

BIOINFORMÁTICA E TÉCNICAS MULTIVARIADAS NA ANÁLISE DA  
DIVERGÊNCIA GENÉTICA EM *Carica papaya* L.

**CIBELLE DEGEL BARBOSA**

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE  
DARCY RIBEIRO

CAMPOS DOS GOYTACAZES - RJ  
NOVEMBRO – 2010

BIOINFORMÁTICA E TÉCNICAS MULTIVARIADAS NA ANÁLISE DA  
DIVERGÊNCIA GENÉTICA EM *Carica papaya* L.

**CIBELLE DEGEL BARBOSA**

Tese apresentada ao Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Doutor em Produção Vegetal.

**Orientador: Prof. Dr. Alexandre Pio Viana**

CAMPOS DOS GOYTACAZES - RJ

NOVEMBRO – 2010

BIOINFORMÁTICA E TÉCNICAS MULTIVARIADAS NA ANÁLISE DA  
DIVERGÊNCIA GENÉTICA EM *Carica papaya* L.

**CIBELLE DEGEL BARBOSA**

Tese apresentada ao Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Doutor em Produção Vegetal.

Aprovado em 18 de novembro de 2010.

Comissão Examinadora:

---

Messias Gonzaga Pereira (Ph. D., Melhoramento de Plantas) – UENF

---

Antônio Teixeira do Amaral Júnior (D. Sc., Genética e Melhoramento) – UENF

---

Marcos José de Oliveira Fonseca (D. Sc., Produção Vegetal) – EMBRAPA

---

Alexandre Pio Viana (D. Sc., Produção Vegetal) – UENF (Orientador)

## **Agradecimentos**

**“Tu és digno, Senhor e Deus nosso, de receber a glória, a honra e o poder, porque todas as coisas tu criaste, sim, por causa da tua vontade vieram a existir e foram criadas.” (Apocalipse 4.11)**

A Deus, por ter colocado em meu caminho pessoas tão especiais e pelo amparo em todos os momentos;

Aos meus professores que, com imenso carinho, me acolheram e me ensinaram a caminhar nesse momento tão especial de minha vida. À Professora Rosana Rodrigues, ao Professor Messias Pereira, ao Professor Rogério Figueiredo Daher e ao Professor Antônio Teixeira do Amaral Júnior, que tanto contribuíram para o meu crescimento, com valiosas orientações. Aos demais professores pelos ensinamentos concedidos;

Aos colegas de pós-graduação que tanto me ajudaram, Leandro Simões Azeredo Gonçalves e Silvana Silva Red Quintal. E a todos os outros que estiveram ao meu lado durante esses anos;

A Universidade Estadual do Norte Fluminense Darcy Ribeiro, pela oportunidade concedida para a realização deste curso;

E especialmente ao meu orientador, Professor Alexandre Pio Viana, pela oportunidade, confiança, apoio e compreensão.

## SUMÁRIO

INTRODUÇÃO .....	1
REVISÃO DA LITERATURA .....	4
1. Aspectos gerais da cultura do mamão .....	4
1.1. Aspectos botânicos .....	6
1.2. Genótipos de mamoeiro .....	7
2. Aspectos do melhoramento genético do mamoeiro .....	8
3. Análise da diversidade de recursos genéticos .....	10
3.1. Análise por componentes principais .....	11
3.2. Análises de agrupamento .....	13
3.3. Análise discriminante de Anderson .....	14
3.4. Bioinformática.....	15
3.4.1. Redes neurais artificiais .....	16
MATERIAL E MÉTODOS .....	20
1. Área Experimental.....	20
2. Material Genético .....	21
3. Caracteres analisados.....	24
4. Análise dos dados .....	25
4.1. Distância Euclidiana .....	25
4.2. Componentes principais .....	26
4.3. Método de otimização Tocher .....	28
4.4. Análise discriminante de Anderson .....	29

4.5. Rede neural artificial.....	30
4.5.1. Processo Competitivo.....	31
4.5.2. Processo Adaptativo.....	34
4.6. Gráfico Boxplot.....	35
RESULTADOS .....	37
1. Componentes principais .....	37
1.1. Primeira época .....	37
1.2. Segunda época .....	41
1.3. Média das épocas .....	45
2. Método de otimização Tocher .....	48
2.1. Primeira época .....	48
2.2. Segunda época .....	52
2.3. Média das épocas .....	55
3. Rede neural artificial.....	59
3.1. Determinação do número ótimo de grupos .....	59
3.2. Primeira época .....	60
3.3. Segunda época .....	62
3.4. Média das épocas .....	64
3.5. Gráficos Boxplot .....	65
4. Análise discriminante de Anderson .....	69
4.1. Primeira época .....	69
4.2. Segunda época .....	71
4.3. Média das épocas .....	74
DISCUSSÃO .....	77
RESUMO E CONCLUSÕES .....	83
REFERÊNCIAS BIBLIOGRÁFICAS .....	85
APÊNDICE 1 .....	100
ANEXO 1 .....	104

