*, por exemplo,  $(AB)$ , e a distância entre  $(AB)$  a algum outro objeto  $C$  é calculada pela expressão:

$$d_{(AB),C} = \min\{d_{AC}, d_{BC}\} \quad (2.26)$$

### *Ligação completa (vizinho mais distante)*

Este procedimento de ligação é muito semelhante ao processo da ligação simples, porém, ao invés de considerar a menor distância (ou maior similaridade), a formação de grupos é determinada pela distância máxima entre dois objetos. Considerando que  $(AB)$  forme um grupo, sua distância até outro grupo hipotético  $C$  é expressa por:

$$d_{(AB),C} = \max\{d_{AC}, d_{BC}\} \quad (2.27)$$

### *Ligação Média*

Neste procedimento, a distância considerada é a distância média entre todos os pares possíveis de objetos. Então, a partir da matriz de distâncias  $D = (d_{ij})$  agrupam-se os elementos com menor distância, por exemplo,  $A$  e  $B$  formam um grupo se  $d_{AB} = \min\{d_{ij}\}$ ,  $\forall i, j$  e a distância de  $(AB)$  à  $C$  é expressa por:

$$d_{(AB),C} = \frac{\sum_i \sum_j d_{ij}}{n_{(AB)} n_C} \quad (2.28)$$

em que  $n_{(AB)}$  e  $n_C$  representam as quantidades de elementos presentes em cada grupo.

### *Método de Ward*

Ward em 1963 propôs uma técnica que, segundo Dutra *et al.* (200?), busca minimizar a soma das diferenças entre os objetos de cada grupo e o valor médio do grupo. Isto equivale a minimizar o desvio padrão entre os objetos de cada grupo.

Johnson e Wichern (1998, p. 751) descrevem que este método baseia-se na “perda mínima das informações” ao se formar novos grupos, utilizando o critério da soma ao quadrado dos erros (SQE), entre dois agrupamentos, para todas as amostras.

Inicialmente existe um grupo para cada vetor componente do conjunto de dados, fornecendo um erro nulo para cada grupo, pois o vetor que compõe cada grupo representa o vetor médio do grupo. Posteriormente, os grupos são aglutinados 2 a 2, conforme o menor erro do grupo. É necessário um grande esforço computacional caso o tamanho da amostra seja elevado e a cada iteração o número de elementos pertencentes a cada grupo aumenta, elevando o número de cálculos para o erro de cada grupo.

Então, tem-se que:

$$SQE = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i - \bar{x})' (x_i - \bar{x}) \quad (2.29)$$

em que  $\bar{x}$  representa a média das amostras e  $x_i$  uma medida multivariada associada à  $i$ -ésima observação.

Braulio (2005, p.33) esclarece:

Por fim, a decisão do número de classes ou tipos para análise é tomada, geralmente, a partir do exame do dendograma ou árvore hierárquica, onde podem ser lidos os índices de nível (ou índice de similaridade), que são as distâncias euclidianas em que ocorrem as junções dos pontos observados para formar grupos. Um grande salto nesses índices (o que equivale a uma grande distância no dendograma) indica que a agregação reuniu dois grupos muito dissimilares e, em razão disso, deve-se definir o número de grupos anterior a este salto.



### c) Agrupamento não hierárquico

De acordo com Johnson e Wichern (1998, p. 754), as técnicas de agrupamento não hierárquico são utilizadas quando se pretende formar  $k$  grupos de itens ou objetos. Como não é necessário calcular a matriz de distâncias (ou similaridade) nem manter na memória do computador informações básicas durante o processo, estas técnicas têm maior aplicabilidade que as técnicas hierárquicas.

O método das  $k$ -médias, um dos mais populares e utilizado, é composto por três etapas:

- 1) particiona-se, arbitrariamente, as observações em  $k$  grupos;
- 2) realoca-se cada observação no grupo com a média (centróide) mais próximo possível, usualmente considerando a distância Euclidiana, e recalcula-se o centróide para aquele grupo que recebeu uma nova observação e para o grupo que perdeu uma observação;
- 3) repete-se a etapa 2 até que não restem mais alocações para serem feitas.

### 2.2.3 Árvore de Decisão

A árvore de decisão consiste em uma hierarquia de nós internos e externos, os quais são conectados por ramos. O nó interno, também conhecido como nó decisório ou nó intermediário, é a unidade de tomada de decisão que avalia, através de um teste lógico, qual será o próximo nó descendente ou filho. Em contraste, um nó externo (que não apresenta nó descendente), também conhecido como folha ou nó terminal, está associado a um rótulo ou a um valor.

Em geral, o procedimento de uma árvore de decisão é simples. Apresenta-se um conjunto de dados ao nó inicial (ou nó raiz, que também é um nó interno) da árvore; dependendo do resultado do teste lógico usado pelo nó, a árvore ramifica-se para um dos nós filhos e este procedimento é repetido até que um nó terminal seja alcançado. A repetição deste procedimento caracteriza a recursividade da árvore de decisão.

No caso das árvores de decisão binária, cada nó intermediário é dividido exatamente em dois nós descendentes: o nó esquerdo e o nó direito. Quando os

dados satisfazem o teste lógico do nó intermediário, seguem para o nó esquerdo e quando não o satisfazem seguem para o nó direito. Logo, uma decisão é sempre interpretada como verdadeira ou falsa.

O aprendizado de uma árvore de decisão é supervisionado, ou seja, o método aproxima funções alvo de valor discreto, na qual a função aprendida é representada por uma árvore de decisão. As árvores treinadas podem ser representadas como um conjunto de regras “Se - Então” para melhoria da compreensão e interpretação.

Árvores de decisão usadas para problemas de classificação são chamadas de Árvores de Classificação. Nas árvores de classificação cada nó terminal ou folha contém um rótulo que indica a classe predita para um determinado conjunto de dados. Nesse tipo de árvore pode existir dois ou mais nós terminais com a mesma classe.

Árvores de decisão usadas para problemas de regressão são chamadas de Árvores de Regressão. Nas árvores de regressão, cada nó terminal ou folha contém uma constante (geralmente uma média) ou uma equação para o valor previsto de um determinado conjunto de dados (ÁRVORES de decisão, 2007).

### 3 MATERIAL E MÉTODO

#### 3.1 MATERIAL

O conjunto de dados analisado neste trabalho foi cedido pela Embrapa Florestas, situada em Colombo – PR. Foram acrescentadas duas amostras populacionais ao conjunto de dados original, o qual já foi objeto de estudo em outros trabalhos realizados na área da genética de conservação.

Todo o material coletado foi profundamente estudado por pesquisadores da própria Embrapa através da análise de marcadores genéticos, incluindo a análise de segregação genética, de herança e *linkage*, garantindo a qualidade do conjunto de dados e revelando os marcadores bioquímicos empregados. Isso resultou na publicação de artigos especialmente em revistas internacionais. Portanto, neste trabalho não serão abordados aspectos de natureza biológica a respeito do conjunto de dados. A descrição em detalhes dos procedimentos genéticos estão presentes em Sousa *et al.* (2002) e Sousa *et al.* (2005).

##### 3.1.1 Amostras das populações e subpopulações

As árvores amostradas para o presente estudo foram obtidas nas maiores áreas de proteção ambiental localizadas nos estados de São Paulo (parque estadual) e Minas Gerais (área particular e APP – Área de Proteção Permanente), Paraná e Santa Catarina (em parques nacionais), compreendendo as cidades de Campos do Jordão, Irati, Caçador, Camanducaia (APP) e Barbacena (propriedade particular).

Irati está localizada a 590 km ao sul de Campos do Jordão; Caçador está a 700 km ao sul de Campos do Jordão e a 150 km a oeste de Irati. Barbacena está a 389 km de Camanducaia, a 1.101 km de Irati, a 1.241 km de Caçador e a 418 km de Campos do Jordão. O Quadro 3.1 fornece todas as distâncias entre as regiões de estudo.

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	590			
Caçador	700	150		
Camanducaia	107	710	851	
Barbacena	418	1.101	1.241	389

QUADRO 3.1 – DISTÂNCIAS (EM KM) ENTRE AS REGIÕES DE ESTUDO

FONTE: O autor (2008)



FIGURA 3.1 – MAPA DO BRASIL APRESENTANDO AS 5 REGIÕES DE ESTUDO

FONTE: O autor (2008)

A região de Campos do Jordão está localizada no estado de São Paulo a 1.800 m de altitude, onde o clima é mesotérmico subtropical com verões frios – Cfb, de acordo com a classificação de Köppen. O Parque Estadual possui 8.223 ha de área dos quais 5.655 ha são florestas e o restante são pastagens nativas. Fora dele, a vegetação tem sofrido grande influência da ação humana.

A segunda região, a Floresta Nacional de Irati, está localizada no estado do Paraná a 880 m de altitude. Possui as mesmas características climáticas de Campos do Jordão, com uma precipitação anual média de 1.442 mm. Sua ocupação é de 3.495 ha, sendo 36% de floresta natural e apenas 2,4% de área de reflorestamento.

A Floresta Natural de Caçador, a terceira região, está localizada no estado de Santa Catarina a 1.100 m de altitude numa área de aproximadamente 1.600 ha. O clima é considerado Cfb (de acordo com Köppen) com dias de frio intenso durante o inverno. Nesta região, a vegetação é chamada “Floresta de Pinheiro Brasileiro”, sendo a *Araucaria angustifolia* a espécie dominante.

A região de Camanducaia está localizada no sul do estado de Minas Gerais, em um vale que se abre para dar passagem ao rio Camanducaia, na encosta da Serra da Mantiqueira. A 1.200 m de altitude, o clima é frio e seco (Cwb, conforme Köppen) com temperatura média anual de 19,2°C e precipitação anual média de 1.700 mm.

A quinta e última região, Barbacena, também chamada de “A Cidade das Rosas”, está localizada no estado de Minas Gerais, ocupando uma área de aproximadamente 1.400 m<sup>2</sup>. Situada na Serra da Mantiqueira, região popularmente conhecida como Campo das Vertentes, apresenta clima tropical de altitude (Cwb), a 1.164 m do mar, com temperatura anual média de 17°C e precipitação anual média de 900 mm.

Em cada população foram coletadas amostras em grandes áreas florestais e em locais de menor incidência (ilhas de floresta ou capões) que variam de 12 a 60 ha. A menor distância entre as árvores amostradas nessas regiões foi de, no mínimo, 3 km uma da outra, formando parcelas amostrais de 1 ha nas grandes florestas e 0,3 ha em cada ilha de floresta.

Em cada uma das duas subpopulações de Campos do Jordão (Floresta Intocada e Floresta Explorada) foram amostradas 35 árvores, totalizando 70 árvores. Já em Irati e Caçador foram amostradas 10 árvores para cada subpopulação, totalizando 50 árvores em Irati e 60 em Caçador. Barbacena não foi dividida em subpopulações porque houve dificuldade em encontrar povoamentos definidos; em Camanducaia, as sementes foram coletadas em APPs, não tendo sido dividida em subpopulações.

O Quadro 3.2 fornece mais algumas informações sobre as amostras.

População	Localidade / Município	Latitude	Longitude	Subpopulação	Árvores amostradas
1	Campos do Jordão SP	22°44'	43°44'	Floresta Intocada	35
				Floresta Explorada	35
2	Irati PR	25°30'	50°36'	Floresta Intocada	10
				Capão 1	10
				Capão 2	10
				Floresta Explorada	10
3	Caçador SC	26°50'	51°00'	Floresta Intocada	10
				Capão 1	10
				Capão 2	10
				Capão 3	10
				Capão 4	10
4	Camanducaia MG	22°45'	46°08'		70
5	Barbacena MG	21°13'	43°46'		70

QUADRO 3.2 – DESCRIÇÃO DO CONJUNTO DE DADOS

FONTE: O autor (2008)

Algumas árvores foram excluídas das análises devido a problemas ocorridos durante os procedimentos de laboratório, ou ainda, porque alguns zimogramas não puderam ser interpretados. Em Irati, nas subpopulações de Capão 1 e Floresta Plantada, uma árvore de cada foi excluída. As amostras da população de Barbacena, armazenadas em *freezer* (-18°C) para análise sofreram uma queda na atividade metabólica decorrente do descongelamento causado por queda na energia elétrica. Isso fez com que os dados de 38 árvores fossem excluídos por dificuldade de interpretação.

### 3.1.2 Eletroforese em gel

O procedimento de eletroforese foi feito partindo-se de um conjunto de 20 sementes de cada uma das árvores de Campos do Jordão e 8 sementes das árvores das demais populações. A razão para a análise de 20 sementes para Campos do Jordão deve-se à análise de herança. As 8 sementes foram utilizadas para as demais populações e subpopulações pois esse é o número que seguramente estima o genótipo nuclear.

As sementes foram armazenadas a  $-30^{\circ}\text{C}$  até a extração da enzima para a eletroforese. Embriões e gametófitos femininos foram homogeneizados separadamente em uma extração tampão contendo 5 g de Sacarose, 150 mg de DTT (Dithiothreitol) e 3 g de PVP (Polivinilpirrolidona) dissolvido em 100 ml de 0,1 M Tris-HCl tampão, num pH 7,5. Extratos da enzima obtida do gametófito feminino e do embrião foram posicionados de forma adjacente no gel. O gel de amido horizontal (10,5% de amido e 2,5 – 3,5% sacarose) foi preparado seguindo Feret e Bergmann (1976), Conkle *et al.* (1982) e Liengsiri *et al.* (1990). O sistema tampão de eletroforese utilizado foi o ASHTON ajustado para pH 8,6 para *GOT* e *PGM*, e o sistema TRIS-Citrato ajustado para pH 7,3 para as outras enzimas. Os géis correram a 80 mA para o sistema ASHTON e a 180 mA para o sistema TRIS-Citrato.

Foram examinados 6 *locus*: *GOT-B*, *GOT-C*, *PGM-A*, *SKDH-B*, *MDH-B* e *6-PGDH-B*. No total foram preparados 2.280 zimogramas. Os detalhes sobre a metodologia da eletroforese de embriões e gametófitos e a interpretação genética dos zimogramas estão presentes em Sousa *et al.* (2002).

### 3.2 MÉTODO

Embora exista uma grande quantidade de marcadores genéticos, o princípio de análise desses marcadores é simples: marcadores comuns significam semelhança genética e marcadores não comuns significam diferenças genéticas. As semelhanças e diferenças genéticas avaliadas em um banco de germoplasma geram grande quantidade de informações sobre a diversidade genética.

A primeira etapa do processo de análise de dados de marcadores genéticos é a codificação em dados binários (no caso de marcadores dominantes) ou dados de coincidência alélica (no caso de marcadores co-dominantes). Para os dados de marcadores dominantes, geralmente, a codificação é: 1, para representar a presença do marcador e 0, para representar a ausência do marcador. Já para os dados de marcadores co-dominantes, geralmente, a codificação é: 1, para alelos comuns no *locus*,  $\frac{1}{2}$ , para um alelo comum no *locus* e 0, para a ausência de alelos comuns no *locus*.

A adequada aplicação dos métodos de Análise Multivariada exigiu tal codificação. Cada *locus* analisado é formado por dois alelos (o primeiro da mãe e o segundo do pai), pois nesse caso trata-se de genótipos ordenados e estes podem assumir os valores 1, 2 ou 3, de acordo com a interpretação obtida no gel de amido. A Tabela 3.1 descreve as 5 primeiras observações do conjunto de dados:

TABELA 3.1 – CONJUNTO DE DADOS ANALISADO (5 PRIMEIRAS OBSERVAÇÕES)

POP	SUBPOP	ARV	GOT-B		GOT-C		PGMA-A		MDH-B		SKDH-B		6PGDH-B	
			AL1	AL1	AL2	AL2	AL3	AL3	AL4	AL4	AL5	AL5	AL6	AL6
1	1	1	2	2	2	2	1	1	1	2	1	2	2	2
1	1	2	2	2	2	2	1	1	1	2	2	2	2	3
1	1	3	2	2	2	2	1	1	1	2	2	2	2	2
1	1	4	1	2	2	2	1	1	1	2	1	2	2	2
1	1	5	1	1	2	2	1	1	2	2	2	2	2	3

FONTE: O autor (2008)

em que:

POP: representa o número da população

SUBPOP: representa o número da subpopulação

ARV: indica o número da árvore amostrada

AL1, ..., AL6: representam os alelos presentes em cada *locus*

Após a codificação, cada *locus* passou a assumir o valor 1 (para alelos comuns em um *locus*) ou 0 (para alelos diferentes). Com isso pode-se representar a presença ou ausência de homozigose, respectivamente. As mesmas 5 primeiras observações do conjunto de dados original foram reescritas conforme a Tabela 3.2.



TABELA 3.2 – CONJUNTO DE DADOS CODIFICADO (5 PRIMEIRAS OBSERVAÇÕES)

POP	SUBPOP	ARV	GOT-B	GOT-C	PGMA-A	MDH-B	SKDH-B	6PGDH-B
1	1	1	1	1	1	0	0	1
1	1	2	1	1	1	0	1	0
1	1	3	1	1	1	0	1	1
1	1	4	0	1	1	0	0	1
1	1	5	1	1	1	1	1	0

FONTE: O autor (2008)

A segunda etapa na análise dos dados consiste em utilizar as informações codificadas para estimar os índices de similaridade ou de distância genética entre pares de observações. De acordo com a literatura existente sobre divergência genética, os coeficientes de similaridade mais utilizados e que foram analisados neste trabalho estão apresentados a seguir.

COEFICIENTES DE SIMILARIDADE	EXPRESSÃO
Jaccard (1912)	$\frac{a}{a+b+c}$
Sokal & Sneath (1963)	$\frac{a}{a+2(b+c)}$
Southwood (1978)	$\frac{\frac{a}{2}(2a+b+c)}{(a+b)(a+c)}$
Gower & Legendre (1986)	$\frac{a-(b+c)+d}{a+b+c+d}$
Holliday <i>et al.</i> (2002)	$\frac{ad-bc}{\sqrt{a+b+c+d(a+b)(a+c)}}$
Koleff <i>et al.</i> (2003)	$\log(2a+b+c) - \left( \frac{1}{2a+b+c} 2a \log 2 \right) - \left( \frac{1}{2a+b+c} ((a+b)\log(a+b) + (a+c)\log(a+c)) \right)$

QUADRO 3.3 – COEFICIENTES DE SIMILARIDADE UTILIZADOS NO TRABALHO

FONTE: O autor (2008)

A partir desses índices foram calculadas as respectivas matrizes de similaridade para as observações. Essas matrizes serviram como fonte de informação para as análises de agrupamento, de dispersão dos dados e construção dos dendogramas. De acordo com a literatura sobre divergência genética, os

métodos de ligação mais comuns são: método do vizinho mais próximo, método da ligação média e método do vizinho mais distante. Neste trabalho aplicou-se o método do vizinho mais próximo, o método da ligação média e o método de Ward. Dias (1998) descreve a aplicação dessas metodologias relatando as principais vantagens e desvantagens de cada uma nesse contexto.

Após esse procedimento, as análises foram conduzidas considerando 3 avaliações distintas.

Na primeira avaliação, as árvores presentes em cada amostra da população (e subpopulação) foram consideradas repetições. Por exemplo, a amostra da população de Campos do Jordão contém 70 repetições e as amostras das subpopulações de Caçador possuem 10 repetições cada. A mesma idéia é válida para as demais. Com isso foram formados vetores de observações.

Como a quantidade de árvores amostradas foi diferente em cada população, algumas árvores foram selecionadas de forma aleatória para que todos os vetores ficassem com a mesma dimensão. Levando-se em conta a menor amostra populacional (no caso Barbacena com 32 árvores), foram consideradas apenas 32 árvores de cada amostra populacional. O mesmo procedimento foi realizado para as subpopulações.

Na segunda avaliação, buscou-se representar as amostras das populações e subpopulações através de uma medida resumo. Tal representação deu-se através do vetor de médias amostral.

Finalmente, na terceira avaliação, as árvores foram comparadas uma a uma, tanto para as amostras das populações quanto para as subpopulações. Cada árvore recebeu uma numeração adequada, de acordo com a sua origem.

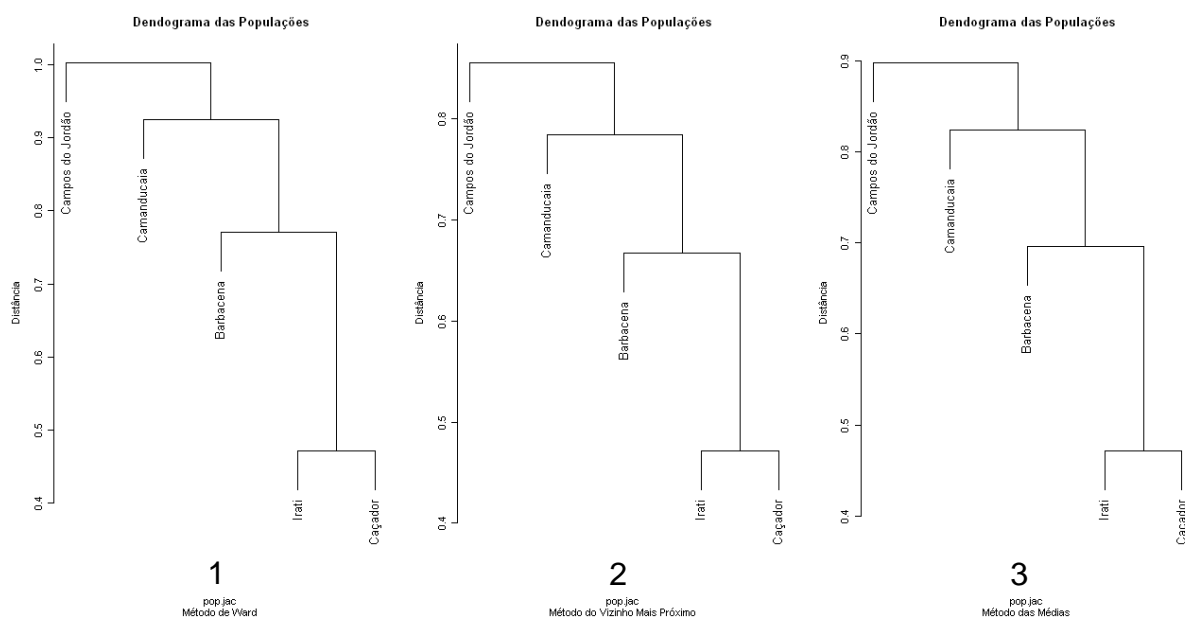
## 4 RESULTADOS E DISCUSSÃO

### 4.1 PRIMEIRA SITUAÇÃO DE ANÁLISE

#### i) Amostras das populações

Para a entrada dos dados no *software R*, os vetores de observações das amostras das populações possuem dimensão 192. A cada 6 características tem-se uma árvore. A partir da comparação desses vetores foram construídas as matrizes de similaridades considerando os 6 coeficientes de similaridade propostos e, com bases nelas, dendogramas, considerando os 3 métodos de ligação. Os resultados obtidos são apresentados nas figuras a seguir.

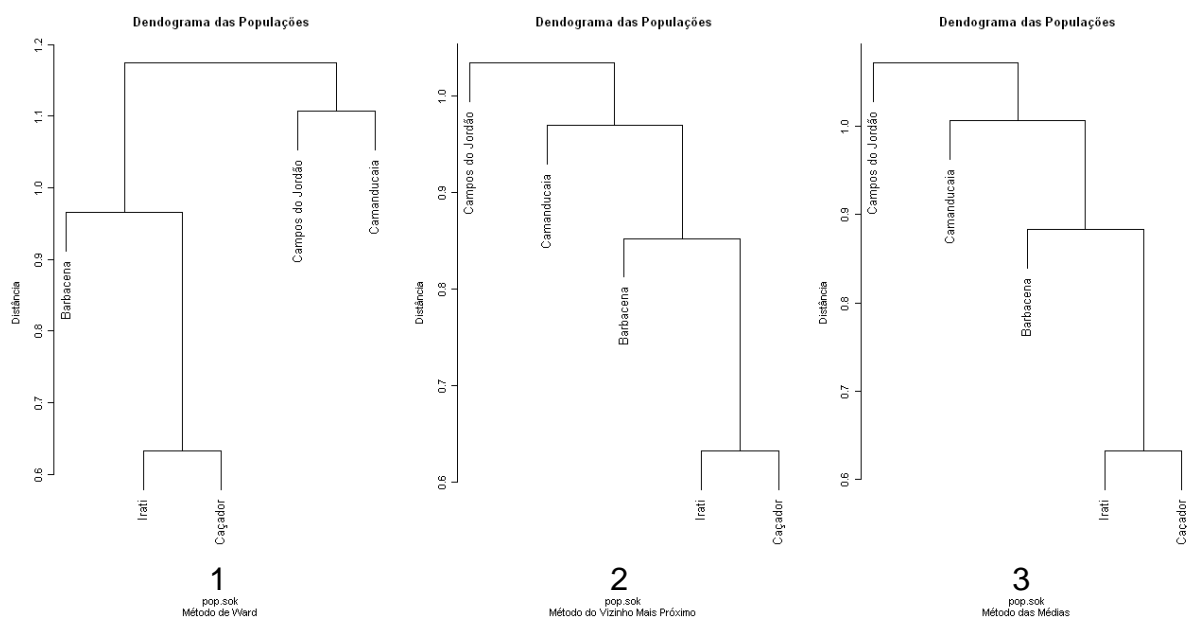
A Figura 4.1 representa os dendogramas das cinco amostras populacionais construídos pelos três métodos de ligação citados. Observa-se que a forma dos agrupamentos é a mesma, exceto na escala de distância. Verifica-se que a população de Irati é muito semelhante à de Caçador porque se agruparam com menor distância (0,4714). Barbacena foi a que mais se assemelhou ao grupo de Irati e Caçador; Campos do Jordão se assemelhou mais a Camanducaia nos três dendogramas, porém apenas um agrupamento pode ser observado.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.1 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO

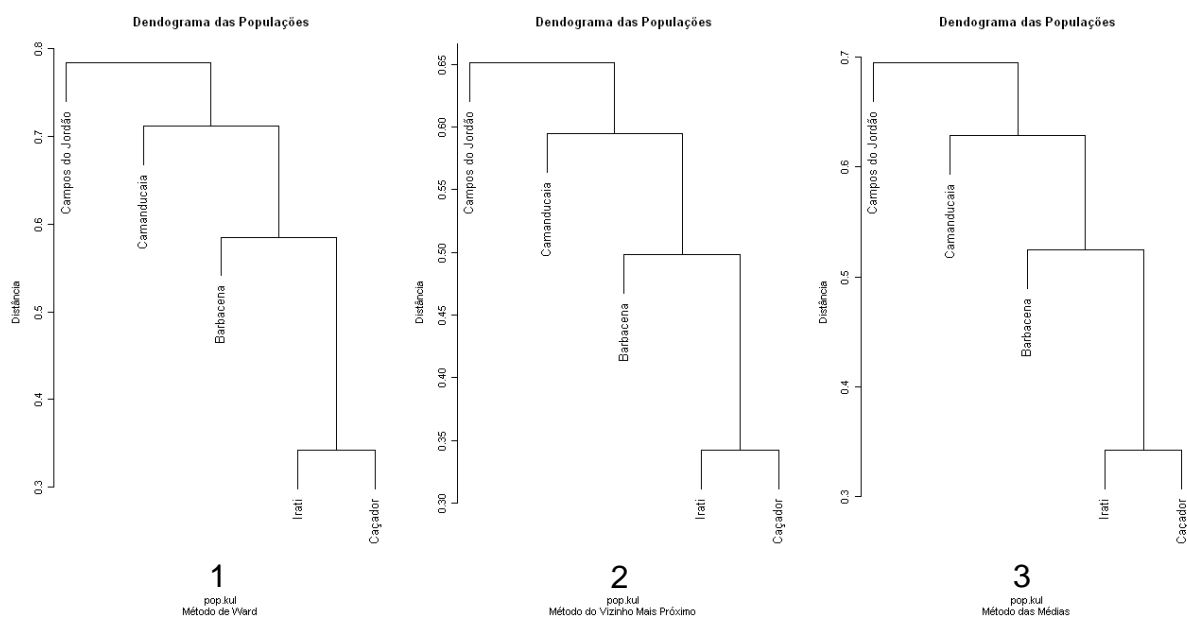
Na Figura 4.2 observa-se que o método de ligação de Ward proporciona o agrupamento das populações de Irati e Caçador (o mais semelhante com distância de 0,6324), e também o agrupamento das populações de Campos do Jordão e Camanducaia (com uma distância de 1,1071). Com os outros dois métodos de ligação, estas duas populações possuem uma distância pequena, mas não chegam a formar um agrupamento. Estes dendogramas apresentam a mesma forma da Figura 4.1, apenas com uma pequena diferença na escala das distâncias. Nos três dendogramas, Barbacena é mais semelhante ao grupo mais homogêneo.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.2 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO

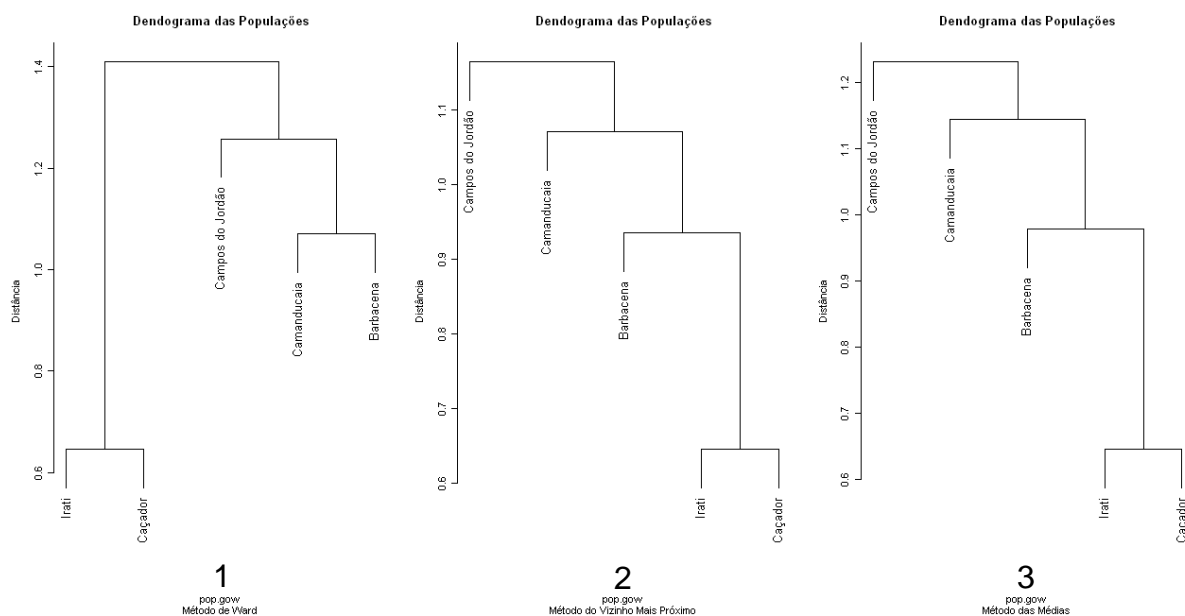
Os dendrogramas apresentados na Figura 4.3 permanecem com a mesma forma daqueles apresentados nas Figuras 4.1 e 4.2. O grupo formado por Irati e Caçador mantém-se o mais homogêneo porque a distância entre essas populações continua sendo a menor (0,3421).



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.3 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO

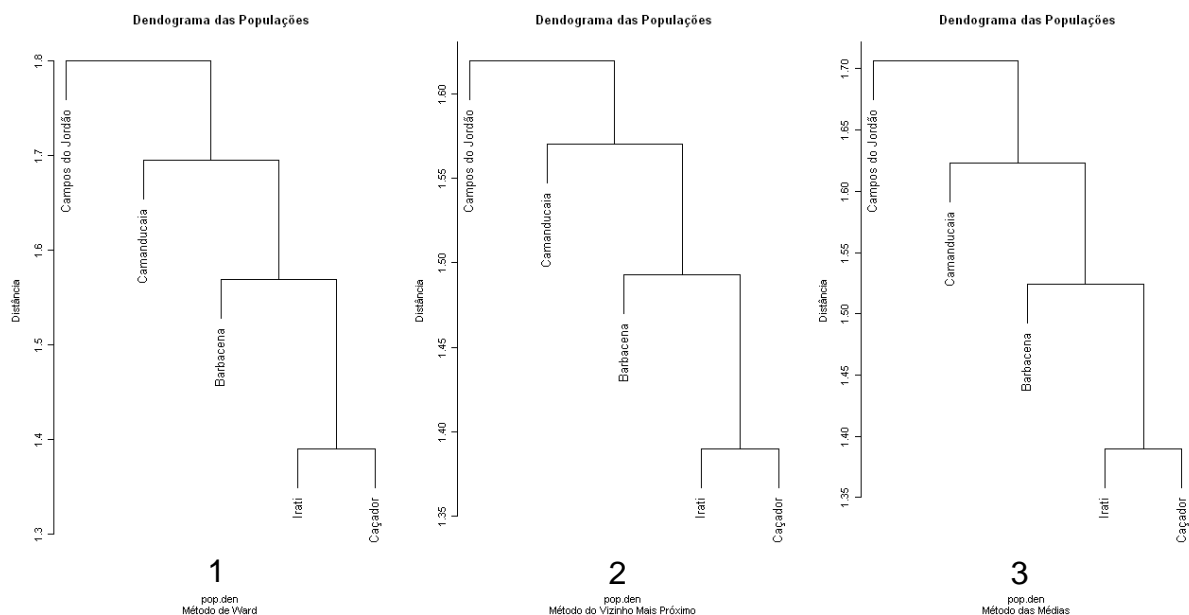
Na Figura 4.4 o método de ligação de Ward apresenta um dendograma com um grupo a mais que os outros. Este é formado pelas populações de Camanducaia e Barbacena, as quais são semelhantes nos outros dendogramas, porém não chegam a formar um agrupamento. Nos três dendogramas, Irati e Caçador formam mais uma vez o grupo com menor distância (0,6454).



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.4 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO

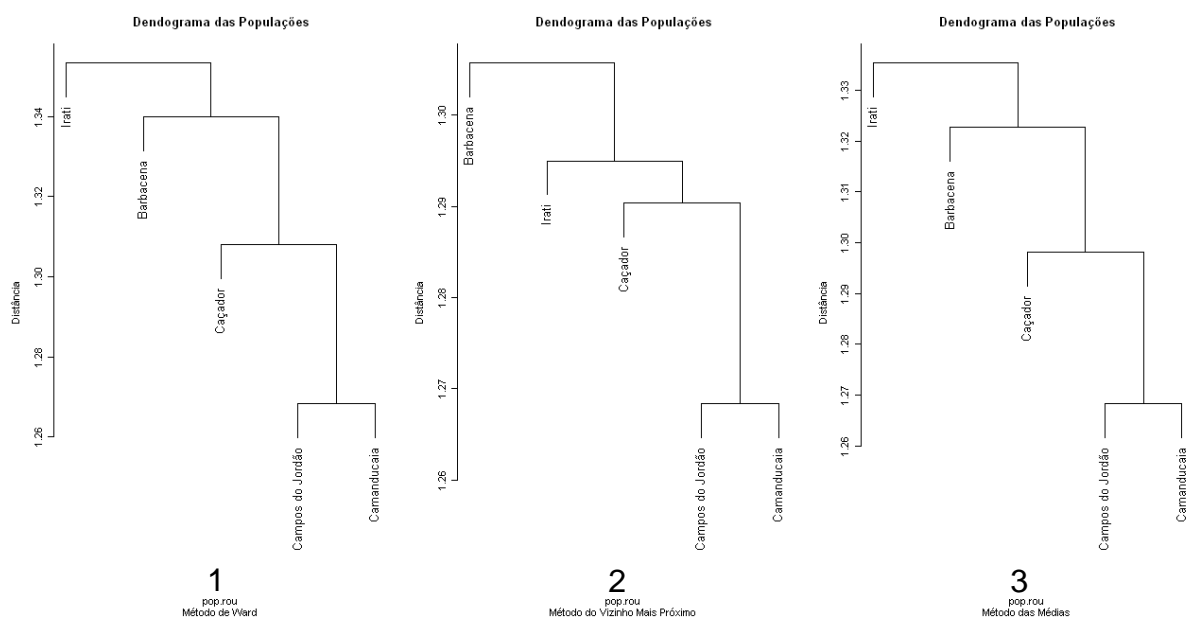
Os dendogramas da Figura 4.5 apresentam a mesma formação dos dendogramas obtidos nas Figuras 4.1 e 4.3.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.5 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO

Os dendogramas da Figura 4.6 não apresentam a mesma forma dos anteriores. O grupo mais semelhante, agora, é formado pelas populações de Campos do Jordão e Camanducaia, enquanto que nos dendogramas anteriores estas populações eram apenas semelhantes entre si e as mais diferentes das demais. Por não representar de forma satisfatória os agrupamentos, este coeficiente não será utilizado na construção dos demais dendogramas desta etapa.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.6 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE KOLEFF E 3 MÉTODOS DE AGRUPAMENTO

Analisando todos dendogramas obtidos até agora, verifica-se claramente o agrupamento das populações de Irati e Caçador, para qualquer coeficiente de similaridade e método de ligação. Em todos os casos a população de Barbacena é a que mais se assemelha a esse grupo, seguido de Camanducaia e Campos do Jordão, respectivamente. Estas duas são as que mais se diferenciam do restante, porém são semelhantes entre si. Geograficamente estas três populações estão muito próximas umas das outras: Barbacena e Camanducaia localizam-se em um lado da serra da Mantiqueira, enquanto que Campos do Jordão localiza-se do outro lado.

O método de Ward apresentou algumas pequenas variações na forma dos dendogramas ao agrupar as populações de Camanducaia e Barbacena, mas com os outros dois métodos de ligação estas duas populações apresentam-se bem semelhantes.

O processo de amostragem aleatória das árvores foi repetido para construção de novos dendogramas a fim de compará-los com os já obtidos. Os resultados (em Apêndices) foram os mesmos: o grupo mais semelhante continua sendo formado pelas populações de Irati e Caçador e as demais populações permaneceram com as mesmas distâncias. Isso garante que a estrutura da variabilidade genética não é afetada pelas árvores que fazem parte da amostra.

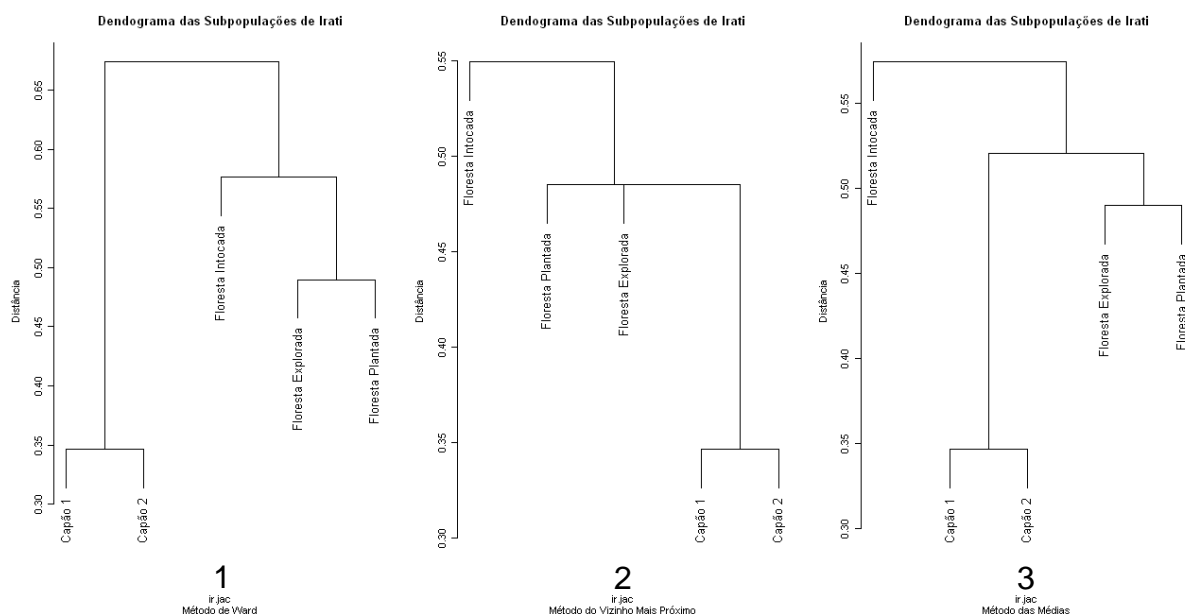


## ii) Subpopulações

No caso das subpopulações, somente as de Irati e Caçador foram comparadas entre si. As demais não foram comparadas devido ao pequeno número de subpopulações (como Campos do Jordão com duas) ou por não tê-las (como Camanducaia e Barbacena).

Em Irati, as subpopulações de Capão 1 e Floresta Intocada apresentam a menor quantidade de árvores (9 em cada). Por isso, uma amostragem aleatória foi realizada nas demais subpopulações para que todos os vetores de observações ficassem com dimensão 54. Os resultados da comparação entre eles são apresentados nas figuras a seguir.

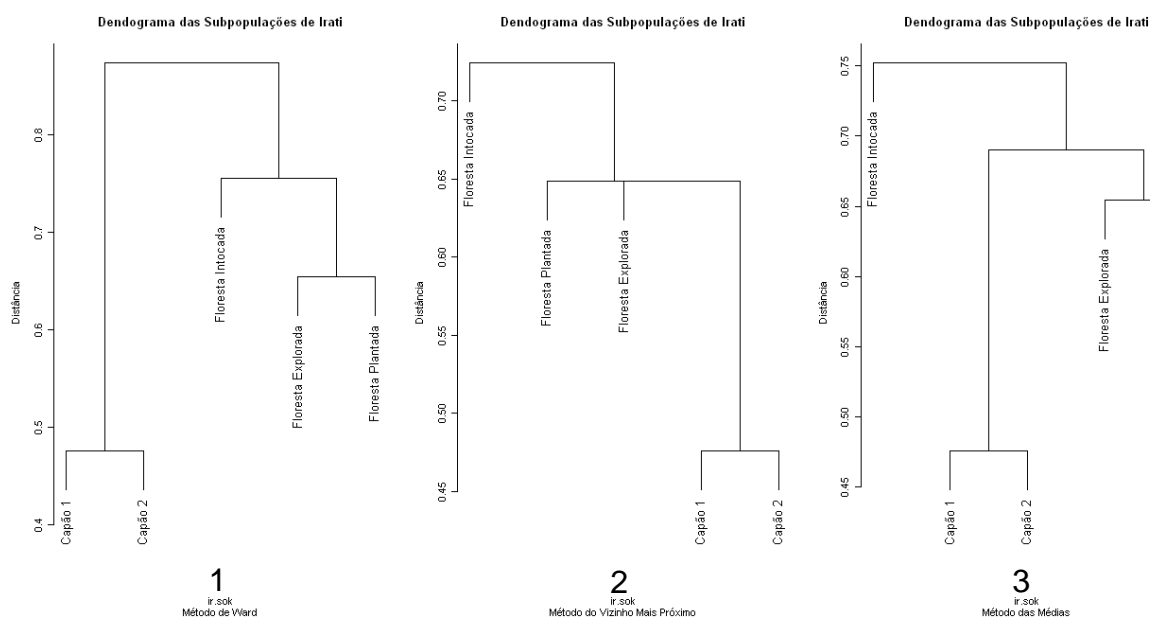
A Figura 4.7 mostra os dendogramas construídos com o coeficiente de similaridade de Jaccard. Pode-se observar que as subpopulações Capão 1 e Capão 2 formam o grupo de menor distância (0,3464). O segundo grupo mais homogêneo é formado pela Floresta Explorada e pela Floresta Plantada com uma distância de 0,4898. A subpopulação da Floresta Intocada foi a menos semelhante nos três dendogramas.



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.7 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO

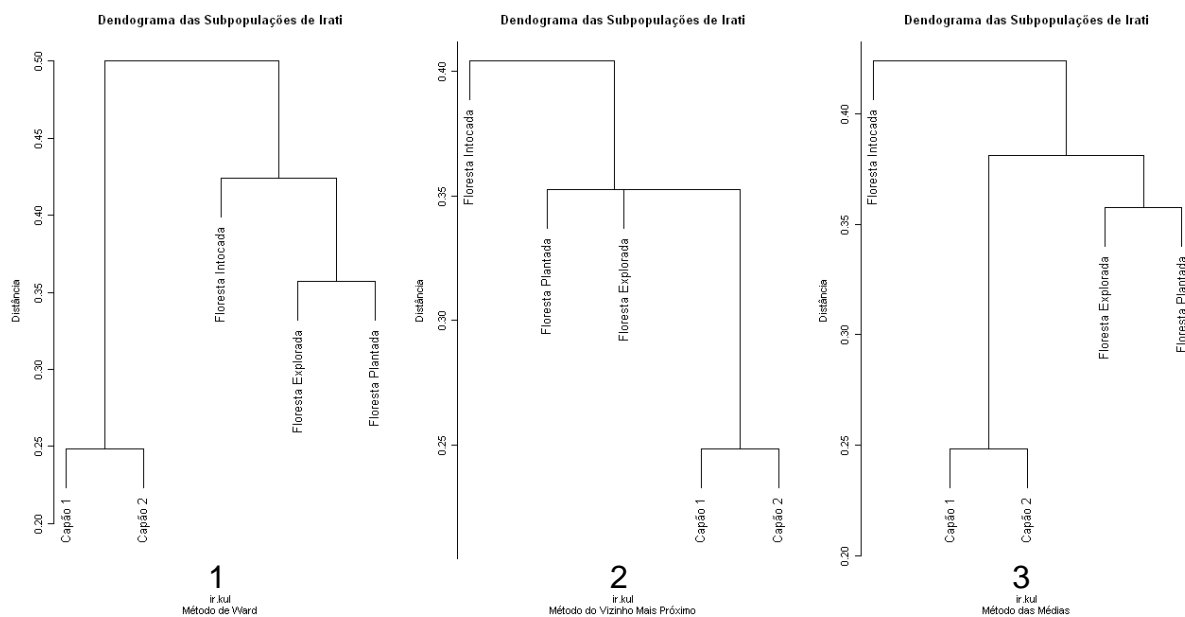
Na Figura 4.8 observa-se exatamente os mesmos agrupamentos dos dendogramas da Figura 4.7, apenas com diferença na escala de distância. O método de ligação de Ward apresenta um dendograma diferente dos demais: enquanto nele o grupo formado pela Floresta Explorada e pela Floresta Plantada se assemelha à Floresta Intocada para depois se unir ao grupo de Capão 1 e Capão 2, nos outros métodos de ligação o grupo de Floresta Explorada e Floresta Plantada se assemelha ao grupo de Capão 1 e Capão 2 para depois se unir à Floresta Intocada.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

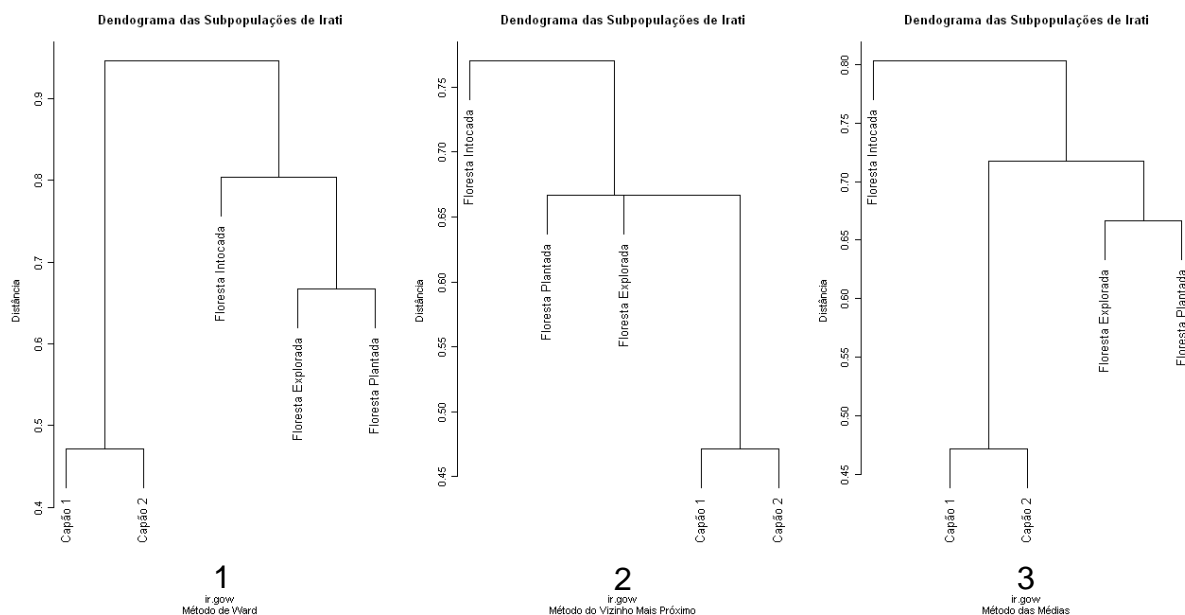
FIGURA 4.8 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO

Conforme a Figura 4.9, os três dendogramas gerados apresentam a mesma forma dos dendogramas das Figuras 4.7 e 4.8. Apenas neste caso a distância dos agrupamentos é um pouco menor: Capão 1 e Capão 2 possui uma distância de 0,2482 e o grupo de Floresta Explorada e Floresta Plantada apresenta a distância de 0,3572.

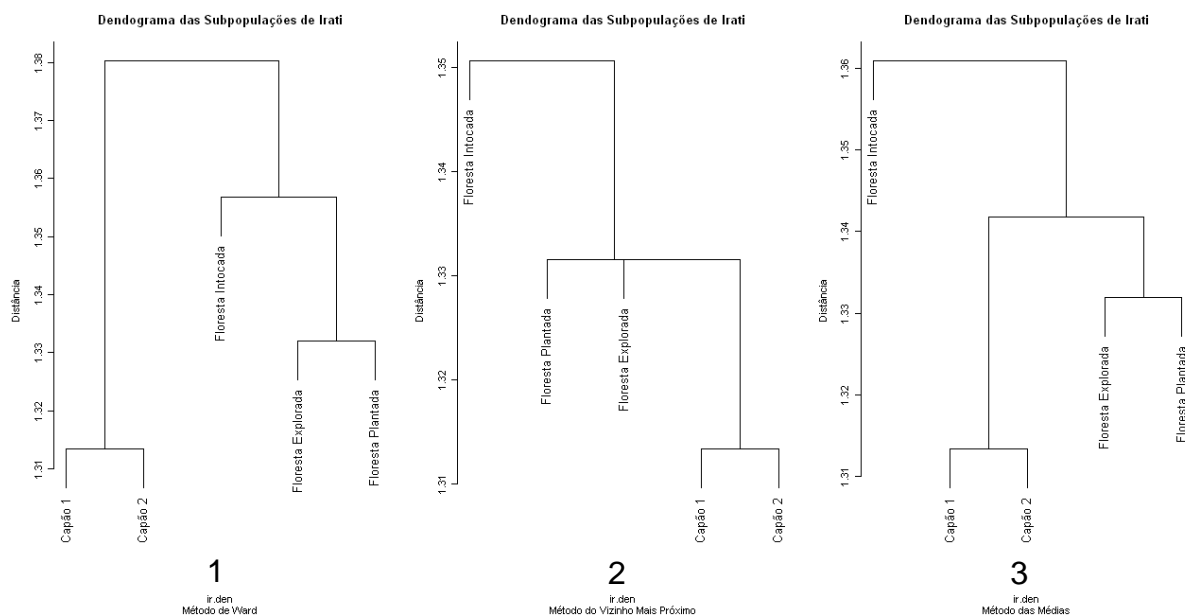


(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.  
 FIGURA 4.9 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO

Os dendogramas da Figura 4.10 permanecem com a mesma forma em relação aos dendogramas obtidos anteriormente (Figura 4.9).



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.  
 FIGURA 4.10 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.11 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO

Comparando os resultados obtidos desde a Figura 4.7 até a Figura 4.11 verificam-se dois agrupamentos: o primeiro, formado pelas subpopulações de Capão 1 e Capão 2 é o mais semelhante, e o segundo é formado pela Floresta Explorada e pela Floresta Plantada, considerando os coeficientes de similaridade e métodos de ligação propostos. A subpopulação menos semelhante das demais é a Floresta Intocada (resultado esperado), o que garante a identidade de sua variabilidade genética natural.

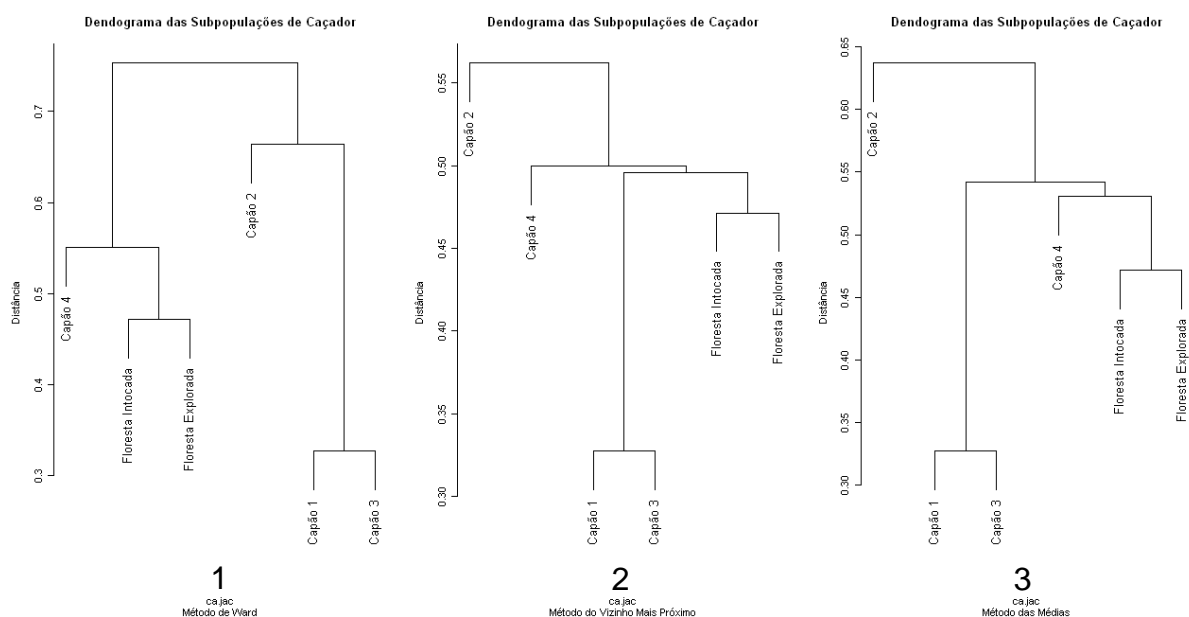
É importante observar que a Floresta Explorada possui as características genéticas semelhantes à Floresta Plantada porque no plantio das árvores (quer sejam oriundas de áreas mais ou menos conservadas) é desconsiderada a base genética. Com isso, a variabilidade genética de uma subpopulação é transmitida diretamente para a outra.

Cabe ressaltar que os coeficientes de similaridades utilizados até agora são baseados em empates (1-1 e 0-0) e desempates (1-0 e 0-1) presentes na comparação dos vetores de observações. Isto significa que a posição das árvores em um determinado vetor pode alterar essas quantidades e resultar em diferentes agrupamentos. Para verificar essa suposição, os vetores de observações foram reordenados e novos dendrogramas foram gerados (resultados em Apêndices).

Os dendogramas obtidos sob a condição de reorganização dos vetores mantêm a formação do grupo mais homogêneo formado pelas subpopulações de Capão 1 e Capão 2. Agora, as subpopulações de Floresta Plantada, Floresta Explorada e Floresta Intocada formam um novo agrupamento, enquanto que antes esta última subpopulação não pertencia a este grupo, mas assim mesmo era muito semelhante a ele. De uma forma geral, a estrutura dos agrupamentos não é alterada pela posição das árvores nos vetores de observação das subpopulações de Irati.

Em Caçador, todas as subpopulações contêm 10 árvores e, portanto, os vetores de observações apresentam dimensão 60. Os resultados da comparação desses vetores são apresentados nos dendogramas a seguir.

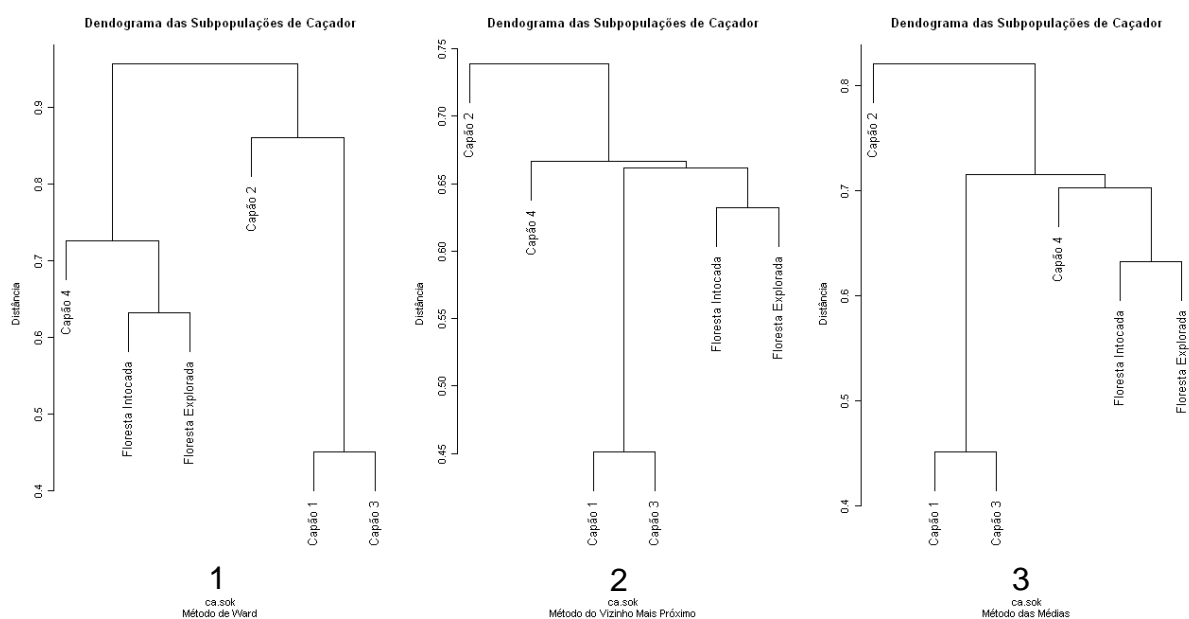
Conforme a Figura 4.12, as subpopulações de Capão 1 e Capão 3 são as mais semelhantes (com distância de 0,3273). No dendograma com o método de ligação de Ward o agrupamento formado pelas subpopulações de Floresta Intocada e Floresta Plantada se assemelha à subpopulação de Capão 4, que por sua vez se une à Capão 2, subpopulação mais semelhante ao grupo de Capão 1 e Capão 3. Com o método do vizinho mais próximo o grupo de Capão 1 e Capão 3 se assemelha ao grupo da Floresta Intocada e Floresta Explorada, para depois se unir à Capão 4 e finalmente à Capão 2. Já o método de ligação média gera um dendograma onde o grupo de Capão 1 e Capão 3 se assemelha à Capão 4, que por sua vez se assemelha ao grupo da Floresta Intocada e Floresta Explorada.