## ÍNDICE DE FIGURAS

- Figura 1. Arquitetura da rede neural artificial Kohonen composta por  $n$  acessos e  $m$  caracteres na camada de entrada, e por  $k$  neurônios na camada competitiva.....31
- Figura 2. Dispersão dos 50 acessos de mamoeiro (*Carica papaya* L.) em dois componentes principais, com base nas distâncias Euclidianas médias, estimadas a partir de 9 caracteres quantitativos, na primeira época. A descrição dos acessos encontra-se na seção 2 do capítulo Material e Métodos. ....39
- Figura 3. Dispersão dos 48 acessos de mamoeiro (*Carica papaya* L.) em dois componentes principais, com base nas distâncias Euclidianas médias, estimadas a partir de 9 caracteres quantitativos, na segunda época. A descrição dos acessos encontra-se na seção 2 do capítulo Material e Métodos. ....43
- Figura 4. Dispersão dos 48 acessos de mamoeiro (*Carica papaya* L.) em dois componentes principais, com base nas distâncias Euclidianas médias, estimadas a partir de 9 caracteres quantitativos, na média das épocas. A descrição dos acessos encontra-se na seção 2 do capítulo Material e Métodos. ....47
- Figura 5. Boxplot dos valores mínimos, máximos e mediana para os três grupos propostos pela rede neural artificial, na média das épocas, a partir de 9

caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.).....	66
--	----

## ÍNDICE DE QUADROS

Quadro 1. Cinquenta acessos avaliados e mantidos na coleção de germoplasma Caliman/UENF da espécie <i>Carica papaya</i> L.....	21
Quadro 2. Estimativa das variâncias, autovalores (AV), associados aos componentes principais (CP), e respectivos coeficientes de ponderação (autovetores) de 9 caracteres quantitativos avaliados em 50 acessos de mamoeiro ( <i>Carica papaya</i> L.), na primeira época.....	38
Quadro 3. Estimativa das variâncias, autovalores (AV), associados aos componentes principais (CP), e respectivos coeficientes de ponderação (autovetores) de 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na segunda época.....	42
Quadro 4. Estimativa das variâncias, autovalores (AV), associados aos componentes principais (CP), e respectivos coeficientes de ponderação (autovetores) de 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na média das épocas.....	45
Quadro 5. Grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na primeira época, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos avaliados em 50 acessos.....	49
Quadro 6. Estimativa das distâncias genéticas intragrupos e intergrupos dos grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na primeira época, formados através do método de agrupamento de Tocher, a partir da distância	

Euclidiana, baseando-se em 9 caracteres quantitativos e avaliados em 50 acessos.....	51
Quadro 7. Grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na segunda época, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.....	52
Quadro 8. Estimativa das distâncias genéticas intragrupos e intergrupos dos grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na segunda época, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos e avaliados em 48 acessos.....	55
Quadro 9. Grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na média das épocas, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.....	56
Quadro 10. Estimativa das distâncias genéticas intragrupos e intergrupos dos grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na média das épocas, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos e avaliados em 48 acessos.....	59
Quadro 11. Resultado da análise discriminante de Anderson, segundo percentual de classificação correta dos grupos propostos pela rede neural artificial, para <i>Carica papaya</i> L.....	60
Quadro 12. Grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na primeira época, formados através da rede neural artificial, a partir das médias originais, baseando-se em 9 caracteres quantitativos avaliados em 50 acessos.....	61
Quadro 13. Grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.), na segunda época, formados através da rede neural artificial, a partir das médias originais, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.....	63
Quadro 14. Grupos dos acessos de mamoeiro ( <i>Carica papaya</i> L.) formados através da rede neural artificial, a partir das médias das épocas, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.....	65

Quadro 15. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela técnica de componentes principais, considerando 9 caracteres quantitativos avaliados em 50 acessos de mamoeiro ( <i>Carica papaya</i> L.), na primeira época. ....	70
Quadro 16. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pelo método de otimização Tocher, considerando 9 caracteres quantitativos avaliados em 50 acessos de mamoeiro ( <i>Carica papaya</i> L.), na primeira época. ....	70
Quadro 17. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela rede neural artificial, considerando 9 caracteres quantitativos avaliados em 50 acessos de mamoeiro ( <i>Carica papaya</i> L.), na primeira época. ....	71
Quadro 18. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela técnica de componentes principais, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na segunda época.....	72
Quadro 19. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pelo método de otimização Tocher, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na segunda época.....	73
Quadro 20. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela rede neural artificial, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na segunda época. ....	73
Quadro 21. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela técnica de componentes principais, considerando	