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.12 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO

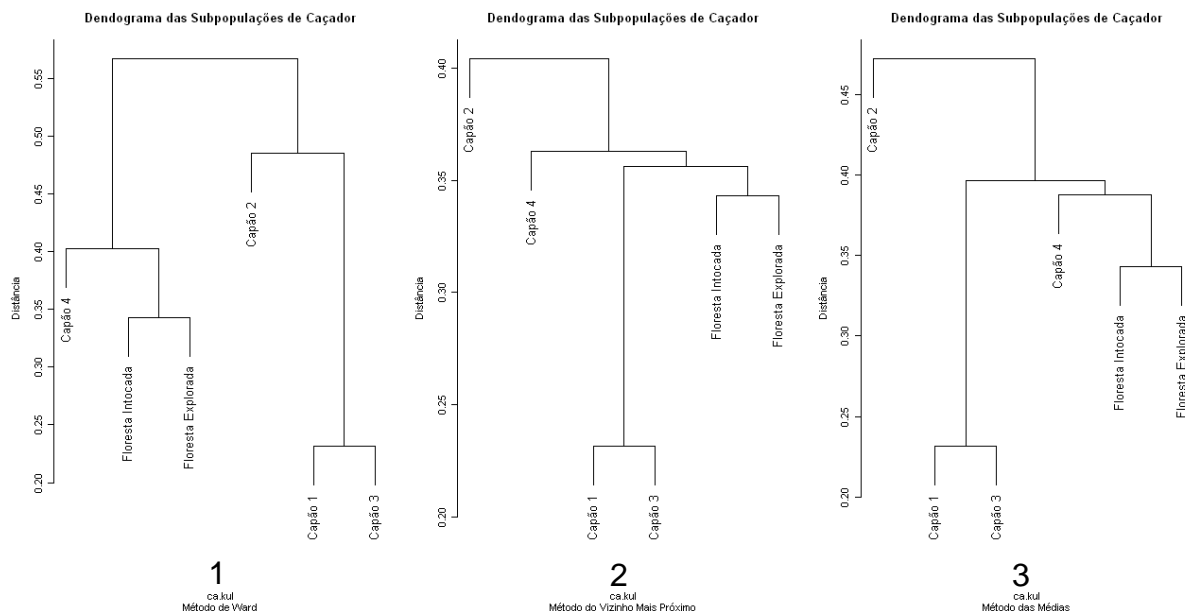
A Figura 4.13 apresenta os dendrogramas com a mesma forma da Figura 4.12, conforme cada método de ligação. Na comparação entre eles, os dois grupos mais semelhantes (Capão 1 e Capão 3; Floresta Intocada e Floresta Explorada) permanecem com a mesma formação. Apenas a escala de distância muda um pouco em relação à Figura 4.12.



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.13 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO

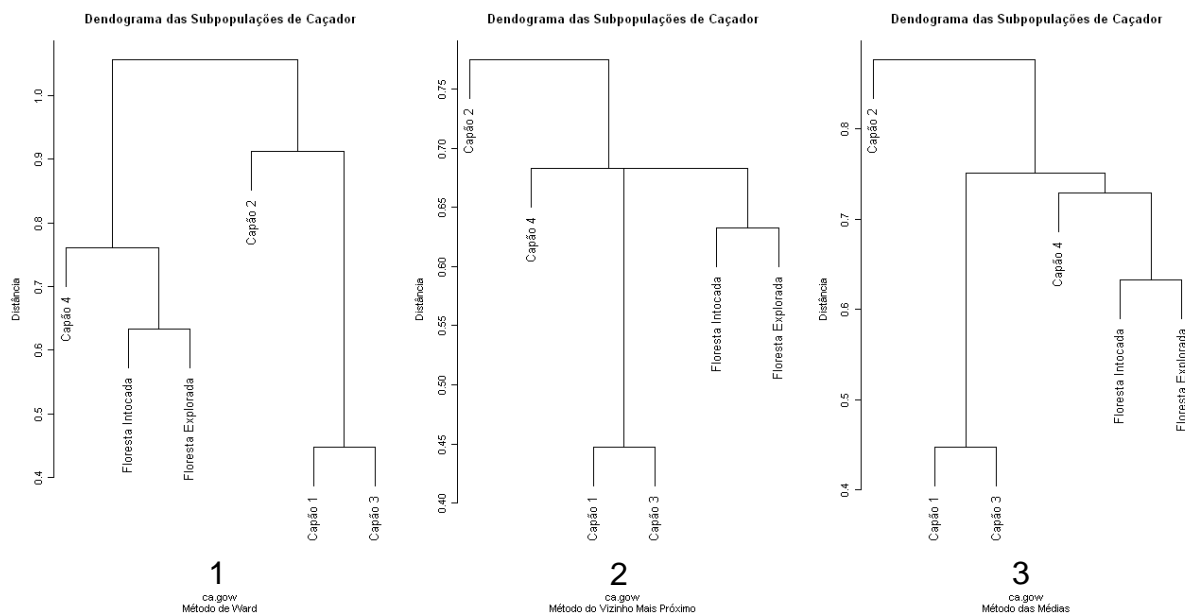
A escala da distância apresentada na Figura 4.14 é um pouco menor que a da escala dos dendogramas observados nas figuras anteriores, mas a forma dos agrupamentos permanece a mesma.



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

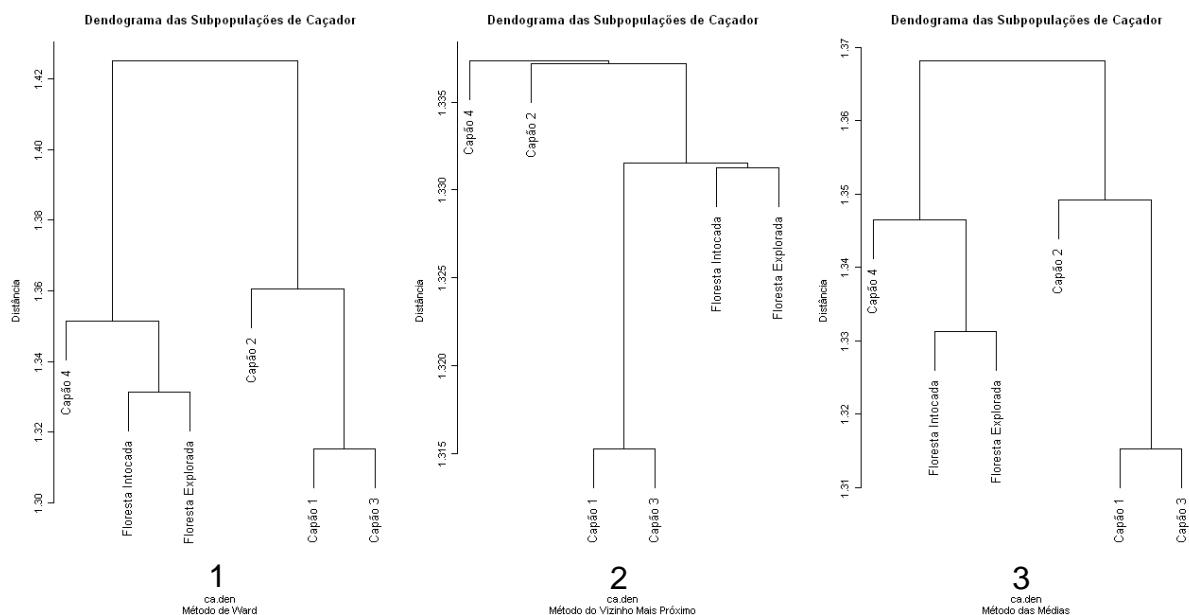
FIGURA 4.14 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO

Os dendogramas da Figura 4.15 apresentam os mesmo agrupamentos observados até agora, respectivamente aos métodos de ligação utilizados.



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.15 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.16 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO

De acordo com os dendogramas observados da Figura 4.12 até a Figura 4.16, verifica-se que as subpopulações de Capão 1 e Capão 3 formam o grupo mais semelhante, para qualquer coeficiente de similaridade e método de ligação utilizado. O outro grupo mais semelhante é composto pelas subpopulações de Floresta Intocada e Floresta Explorada, sendo Capão 4 a mais semelhante a este grupo.

Mais uma vez a posição das árvores foi alterada em cada vetor de observações e novos dendogramas foram gerados (em Apêndices). Os resultados foram similares: os dendogramas apresentaram um agrupamento formado pelas subpopulações de Capão 1 e Capão 3 (observado nas figuras anteriores) seguido pela Floresta Intocada e Floresta Explorada, as quais eram apresentadas, até agora, em um outro grupo distinto. Capão 2 continuou sendo a subpopulação menos semelhante das demais e apenas o coeficiente de similaridade de Holliday apresentou agrupamentos bem diferentes dos demais.

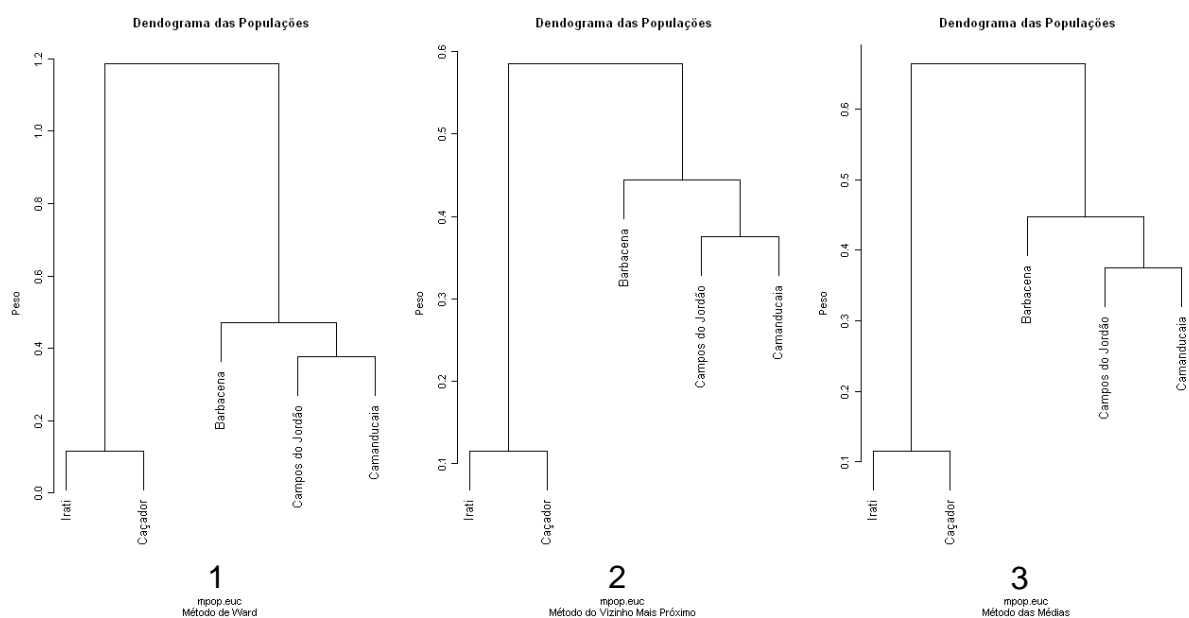


## 4.2 SEGUNDA SITUAÇÃO DE ANÁLISE

### i) Amostras das populações

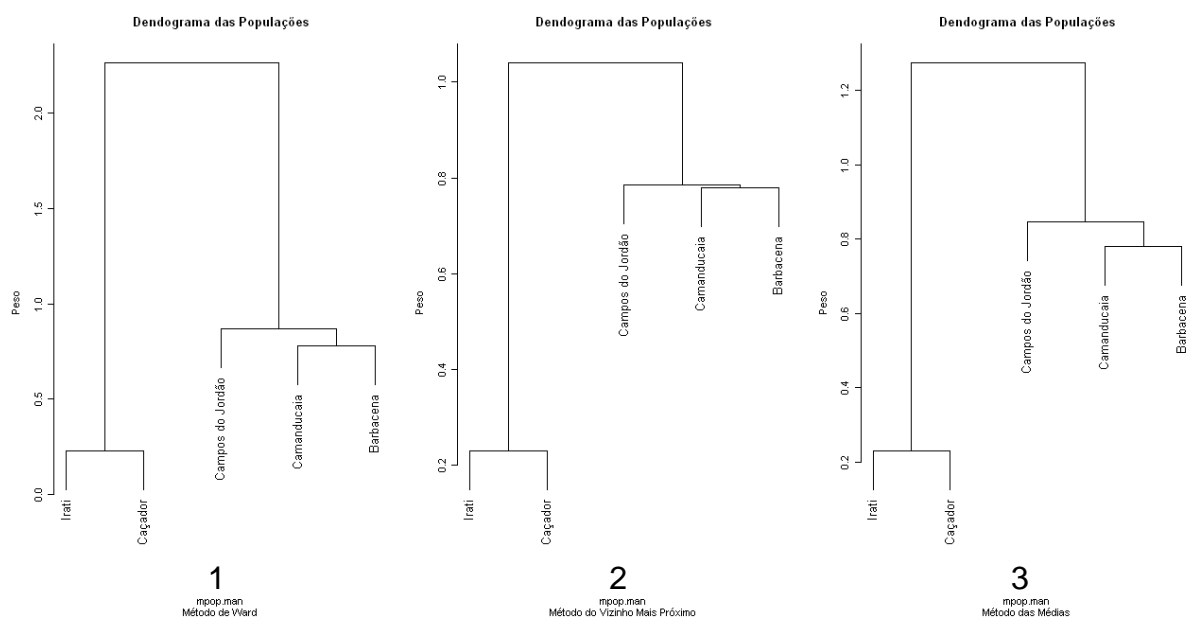
Neste estágio, os vetores de médias das amostras populacionais assumem valores diferentes de 0 ou 1. Usando a distância Euclidiana e a de Manhattan obtém-se as respectivas matrizes de distância e posteriormente os dendogramas, mantendo os três métodos de ligação. Os dendogramas correspondentes são apresentados nas próximas figuras.

Como pode ser observado na Figura 4.17, os três dendogramas apresentam a mesma formação. O agrupamento mais homogêneo é composto pelas populações de Irati e Caçador, pois apresentam a menor distância (0,1152). Outro grupo observado dá-se pela união das populações de Campos do Jordão e Camanducaia, com distância de 0,3755. Barbacena, apesar de não ter se unido a nenhum dos dois grupos formados, é mais semelhante a este último grupo.



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.17 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO A DISTÂNCIA EUCLIDIANA E 3 MÉTODOS DE AGRUPAMENTO



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.18 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO A DISTÂNCIA DE MANHATTAN E 3 MÉTODOS DE AGRUPAMENTO

Ao utilizar tanto a distância Euclidiana quanto a de Manhattan, os resultados foram semelhantes entre si e semelhantes aos já obtidos na primeira situação de análise. Irati e Caçador continuam a formar o grupo mais semelhante e Campos do Jordão, Camanducaia e Barbacena formam outro grupo.

Para validar os agrupamentos formados, foi realizado um teste  $T^2$  de Hotelling comparando se existe diferença significativa entre o vetor médio do grupo de Irati e Caçador com o vetor médio do grupo de Campos do Jordão, Camanducaia e Barbacena. A questão da Gaussianidade dos vetores aleatórios observados foi verificada aplicando-se o teste baseado em uma estatística qui-quadrado descrito no Capítulo 2. Aceitando-se a premissa de igualdade das matrizes de covariâncias, o teste foi realizado ao nível de 95% de confiança. Como o p-valor encontrado é inferior a 0,05 rejeita-se a hipótese nula; logo, existe diferença significativa entre os vetores médios dos grupos observados, porém com a ressalva de que a premissa de igualdade das matrizes de covariância não pode ser avaliada devido à natureza binária dos dados.

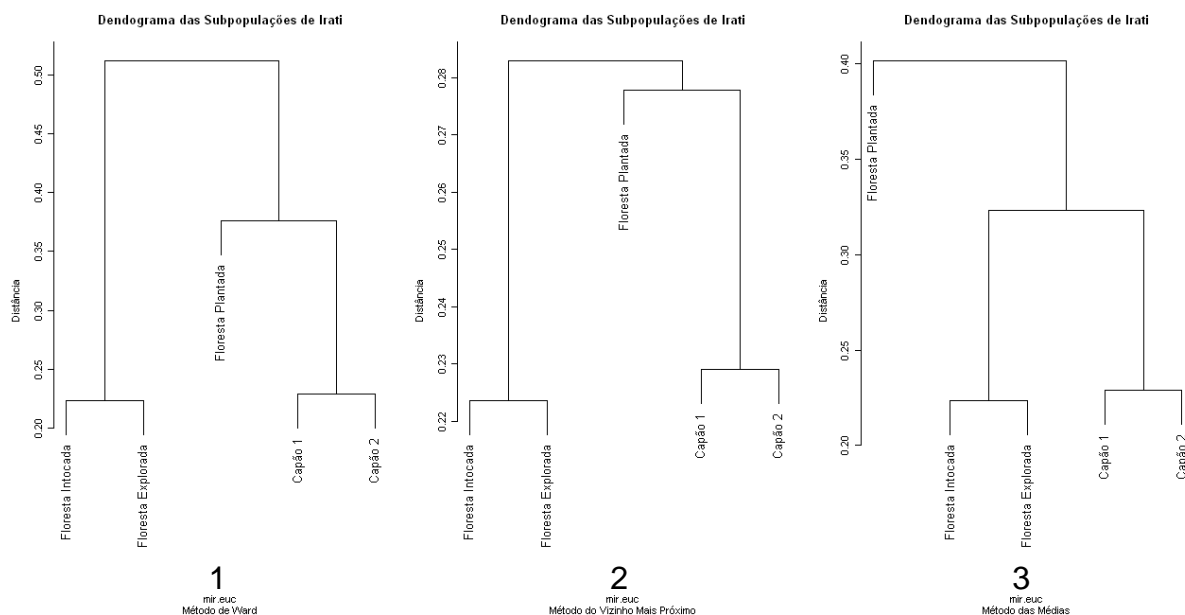
Além dessa comparação, também foi realizada uma análise dentro dos grupos. Entre Irati e Caçador, o teste  $T^2$  de Hotelling forneceu um p-valor inferior a 0,05, indicando que existem evidências de que os vetores médios dessas populações diferem significativamente, porém com a mesma ressalva da igualdade

das matrizes de covariância. Já a MANOVA (para o grupo de Campos do Jordão, Camanducaia e Barbacena) forneceu um p-valor superior a 0,05, indicando que não existe diferença significativa entre os vetores médios dessas populações. Para tornar a realização do teste  $T^2$  de Hotteling possível, neste último caso, foi necessário utilizar a pseudo-inversa da matriz de covariância conjunta através da transformação de Moore-Penrose, já que a matriz  $S_p$  é singular.

## ii) Subpopulações

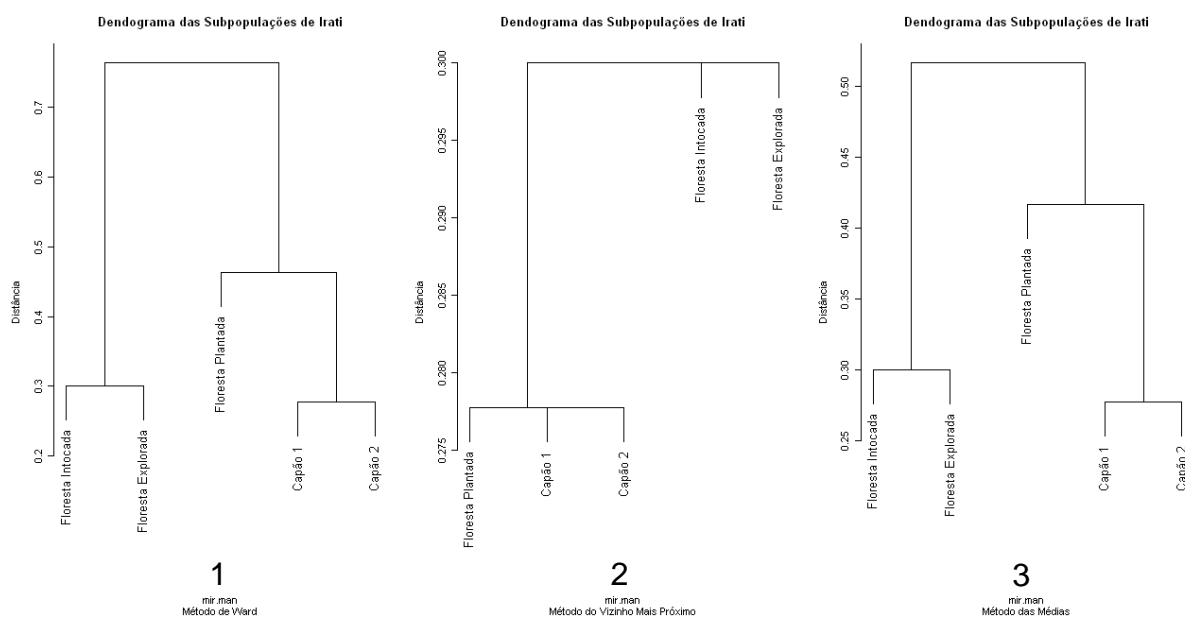
Os dendogramas obtidos na comparação dos vetores de médias subpopulacionais de Irati são apresentados a seguir.

De acordo com a Figura 4.19 observa-se que as subpopulações de Floresta Intocada e Floresta Explorada são as mais semelhantes e formam um agrupamento, uma vez que apresentam a menor distância (0,2236). Capão 1 e Capão 2 formam o segundo grupo com menor distância (0,2290). O dendograma gerado com o método de ligação média apresenta uma pequena diferença dos demais: os dois maiores grupos unem-se para somente depois se unirem à subpopulação de Floresta Plantada, enquanto que nos outros dendogramas a Floresta Plantada primeiro se une ao grupo de Capão 1 e Capão 2 para depois se unir ao grupo de Floresta Intocada e Floresta Explorada.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.  
 FIGURA 4.19 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO A DISTÂNCIA EUCLIDIANA E 3 MÉTODOS DE AGRUPAMENTO

Os dois agrupamentos formados na Figura 4.20 foram os mesmos observados na primeira situação de análise, mas a semelhança da subpopulação de Floresta Plantada ao grupo dos Capões 1 e 2 não se verificou. Na primeira situação de análise ela fazia parte do agrupamento com a Floresta Explorada.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.  
 FIGURA 4.20 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO A DISTÂNCIA DE MANHATTAN E 3 MÉTODOS DE AGRUPAMENTO

De acordo com as Figuras 4.19 e 4.20, os dendogramas mantiveram a formação dos dois agrupamentos mais semelhantes observados na primeira situação de análise, mesmo com uma pequena variação na escala da distância.

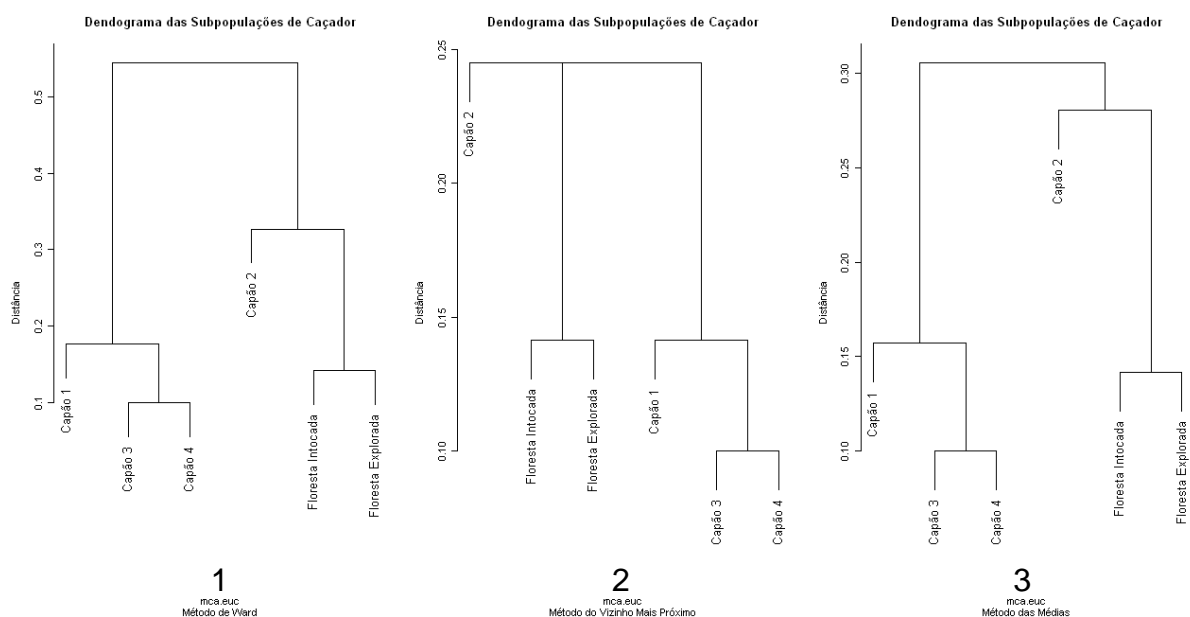
O vetor médio das subpopulações de Floresta Intocada e Floresta Explorada foi comparado ao vetor médio das subpopulações de Capão 1, Capão 2 e Floresta Plantada através do teste  $T^2$  de Hotteling, com 95% de confiança. O p-valor do teste encontrado é inferior a 0,05, indicando uma diferença significativa entre eles, porém com a ressalva de que a premissa de igualdade das matrizes de covariância não pode ser avaliada.

Na comparação das subpopulações de Capão 1 e Capão 2 o p-valor do teste é superior a 0,05. Logo, não existe diferença significativa entre os vetores médios dessas subpopulações. Na comparação do outro grupo formado pelas subpopulações de Floresta Intocada e Floresta Explorada, o teste  $T^2$  de Hotteling também forneceu um p-valor superior a 0,05 e mais uma vez a hipótese nula não foi rejeitada.

Em todos esses testes foi utilizada a pseudo-inversa da matriz de covariância conjunta através da transformação de Moore-Penrose.

Para as subpopulações de Caçador, os dendogramas são apresentados a seguir.

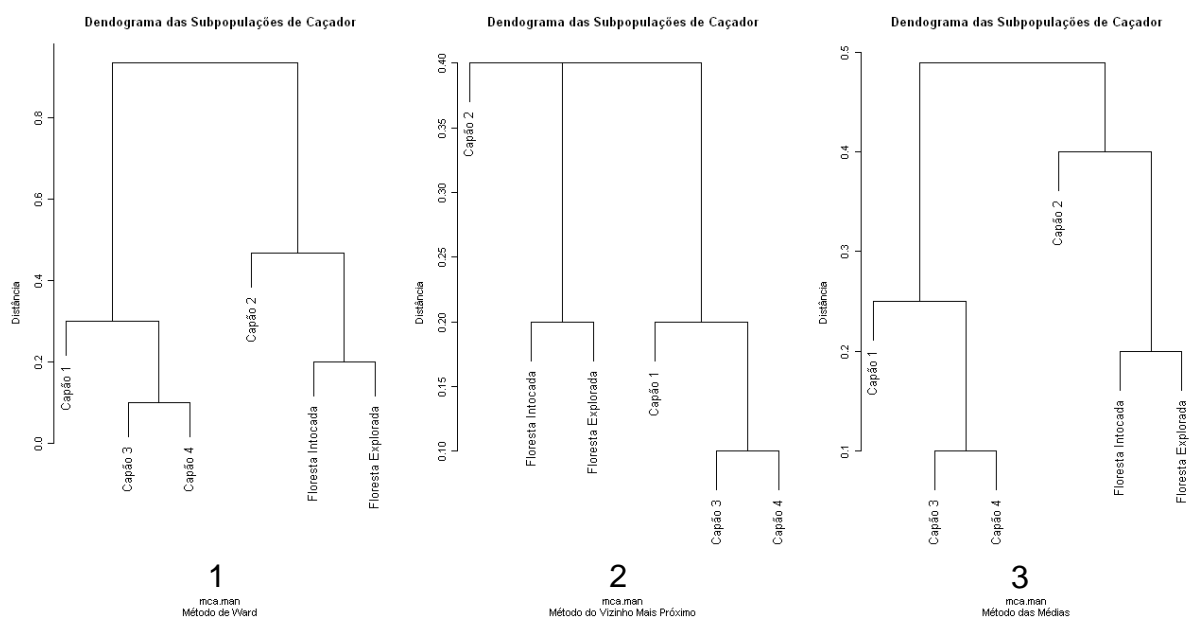
De acordo com os dendogramas da Figura 4.21 verifica-se que a subpopulação de Capão 1 se assemelha ao grupo de menor distância formado por Capão 3 e Capão 4, enquanto que na primeira situação de análise Capão 1 e Capão 3 formavam o grupo mais semelhante. O segundo grupo mais homogêneo é formado pelas subpopulações de Floresta Intocada e Floresta Explorada e este resultado também foi observado na primeira situação de análise, porém a semelhança de Capão 1 ao grupo mais homogêneo e Capão 2 ao segundo mais homogêneo não foi observada.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.21 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO A DISTÂNCIA EUCLIDIANA E 3 MÉTODOS DE AGRUPAMENTO

Conforme a comparação dos dendogramas das Figuras 4.21 e 4.22, nem todos os agrupamentos obtidos na primeira situação de análise foram os mesmos observados nesta etapa.



(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.22 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO A DISTÂNCIA DE MANHATTAN E 3 MÉTODOS DE AGRUPAMENTO

Mais uma vez a validação dos agrupamentos encontrados deu-se através do teste  $T^2$  de Hotelling, utilizando a pseudo-inversa da matriz de covariância conjunta. O vetor médio das subpopulações de Capão 1, Capão 3 e Capão 4 foi comparado ao vetor médio das subpopulações de Capão 2, Floresta Intocada e Floresta Explorada. O p-valor do teste encontrado é inferior a 0,05, indicando uma diferença significativa entre eles, porém com a ressalva de que a premissa de igualdade das matrizes de covariância não pode ser avaliada.

Na comparação dentro dos grupos, tanto o primeiro teste (realizado para comparar as subpopulações de Capão 3 e Capão 4) quanto o segundo (realizado para comparar as subpopulações de Floresta Intocada e Floresta Explorada) forneceram um p-valor maior que 0,05, logo as hipóteses nulas não foram rejeitadas e, portanto, os vetores de médias das subpopulações de cada grupo não diferem entre si.

#### 4.3 TERCEIRA SITUAÇÃO DE ANÁLISE

##### i) Amostras das populações

No mapeamento da divergência genética entre as amostras das populações, as árvores foram numeradas de acordo com a população a que pertenciam, conforme a seguinte tabela:

TABELA 4.1 – NUMERAÇÃO DAS ÁRVORES DAS AMOSTRAS DAS POPULAÇÕES

POPULAÇÃO	NUMERAÇÃO DAS ÁRVORES	NÚMERO DE ÁRVORES
Campos do Jordão	1	70
Irati	2	48
Caçador	3	60
Camanducaia	4	70
Barbacena	5	32

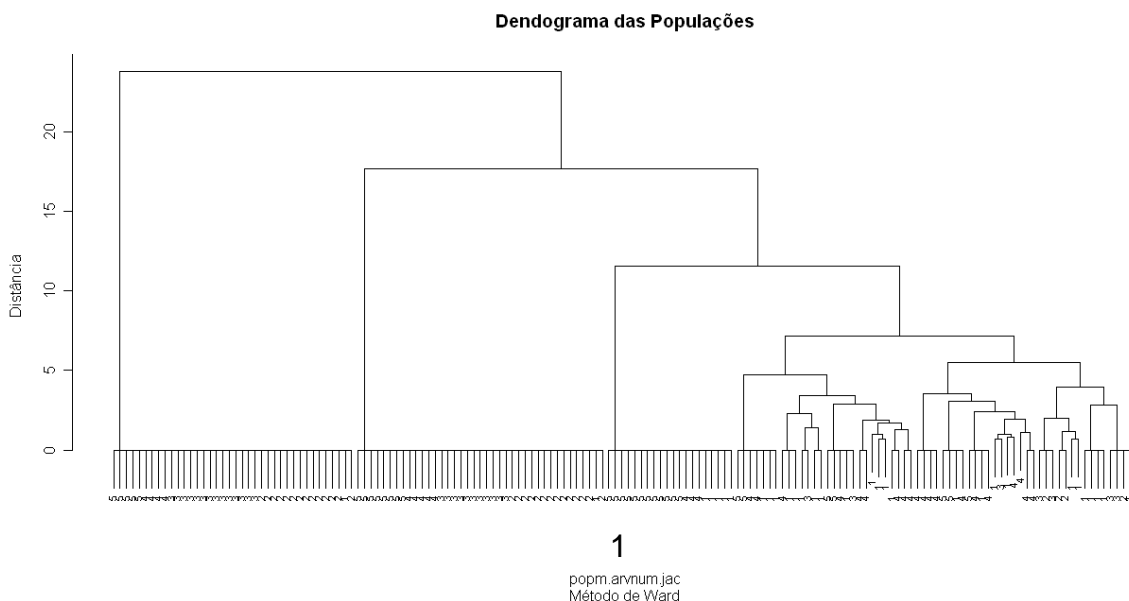
FONTE: O autor (2008)

Foram consideradas 32 árvores sorteadas de forma aleatória em cada amostra populacional a fim de igualar os tamanhos amostrais. Nos dendogramas

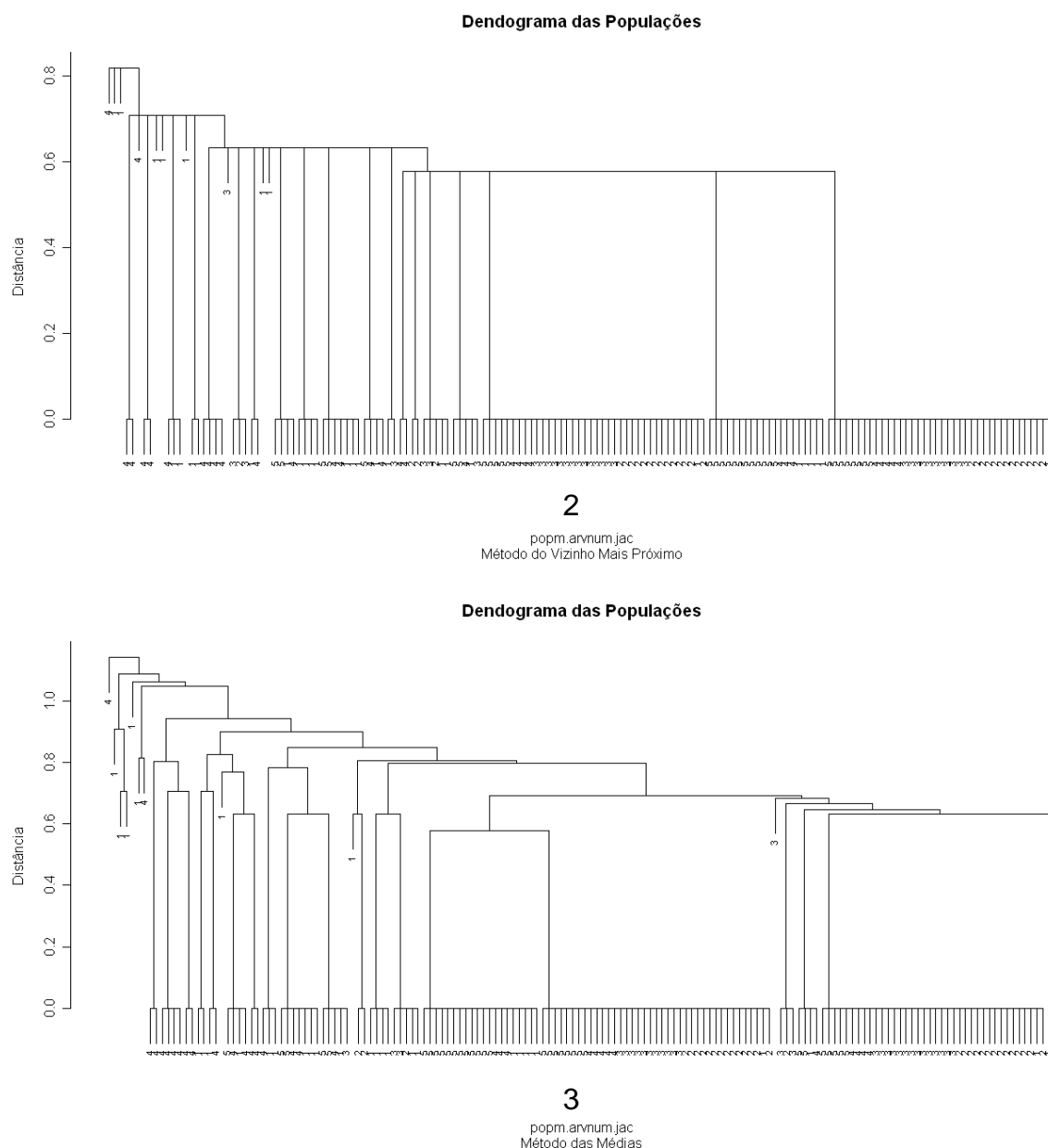
gerados buscou-se visualizar a formação de agrupamentos, identificando a numeração das árvores e suas respectivas populações.

Os dendogramas obtidos nessas condições estão presentes na Figuras 4.23 à 4.28.

Utilizando o coeficiente de Jaccard e os métodos de ligação de Ward e do vizinho mais próximo não é possível distinguir qualquer agrupamento, conforme a Figura 4.23. Somente com o método da ligação média observa-se um agrupamento no canto esquerdo do gráfico formado por algumas árvores de números 1 e 4, ou seja, por árvores de Campos do Jordão e de Camanducaia. Esse agrupamento foi observado na segunda situação de análise e, na primeira situação de análise, essas populações eram muito semelhantes. As demais populações não formam grupos adequadamente indicando que a variabilidade genética entre as árvores é muito grande e, dessa forma, dificultando a distinção das populações mais semelhantes entre si.







(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.23 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO

Os métodos de Ward e do vizinho mais próximo foram desconsiderados nos próximos dendogramas por não apresentarem a formação de agrupamentos.

Conforme a Figura 4.24, é possível observar um agrupamento no canto esquerdo do gráfico formado por árvores de Campos do Jordão e de Camanducaia. Tal dendograma é muito parecido com o da Figura 4.23, apesar do coeficiente de similaridade utilizado ser diferente.

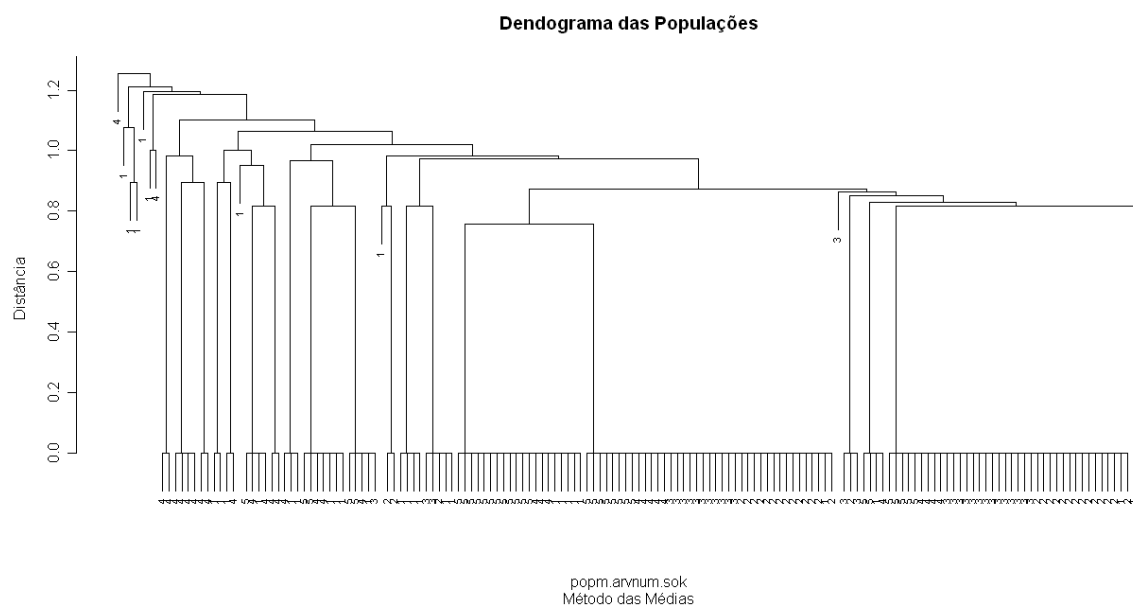


FIGURA 4.24 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH

Mais uma vez, o dendrograma da Figura 4.25 apresenta apenas um agrupamento de árvores numeradas 1 e 4, ou seja, por árvores de Campos do Jordão e de Camanducaia.

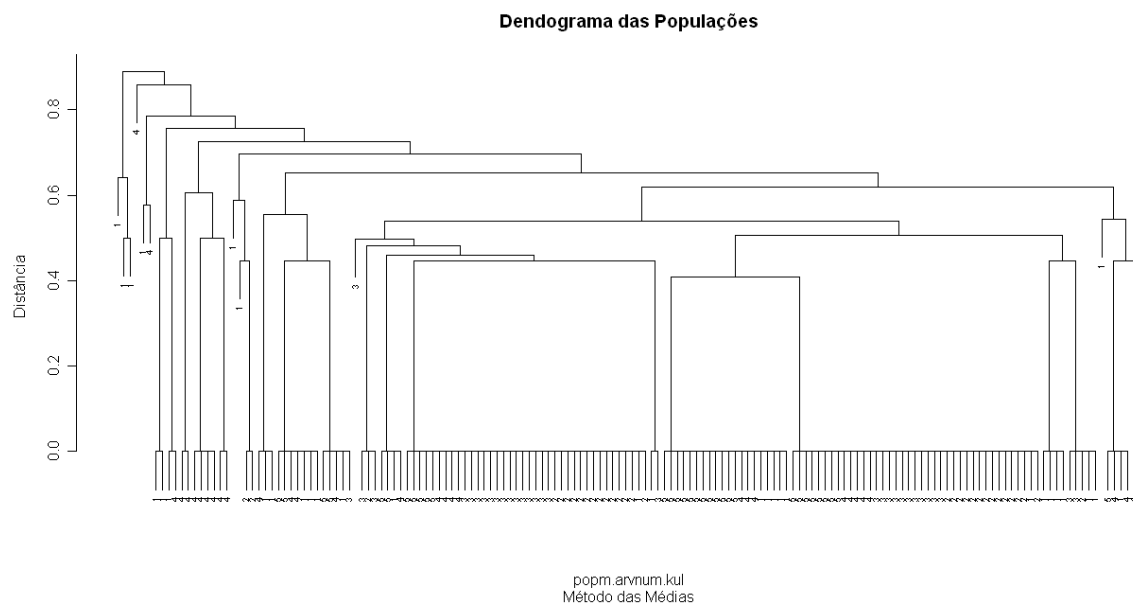


FIGURA 4.25 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE SOUTHWOOD

Observa-se no dendograma da Figura 4.26 que não é possível distinguir qualquer agrupamento; árvores de todas as populações estão espalhadas pelo dendograma. Um pequeno grupo formado por árvores de número 1 localizado no canto direito do gráfico mostra a semelhança entre as árvores da população de Campos do Jordão.

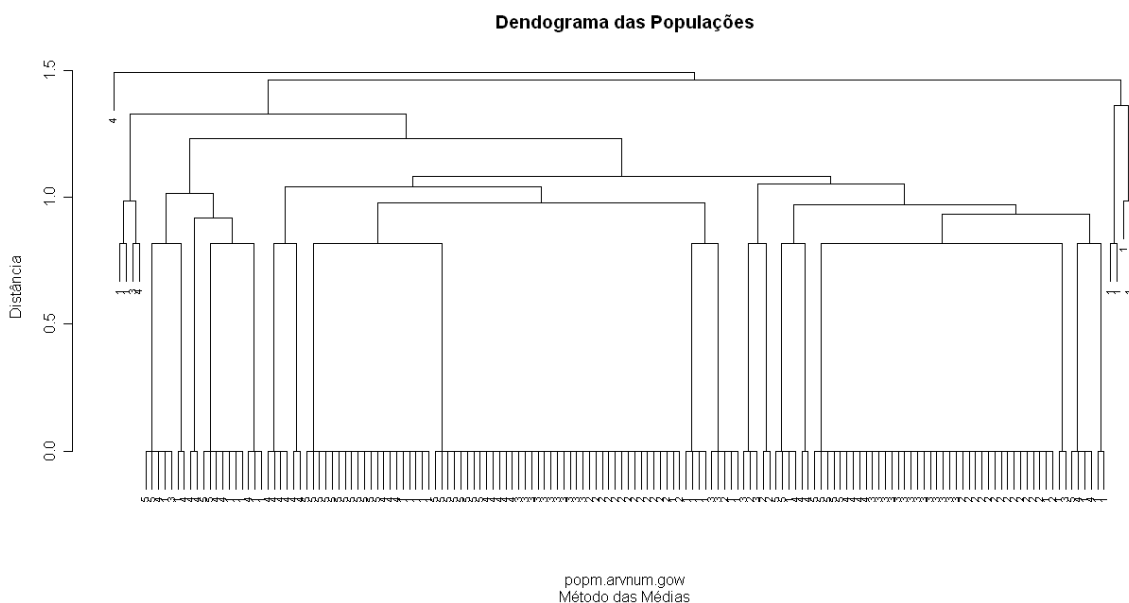


FIGURA 4.26 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE

Utilizando o coeficiente de similaridade de Holliday verifica-se um agrupamento formado pelas árvores de Campos do Jordão, conforme a Figura 4.27. Porém, no restante do dendograma observa-se a presença de árvores de todas as 5 populações, impedindo a distinção de outros agrupamentos.

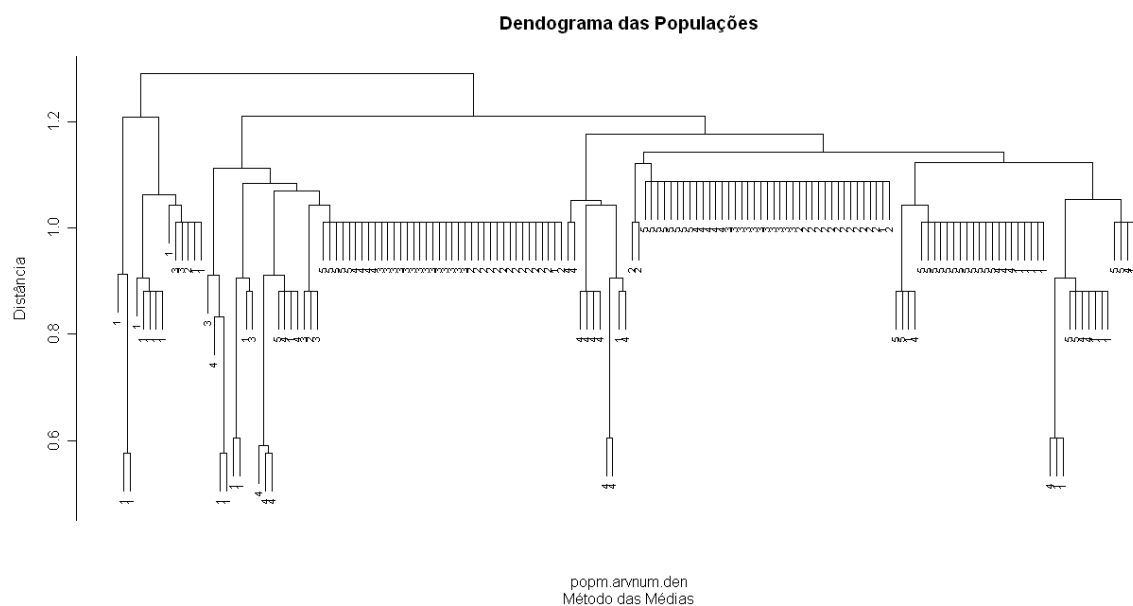


FIGURA 4.27 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE HOLLIDAY

Na Figura 4.28 verifica-se, mais uma vez, um agrupamento formado pelas populações 1 e 4, ou seja, Campos do Jordão e Camanducaia.

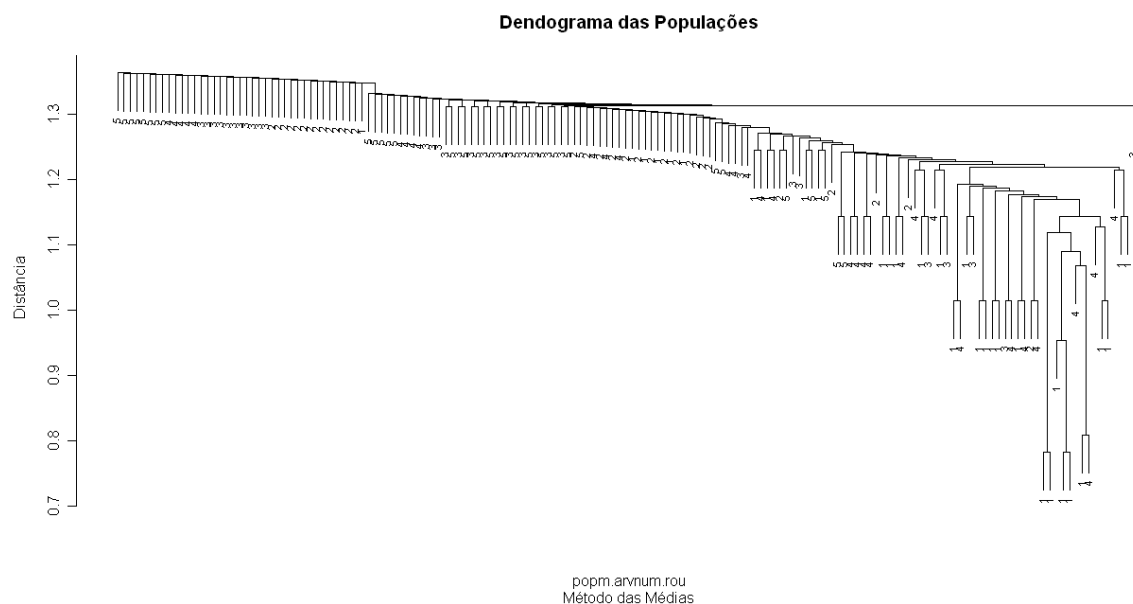


FIGURA 4.28 – DENDOGRAMA DAS POPULAÇÕES CONSIDERANDO O COEFICIENTE DE KOLEFF

De uma forma geral, esta maneira de analisar os dados proporcionou a identificação de apenas um agrupamento formado pelas árvores de Campos do

Jordão e de Camanducaia. Pode-se concluir que estas populações são bem semelhantes entre si e bem diferentes das demais, ou seja, apresentam características genéticas bem distintas as quais devem ser preservadas para evitar sua extinção. As outras populações apresentam uma grande variabilidade genética entre suas respectivas árvores, dificultando a formação de grupos.

## ii) Subpopulações

Para mapear a divergência genética entre as subpopulações de Irati, as árvores foram numeradas da seguinte maneira:

TABELA 4.2 – NUMERAÇÃO DAS ÁRVORES DAS SUBPOPULAÇÕES DE IRATI

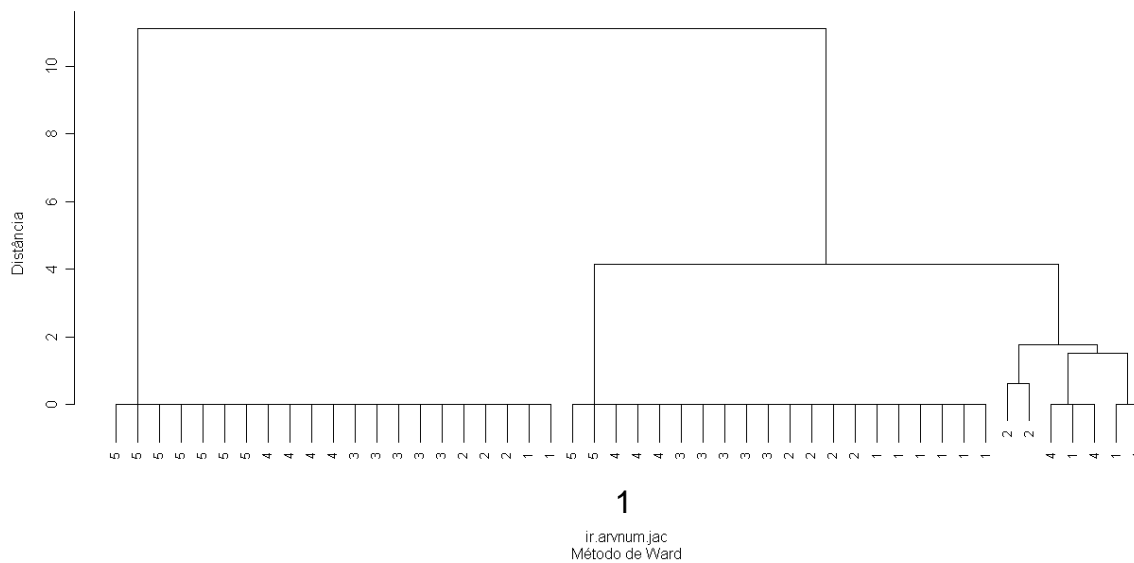
<b>SUBPOPULAÇÃO</b>	<b>NUMERAÇÃO DAS ÁRVORES</b>	<b>NÚMERO DE ÁRVORES</b>
Floresta Intocada	1	10
Capão 1	2	9
Capão 2	3	10
Floresta Explorada	4	10
Floresta Plantada	5	9

FONTE: O autor (2008)

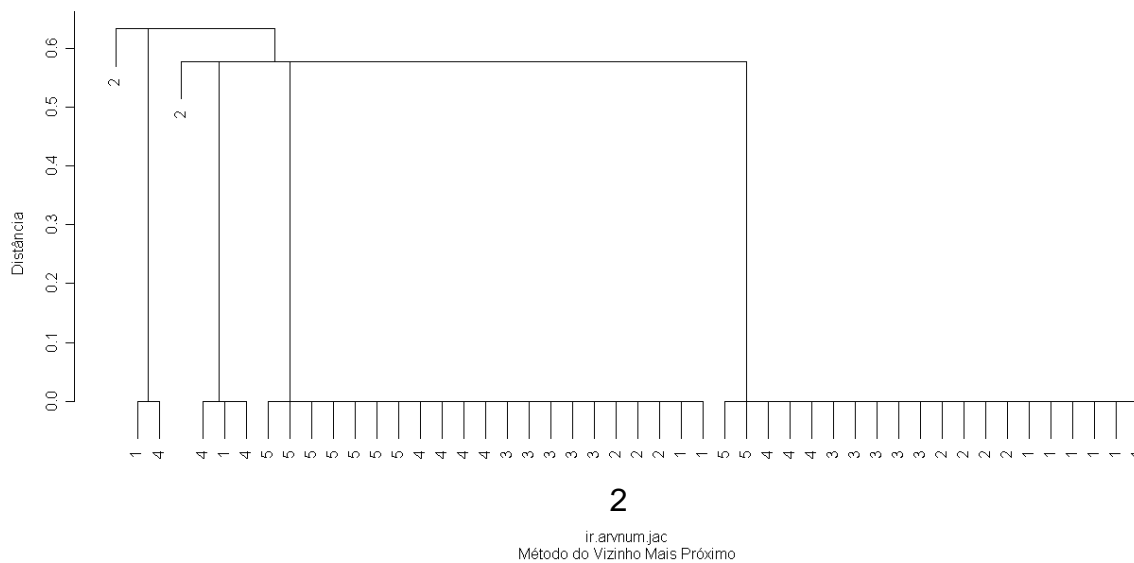
Os dendogramas gerados são apresentados nas figuras a seguir.

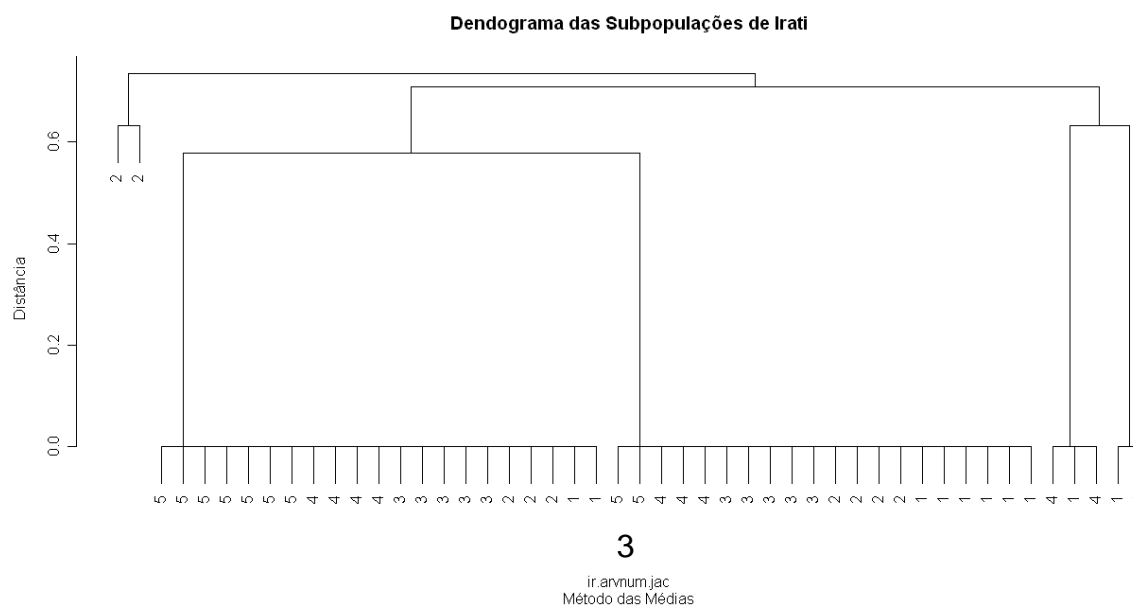
Nos três dendogramas obtidos na Figura 4.29 observa-se um pequeno agrupamento formado pelas árvores de números 1 e 4, ou seja, por árvores das subpopulações de Floresta Intocada e Floresta Explorada. Tal agrupamento foi observado na primeira situação de análise (como o segundo grupo mais homogêneo) e na segunda situação de análise (como o grupo mais homogêneo).

Dendograma das Subpopulações de Irati



Dendograma das Subpopulações de Irati





(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.29 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO

Os métodos de Ward e do vizinho mais próximo foram desconsiderados nos próximos dendogramas não apresentarem a formação de agrupamentos.

Mais uma vez, na Figura 4.30 observa-se um grupo formado pelas árvores das subpopulações de Floresta Intocada e Floresta Explorada; além desse não é possível observar mais nenhum outro agrupamento.

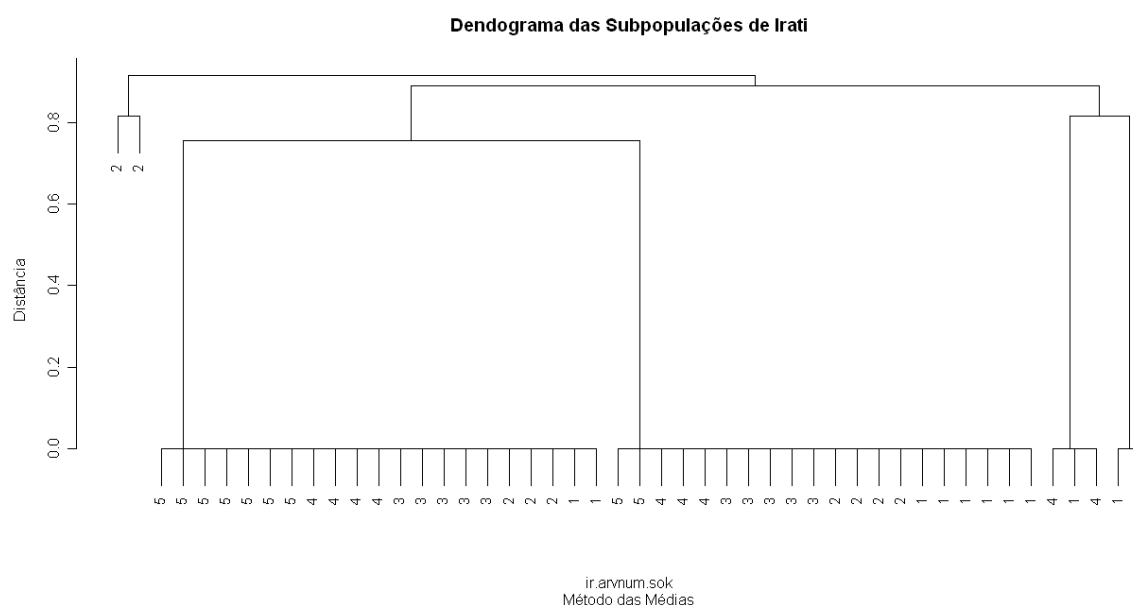


FIGURA 4.30 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH

No lado esquerdo da Figura 4.31 observa-se um agrupamento das árvores de Floresta Intocada e Floresta Explorada, que também foi observado nos dendogramas obtidos nas Figuras 4.29 e 4.30.