9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na média das épocas.....	75
Quadro 22. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pelo método de otimização Tocher, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na média das épocas.....	75
Quadro 23. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela rede neural artificial, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro ( <i>Carica papaya</i> L.), na média das épocas.....	76
Quadro 24. Médias dos acessos avaliados e mantidos na coleção de germoplasma Caliman/UENF da espécie <i>Carica papaya</i> L., na primeira época.....	100
Quadro 25. Médias dos acessos avaliados e mantidos na coleção de germoplasma Caliman/UENF da espécie <i>Carica papaya</i> L., na segunda época.....	102

## RESUMO

A espécie *Carica papaya* L. é amplamente cultivada, porém a cultura tem problemas relacionados com a reduzida disponibilidade de genótipos melhorados para exploração comercial. O estudo da diversidade genética é de fundamental importância na seleção preliminar de acessos com características superiores e a utilização desses materiais com sucesso em programas de melhoramento genético. Dessa forma, o estudo foi realizado com o objetivo geral de avaliar estratégias de análise da diversidade genética em *Carica papaya* L., a partir de acessos do gênero mantidos no banco de germoplasma localizado na Fazenda Caliman Agrícola S/A, no município de Linhares, no Estado do Espírito Santo, em duas diferentes épocas do ano. Nove caracteres quantitativos foram considerados em cinquenta acessos, utilizando-se o delineamento em blocos casualizados com duas repetições. Foi empregada a análise multivariada por meio da técnica de componentes principais e por meio de métodos de agrupamento (Tocher e redes neurais artificiais). A consistência do agrupamento foi verificada através da análise discriminante de Anderson. Observou-se na presente população de estudo, a presença de diversidade genética entre os acessos avaliados. Verificou-se que as técnicas de análise aplicadas foram capazes de identificar a diversidade genética. As informações resultantes são úteis no entendimento da diversidade na coleção e como a mesma está estruturada. Com relação à técnica de componentes principais, a representação gráfica dos escores dos dois primeiros componentes principais demonstra a formação de quatro grupos, na

primeira e segunda época e três grupos, na média das épocas. Os acessos mais dissimilares foram, na primeira época, os acessos 10 - Caliman SG, grupo Solo, e 21 - Mamão Bené, grupo Formosa; na segunda época, os acessos 14 - Baixinho Santa Amália, grupo Solo, e 21 - Mamão Bené, grupo Formosa; na média das épocas, os acessos 10 - Caliman SG, grupo Solo, e 21 - Mamão Bené, grupo Formosa. De acordo com o estudo da contribuição do caráter para a variação acumulada, observou-se que massa de fruto foi o caráter de maior importância no estudo da diversidade genética, nas duas épocas, bem como na média das épocas. Através do método de otimização de Tocher, foram formados sete, dez, e nove grupos, na primeira, segunda e média das épocas, respectivamente. Através da rede neural artificial, foram formados quatro grupos na primeira e segunda épocas e três grupos na média das épocas. Foi possível verificar que a técnica de redes neurais artificiais é similar à técnica de componentes principais e se demonstrou viável na classificação dos acessos. Com relação à análise discriminante de Anderson, ao se considerar o agrupamento proposto pela análise por componentes principais, as taxas de erro aparente foram de 8,00%, 12,50% e 12,50%, para a primeira, segunda e média das épocas, respectivamente. Ao se considerar o agrupamento proposto pelo método de otimização Tocher, as taxas de erro aparente foram de 6,00%, 18,75% e 10,41%, para a primeira, segunda e média das épocas, respectivamente. Ao se considerar o agrupamento proposto pela rede neural artificial, as taxas de erro aparente foram de 8,00%, 8,33% e 6,25%, para a primeira, segunda e média das épocas, respectivamente. A utilização da análise discriminante de Anderson foi considerada viável, como um método para verificar a consistência do agrupamento proposto pelas técnicas utilizadas nesse trabalho.

## ABSTRACT

The sort *Carica papaya* L. is widely cultivated, however the culture has problems connected with the reduced availability of genotypes improved for commercial exploration. The study of the genetic diversity is of basic importance in the preliminary choice of accesses with superior characteristics and the use of these materials with success in programs of genetic improvement. In this form, the study was carried out by the general objective to value strategies of analysis of the genetic diversity in *Carica papaya* L., from accesses of the type maintained in the bank of germoplasm located in the Farm Caliman Agrícola S/A, in the municipal district of Linhares, in the State of the Espírito Santo, in two different seasons. Nine quantitative characters were considered in fifty accesses, using randomized blocks with two repetitions. It was used the multivariate analysis through the technique of principal components and through methods of grouping (Tocher and artificial neural network). The consistency of the grouping was checked through the discriminating analysis of Anderson. It was observed in the present population of study, the presence of genetic diversity between the evaluated accesses. It was checked that the hard-working techniques of analysis were able to identify the genetic diversity. The resultant informations are useful in the understanding of the diversity in the collection and and in the structure. Regarding the technique of principal components, the graphic representation of the scores of the two first principal components demonstrates the formation of four groups, in the first and second time and three groups, on the average of the times. The most dissimilar