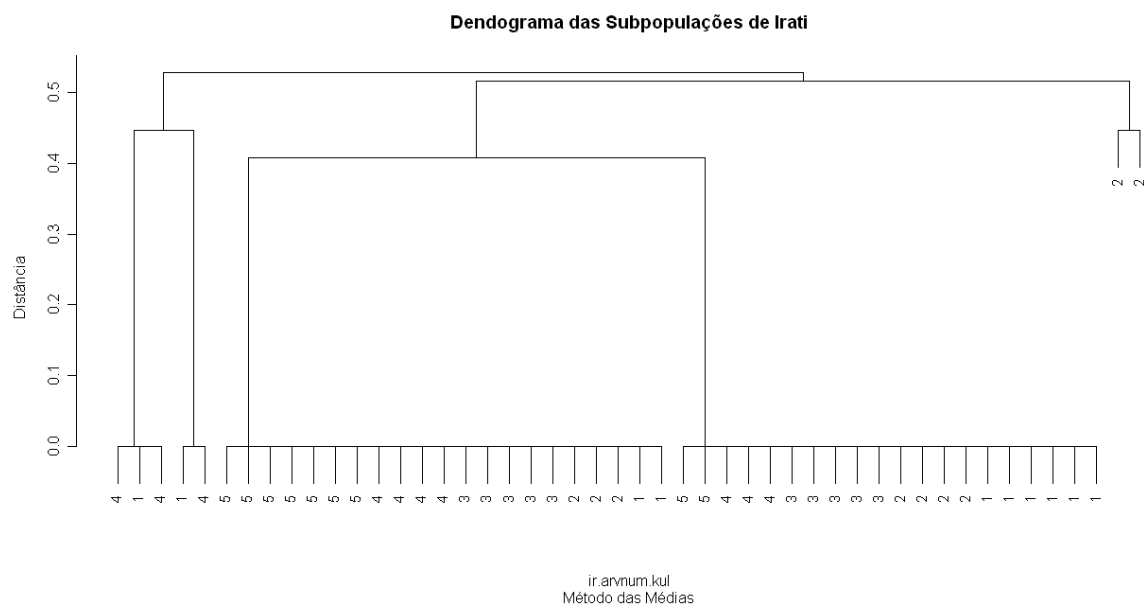


FIGURA 4.31 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE SOUTHWOOD

A forma da Figura 4.32 é semelhante à Figura 4.31, onde o único agrupamento observado é o das árvores de Floresta Intocada e Floresta Explorada. As demais populações não formam grupos.

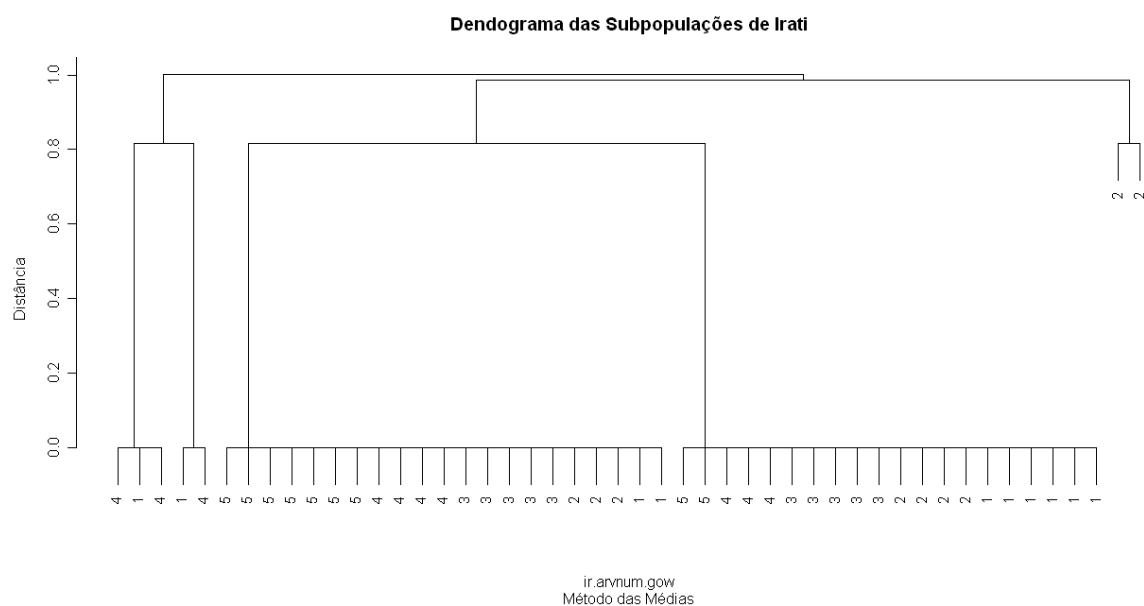


FIGURA 4.32 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE



Na Figura 4.33 verifica-se a presença das árvores das 5 subpopulações em todo o dendograma, o que não permite observar a formação de agrupamentos.

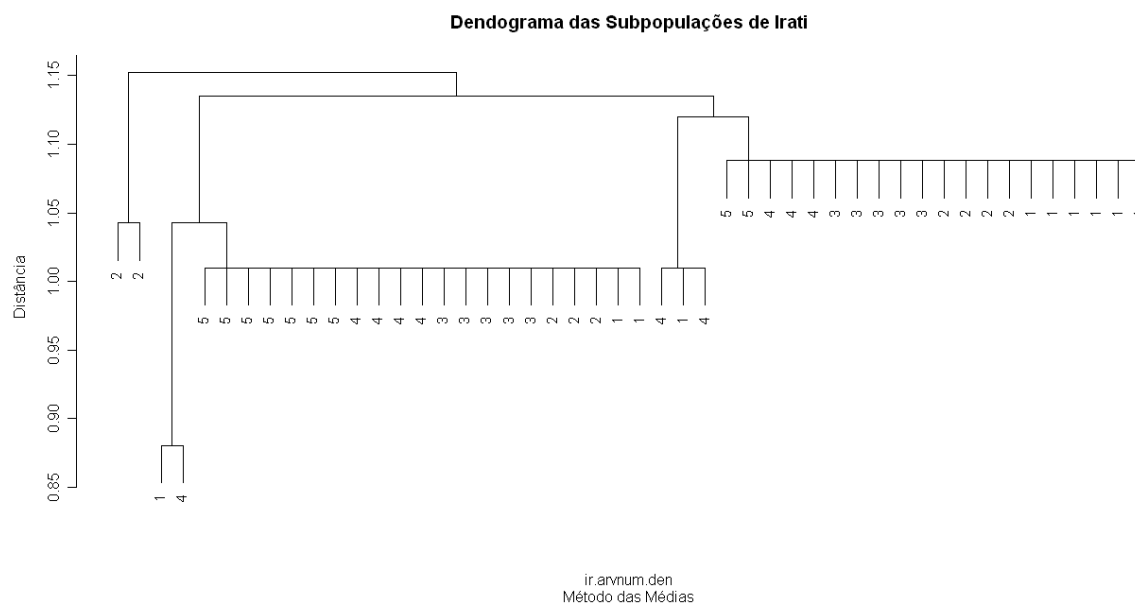


FIGURA 4.33 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE HOLLIDAY

Na Figura 4.34 não é possível observar nenhum agrupamento.

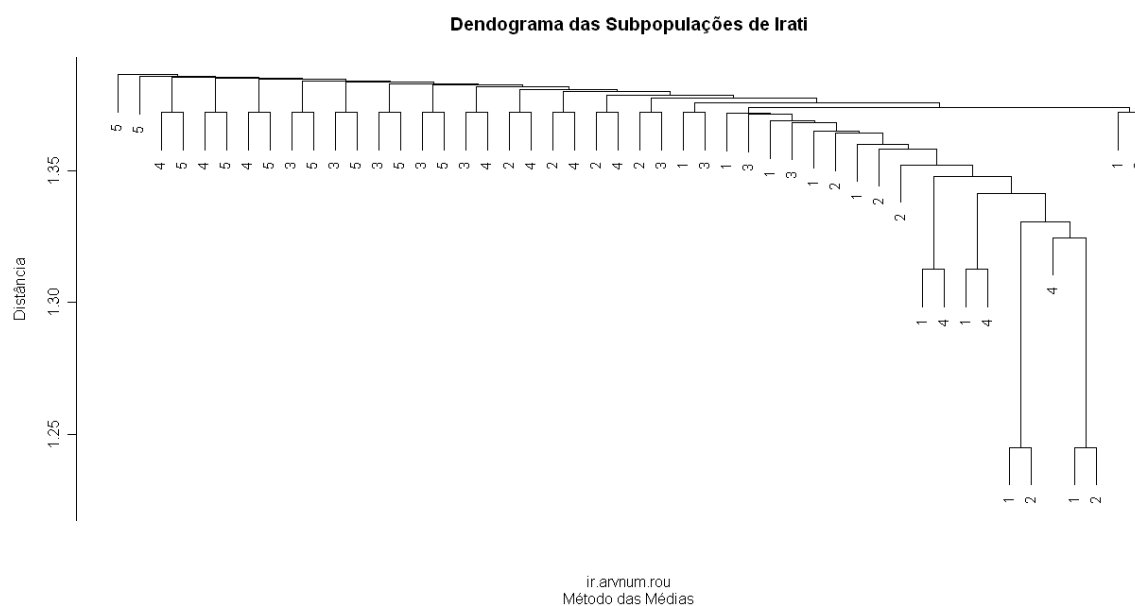


FIGURA 4.34 – DENDOGRAMA DAS SUBPOPULAÇÕES DE IRATI CONSIDERANDO O COEFICIENTE DE KOLEFF

Portanto, de acordo com a análise dos dendogramas obtidos, verifica-se a semelhança entre algumas árvores das subpopulações de Floresta Intocada e Floresta Explorada. Isso foi observado nos dendogramas da primeira e da segunda situação de análise. As demais subpopulações apresentam grande variabilidade genética entre suas árvores de tal maneira que é impossível fazer a distinção de agrupamentos entre elas.

Para mapear a divergência genética entre as árvores das subpopulações de Caçador, as árvores foram numeradas da seguinte maneira:

TABELA 4.3 – NUMERAÇÃO DAS ÁRVORES DAS SUBPOPULAÇÕES DE CAÇADOR

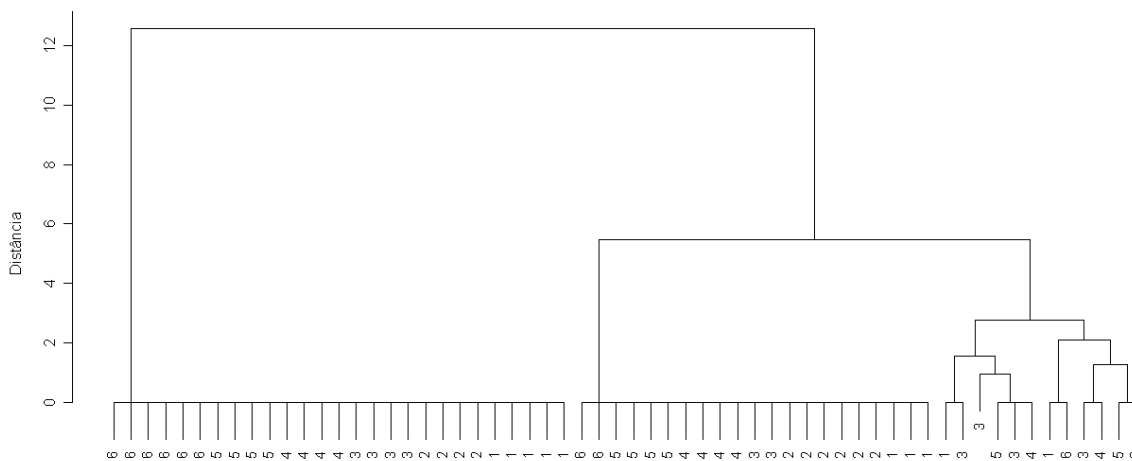
<b>SUBPOPULAÇÃO</b>	<b>NUMERAÇÃO DAS ÁRVORES</b>	<b>NÚMERO DE ÁRVORES</b>
Floresta Intocada	1	10
Capão 1	2	10
Capão 2	3	10
Capão 3	4	10
Capão 4	5	10
Floresta Explorada	6	10

FONTE: O autor (2008)

Os resultados são apresentados a seguir.

Conforme a análise dos dendogramas, obtidos na Figura 4.35, é possível verificar que nenhum agrupamento é formado. Árvores de todas as subpopulações estão dispersas mostrando que a variabilidade genética entre as subpopulações de Caçador é grande, dificultando seu agrupamento.

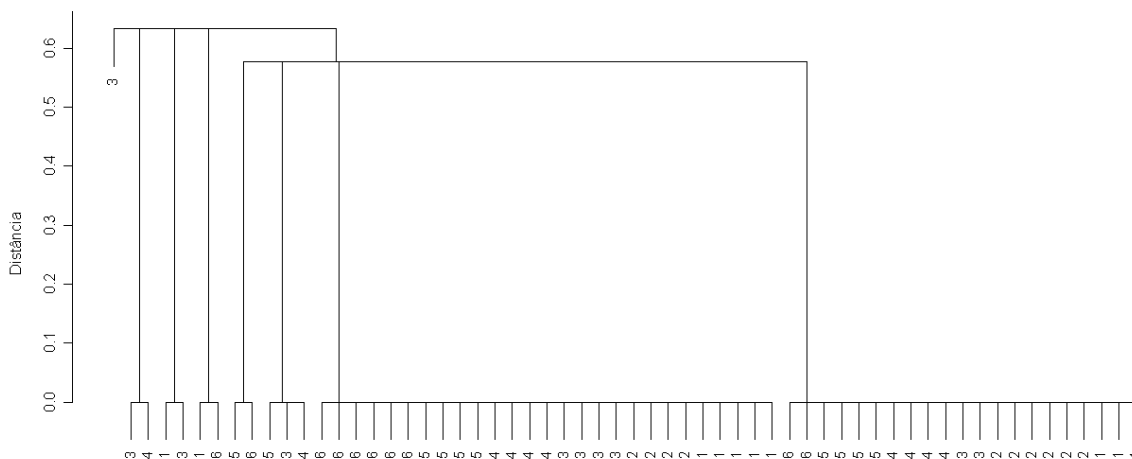
Dendograma das Subpopulações de Caçador



1

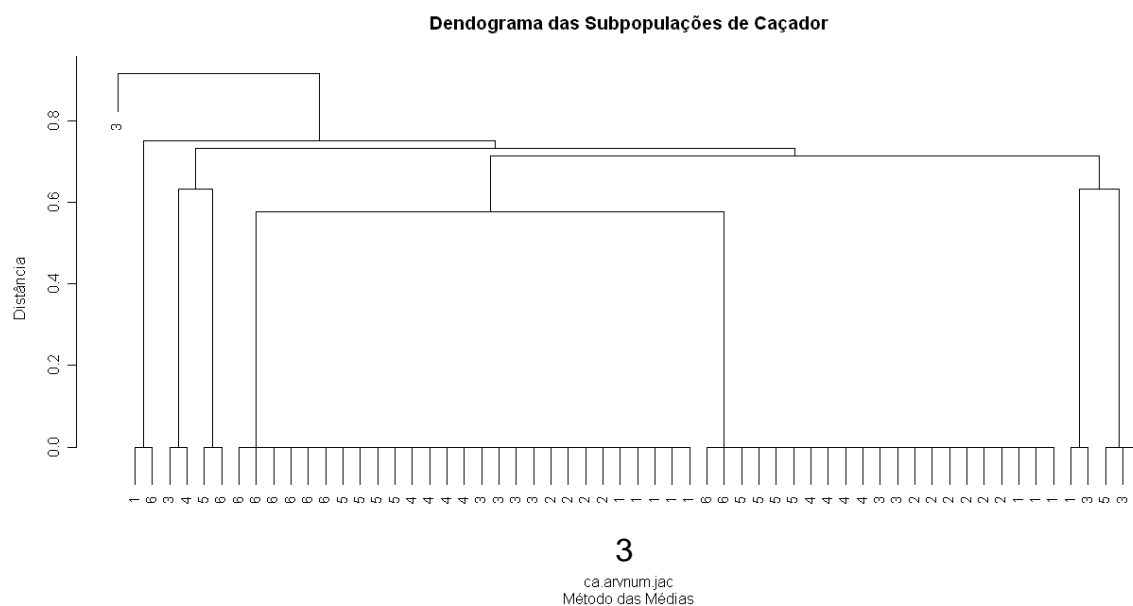
ca.arvnum.jac  
Método de Ward

Dendograma das Subpopulações de Caçador



2

ca.arvnum.jac  
Método do Vizinho Mais Próximo



(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

FIGURA 4.35 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO

Os métodos de Ward e do vizinho mais próximo foram desconsiderados nos próximos dendogramas não apresentarem a formação de agrupamentos.

Nas Figuras 4.36 e 4.37 não se observa a formação de agrupamentos. Verificam-se árvores de todas as subpopulações espalhadas pelos dendogramas, o que não aconteceu com os dendogramas das subpopulações de Irati, pois apesar de haver uma grande dispersão das árvores, um pequeno agrupamento foi formado (pelas subpopulações de Floresta Intocada e Floresta Explorada).

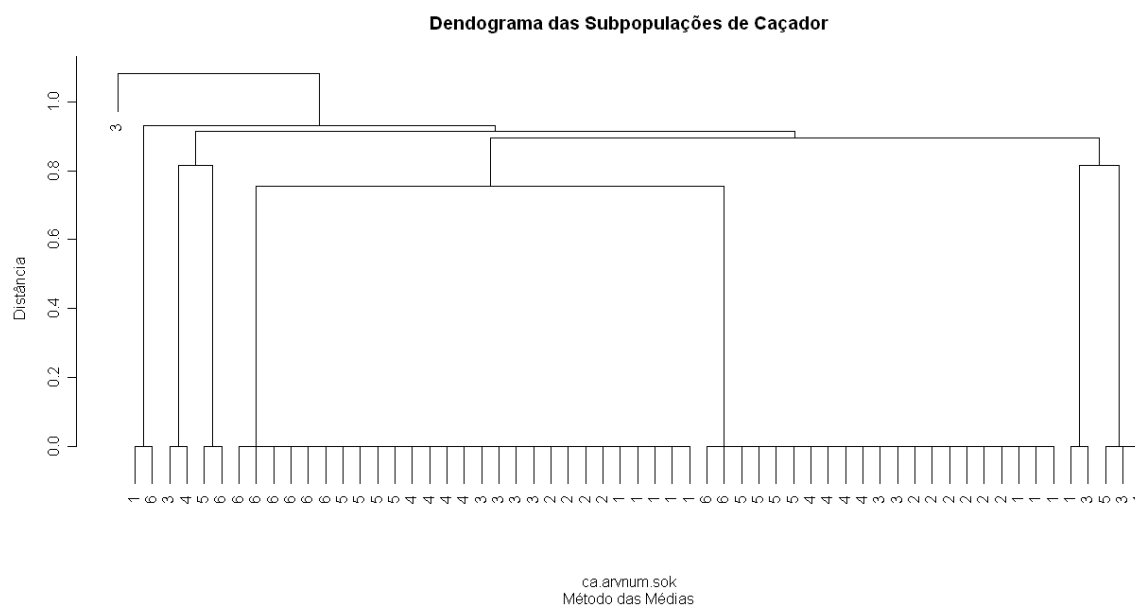


FIGURA 4.36 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH

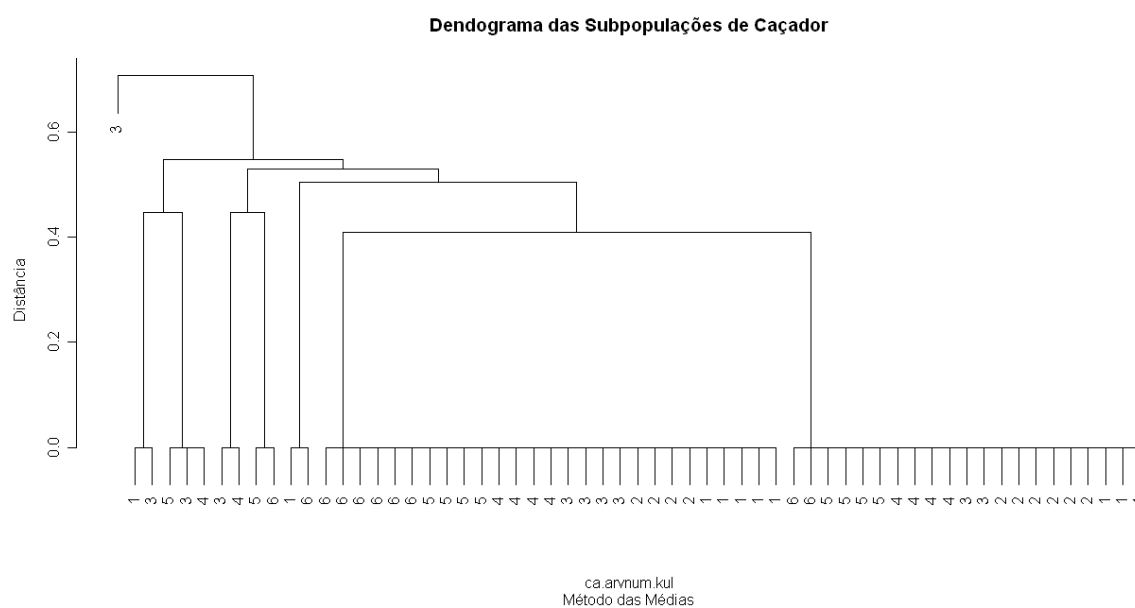


FIGURA 4.37 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOUTHWOOD

Na Figura 4.38 não se observa qualquer agrupamento.



A Figura 4.40 apresenta um dendograma sem a formação de qualquer agrupamento entre as árvores.

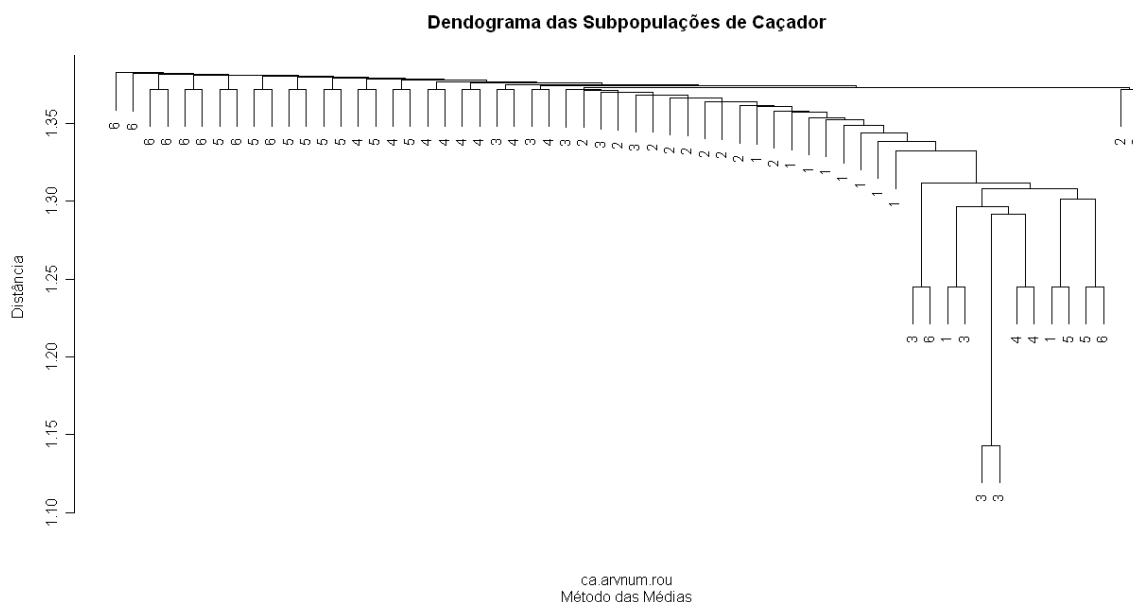


FIGURA 4.40 – DENDOGRAMA DAS SUBPOPULAÇÕES DE CAÇADOR CONSIDERANDO O COEFICIENTE DE KOLEFF

Portanto, utilizando os coeficientes de similaridade citados e analisando seus respectivos dendogramas, pode-se concluir que a comparação das árvores das subpopulações de Caçador não proporciona a formação de agrupamentos, ou seja, a variabilidade genética entre suas árvores é muito pequena.

#### 4.4 ÁRVORES DE CLASSIFICAÇÃO

O conceito de árvores de classificação foi utilizado na tentativa de validar os resultados obtidos nas três situações de análise. As variáveis presentes no conjunto de dados binários foram utilizadas para classificar as amostras das populações e subpopulações.

##### i) Amostras das populações

Considerando a matriz de dados binários, foi gerada a árvore de classificação.

Observa-se na Figura 4.41 que as populações de Camanducaia e de Campos do Jordão foram classificadas corretamente, indicando uma grande semelhança entre elas (resultado observado nas três análises anteriores). Ao se escolher uma árvore qualquer deve-se primeiramente observar o valor da variável *MDH-B* (início da árvore de classificação). Se o valor da variável for 0 deve-se escolher o nó esquerdo, se for 1 escolhe-se o nó direito. Esse processo é repetido de acordo com as variáveis apresentadas na árvore de classificação até se atingir um nó terminal. Irati e Barbacena não são identificadas através da árvore de classificação, ou seja, através das características selecionadas estas populações não conseguem ser classificadas.

Também é possível verificar que a variável mais importante é o *locus MDH-B*, pois é o nó inicial, seguido por *SKDH-B*, *GOT-B* e *PGMA-A*, respectivamente.

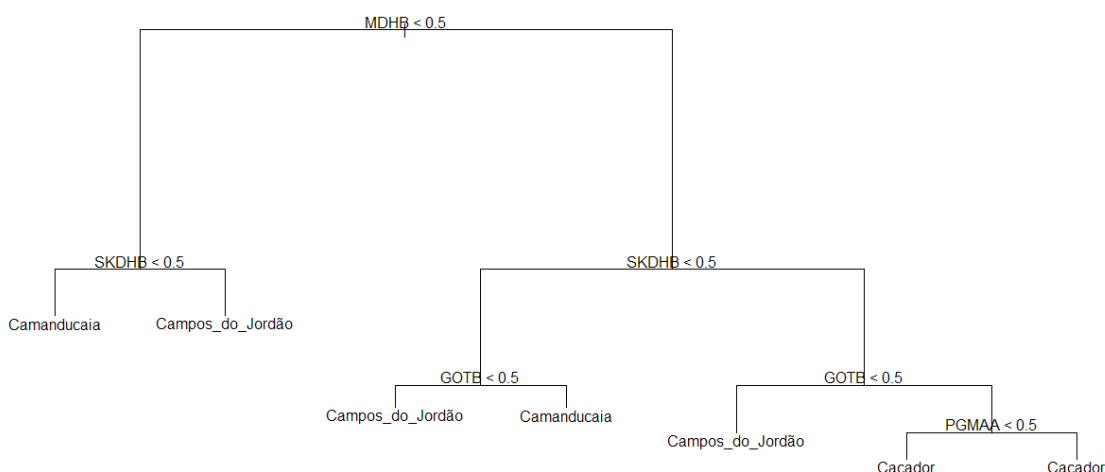


FIGURA 4.41 – ÁRVORE DE CLASSIFICAÇÃO DAS POPULAÇÕES

## ii) Subpopulações

A estrutura da árvore de classificação das subpopulações de Irati é apresentada a seguir.

Pela Figura 4.42, observa-se a classificação de duas subpopulações de Irati: Floresta Intocada e Floresta Explorada, as quais apresentaram grande semelhança



nas análises anteriores. Porém, a classificação de Capão 1 e Capão 2 não é observada.



FIGURA 4.42 – ÁRVORE DE CLASSIFICAÇÃO DAS SUBPOPULAÇÕES DE IRATI

Em Caçador, a árvore de classificação das subpopulações é apresentada a seguir.

Pela Figura 4.43 as subpopulações classificadas corretamente são Capão 2, Floresta Explorada e Capão 1. Porém, as subpopulações mais semelhantes observadas nas análises anteriores não foram observadas aqui. Por isso fica evidente que as árvores de Caçador apresentam uma pequena variação de acordo com as características consideradas.

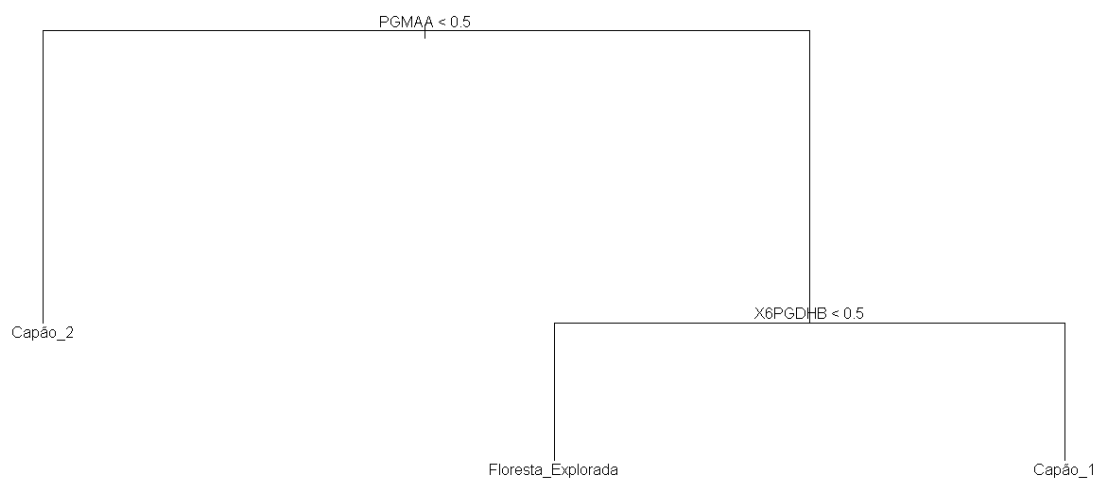


FIGURA 4.43 – ÁRVORE DE CLASSIFICAÇÃO DAS SUBPOPULAÇÕES DE CAÇADOR

## 5 CONSIDERAÇÕES FINAIS E CONCLUSÃO

Neste trabalho buscou-se analisar o conjunto de dados utilizando três situações diferentes, visando caracterizar a variabilidade genética das populações e subpopulações de *Araucaria angustifolia* nas cinco regiões propostas.

A primeira situação de análise mostrou-se muito eficiente. Dos coeficientes utilizados apenas o de Koeff não gerou dendogramas semelhantes aos demais, mostrando-se inapropriado para este tipo de análise. Todos os outros foram coerentes para mapear a divergência genética. Dos métodos de ligação utilizados, todos forneceram resultados satisfatórios principalmente o método do vizinho mais próximo e o da ligação média, pois apresentaram menor variação ao formar os dendogramas a partir de diferentes matrizes de similaridade.

A segunda situação de análise mostrou que, tanto a MANOVA quanto o teste  $T^2$  de Hotteling, puderam ser utilizados para avaliar a qualidade dos agrupamentos formados, porém com algumas restrições devido à natureza binária dos dados. Algumas matrizes de covariância utilizadas nos testes são singulares, necessitando o uso de matrizes pseudo-inversas. Em relação à primeira situação de análise, os dendogramas desta etapa apresentaram a mesma formação dos grupos mais homogêneos, mas também mostraram algumas diferenças em outros agrupamentos.

Na terceira situação de análise os dendogramas obtidos apresentaram apenas os agrupamentos observados nas análises anteriores que não eram os mais homogêneos. Foi necessária uma análise minuciosa dos gráficos para notar tais agrupamentos, pois a existência de uma grande variabilidade genética entre as árvores das populações e subpopulações dificultou a formação de grupos. Contudo, esta maneira de analisar os dados pode ser muito proveitosa nos casos em que as observações de cada amostra apresentem pouca variabilidade e possam formar grupos bem distintos e facilmente observados em dendogramas.

Árvores de classificação confirmaram a semelhança entre algumas populações: Campos do Jordão e Camanducaia, e subpopulações: Floresta Intocada e Floresta Explorada de Irati. Em Caçador as subpopulações que foram classificadas não foram as mesmas agrupadas nas análises anteriores, levando a acreditar que sua variabilidade genética entre árvores é menor.

Os resultados obtidos neste trabalho mostraram a existência de semelhança genética entre algumas populações e entre algumas subpopulações. Por conta disso, os esforços necessários para a preservação das araucárias nas regiões consideradas podem ser concentrados, diminuindo as áreas de conservação em:

- 60% das populações, considerando a adoção de medidas de conservação para a população de: a) Barbacena; b) Irati ou Caçador; c) Campos do Jordão ou Camanducaia;
- 60% das subpopulações de Irati, considerando a adoção de medidas de conservação para as subpopulações de: a) Floresta Plantada; b) Capão 1 ou Capão 2; c) Floresta Explorada ou Floresta Intocada;
- 67% das subpopulações de Caçador, considerando a adoção de medidas de conservação para as subpopulações de: a) Capão 4 e Capão 2; b) Capão 1 ou Capão 3; c) Floresta Intocada ou Floresta Explorada.

Em média, as áreas de preservação podem ser reduzidas para 62% das áreas analisadas neste trabalho.

É importante salientar que Caçador foi a região que apresentou alguns resultados diferentes em cada uma das três situações de análise. Logo, deve ser dada uma atenção especial a esta população.

Todas as metodologias apresentadas para determinar o padrão de variabilidade das amostras das populações de araucária, por marcadores isoenzimáticos, mostraram-se eficientes, cada uma com as suas particularidades. Neste trabalho, em particular, a primeira e a segunda situações de análise proporcionaram os melhores resultados.

Além disso, as regiões mais críticas para conservação observadas foram Barbacena, Floresta Plantada em Irati e os Capões 2 e 4 em Caçador.

Como sugestão para trabalhos futuros relacionados com este tema propõe-se que seja analisada a qualidade das variáveis utilizadas no mapeamento da divergência genética através de métodos estatísticos multivariados como, por exemplo, a Análise de Correlação Canônica ou a Análise Discriminante. Também como sugestão propõe-se o uso de uma metodologia não paramétrica para realizar o mapeamento.

## REFERÊNCIAS

AGUIAR, A. V. de. **Emprego de parâmetros moleculares e quantitativos na conservação e melhoramento de *Eugenia dysenterica* DC.** PPG em Agronomia, Goiânia, 2004.

ANSELMINI, J. I. **Fenologia reprodutiva da *Araucaria angustifolia* (Bert.) O. Ktze, na região de Curitiba – PR.** 52 p. Dissertação (Mestrado em Agronomia) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2005.

ÁRVORES de decisão. Disponível em: <[http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/PRG\\_0599.EXE/3710\\_4.PDF?NrOcoSis=6894&CdLinPrg=pt.](http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/PRG_0599.EXE/3710_4.PDF?NrOcoSis=6894&CdLinPrg=pt.)> Acesso em 10/10/07.

BIODIVERSITY reporting award 2005. Disponível em: <<http://www.biodiversityreporting.org/article.sub?docId=12581&c=Brazil&cRef=Brazil&year=2005&date=August%202004>>. Acesso em 13/08/2007.

BOSCARIOLI, C.; SILVA, L. A. da; HERNANDEZ, E. D. M. **Análise de agrupamentos utilizando mapas auto-organizáveis em agricultura de precisão.** Congresso Brasileiro de Agricultura de Precisão (CONBAP), 2. São Pedro, Escola Superior de Agricultura “Luiz de Queiroz” ESALQ/USP, 2006. 1 arquivo pdf.

BRANDALISE, L. **A determinação de indicadores sócio-econômicos e classificação de um município paranaense através da análise estatística multivariada.** 186 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2003.

BRAULIO, S. N. **Proposta de uma metodologia para a avaliação de imóveis urbanos baseado em métodos estatísticos multivariados.** 131 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2005.

CALDEIRA, M. V. W. *et al.* Relação hipsométrica para *Araucaria angustifolia* (Bert.) O. Ktze na região oeste do estado do paraná. **Revista Acadêmica: ciências agrárias e ambientais.** Curitiba, v.1, n. 2, p. 79-88, abr./jun. 2003. 1 arquivo pdf.

CAMPOS do Jordão – Guia de Hotéis, Pousadas, Restaurantes, Imóveis, Passeios. Disponível em: <<http://www.netcampos.com>>. Acesso em 03/01/07.

CARVALHO, P. E. R. **Espécies florestais brasileiras**: recomendações silviculturais, potencialidade e uso da madeira. Brasília: EMBRAPA-SPI, 1994.

CARVALHO, P. E. R. **Espécies Arbóreas Brasileiras**. EMBRAPA-SPI, 2003.

CHAVES NETO, A. **Análise multivariada aplicada à pesquisa**: notas de aula. Departamento de Estatística, Universidade Federal do Paraná. Curitiba, 1997.

CONKLE, M. T. *et al.* Starch Gel Electrophoresis of Conifer Seeds: a Laboratory Manual. **Pacific Southwest Forest and Range Experimental Station**. 18 p. USDA, 1982.

CONTE, R. **Estrutura genética de populações de *Euterpe edulis* MART. submetidas à ação antrópica utilizando marcadores alozímicos e microssatélites**. 124 p. Tese (Doutorado em Agronomia) – Setor de Ciências Agrárias, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade Federal de São Paulo, Piracicaba, 2004.

DIAS, L. A. S. Análises multidimensionais. In: ALFENAS, A. C. (Ed.). **Eletroforese de isoenzimas e proteínas afins**. Viçosa: Editora UFV, 1998. p. 405 – 475.

DUTRA, R. M. O.; SPERANDIO, M.; COELHO, J. O método ward de agrupamento de dados e sua aplicação em associação com os mapas auto-organizáveis de kohonen. Não paginado. [S.l.], [200?]. 1 arquivo pdf.

ESQUARCINI, R. **Classificação e ranqueamento dos municípios paranaenses segundo suas políticas setoriais através da análise multivariada**. 117 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2005.

EVERITT, B. S. **Cluster analysis**. 3 ed, New York: Edward Arnold, 1993.

FALEIRO, F. G. **Marcadores Genético-Moleculares aplicados a programas de conservação e uso de recursos genéticos**. Embrapa Cerrados, DF, 2007.

FERET, P. P.; BERGMANN, F. Gel electrophoresis of proteins and enzymes. **Modern Methods in Forest Genetics**, Springer, Berlin, Heidelberg, p. 47 – 77, 1976.

FERREIRA, M. A. J. da F. **Utilização das técnicas de marcadores moleculares na genética de populações, na genética quantitativa e no melhoramento de plantas.** Boa Vista, Embrapa Roraima, 2003, 63 p. 1 arquivo pdf.

FRANCISCO, M. R. **Estruturação genética em populações do tangará-dançarino *Chiroxiphia caudata* (Aves, Pipridae) no corredor costeiro da mata atlântica (SP) e sua importância para a conservação.** 103 p. Tese (Doutorado em Ecologia e Recursos Naturais) – Setor de Ciências Biológicas, Universidade Federal de São Carlos, São Carlos, 2005.

GROBE, J. R. **Aplicações da estatística multivariada na análise de resultados em experimentos com solos e animais.** 134 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2005.

INSTITUTO de pesquisas e estudos florestais. Disponível em:  
<<http://www.ipef.br/identificacao/araucaria.angustifolia.asp> >. Acesso em 03/04/2007.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** 2. ed. New Jersey: Prentice Hall, Inc., 1998.

KOCH, Z.; CORRÊA, M. C. **Araucária:** a floresta do Brasil meridional. Curitiba: Olhar Brasileiro, 2002, 148 p.

LABOREAU, L. F. G. & MATOS FILHO, A. **Notas Preliminares sobre a região de Araucária.** Em: Anuário Brasileiro de Economia Florestal, 1. Rio de Janeiro, Brasil, 1948.

LIENGSIRI, C.; PIEWLUANG, C.; BOYTLE, T. J. B. Starch gel electrophoresis of tropical trees. A Manual. The ASEAN – **Canada Forest Tree Seed Centre**, 51 p. Saraburi, 1990.

LIMA, J. D. **A análise econômico-financeira de empresas sob a ótica da estatística multivariada.** Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2002.

MAINIERI, C.; CHIMELO, J.P. **Fichas de características das madeiras brasileiras.** São Paulo: IPT, 1989. 418p.

MAIS interior. Disponível em:

<[http://www.maisinterior.com.br/saocarlos/v4\\_araucaria.asp](http://www.maisinterior.com.br/saocarlos/v4_araucaria.asp)>. Acesso em 03/04/2007.

MANTOVANI, A.; MORELLATO, P. C.; REIS, M. S. DOS. Fenologia reprodutiva e produção de semente em *Araucaria angustifolia* (Bert.) O. Kuntze. **Revista Brasileira de Botânica**, v. 27, n. 4, p. 787 – 796, 2004. 1 arquivo pdf.

MARQUES, M. A. M. **Aplicação da análise multivariada no estudo da infraestrutura dos serviços de saúde dos municípios paranaenses**. 133 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2006.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada** - uma abordagem aplicada. Belo Horizonte: Didática, 2005.

PORTAL São Francisco. Disponível em:

<<http://www.colegiosaofrancisco.com.br/alfa/meio-ambiente-mata-das-araucarias/mata-das-araucarias.php>>. Acesso em 03/08/2007.

REVISTA eletrônica de jornalismo científico (Com Ciência). Disponível em:

<<http://www.comciencia.br/reportagens/biodiversidade/bio09.htm>>. Acesso em 14/08/2007.

SISTEMAS para análises estatísticas. Disponível em:

<<http://www.ufv.br/saeg/saeg50.htm>>. Acesso em 10/05/2007.

SHARMA, S. **Applied multivariate techniques**. South Carolina: John Wiley & Sons, Inc, 1996.

SOUSA, V. A., HATTEMER, H. H.; ROBINSON, I. P. Inheritance and linkage relationships of isozyme variants of *Araucaria angustifolia* (Bert.) O. Ktze. **Silvae Genetica** 51. Frankfurt, v. 5 – 6, p. 191-196, 2002.

SOUSA, V. A., SEBBENN, A. M., HATTEMER, H. H.; ZIEHE, M. Correlated mating in populations of a dioecious brazilian conifer, *Araucaria angustifolia* (Bert.) O. Ktze. **Forest Genetics** 12(2). Göttingen, p. 107-119, 2005.

SOUZA, M. I. F. de. **Análise da diversidade genética de populações de *Araucaria angustifolia* (Bertol.) Kuntze utilizando marcador AFLP**. 97 p. Dissertação (Mestrado em Genética) – Setor de Ciências Biológicas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.



TRINDADE, M. da G. **Estrutura Genética de Populações Naturais de Cagaiteira (*Eugenia dysenterica*) do Nordeste de Goiás, Brasil.** 104 p. Dissertação (Mestrado em Agronomia) – Setor de Ciências Agrárias, Universidade Federal de Goiás, Goiânia, 2001.

WADT, L. H. de O. **Avaliação da divergência genética em coqueiro (*Cocos nucifera* L.) usando marcadores RAPD em amostras de plantas individuais ou compostas.** 92 p. Dissertação (Mestrado em Agronomia) – Centro de Ciências e Tecnologias Agrárias, Universidade Estadual do Norte Fluminense, Rio de Janeiro, 1997.

WADT, L. H. de O. **Estrutura genética de populações naturais de pimenta longa (*Piper hispidinervum* C. DC.) visando seu uso e conservação.** 92 p. Tese Doutorado em Agronomia) – Setor de Ciências Agrárias, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade Federal de São Paulo, Piracicaba, 2001.

48ª Reunião da RBRAS e 10º SEAGRO. Universidade Federal de Lavras, MG. Julho/ 2003.

## DOCUMENTOS CONSULTADOS

ALCOCHETE, A. A. N. de. **Diversidade genética e mapeamento de QTLs do sistema gênico de macho-esterilidade termo-sensível (TGMS) do genoma de arroz (*Oryza sativa* L.)** 145 p. Tese (Doutorado em Biologia Celular) – Setor de Ciências Biológicas, Universidade de Brasília, Brasília, 2005.

ATIVIDADES de apoio em biologia celular e molecular. Disponível em: <[http://educar.sc.usp.br/licenciatura/2003/siteprojeto/2003/6\\_eletroforese\\_gel.pdf](http://educar.sc.usp.br/licenciatura/2003/siteprojeto/2003/6_eletroforese_gel.pdf)>. Acesso em 22/08/2007.

BARROSO, L. P. **Análise Multivariada**. 48ª Reunião da RBRAS e 10º SEAGRO. Lavras – MG. Departamento de Ciências Exatas. Universidade Federal de Lavras, 2005.

DOKI, N. de O. **Desenvolvimento de metodologia estatística para estimação de numero de equipamentos de uso compartilhado**. 153 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2002.

DUNTFMAN, G. H. **Introduction to multivariate analysis**. Beverly Hills: Sage Publications, 1984.

JAIN, A. K.; DUBES, R. C. **Algorithms for Clustering Data**. Prentice Hall, 1988.

KRZANOWSKI, W. J. **Principles of multivariate analysis: a user's perspective**. Oxford: Oxford Science, 1988.

MARCADORES moleculares e o melhoramento genético de espécies florestais. Disponível em <<http://www.ufv.br/dbg/bioano01/div39.htm>>. Acesso em 7/02/2007.

MARCADORES RAPD na análise da diversidade genética de isolados de *Acremonium strictum*. Disponível em: <<http://www.scielo.br/pdf/fb/v29n6/a09v29n6.pdf>>. Acesso em 7/02/2007.

METZ, J.; MONARD, M. C. **Estudo e análise das diversas representações e estruturas de dados utilizadas nos algoritmos de *clustering* hierárquicos**. São Carlos: jan/2006. 38 p. Disponível em: <[http://www.icmc.usp.br/~biblio/download/RT\\_269.pdf](http://www.icmc.usp.br/~biblio/download/RT_269.pdf)> Acesso em 11/02/2007.

MOURÃO Jr, M. **Análise de agrupamento: propriedades e aplicações**. Embrapa Roraima, 2005.

PORTAL flora e fauna. Disponível em:  
<<http://www.floraefauna.com/plantamesjunho.htm>>. Acesso em 03/08/2007.

WEIR, B. S. **Genetic data analysis: meyhods for discrete population genetic data**. Suderland: North Carolina State University / Sinauer Associates, 1996.

WIKIPIDEA. Diksponível em: <[http://pt.wikipedia.org/wiki/Eletroforese\\_em\\_Gel](http://pt.wikipedia.org/wiki/Eletroforese_em_Gel)>. Acesso em 22/08/2007.

## **APÊNDICES**





## b) 2ª SITUAÇÃO DE ANÁLISE

## i) Amostras das populações

POPULAÇÃO	GOT-B	GOT-C	PGMA-A	MDH-B	SKDH-B	6PGDH-B
Campos do Jordão	0,62857	0,94286	0,70000	0,47143	0,78571	0,65714
Irati	1,00000	0,89583	0,95833	1,00000	1,00000	0,50000
Caçador	0,93333	0,95000	0,90000	1,00000	1,00000	0,45000
Camanducaia	0,78571	0,97143	0,94286	0,52857	0,57143	0,57143
Barbacena	0,87500	1,00000	1,00000	0,50000	0,96875	0,75000

QUADRO A.4 – VETORES DE MÉDIAS AMOSTRAIS DAS 5 POPULAÇÕES UTILIZADOS NA 2ª SITUAÇÃO DE ANÁLISE

POPULAÇÃO	OBSERVAÇÕES DENTRO DA ELIPSE (%)
Campos do Jordão	55,71
Irati	-
Caçador	-
Camanducaia	71,43
Barbacena	-
5 populações conjuntamente	61,07

QUADRO A.5 – TESTE DE NORMALIDADE DOS DADOS DA MANOVA (POPULAÇÕES)

## ii) Subpopulações

POPULAÇÃO	SUBPOPULAÇÃO	GOT-B	GOT-C	PGMA-A	MDH-B	SKDH-B	6PGDH-B
Irati	Floresta Intocada	1,00000	0,80000	1,00000	1,00000	1,00000	0,70000
	Capão 1	1,00000	1,00000	0,77778	1,00000	1,00000	0,55556
	Capão 2	1,00000	1,00000	1,00000	1,00000	1,00000	0,50000
	Floresta Explorada	1,00000	0,70000	1,00000	1,00000	1,00000	0,50000
	Floresta Plantada	1,00000	1,00000	1,00000	1,00000	1,00000	0,22222

QUADRO A.6 – VETORES DE MÉDIAS AMOSTRAIS DAS SUBPOPULAÇÕES DE IRATI UTILIZADOS NA 2ª SITUAÇÃO DE ANÁLISE

<b>POPULAÇÃO</b>	<b>SUBPOPULAÇÃO</b>	<b>GOT-B</b>	<b>GOT-C</b>	<b>PGMA-A</b>	<b>MDH-B</b>	<b>SKDH-B</b>	<b>6PGDH-B</b>
Caçador	Floresta Intocada	1,00000	0,90000	0,90000	1,00000	1,00000	0,30000
	Capão 1	1,00000	1,00000	1,00000	1,00000	1,00000	0,60000
	Capão 2	0,90000	0,90000	0,70000	1,00000	1,00000	0,40000
	Capão 3	0,90000	1,00000	0,90000	1,00000	1,00000	0,50000
	Capão 4	0,90000	1,00000	0,90000	1,00000	1,00000	0,60000
	Floresta Explorada	0,90000	0,90000	1,00000	1,00000	1,00000	0,30000

QUADRO A.7 – VETORES DE MÉDIAS AMOSTRAIS DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZADOS NA 2ª SITUAÇÃO DE ANÁLISE



## APÊNDICE 2

### MATRIZES DE DISTÂNCIA

#### a) 1ª SITUAÇÃO DE ANÁLISE

##### i) Amostras das Populações

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	0,8932370			
Caçador	0,9004503	0,4714045		
Camanducaia	0,9400643	0,8397995	0,8465090	
Barbacena	0,8545976	0,6666667	0,7254763	0,7839165

QUADRO A.8 – MATRIZ DE DISTÂNCIA DAS POPULAÇÕES UTILIZANDO O COEFICIENTE DE JACCARD

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	1,0680283			
Caçador	1,0741723	0,6324555		
Camanducaia	1,1071614	1,0211766	1,0271894	
Barbacena	1,0343882	0,8528029	0,9128709	0,9696241

QUADRO A.9 – MATRIZ DE DISTÂNCIA DAS POPULAÇÕES UTILIZANDO O COEFICIENTE DE SOKAL & SNEATH

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	0,6811816			
Caçador	0,6942207	0,3421413		
Camanducaia	0,7511895	0,6406605	0,6513787	
Barbacena	0,6511037	0,4984461	0,5502553	0,5946003

QUADRO A.10 – MATRIZ DE DISTÂNCIA DAS POPULAÇÕES UTILIZANDO O COEFICIENTE DE SOUTHWOOD

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	1,2500000			
Caçador	1,2500000	0,6454972		
Camanducaia	1,2583057	1,1814539	1,1814539	
Barbacena	1,1636867	0,9354143	1,0206207	1,0704360

QUADRO A.11 – MATRIZ DE DISTÂNCIA DAS POPULAÇÕES UTILIZANDO O COEFICIENTE DE GOWER & LEGENDRE

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	1,6611670			
Caçador	1,7021710	1,3896820		
Camanducaia	1,8420290	1,6347490	1,6641320	
Barbacena	1,6194260	1,4928620	1,5553290	1,5701330

QUADRO A.12 – MATRIZ DE DISTÂNCIA DAS POPULAÇÕES UTILIZANDO O COEFICIENTE DE HOLLIDAY

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	1,2949950			
Caçador	1,2903800	1,3851950		
Camanducaia	1,2684150	1,3093470	1,3058320	
Barbacena	1,3056980	1,3518870	1,3379470	1,3244360

QUADRO A.13 – MATRIZ DE DISTÂNCIA DAS POPULAÇÕES UTILIZANDO O COEFICIENTE DE KOLEFF

## ii) Subpopulações

	Floresta Intocada	Capão 1	Capão 2	Floresta Explorada
Capão 1	0,6382847			
Capão 2	0,5494423	0,3464102		
Floresta Explorada	0,5547002	0,5883484	0,4850713	
Floresta Plantada	0,5547002	0,5239368	0,4850713	0,4898979

QUADRO A.14 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE IRATI UTILIZANDO O COEFICIENTE DE JACCARD

	Floresta Intocada	Capão 1	Capão 2	Floresta Explorada
Capão 1	0,8227534			
Capão 2	0,7242860	0,4758310		
Floresta Explorada	0,7302967	0,7682213	0,6488857	
Floresta Plantada	0,7302967	0,6948083	0,6488857	0,6546537

QUADRO A.15 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE IRATI UTILIZANDO O COEFICIENTE DE SOKAL & SNEATH

	Floresta Intocada	Capão 1	Capão 2	Floresta Explorada
Capão 1	0,4760416			
Capão 2	0,4040610	0,2482935		
Floresta Explorada	0,4072721	0,4350552	0,3524001	
Floresta Plantada	0,4072721	0,3836185	0,3524001	0,3572948

QUADRO A.16 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE IRATI UTILIZANDO O COEFICIENTE DE SOUTHWOOD

	Floresta Intocada	Capão 1	Capão 2	Floresta Explorada
Capão 1	0,9026709			
Capão 2	0,7698004	0,4714045		
Floresta Explorada	0,7698004	0,8164966	0,6666667	
Floresta Plantada	0,7698004	0,7200823	0,6666667	0,6666667

QUADRO A.17 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE IRATI UTILIZANDO O COEFICIENTE DE GOWER & LEGENDRE

	Floresta Intocada	Capão 1	Capão 2	Floresta Explorada
Capão 1	1,3881800			
Capão 2	1,3540190	1,3133980		
Floresta Explorada	1,3506250	1,3626250	1,3315950	
Floresta Plantada	1,3506250	1,3414340	1,3315950	1,3319760

QUADRO A.18 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE IRATI UTILIZANDO O COEFICIENTE DE HOLLIDAY

	Floresta Intocada	Capão 1	Capão 2	Capão 3	Capão 4
Capão 1	0,4955946				
Capão 2	0,6546537	0,5619515			
Capão 3	0,5923489	0,3273268	0,5976143		
Capão 4	0,5619515	0,5207556	0,6695341	0,5570860	
Floresta Explorada	0,4714045	0,4955946	0,7008766	0,5923489	0,5000000

QUADRO A.19 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZANDO O COEFICIENTE DE JACCARD

	Floresta Intocada	Capão 1	Capão 2	Capão 3
Capão 1	0,6614378			
Capão 2	0,8401681	0,7385489		
Capão 3	0,7726674	0,4509876	0,7784989	
Capão 4	0,7385489	0,6910947	0,8558004	0,7330167
Floresta Explorada	0,6324555	0,6614378	0,8881061	0,7726674

QUADRO A.20 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZANDO O COEFICIENTE DE SOKAL & SNEATH

	Floresta Intocada	Capão 1	Capão 2	Capão 3	Capão 4
Capão 1	0,3560206				
Capão 2	0,4891786	0,4040610			
Capão 3	0,4377658	0,3145500	0,4396589		
Capão 4	0,4122313	0,3805802	0,4982964	0,4099564	
Floresta Explorada	0,3429972	0,3560206	0,5284995	0,4377658	0,3630541

QUADRO A.21 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZANDO O COEFICIENTE DE SOUTHWOOD

	Floresta Intocada	Capão 1	Capão 2	Capão 3	Capão 4
Capão 1	0,6831301				
Capão 2	0,8944272	0,7745967			
Capão 3	0,8164966	0,4472136	0,8164966		
Capão 4	0,7745967	0,7302967	0,9309493	0,7745967	
Floresta Explorada	0,6324555	0,6831301	0,9660918	0,8164966	0,6831301

QUADRO A.22 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZANDO O COEFICIENTE DE GOWER & LEGENDRE

	Floresta Intocada	Capão 1	Capão 2	Capão 3	Capão 4
Capão 1	1,3315120				
Capão 2	1,3902720	1,3371800			
Capão 3	1,3678000	1,3152390	1,3612740		
Capão 4	1,3555950	1,3513970	1,3945560	1,3602880	
Floresta Explorada	1,3312320	1,3315120	1,4178610	1,3678000	1,3373240

QUADRO A.23 – MATRIZ DE DISTÂNCIA DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZANDO O COEFICIENTE DE HOLLIDAY

## b) 2ª SITUAÇÃO DE ANÁLISE

### i) Amostras das Populações

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	0,7462613			
Caçador	0,7079146	0,1152461		
Camanducaia	0,3755227	0,6803643	0,6668970	
Barbacena	0,4437743	0,5845420	0,5973935	0,4501340

QUADRO A.24 – MATRIZ DE DISTÂNCIAS DAS POPULAÇÕES UTILIZANDO A DISTÂNCIA EUCLIDIANA

	Campos do Jordão	Irati	Caçador	Camanducaia
Irati	1,57679			
Caçador	1,46190	0,22917		
Camanducaia	0,78570	1,27679	1,23334	
Barbacena	0,90804	1,05209	1,03958	0,77946

QUADRO A.25 – MATRIZ DE DISTÂNCIAS DAS POPULAÇÕES UTILIZANDO A DISTÂNCIA DE MANHATTAN

## ii) Subpopulações

	Floresta Intocada	Capão 1	Capão 2	Floresta Explorada
Capão 1	0,3320311			
Capão 2	0,2828427	0,2290603		
Floresta Explorada	0,2236068	0,3774502	0,3000000	
Floresta Plantada	0,5179515	0,4006211	0,2777800	0,4088542

QUADRO A.26 – MATRIZ DE DISTÂNCIAS DAS SUBPOPULAÇÕES DE IRATI UTILIZANDO A DISTÂNCIA EUCLIDIANA

	Floresta Intocada	Capão 1	Capão 2	Floresta Explorada
Capão 1	0,56666			
Capão 2	0,40000	0,27778		
Floresta Explorada	0,30000	0,57778	0,30000	
Floresta Plantada	0,67778	0,55556	0,27778	0,57778

QUADRO A.27 – MATRIZ DE DISTÂNCIAS DAS SUBPOPULAÇÕES DE IRATI UTILIZANDO A DISTÂNCIA DE MANHATTAN

	Floresta Intocada	Capão 1	Capão 2	Capão 3	Capão 4
Capão 1	0,3316625				
Capão 2	0,2449490	0,3872983			
Capão 3	0,2449490	0,1732051	0,2449490		
Capão 4	0,3316625	0,1414214	0,3000000	0,1000000	
Floresta Explorada	0,1414214	0,3316625	0,3162278	0,2449490	0,3316625

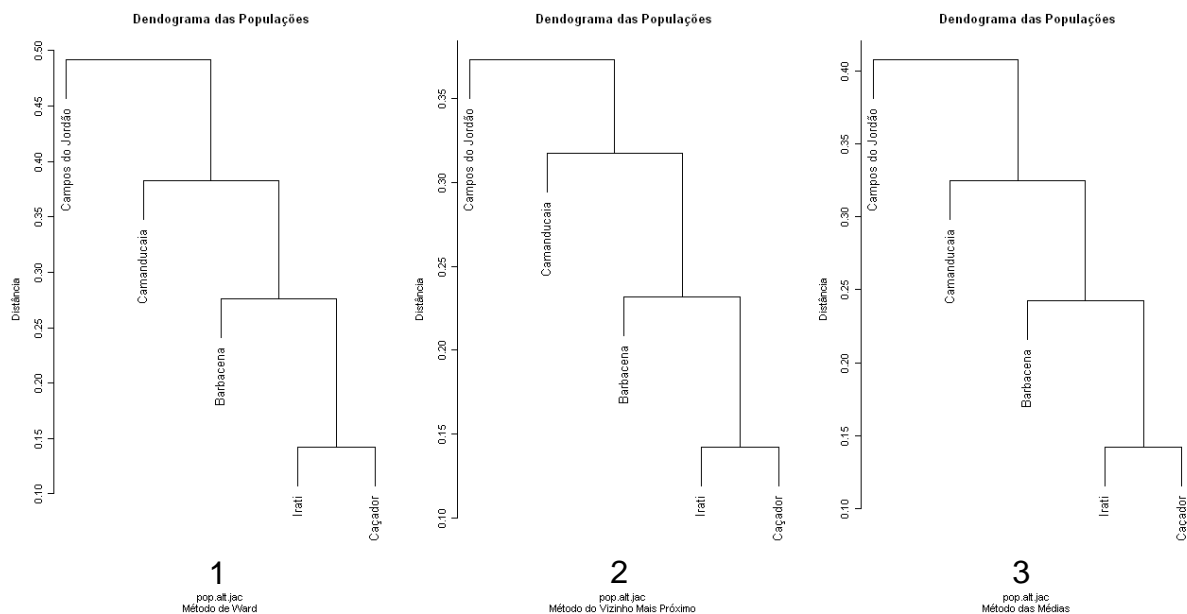
QUADRO A.28 – MATRIZ DE DISTÂNCIAS DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZANDO A DISTÂNCIA EUCLIDIANA

	Floresta Intocada	Capão 1	Capão 2	Capão 3	Capão 4
Capão 1	0,5				
Capão 2	0,4	0,7			
Capão 3	0,4	0,3	0,4		
Capão 4	0,5	0,2	0,5	0,1	
Floresta Explorada	0,2	0,5	0,4	0,4	0,5