accesses were, in the first time, the accesses 10 - Caliman SG, Solo group, and 21 - Mamão Bené, Formosa group; in the second time, the accesses 14 - Baixinho Santa Amália, Solo group, and 21 - Mamão Bené, Formosa group; on the average of the times, the accesses 10 - Caliman SG, Solo group, and 21 - Papaya Bené, Formosa group. In accordance with the study of the contribution of the character for the accrued variation, it was noticed that mass of the fruit was the character of bigger importance in the study of the genetic diversity, in two times, as well as on the average of the times. Through the method of optimization of Tocher, were formed seven, ten, and nine groups, in first, second and average of the times, respectively. Through the artificial neural network, were formed four groups in first and second times and three groups on the average of the times. It was possible to check that the technique of artificial neural network is similar to the technique of principal components and it was viable in the classification of the accesses. Regarding the discriminating analysis of Anderson, considering the grouping proposed by the analysis for principal components, the taxes of apparent mistake were 8.00%, 12.50% and 12.50%, for first, second and average of the times, respectively. Considering the grouping proposed by the method of optimization of Tocher, the taxes of apparent mistake were 6.00%, 18.75% and 10.41%, for first, second and average of the times, respectively. Considering the grouping proposed by the artificial neural network, the taxes of apparent mistake were 8.00%, 8.33% and 6.25%, for first, second and average of the times, respectively. The use of the discriminating analysis of Anderson was considered viable, as a method to check the consistency of the grouping proposed by the techniques used in this work.

## INTRODUÇÃO

O melhoramento de plantas é formado por um conjunto valioso de estratégias utilizadas para o aumento da produtividade de forma sustentável e ecologicamente equilibrada. Inicialmente era uma arte, pois desde a época dos primeiros agricultores as sementes dos tipos mais desejáveis eram separadas para a propagação da espécie.

A análise da diversidade genética se destina à identificação de genitores adequados à obtenção de híbridos com maior efeito heterótico e que proporcionem maior segregação em recombinações, possibilitando uma maior probabilidade de aparecimento de genótipos superiores (Cruz e Carneiro, 2003).

No estudo da diversidade genética não existem métodos únicos para se atingir objetivos específicos. O melhorista deve, de forma crítica, avaliar cada situação e otimizar os recursos disponíveis para alcançar os objetivos dentro da melhor relação custo-benefício (Borém e Miranda, 2005).

A avaliação da performance relativa de centenas ou milhares de plantas geneticamente distintas só se tornou possível com o desenvolvimento de técnicas experimentais e de análises estatísticas, que permitem afirmar, com certo nível de confiança, entre outras coisas, que determinados indivíduos são superiores aos demais (Barros *et al.*, 2005). Independente da cultura, o sucesso do melhoramento vegetal depende da identificação e utilização da variabilidade genética existente nas populações, que permitam alcançar o principal objetivo, que é a identificação de genótipos superiores (Borém e Miranda, 2005).

A variabilidade genética detectada indica a possibilidade de seleção preliminar de acessos/plantas com características superiores e a utilização desses materiais com sucesso em programas de melhoramento genético (Dantas e Morales, 1996).

De maneira geral a variabilidade genética pode ser conceituada como quaisquer diferenças entre espécies ou indivíduos dentro da espécie. O conhecimento da divergência genética entre um grupo de parentais é importante no melhoramento, sobretudo para identificar combinações híbridas de maior efeito heterótico (Carvalho *et al.*, 2003). As combinações híbridas de maior efeito heterótico aumentam a probabilidade de se recuperar genótipos superiores nas gerações segregantes. Segundo Fehr (1987), o uso de parentais com insuficiente divergência genética, na formação de populações para hibridação, reduz a variabilidade genética quanto aos caracteres quantitativos, e o emprego de parentais com ampla divergência genética deve ser utilizado para produção de melhores tipos segregantes.

Os recursos genéticos de fruteiras tropicais estão sendo utilizados em programas de melhoramento genético e pesquisas correlatas, aproveitando-se o fato de o Brasil ser um dos mais importantes centros de diversidade genética de algumas espécies frutíferas tropicais (Ferreira, 1997), como o mamão, o abacaxi, o caju, a goiaba e o maracujá. Os trabalhos com mamoeiro nesta área têm sido relativamente recentes, verificando-se que, pela estreita base genética, com poucos genótipos disponíveis, o melhoramento genético da cultura torna-se restrito (Pereira *et al.*, 2010; Quintal *et al.*, 2007a, 2007b).

Apesar da facilidade de cultivo e do grande consumo de mamão, a cultura tem problemas relacionados com pragas, doenças, características agrônomicas e reduzida disponibilidade de genótipos melhorados para exploração comercial (Pereira *et al.*, 2010; Foltran *et al.*, 1993), com produtividades limitadas. Com a perspectiva de aumento de produtividade, o melhoramento genético do mamoeiro requer melhor aproveitamento dos métodos de melhoramento, por meio de seleção de variedades agronomicamente superiores e síntese de linhagens e, ou, híbridos resistentes às doenças e pragas, com características comerciais favoráveis (Dantas *et al.*, 2002).

A produção do mamoeiro é contínua durante todo o ciclo produtivo, porém devido ao efeito da temperatura sobre o crescimento e maturação dos frutos, observa-se a ocorrência de sazonalidade de produção (Marin e Silva, 1996).

Alguns autores vêm relatando as conseqüências decorrentes das diferentes épocas de desenvolvimento dos frutos de mamão durante o ano, tendo as características de qualidade como sólidos solúveis, acidez titulável, vitamina C, dentre outros atributos qualitativos, alteradas de acordo com as condições edafoclimáticas de desenvolvimento dos frutos (Oliveira, 1999; Yamanishi *et al.*, 2006). Dessa forma, ressalta-se a importância de análises que possam estudar a diversidade genética em dados coletados em vários ambientes ou épocas do ano.

O estudo foi realizado com o objetivo geral de avaliar estratégias de análise da diversidade genética em *Carica papaya* L., a partir de acessos da espécie mantidos no banco de germoplasma localizado na Fazenda Caliman Agrícola S/A, no município de Linhares, no Estado do Espírito Santo, em duas diferentes épocas do ano.

Considerou-se a utilização de informações das variações físico-químicas e morfológicas, dos 50 acessos que formam o banco. Após a coleta, os dados foram submetidos à análise multivariada como forma de estudar a diversidade genética entre os genótipos. Os objetivos específicos são:

a) Estudar a diversidade genética com o auxílio de técnicas multivariadas, utilizar os métodos disponíveis no programa computacional Genes (Cruz, 2006), sendo: análise por componentes principais, método de otimização Tocher (Rao, 1952) e análise discriminante de Anderson (1958);

b) Implementar uma rede neural artificial (programa computacional), segundo modelo proposto por Kohonen (1982), que através de extração de características da base de dados, proporá uma classificação para os acessos estudados;

c) Avaliar a eficiência do programa computacional implementado através da comparação de métodos e resultados.

## REVISÃO DA LITERATURA

### 1. Aspectos gerais da cultura do mamão

A espécie *Carica papaya* L. foi encontrada pelos espanhóis no Panamá. Após a descoberta do Novo Mundo, o mamoeiro foi amplamente distribuído nos trópicos, inicialmente por navegadores espanhóis e portugueses, e, posteriormente, por mercadores árabes (Dantas, 2000a).

Segundo Dantas *et al.* (2002), o local de origem do mamoeiro ainda é muito discutido, pois alguns autores mencionam o sul do México, outros citam as terras baixas da América Central e as Antilhas.

Atualmente considera-se que o centro de origem da espécie é, muito provavelmente, o noroeste da América do Sul, vertente oriental dos Andes, ou mais precisamente, a parte alta da Bacia Amazônica, onde sua diversidade genética é máxima (Martins e Costa, 2003).

Contudo, Lorenzi *et al.* (2006) consideram como centro de origem e dispersão a América tropical como um todo.

As primeiras sementes de mamoeiro, que saíram de seu centro de origem, foram levadas para o Panamá, São Domingos, outras ilhas do Caribe e regiões da América do Sul (Dantas *et al.*, 2003).

A primeira menção sobre a existência de mamoeiro foi feita na Europa, em 1535, pelo cronista espanhol G. H. de Oviedo em seu livro "*The general and*

*natural history of the Indies*”, onde informava ter encontrado plantas desse gênero crescendo entre o sul do México e o norte da Nicarágua (Bruckner, 2002).

No final do século XVI, o mamoeiro foi introduzido no Brasil pelos portugueses. Em pouco tempo se espalhou pelo território nacional. Em 1587, na Bahia, o padre Gabriel Soares de Sousa referiu-se ao mamoeiro em seu livro “Notícias do Brasil”, afirmando que o mesmo teria sido introduzido na Bahia via Pernambuco (Silva e Tassara, 2005).

O mamoeiro é uma das espécies frutíferas mais cultivadas em todo o mundo, em uma faixa equatorial que se estende a 32º de latitude Norte e Sul (Dantas, 2000a). Possui sabor agradável, valor nutritivo e propriedades medicinais.