QUADRO A.29 – MATRIZ DE DISTÂNCIAS DAS SUBPOPULAÇÕES DE CAÇADOR UTILIZANDO A DISTÂNCIA DE MANHATTAN

## APÊNDICE 3

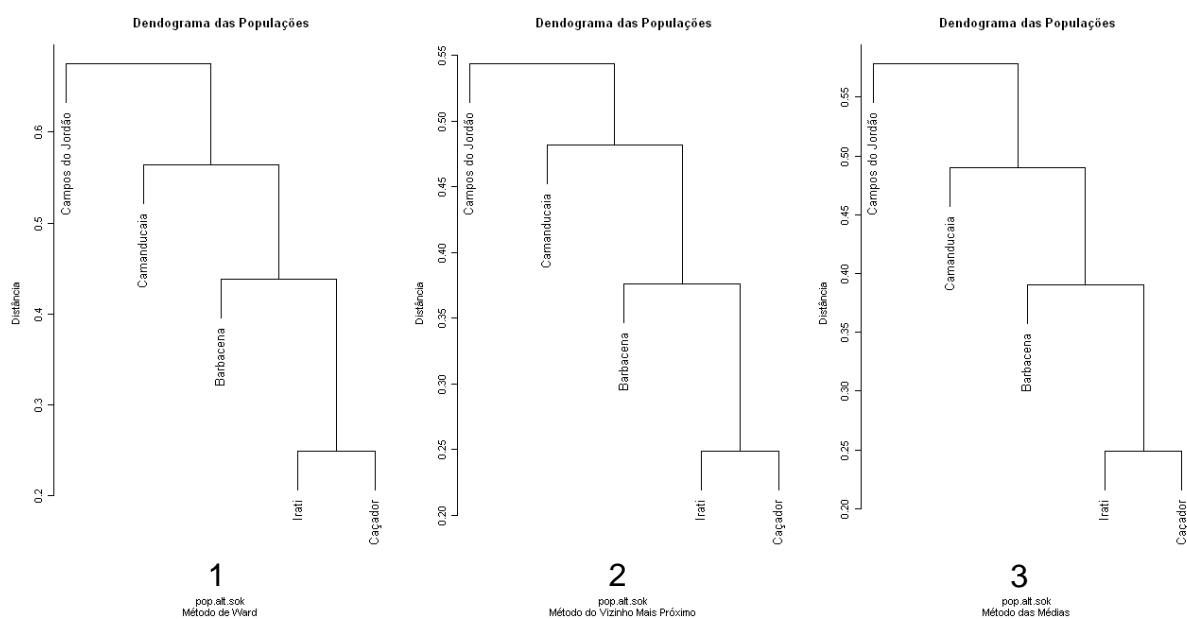
### DENDOGRAMAS DAS POPULAÇÕES



\* Nova amostragem

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

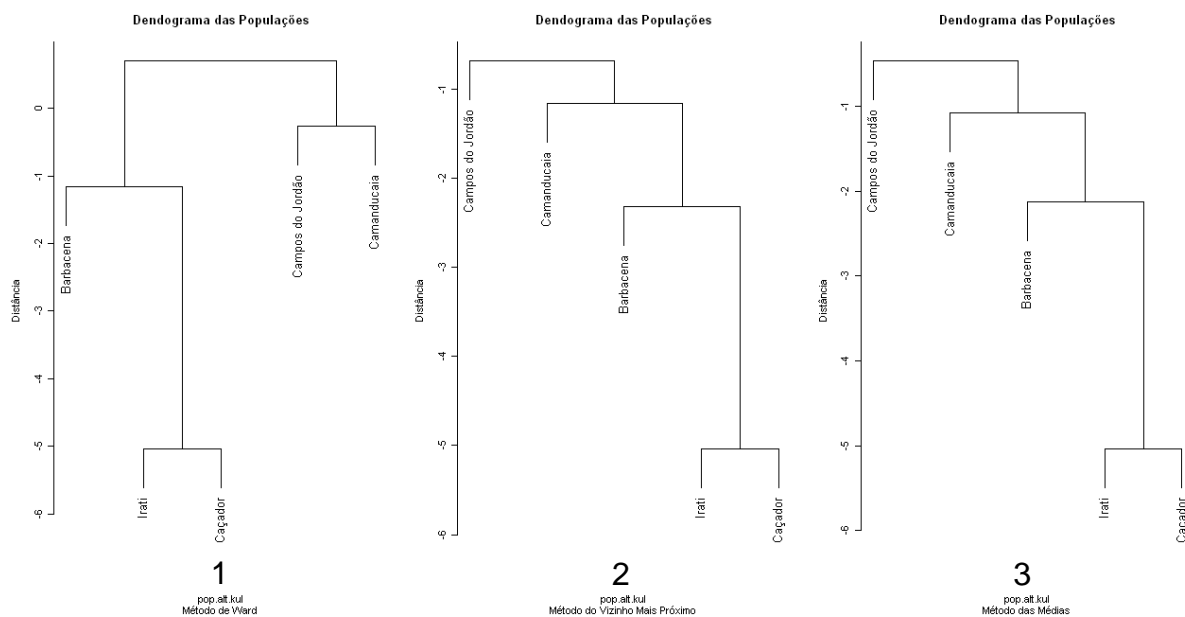
FIGURA A.1 – DENDOGRAMA DAS POPULAÇÕES\* CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO



\* Nova amostragem

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

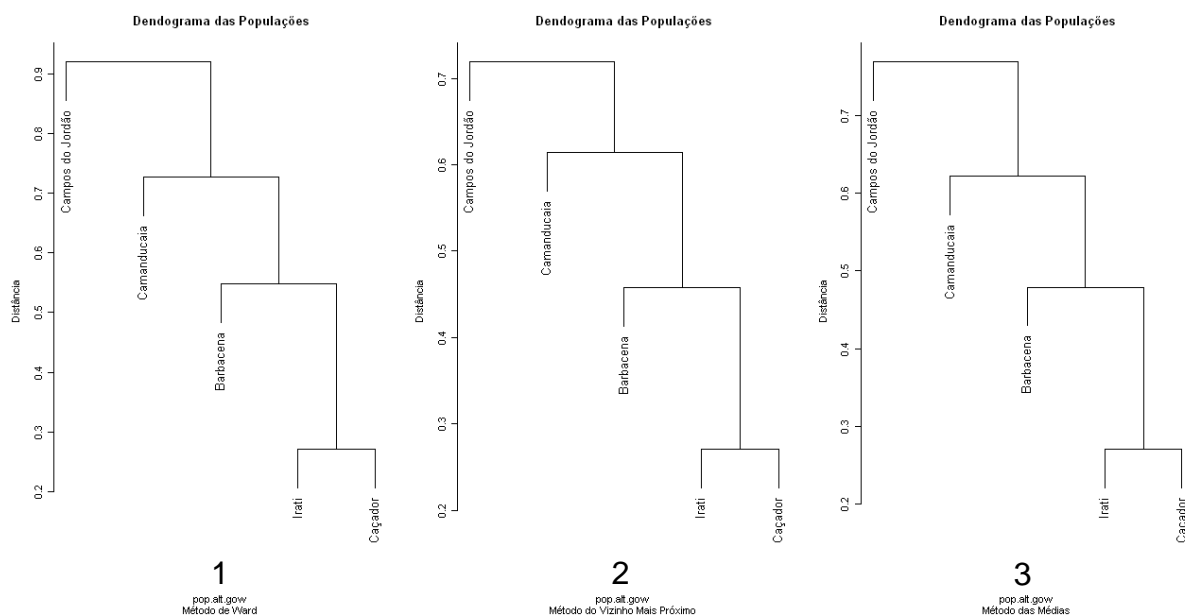
FIGURA A.2 – DENDOGRAMA DAS POPULAÇÕES\* CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO



\* Nova amostragem

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

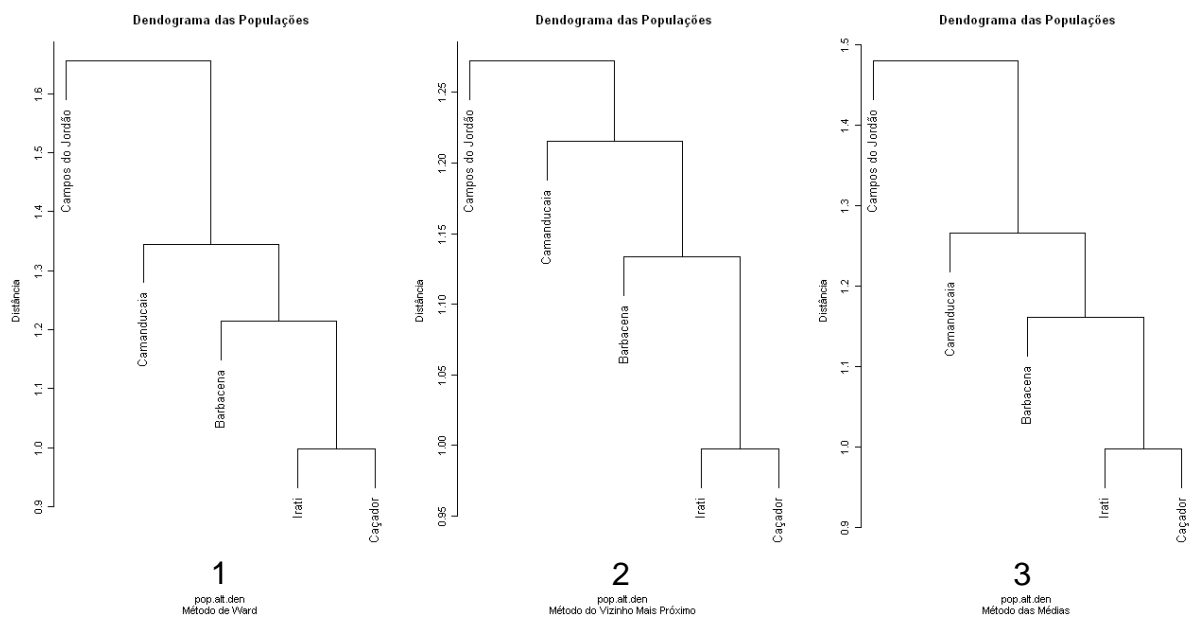
FIGURA A.3 – DENDOGRAMA DAS POPULAÇÕES\* CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO



\* Nova amostragem

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA A.4 – DENDOGRAMA DAS POPULAÇÕES\* CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO



\* Nova amostragem

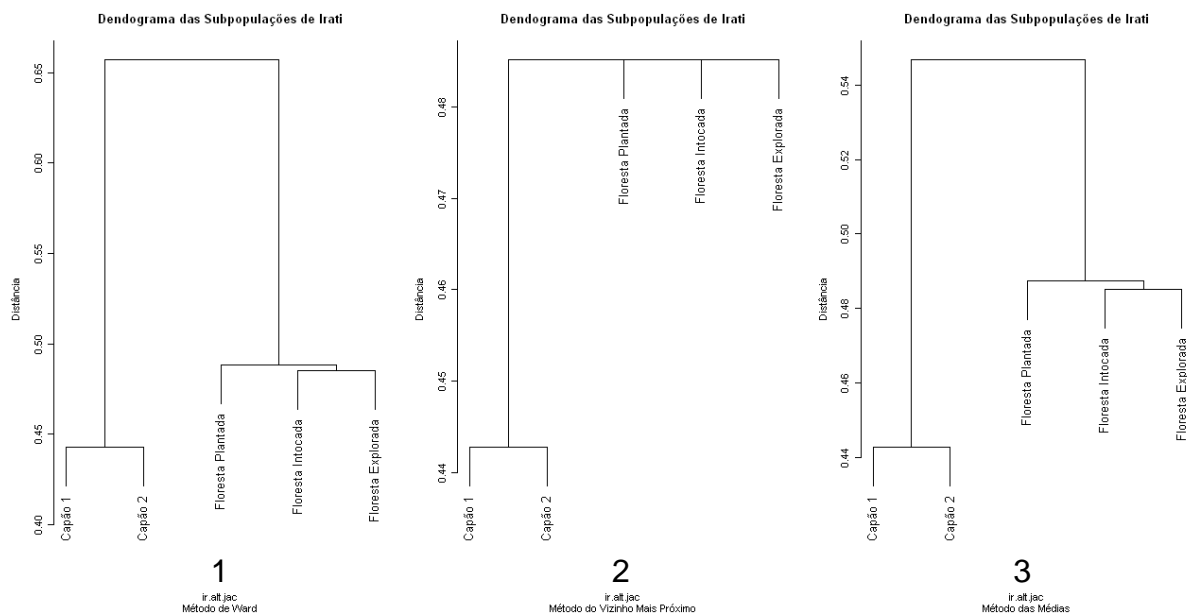
(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA A.5 – DENDOGRAMA DAS POPULAÇÕES\* CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO



## APÊNDICE 4

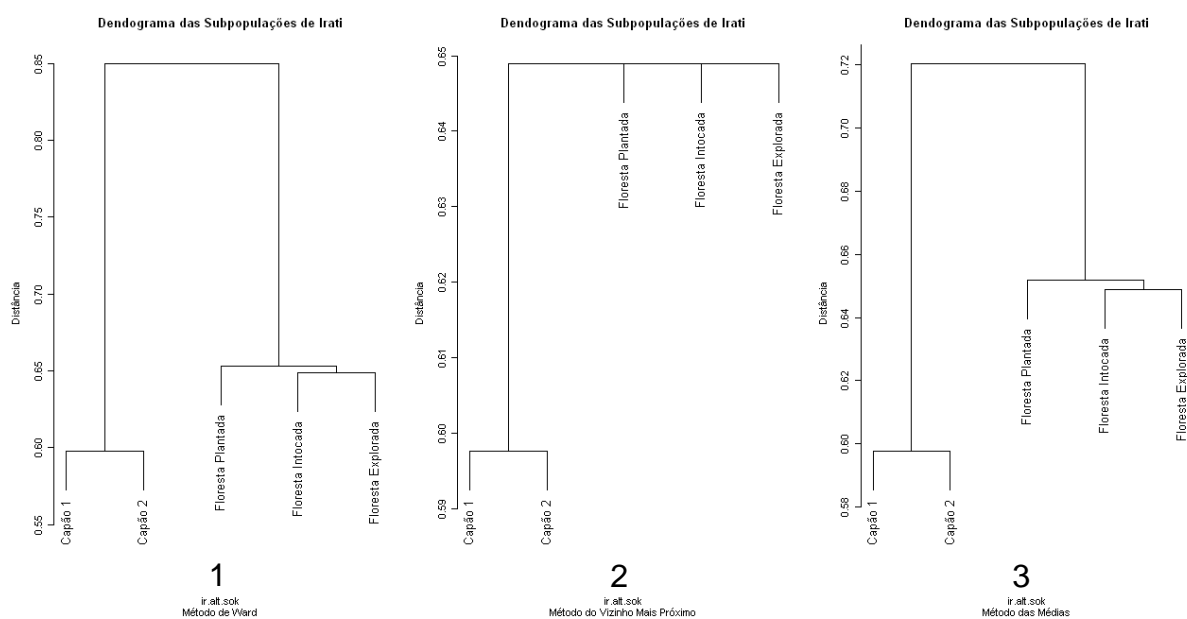
### DENDOGRAMAS DAS SUBPOPULAÇÕES



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

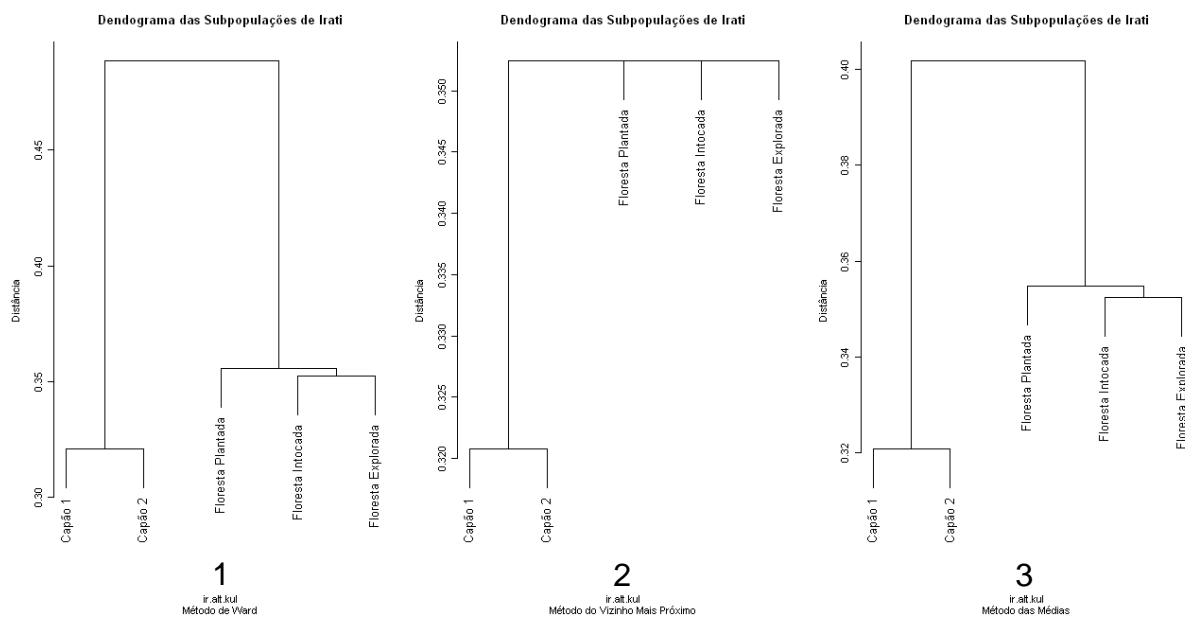
FIGURA A.6 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE IRATI CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

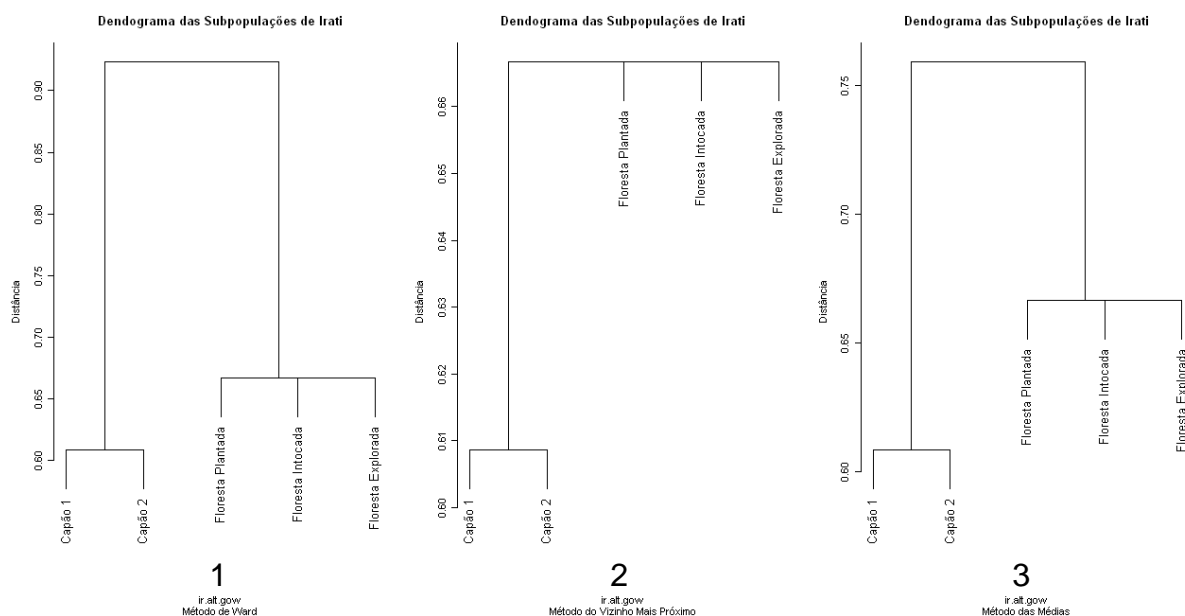
FIGURA A.7 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE IRATI CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

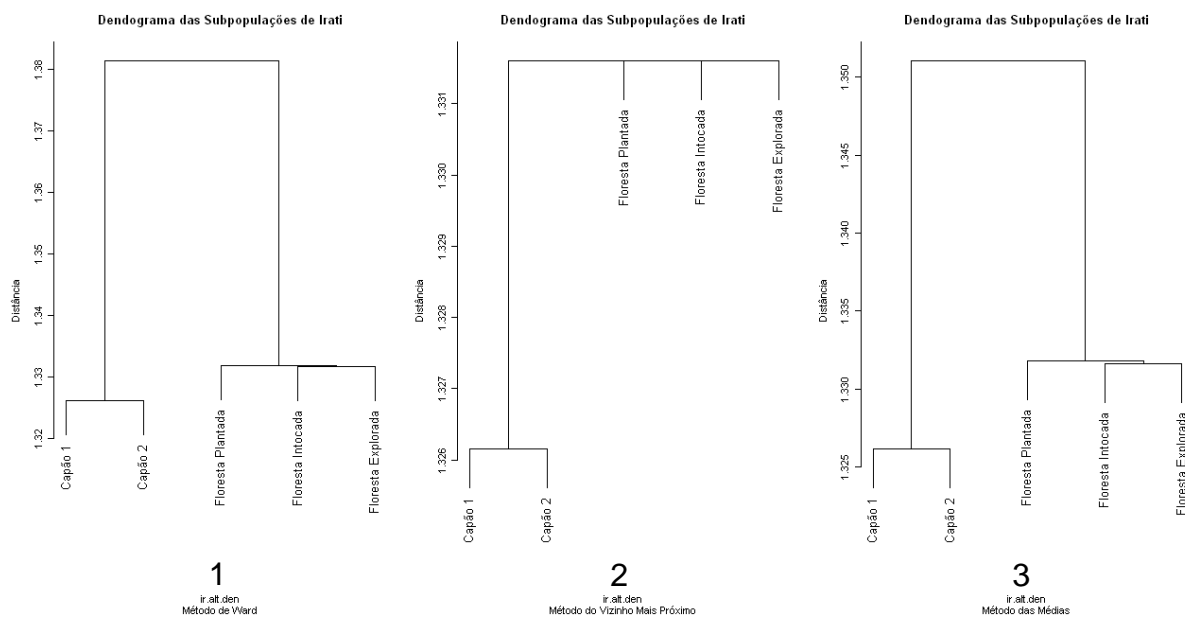
FIGURA A.8 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE IRATI CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

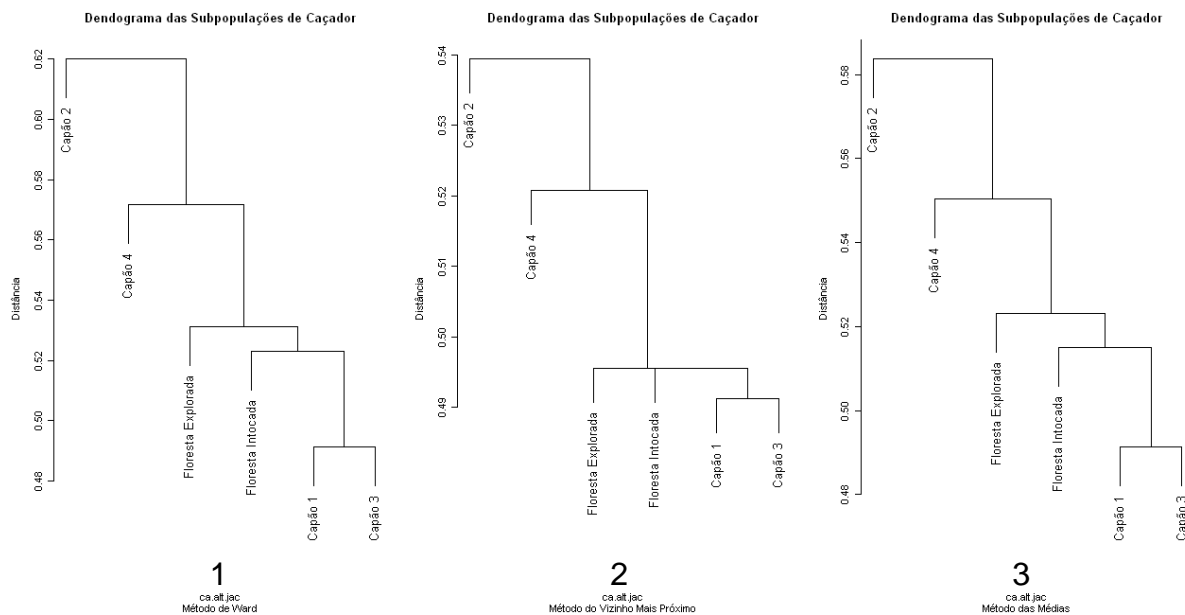
FIGURA A.9 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE IRATI CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

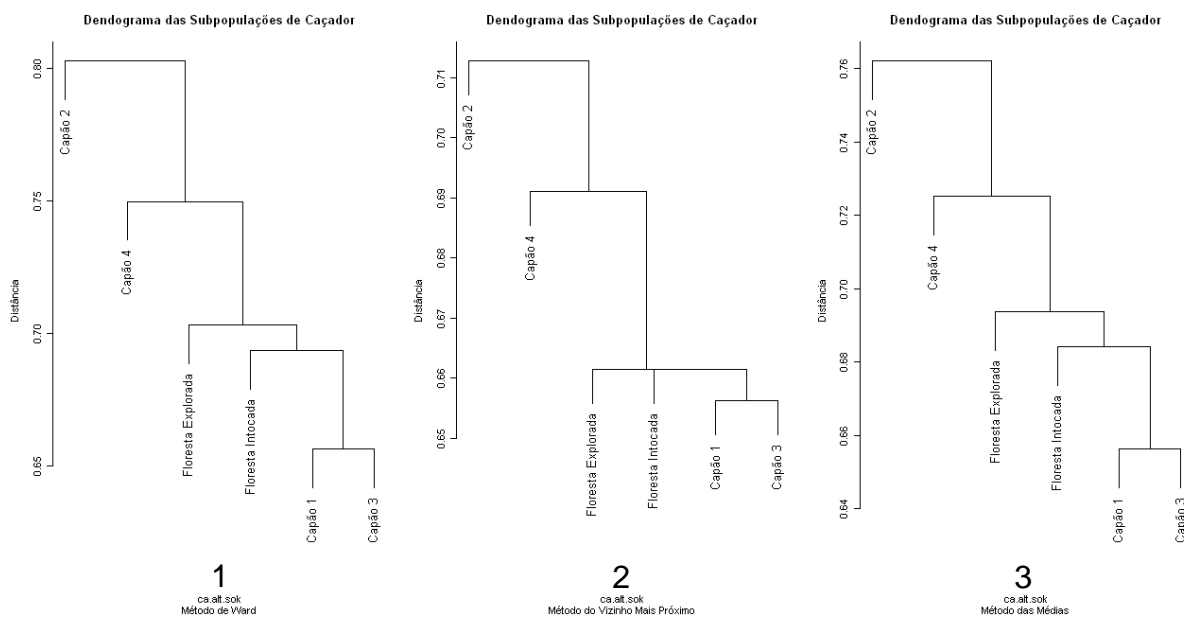
FIGURA A.10 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE IRATI CONSIDERANDO O COEFICIENTE DE HOLLIDAY E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinho Mais Próximo. (3) Método da Ligação Média.