Por ser uma planta tropical, o mamoeiro encontra excelentes condições de desenvolvimento em várias regiões do Brasil. Dentre os Estados com maiores volumes de produção destacam-se a Bahia, responsável pela produção de 902.525 toneladas de frutos em 16.684 hectares plantados, e o Espírito Santo, com a produção de 630.124 toneladas de frutos em 7.796 hectares plantados. O Brasil destaca-se como um dos principais produtores mundiais de mamão, com rendimento de 51,67 t/ha em 2008 (IBGE, 2008).

O cultivo desta fruteira abastece os mercados locais e de exportação de frutas frescas, constituindo importante fonte de fornecimento de papaína, empregada para os mais variados usos nas indústrias têxteis, farmacêutica, de alimentos e de cosméticos (Oliveira *et al.*, 1994).

A cultura do mamoeiro no Brasil se sustenta em estreita base genética, sendo bastante limitado o número de cultivares plantadas nas principais regiões produtoras (Pereira *et al.*, 2010; Quintal *et al.*, 2007a, 2007b). Fato que se comprova na medida em que o melhoramento genético do mamoeiro é desenvolvido apenas dentro de uma espécie (*Carica papaya* L.) e dentro, somente, dos grupos de mamoeiros hermafroditas (grupos Solo e Formosa), que são os de interesse comercial.

## 1.1. Aspectos botânicos

O mamoeiro (*Carica papaya* L.) é uma planta herbácea, pertence à classe Dicotiledônea, subclasse Archiclamydeae, ordem Violales, família Caricaceae. A família Caricaceae compreendia 31 espécies distribuídas em três gêneros (*Carica*, *Jacaratia* e *Jarilla*) da América Tropical e um quarto gênero, *Cylicomorpha*, da África Equatorial (Nakasone e Paull, 1998). Em 2000, Badillo promoveu a reabilitação de *Vasconcellea* à categoria de gênero, com base em características morfológicas, taxonômicas e dados moleculares (Badillo, 2000). Dessa forma, o gênero *Carica* é atualmente monoespecífico, composto apenas pela espécie *Carica papaya*, enquanto *Vasconcellea*, que compreende 21 espécies, é o maior dentro da família Caricaceae (Badillo, 2000).

Os gêneros da família Caricaceae são nativos das regiões tropicais e subtropicais das Américas e da África, com plantas arbustivas, providas de látex branco (Joly, 1993). O gênero *Carica* é diplóide com  $2n = 2x = 18$ , ou seja, número básico de cromossomos haplóides ou genoma  $n = 9$  (Storey, 1941). A espécie *Carica papaya* L. é polígama (Badillo, 1993).

O mamoeiro cultivado (*Carica papaya* L.) possui sistema radicular pivotante, com raiz principal bastante desenvolvida, de coloração branco-cremosa (Dantas e Neto, 2000).

O caule é cilíndrico, com 10cm a 30cm de diâmetro, herbáceo, fistuloso, ereto, de coloração verde-clara no ápice e verde-grisácea a acinzentada na base. As folhas são grandes, glabras, com longos pecíolos fistulosos, verde-pálido. As flores são divididas em três tipos diferentes: masculinas, femininas e hermafroditas (Dantas e Neto, 2000).

O fruto é uma baga de forma variável de acordo com o tipo de flor, podendo ser arredondado, oblongo, alongado, cilíndrico e piriforme. A casca é fina e lisa, de coloração amarelo-clara a alaranjada, protegendo uma polpa com coloração que pode variar de amarela, rosada a alaranjada. As sementes são pequenas, arredondadas rugosas e recobertas por uma camada mucilagínosa, com coloração de acordo com a variedade (Luna, 1980).

## 1.2. Genótipos de mamoeiro

Conforme as características dos frutos, os mamoeiros hermafroditas podem ser classificados em dois grupos distintos, o grupo Solo e o grupo Formosa.

Até meados da década de 70, predominavam no Brasil cultivos de mamoeiros dióicos ou comuns. Praticamente não existiam variedades comerciais para plantio, visto que as sementes utilizadas tinham elevado grau de segregação (Marin e Ruggiero, 1988).

A partir de 1976, a cultura retomou sua importância econômica para o Brasil, principalmente devido à introdução de cultivares havaianas do grupo Solo e de híbridos chineses do grupo Formosa, notadamente nos Estados do Pará, Bahia e Espírito Santo. Vale ressaltar que a simples introdução de cultivares do grupo Solo provocou uma significativa expansão da comercialização do fruto, devido à sua grande aceitação tanto no mercado interno, quanto para exportação (Marin *et al.*, 1994).

De uma forma geral, os frutos do grupo Formosa são indicados para a comercialização no mercado interno, enquanto que os frutos do grupo Solo são adequados para comercialização tanto no mercado interno, quanto no mercado externo. No mercado externo, os frutos com peso médio variando de 350g a 500g alcançam maiores cotações de preços (Souza, 2000).

As cultivares do grupo Solo são, em sua maioria, genótipos uniformes, constituídas de linhagens puras fixadas por sucessivas gerações de autofecundação (Andreotti, 1998). No grupo Solo encontra-se a maioria das cultivares de mamão utilizadas no mundo (Dantas, 2000b).

O grupo Formosa abrange híbridos  $F_1$ . O híbrido do grupo Formosa que tem sido mais utilizado comercialmente é o 'Tainung nº 1', por possuir características superiores, sendo produtivo, resistente ao transporte, apresentar boa vida útil, ótimo sabor, entre outras.

No Brasil, dos problemas relacionados com a cultura do mamão, ressalta-se a limitação de alternativas quanto à escolha de cultivares e/ou híbridos comerciais para o plantio que atendam tanto as exigências do mercado nacional como internacional. Aliado a isto, o elevado preço das sementes híbridas dos mamoeiros do grupo Formosa tem levado muitos fruticultores a utilizar plantios

sucessivos com as gerações  $F_2$ ,  $F_3$  e  $F_4$ , acarretando inúmeros problemas, sobretudo com a perda de vigor e segregação para o formato do fruto (Marin *et al.*, 2005).

Com base nesses aspectos, a Universidade Estadual do Norte Fluminense, em parceria com a Caliman Agrícola S/A, lançou o primeiro híbrido de mamão brasileiro, o 'UENF/Caliman01' (Martins e Costa, 2003), obtido do cruzamento entre um progenitor do Grupo Formosa e um progenitor do Grupo Solo.

## **2. Aspectos do melhoramento genético do mamoeiro**

A sustentabilidade da cultura do mamoeiro depende basicamente de novos genótipos com características agrônômicas superiores, frutos que atendam as exigências dos consumidores, dos mercados interno e externo, com maior longevidade pós-colheita, e que possuam resistência às principais pragas e doenças que atacam a cultura (Cattaneo, 2003).

O interesse na seleção e preservação dos genótipos de mamão de superior qualidade se deve ao seu reconhecido valor alimentício e medicinal e à importância socioeconômica da cultura para o País, por gerar grande número de empregos e proporcionar substancial valor de produção, excedentes de divisas e contribuição fiscal (Martins e Costa, 2003).

A base para os trabalhos de melhoramento genético do mamoeiro em vários países tem sido os estudos genéticos feitos por Hofmeyer (1938), Storey (1938) e Awada (1953).

O melhoramento genético do mamoeiro em diversas partes do mundo estava voltado basicamente para a obtenção de cultivares endógamas com características específicas, visando atender às demandas dos usuários do produto (Sampaio *et al.*, 1983).

De acordo com Nakasone (1980), no Havaí os trabalhos de melhoramento genético não exploravam o efeito da heterose no mamoeiro, devido ao fato de as linhagens oriundas do grupo Solo não possuírem caracteres contrastantes. Porém, Cattaneo (2001, 2007), comprovou que a utilização de hibridações, como estratégia para aumento da produtividade, é totalmente viável.

O melhoramento genético do mamoeiro no Brasil vem sendo executado por importantes instituições. Os recursos genéticos disponíveis no País são satisfatórios, necessitando, para seu enriquecimento, da realização de coletas e novas introduções, visando ampliar as variabilidades intra e interespecífica disponíveis, o que certamente contribuirá para o desenvolvimento de variedades e híbridos com as características desejáveis, resistentes às pragas e doenças (Bruckner, 2002).

A Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), fundada em 1993, iniciou seus trabalhos de pesquisa com a cultura de mamão em 1996. Trabalhos publicados envolvem o melhoramento genético apoiado em procedimentos clássicos e assistido por marcadores de DNA (Silva *et al.*, 2007a, 2007b, 2007c, 2007d, 2008; Ramos *et al.*, 2007a, 2007b, 2007c).

Nos programas de melhoramento de plantas, a informação quanto à diversidade e à divergência genética, dentro de uma espécie, é essencial para o uso racional dos recursos genéticos. Os estudos sobre a diversidade genética nas coleções de germoplasma podem ser realizados a partir de caracteres morfológicos de natureza qualitativa ou quantitativa. E neste caso várias técnicas estatísticas podem ser utilizadas na predição da diversidade presente.

Para a cultura do mamoeiro, estudos desta natureza são relativamente escassos. Dentre as pesquisas realizadas na UENF, trabalhos têm sido conduzidos focando a avaliação de populações segregantes, via estimação de parâmetros genéticos, uso de índices de seleção e estimativa de correlações entre caracteres relacionados a atributos produtivos e de qualidade de fruto (Silva *et al.*, 2007a, 2007b, 2007c, 2007d, 2008; Ramos *et al.*, 2007a, 2007b, 2007c).

Considerando os resultados obtidos pelos vários programas de melhoramento até então desenvolvidos, em Bruckner (2002) uma das ações propostas consiste em revisar os métodos clássicos de melhoramento e ampliar os trabalhos com a aplicação de técnicas como bioinformática e biotecnologia, associando essas técnicas aos métodos tradicionais de melhoramento.