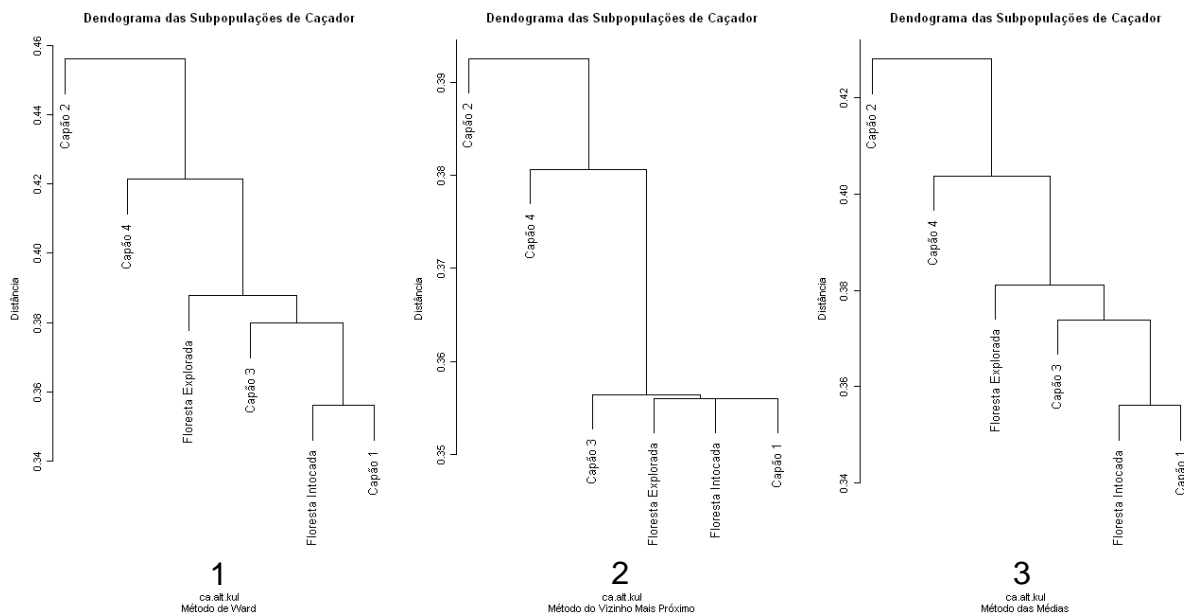
FIGURA A.11 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE CAÇADOR CONSIDERANDO O COEFICIENTE DE JACCARD E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

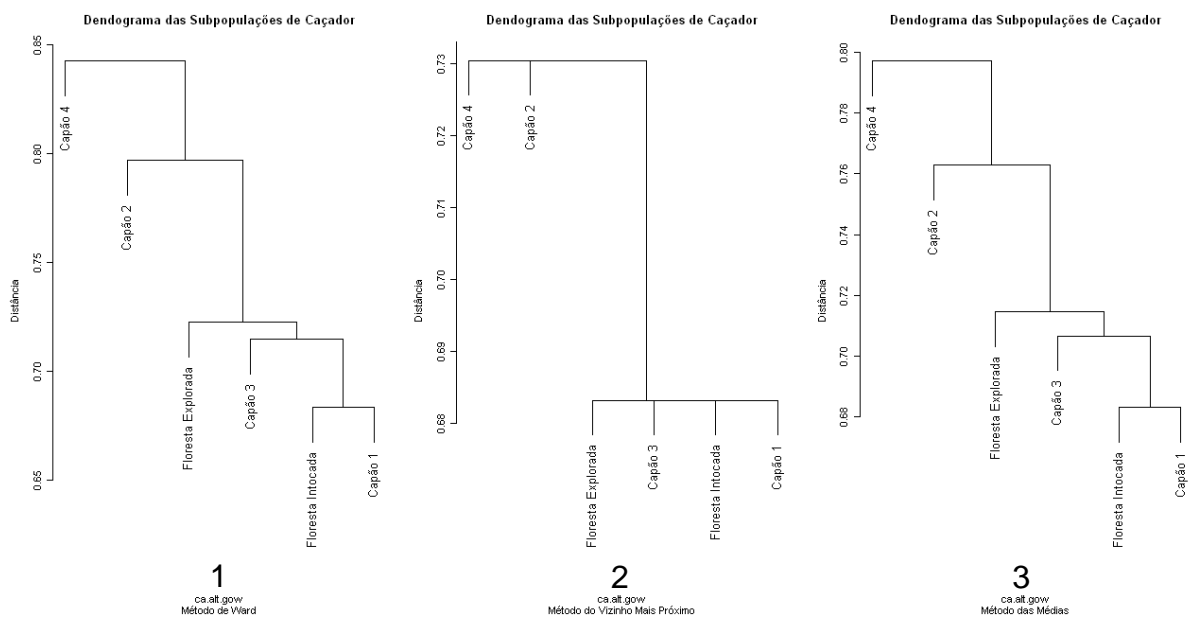
FIGURA A.12 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOKAL & SNEATH E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

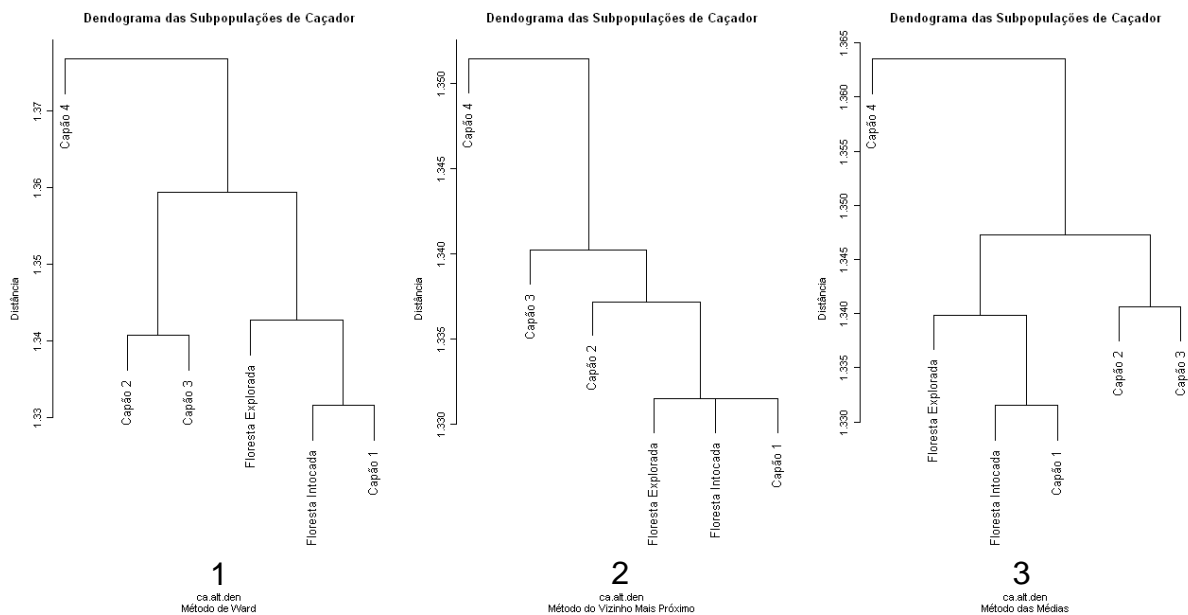
FIGURA A.13 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE CAÇADOR CONSIDERANDO O COEFICIENTE DE SOUTHWOOD E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA A.14 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE CAÇADOR CONSIDERANDO O COEFICIENTE DE GOWER & LEGENDRE E 3 MÉTODOS DE AGRUPAMENTO



\*\* Vetores reordenados

(1) Método de Ward. (2) Método do Vizinheiro Mais Próximo. (3) Método da Ligação Média.

FIGURA A.15 – DENDOGRAMA DAS SUBPOPULAÇÕES\*\* DE CAÇADOR CONSIDERANDO O COEFICIENTE DE HOLLIDAY (2002) E 3 MÉTODOS DE AGRUPAMENTO

## APÊNDICE 5

### COMANDOS UTILIZADOS NO PROGRAMA R

#### a) 1ª SITUAÇÃO DE ANÁLISE

##### i) Amostras das Populações

```
# Construção dos dendogramas
require(amap)
require(fpc)
require(simba)
cj<-c(1,,,,,1)
ir<-c(1,,,,,0)
ca<-c(1,,,,,0)
cam<-c(1,,,,,1)
ba<-c(1,,,,,1)
# Formando a matriz de dados:
pop<-matrix(c(cj,ir,ca,cam,ba),nrow=5,byrow=T,dimnames=list(c("Campos do
Jordão","Irati","Caçador","Camanducaia","Barbacena"),c()))
#Calculando a matriz de similaridade utilizando Jaccard
pop.jac<-sim(pop,method="jaccard")
# Transformando a matriz de similaridade em matriz de distância
pop.jac=sqrt(2*(1-pop.jac))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.jac,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Populações",
sub="Método de Ward")
plot(hclust(pop.jac,method="single"), cex=1.2, ylab="Distância", main="Dendograma das Populações",
sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.jac,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
pop.sok<-sim(pop,method="sokal1")
# Transformando a matriz de similaridade em matriz de distância
pop.sok=sqrt(2*(1-pop.sok))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.sok,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Populações",
sub="Método de Ward")
plot(hclust(pop.sok,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.sok,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood
pop.kul<-sim(pop,method="kulcz2insky")
# Transformando a matriz de similaridade em matriz de distância
pop.kul=sqrt(2*(1-pop.kul))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.kul,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Populações",
sub="Método de Ward")
```

```

plot(hclust(pop.kul,method="single"), cex=1.2, ylab="Distância", main="Dendograma das Populações",
sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.kul,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
pop.gow<-sim(pop,method="gower")
# Transformando a matriz de similaridade em matriz de distância
pop.gow=sqrt(2*(1-pop.gow))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.gow,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(pop.gow,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.gow,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
pop.den<-sim(pop,method="dennis")
# Transformando a matriz de similaridade em matriz de distância
pop.den=sqrt(2*(1-pop.den))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.den,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Populações",
sub="Método de Ward")
plot(hclust(pop.den,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.den,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Koeff
pop.rou<-sim(pop,method="rout2ledge")
# Transformando a matriz de similaridade em matriz de distância
pop.rou=sqrt(2*(1-pop.rou))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.rou,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Populações",
sub="Método de Ward")
plot(hclust(pop.rou,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.rou,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
# Alterando a posição das árvores dentro dos vetores
cj.alt<-c(0,,,,,1)
ir.alt<-c(1,,,,,0)
ca.alt<-c(1,,,,,1)
cam.alt<-c(0,,,,,0)
ba.alt<-c(1,,,1,)
# Formando a matriz de dados alterados:
pop.alt<-matrix(c(cj,alt,ir,alt,ca,alt,cam,alt,ba,alt),nrow=5,byrow=T,dimnames=list(c("Campos do
Jordão","Iratí","Caçador","Camanducaia","Barbacena"),c()))
#Calculando a matriz de similaridade utilizando Jaccard
pop.alt.jac<-sim(pop.alt,method="jaccard")
# Transformando a matriz de similaridade em matriz de distância
pop.alt.jac=sqrt(2*(1-pop.alt.jac))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.alt.jac,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(pop.alt.jac,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")

```

```

plot(hclust(pop.alt.jac,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
pop.alt.sok<-sim(pop.alt,method="sokal1")
# Transformando a matriz de similaridade em matriz de distância
pop.alt.sok=sqrt(2*(1-pop.alt.sok))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.alt.sok,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(pop.alt.sok,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.alt.sok,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood
pop.alt.kul<-sim(pop.alt,method="kulcz2insky")
# Transformando a matriz de similaridade em matriz de distância
pop.alt.kul=sqrt(2*(1-pop.alt.kul))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.alt.kul,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(pop.alt.kul,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.alt.kul,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
pop.alt.gow<-sim(pop.alt,method="gower")
# Transformando a matriz de similaridade em matriz de distância
pop.alt.gow=sqrt(2*(1-pop.alt.gow))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.alt.gow,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(pop.alt.gow,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.alt.gow,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
pop.alt.den<-sim(pop.alt,method="dennis")
# Transformando a matriz de similaridade em matriz de distância
pop.alt.den=sqrt(2*(1-pop.alt.den))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(pop.alt.den,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(pop.alt.den,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(pop.alt.den,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")

```

## ii) Subpopulações

```

# IRATI

# Construção dos dendogramas
require(AMAP)
require(fpc)
require(simba)

```



```

ir_int<-c(1,,,,1)
ir_is1<-c(1,,,,1)
ir_is2<-c(1,,,,1)
ir_exp<-c(1,,,,0)
ir_pla<-c(1,,,,0)
# Formando a matriz de dados:
ir<-matrix(c(ir_int,ir_is1,ir_is2,ir_exp,ir_pla),nrow=5,byrow=T,dimnames=list(c("Floresta
Intocada","Capão 1","Capão 2","Floresta Explorada","Floresta Plantada"),c()))
#Calculando a matriz de similaridade utilizando Jaccard
ir.jac<-sim(ir,method="jaccard")
# Transformando a matriz de similaridade em matriz de distância
ir.jac=sqrt(2*(1-ir.jac))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.jac,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método de Ward")
plot(hclust(ir.jac,method="single"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.jac,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
ir.sok<-sim(ir,method="sokal1")
# Transformando a matriz de similaridade em matriz de distância
ir.sok=sqrt(2*(1-ir.sok))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.sok,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método de Ward")
plot(hclust(ir.sok,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.sok,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood
ir.kul<-sim(ir,method="kulcz2insky")
# Transformando a matriz de similaridade em matriz de distância
ir.kul=sqrt(2*(1-ir.kul))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.kul,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método de Ward")
plot(hclust(ir.kul,method="single"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.kul,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
ir.gow<-sim(ir,method="gower")
# Transformando a matriz de similaridade em matriz de distância
ir.gow=sqrt(2*(1-ir.gow))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.gow,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método de Ward")
plot(hclust(ir.gow,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.gow,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
ir.den<-sim(ir,method="dennis")
# Transformando a matriz de similaridade em matriz de distância
ir.den=sqrt(2*(1-ir.den))

```

```

# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.den,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método de Ward")
plot(hclust(ir.den,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.den,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
# Alterando a posição das árvores dentro dos vetores
ir_int.alt<-c(1,,,,0)
ir_is1.alt<-c(1,,,,1)
ir_is2.alt<-c(1,,,,1)
ir_exp.alt<-c(1,,,,0)
ir_pla.alt<-c(1,,,,0)
# Formando a matriz de dados:
ir.alt<-
matrix(c(ir_int.alt,ir_is1.alt,ir_is2.alt,ir_exp.alt,ir_pla.alt),nrow=5,byrow=T,dimnames=list(c("Floresta
Intocada","Capão 1","Capão 2","Floresta Explorada","Floresta Plantada"),c()))
#Calculando a matriz de similaridade utilizando Jaccard
ir.alt.jac<-sim(ir.alt,method="jaccard")
# Transformando a matriz de similaridade em matriz de distância
ir.alt.jac=sqrt(2*(1-ir.alt.jac))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.alt.jac,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(ir.alt.jac,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.alt.jac,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
ir.alt.sok<-sim(ir.alt,method="sokal1")
# Transformando a matriz de similaridade em matriz de distância
ir.alt.sok=sqrt(2*(1-ir.alt.sok))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.alt.sok,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(ir.alt.sok,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.alt.sok,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood
ir.alt.kul<-sim(ir.alt,method="kulcz2insky")
# Transformando a matriz de similaridade em matriz de distância
ir.alt.kul=sqrt(2*(1-ir.alt.kul))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.alt.kul,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(ir.alt.kul,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.alt.kul,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
ir.alt.gow<-sim(ir.alt,method="gower")
# Transformando a matriz de similaridade em matriz de distância
ir.alt.gow=sqrt(2*(1-ir.alt.gow))
# Gerando os dendogramas
par(mfrow=c(1,3))

```

```

plot(hclust(ir.alt.gow,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(ir.alt.gow,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.alt.gow,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
ir.alt.den<-sim(ir.alt,method="dennis")
# Transformando a matriz de similaridade em matriz de distância
ir.alt.den=sqrt(2*(1-ir.alt.den))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ir.alt.den,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(ir.alt.den,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.alt.den,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")

```

## # CAÇADOR

```

# Construção dos dendogramas
require(amac)
require(fpc)
require(simba)
ca_int<-c(1,,,,,0)
ca_is1<-c(1,,,,,1)
ca_is2<-c(1,,,,,1)
ca_is3<-c(1,,,,,1)
ca_is4<-c(1,,,,,0)
ca_exp<-c(1,,,,,0)
# Formando a matriz de dados:
ca<-matrix(c(ca_int,ca_is1,ca_is2,ca_is3,ca_is4,ca_exp),nrow=6,byrow=T,dimnames=list(c("Floresta
Intocada","Capão 1","Capão 2","Capão 3","Capão 4","Floresta Explorada"),c()))
#Calculando a matriz de similaridade utilizando Jaccard
ca.jac<-sim(ca,method="jaccard")
# Transformando a matriz de similaridade em matriz de distância
ca.jac=sqrt(2*(1-ca.jac))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.jac,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Caçador", sub="Método de Ward")
plot(hclust(ca.jac,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(ca.jac,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
ca.sok<-sim(ca,method="sokal1")
# Transformando a matriz de similaridade em matriz de distância
ca.sok=sqrt(2*(1-ca.sok))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.sok,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.sok,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(ca.sok,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood

```

```

ca.kul<-sim(ca,method="kulcz2insky")
# Transformando a matriz de similaridade em matriz de distância
ca.kul=sqrt(2*(1-ca.kul))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.kul,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das Subpopulações
de Caçador", sub="Método de Ward")
plot(hclust(ca.kul,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinheiro Mais Próximo")
plot(hclust(ca.kul,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
ca.gow<-sim(ca,method="gower")
# Transformando a matriz de similaridade em matriz de distância
ca.gow=sqrt(2*(1-ca.gow))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.gow,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.gow,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinheiro Mais Próximo")
plot(hclust(ca.gow,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
ca.den<-sim(ca,method="dennis")
# Transformando a matriz de similaridade em matriz de distância
ca.den=sqrt(2*(1-ca.den))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.den,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.den,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinheiro Mais Próximo")
plot(hclust(ca.den,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
# Alterando a posição das árvores dentro dos vetores
ca_int.alt<-c(1,,,,,0)
ca_is1.alt<-c(1,,,,,1)
ca_is2.alt<-c(1,,,,,0)
ca_is3.alt<-c(1,,,,,0)
ca_is4.alt<-c(1,,,,,0)
ca_exp.alt<-c(1,,,,,0)
# Formando a matriz de dados:
ca.alt<-
matrix(c(ca_int.alt,ca_is1.alt,ca_is2.alt,ca_is3.alt,ca_is4.alt,ca_exp.alt),nrow=6,byrow=T,dimnames=lis
t(c("Floresta Intocada","Capão 1","Capão 2","Capão 3","Capão 4","Floresta Explorada"),c()))
#Calculando a matriz de similaridade utilizando Jaccard
ca.alt.jac<-sim(ca.alt,method="jaccard")
# Transformando a matriz de similaridade em matriz de distância
ca.alt.jac=sqrt(2*(1-ca.alt.jac))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.alt.jac,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.alt.jac,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinheiro Mais Próximo")
plot(hclust(ca.alt.jac,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
ca.alt.sok<-sim(ca.alt,method="sokal1")

```

```

# Transformando a matriz de similaridade em matriz de distância
ca.alt.sok=sqrt(2*(1-ca.alt.sok))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.alt.sok,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.alt.sok,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(ca.alt.sok,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood
ca.alt.kul<-sim(ca.alt,method="kulcz2insky")
# Transformando a matriz de similaridade em matriz de distância
ca.alt.kul=sqrt(2*(1-ca.alt.kul))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.alt.kul,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.alt.kul,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(ca.alt.kul,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
ca.alt.gow<-sim(ca.alt,method="gower")
# Transformando a matriz de similaridade em matriz de distância
ca.alt.gow=sqrt(2*(1-ca.alt.gow))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.alt.gow,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.alt.gow,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(ca.alt.gow,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
ca.alt.den<-sim(ca.alt,method="dennis")
# Transformando a matriz de similaridade em matriz de distância
ca.alt.den=sqrt(2*(1-ca.alt.den))
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(ca.alt.den,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.alt.den,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(ca.alt.den,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")

```

## b) 2ª SITUAÇÃO DE ANÁLISE

### i) Amostras das Populações

```

require(amap)
require(fpc)
require(simba)
mcj<-c(0,62857,0,94286,0,70000,0,47143,0,78571,0,65714)
mir<-c(1,00000,0,89583,0,95833,1,00000,1,00000,0,50000)
mca<-c(0,93333,0,95000,0,90000,1,00000,1,00000,0,45000)
mcam<-c(0,78571,0,97143,0,94286,0,52857,0,57143,0,57143)

```

```

mba<-c(0,87500,1,00000,1,00000,0,50000,0,96875,0,75000)
# Formando a matriz de dados:
mpop<-matrix(c(mcj,mir,mca,mcam,mba),nrow=5,byrow=T,dimnames=list(c("Campos do
Jordão", "Irati", "Caçador", "Camanducaia", "Barbacena"),c()))
#Calculando a matriz de distância utilizando a distância Euclidiana
mpop.euc<-dist(mpop,method="euclidean")
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(mpop.euc,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(mpop.euc,method="single"), cex=1.2, ylab=" Distância ", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(mpop.euc,method="average"), cex=1.2, ylab=" Distância ", main="Dendograma das
Populações", sub="Método das Médias")
#Calculando a matriz de distância utilizando a distância de Manhattan
mpop.man<-dist(mpop,method="manhattan")
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(mpop.man,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método de Ward")
plot(hclust(mpop.man,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método do Vizinho Mais Próximo")
plot(hclust(mpop.man,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Populações", sub="Método das Médias")

```

## ii) Subpopulações

# IRATI

```

require (amap)
require(fpc)
require(simba)
mir_int<-c(1,00000,0,80000,1,00000,1,00000,1,00000,0,70000)
mir_is1<-c(1,00000,1,00000,0,77778,1,00000,1,00000,0,55556)
mir_is2<-c(1,00000,1,00000,1,00000,1,00000,1,00000,0,50000)
mir_exp<-c(1,00000,0,70000,1,00000,1,00000,1,00000,0,50000)
mir_pla<-c(1,00000,1,00000,1,00000,1,00000,1,00000,0,22222)
# Formando a matriz de dados:
mir<-matrix(c(mir_int,mir_is1,mir_is2,mir_exp,mir_pla),nrow=5,byrow=T,dimnames=list(c("Floresta
Intocada", "Capão 1", "Capão 2", "Floresta Explorada", "Floresta Plantada"),c()))
#Calculando a matriz de distância utilizando a distância Euclidiana
mir.euc<-dist(mir,method="euclidean")
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(mir.euc,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(mir.euc,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(mir.euc,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de distância utilizando a distância de Manhattan
mir.man<-dist(mir,method="manhattan")
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(mir.man,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(mir.man,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método do Vizinho Mais Próximo")

```

```
plot(hclust(mir.man,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
```

```
# CAÇADOR
```

```
require (amap)
require(fpc)
require(simba)
mca_int<-c(1,00000,0,90000,0,90000,1,00000,1,00000,0,30000)
mca_is1<-c(1,00000,1,00000,1,00000,1,00000,1,00000,0,60000)
mca_is2<-c(0,90000,0,90000,0,70000,1,00000,1,00000,0,40000)
mca_is3<-c(0,90000,1,00000,0,90000,1,00000,1,00000,0,50000)
mca_is4<-c(0,90000,1,00000,0,90000,1,00000,1,00000,0,60000)
mca_exp<-c(0,90000,0,90000,1,00000,1,00000,1,00000,0,30000)
# Formando a matriz de dados:
mca<-
matrix(c(mca_int,mca_is1,mca_is2,mca_is3,mca_is4,mca_exp),nrow=6,byrow=T,dimnames=list(c("Fl
oresta Intocada","Capão 1","Capão 2","Capão 3","Capão 4","Floresta Explorada"),c()))
#Calculando a matriz de distância utilizando a distância Euclidiana
mca.euc<-dist(mca,method="euclidean")
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(mca.euc,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(mca.euc,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(mca.euc,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de distância utilizando a distância de Manhattan
mca.man<-dist(mca,method="manhattan")
# Gerando os dendogramas
par(mfrow=c(1,3))
plot(hclust(mca.man,method="ward"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(mca.man,method="single"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(mca.man,method="average"), cex=1.2, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
```

c) 3ª SITUAÇÃO DE ANÁLISE

i) Amostras das Populações

```
require (amap)
require(fpc)
require(simba)
a1<-c(1,1,1,0,1,1)
...
e160<-c(1,1,1,0,1,1)
# Formando a matriz de dados:
pop.arvnum<-
matrix(c(a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a12,a13,a14,a15,a16,a17,a18,a19,a20,a21,a22,a23,a24
,a25,a26,a27,a28,a29,a30,a31,a32,b33,b34,b35,b36,b37,b38,b39,b40,b41,b42,b43,b44,b45,b46,b47,
b48,b49,b50,b51,b52,b53,b54,b55,b56,b57,b58,b59,b60,b61,b62,b63,b64,c65,c66,c67,c68,c69,c70,c
71,c72,c73,c74,c75,c76,c77,c78,c79,c80,c81,c82,c83,c84,c85,c86,c87,c88,c89,c90,c91,c92,c93,c94,
c95,c96,d97,d98,d99,d100,d101,d102,d103,d104,d105,d106,d107,d108,d109,d110,d111,d112,d113,d
114,d115,d116,d117,d118,d119,d120,d121,d122,d123,d124,d125,d126,d127,d128,e129,e130,e131,e
```





```

ia1<-c(1,1,1,1,1,0)
...
ie48<-c(1,1,1,1,1,0)
# Formando a matriz de dados:
ir.arvnum<-
matrix(c(ia1,ia2,ia3,ia4,ia5,ia6,ia7,ia8,ia9,ia10,ib11,ib12,ib13,ib14,ib15,ib16,ib17,ib18,ib19,ic20,ic21,ic
22,ic23,ic24,ic25,ic26,ic27,ic28,ic29,id30,id31,id32,id33,id34,id35,id36,id37,id38,id39,ie40,ie41,ie42,ie
43,ie44,ie45,ie46,ie47,ie48),nrow=48,byrow=T,dimnames=list(c(1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,
3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,4,5,5,5,5,5,5,5,5,5,5),c()))
#Calculando a matriz de similaridade utilizando Jaccard
ir.arvnum.jac<-sim(ir.arvnum,method="jaccard")
# Transformando a matriz de similaridade em matriz de distância
ir.arvnum.jac=sqrt(2*(1-ir.arvnum.jac))
#Gerando os dendogramas
plot(hclust(ir.arvnum.jac,method="ward"), cex=0.8, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método de Ward")
plot(hclust(ir.arvnum.jac,method="single"), ylab="Distância", main="Dendograma das Subpopulações
de Irati", sub="Método do Vizinho Mais Próximo")
plot(hclust(ir.arvnum.jac,method="average"), ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
ir.arvnum.sok<-sim(ir.arvnum,method="sokal1")
# Transformando a matriz de similaridade em matriz de distância
ir.arvnum.sok=sqrt(2*(1-ir.arvnum.sok))
#Gerando os dendogramas
plot(hclust(ir.arvnum.sok,method="average"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood
ir.arvnum.kul<-sim(ir.arvnum,method="kulcz2insky")
# Transformando a matriz de similaridade em matriz de distância
ir.arvnum.kul=sqrt(2*(1-ir.arvnum.kul))
# Gerando os dendogramas
plot(hclust(ir.arvnum.kul,method="average"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
ir.arvnum.gow<-sim(ir.arvnum,method="gower")
# Transformando a matriz de similaridade em matriz de distância
ir.arvnum.gow=sqrt(2*(1-ir.arvnum.gow))
# Gerando os dendogramas
plot(hclust(ir.arvnum.gow,method="average"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
ir.arvnum.den<-sim(ir.arvnum,method="dennis")
# Transformando a matriz de similaridade em matriz de distância
ir.arvnum.den=sqrt(2*(1-ir.arvnum.den))
# Gerando os dendogramas
plot(hclust(ir.arvnum.den,method="average"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Koeff
ir.arvnum.rou<-sim(ir.arvnum,method="rou2ledge")
# Transformando a matriz de similaridade em matriz de distância
ir.arvnum.rou=sqrt(2*(1-ir.arvnum.rou))
# Gerando os dendogramas
plot(hclust(ir.arvnum.rou,method="average"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Irati", sub="Método das Médias")

# CAÇADOR

require (amap)

```

```

require(fpc)
require(simba)
ca1<-c(1,1,1,1,1,0)
...
cf60<-c(1,1,1,1,1,0)
# Formando a matriz de dados:
ca.arvnum<-
matrix(c(ca1,ca2,ca3,ca4,ca5,ca6,ca7,ca8,ca9,ca10,cb11,cb12,cb13,cb14,cb15,cb16,cb17,cb18,cb19
,cb20,cc21,cc22,cc23,cc24,cc25,cc26,cc27,cc28,cc29,cc30,cd31,cd32,cd33,cd34,cd35,cd36,cd37,cd
38,cd39,cd40,ce41,ce42,ce43,ce44,ce45,ce46,ce47,ce48,ce49,ce50,cf51,cf52,cf53,cf54,cf55,cf56,cf5
7,cf58,cf59,cf60),nrow=60,byrow=T,dimnames=list(c(1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,
3,3,3,3,4,4,4,4,4,4,4,4,4,5,5,5,5,5,5,5,5,5,6,6,6,6,6,6,6,6),c()))
#Calculando a matriz de similaridade utilizando Jaccard
ca.arvnum.jac<-sim(ca.arvnum,method="jaccard")
#Transformando a matriz de similaridade em matriz de distância
ca.arvnum.jac=sqrt(2*(1-ca.arvnum.jac))
#Gerando os dendogramas
plot(hclust(ca.arvnum.jac,method="ward"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método de Ward")
plot(hclust(ca.arvnum.jac,method="single"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método do Vizinho Mais Próximo")
plot(hclust(ca.arvnum.jac,method="average"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Sokal
ca.arvnum.sok<-sim(ca.arvnum,method="sokal1")
#Transformando a matriz de similaridade em matriz de distância
ca.arvnum.sok=sqrt(2*(1-ca.arvnum.sok))
# Gerando os dendogramas
plot(hclust(ca.arvnum.sok,method="average"), ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Southwood
ca.arvnum.kul<-sim(ca.arvnum,method="kulcz2insky")
#Transformando a matriz de similaridade em matriz de distância
ca.arvnum.kul=sqrt(2*(1-ca.arvnum.kul))
# Gerando os dendogramas
plot(hclust(ca.arvnum.kul,method="average"), ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Gower
ca.arvnum.gow<-sim(ca.arvnum,method="gower")
#Transformando a matriz de similaridade em matriz de distância
ca.arvnum.gow=sqrt(2*(1-ca.arvnum.gow))
# Gerando os dendogramas
plot(hclust(ca.arvnum.gow,method="average"), ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Holliday
ca.arvnum.den<-sim(ca.arvnum,method="dennis")
#Transformando a matriz de similaridade em matriz de distância
ca.arvnum.den=sqrt(2*(1-ca.arvnum.den))
# Gerando os dendogramas
plot(hclust(ca.arvnum.den,method="average"), ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")
#Calculando a matriz de similaridade utilizando Koeff
ca.arvnum.rou<-sim(ca.arvnum,method="rout2ledge")
#Transformando a matriz de similaridade em matriz de distância
ca.arvnum.rou=sqrt(2*(1-ca.arvnum.rou))
# Gerando os dendogramas
plot(hclust(ca.arvnum.rou,method="average"), cex=0.9, ylab="Distância", main="Dendograma das
Subpopulações de Caçador", sub="Método das Médias")

```

## d) ÁRVORES DE DECISÃO

## i) Amostras das Populações

```
require(tree)
# Leitura do conjunto de dados (indicar o diretório correto)
dados<-read.table("pop arvore decisao.txt",h=T)
pop.decisao=tree(factor(POP)~GOTB+GOTC+PGMAA+MDHB+SKDHB+X6PGDHB,data=dados)
plot(pop.decisao)
text(pop.decisao, main="Árvore de Decisão para as populações")
subpop.decisao=tree(factor(SUBPOP)~GOTB+GOTC+PGMAA+MDHB+SKDHB+X6PGDHB,data=da
dos)
plot(subpop.decisao)
text(subpop.decisao, main="Árvore de Decisão para as subpopulações")
```

## ii) Subpopulações

## # IRATI

```
# Leitura do conjunto de dados (indicar o diretório correto)
ir<-read.table("ir arvore decisao.txt",h=T)
ir.decisao=tree(factor(SUBPOP)~GOTB+GOTC+PGMAA+MDHB+SKDHB+X6PGDHB,data=ir)
plot(ir.decisao)
text(ir.decisao, main="Árvore de Decisão para as subpopulações de Irati")
```

## # CAÇADOR

```
# Leitura do conjunto de dados (indicar o diretório correto)
ca<-read.table("ca arvore decisao.txt",h=T)
ca.decisao=tree(factor(SUBPOP)~GOTB+GOTC+PGMAA+MDHB+SKDHB+X6PGDHB,data=ca)
plot(ca.decisao)
text(ca.decisao, main="Árvore de Decisão para as subpopulações de Caçador")
```