O desenvolvimento das diversas técnicas de bioinformática e biotecnologia, como ferramentas importantes em programas de melhoramento e propagação do mamoeiro, deverá reduzir o tempo de obtenção de novas variedades.

### 3. Análise da diversidade de recursos genéticos

A divergência genética tem sido avaliada por meio de técnicas biométricas, baseadas na quantificação da heterose, ou por processos preditivos. Os métodos fundamentados em modelos biométricos, que se destinam à avaliação da divergência dos progenitores, avaliam tanto a capacidade específica quanto a heterose manifestada nos híbridos (Cruz *et al.*, 2004).

São métodos preditivos aqueles que tomam por base as diferenças morfológicas, fisiológicas, entre outras, apresentadas pelos progenitores na determinação da divergência, que é geralmente quantificada por uma medida de dissimilaridade, por exemplo, distâncias Euclidianas ou de Mahalanobis (Cruz e Carneiro, 2003).

Quanto à indicação de cruzamentos, a análise multivariada possibilita a predição da heterose (Sudre *et al.*, 2005), servindo como base para a recomendação de cruzamentos. Porém, as possibilidades são muitas, desde que se obedeça aos princípios de se cruzar os acessos mais distantes geneticamente e com as melhores características agrônômicas desejadas.

Em outras espécies de plantas perenes, as abordagens multivariadas têm sido utilizadas com sucesso no estudo da diversidade genética presente em populações de trabalho, utilizadas na caracterização da estrutura genética dos indivíduos amostrados, e tem sido utilizada como critério e indicadores para a seleção de genótipos promissores para a utilização em programas de melhoramento genético, além da conservação de germoplasma das referidas espécies (Silva *et al.*, 2009; Oliveira, *et al.*, 2007; Arriel *et al.*, 2006; Oliveira *et al.*, 2006; Viana *et al.*, 2006; Melo Júnior *et al.*, 2004; Teles *et al.*, 2003; Paiva *et al.*, 2001).

Na predição da divergência genética, vários métodos multivariados podem ser aplicados. Dentre eles, citam-se os métodos aglomerativos, a análise por componentes principais, a análise discriminante de Anderson e as técnicas de bioinformática. A escolha do método mais adequado tem sido determinada pela precisão desejada pelo pesquisador, pela facilidade da análise e pela forma como os dados foram obtidos (Cruz e Carneiro, 2003).

As técnicas de análise de agrupamento têm como objetivo dividir um grupo original de observações em vários grupos, seguindo algum critério de similaridade ou dissimilaridade (Cruz *et al.*, 2004).

A análise de agrupamento pode ser complementada com a análise de componentes principais, cujo objetivo é construir, mediante processo matemático, um conjunto menor de combinações lineares das variáveis originais que preserve a maior parte da informação fornecida por essas variáveis (Cruz, 1990).

A técnica multivariada denominada Função Discriminante, proposta por Anderson (1958), tem por finalidade classificar um ou mais indivíduos em populações já conhecidas. É utilizada para verificar a consistência de um determinado agrupamento, conforme descrito por Cruz e Carneiro (2003).

A bioinformática é responsável pela aplicação de técnicas computacionais e matemáticas na análise biológica. Como técnicas de bioinformática podem-se citar as redes neurais artificiais, os algoritmos genéticos, entre outras.

As técnicas de bioinformática se enquadram no contexto da fruticultura em qualquer situação na qual se deseja prever algo, reconhecer algum padrão, ou como técnicas de análise de agrupamento (Ruggiero *et al.*, 2003).

### **3.1. Análise por componentes principais**

A análise por componentes principais consiste em transformar um conjunto original de variáveis em outro conjunto de dimensão equivalente, mas com propriedades importantes, que são de grande interesse em certos estudos de melhoramento. Cada componente principal é uma combinação linear das variáveis originais. Além disso, são independentes entre si e estimados com o propósito de reter, em ordem de estimação, o máximo da informação, em termos de variação total, contida nos dados iniciais (Cruz e Carneiro, 2003).

A viabilidade de utilização dos componentes principais em estudos sobre divergência genética dependerá da possibilidade de resumir o conjunto de variáveis originais em poucos componentes, o que significará ter uma boa aproximação do comportamento dos indivíduos (progenitores).

Com base no princípio de que a importância ou variância dos componentes principais decresce do primeiro para o último, tem-se que os últimos componentes explicam uma fração muito pequena da variância total. Então, a

variável de maior coeficiente no componente de menor autovalor dever ser a menos importante para explicar a variância total e, portanto, passível de descarte.

Além de possibilitar o estudo da diversidade genética de um grupo de progenitores, a técnica dos componentes principais tem a vantagem de possibilitar a avaliação da importância de cada caráter estudado sobre a variação total disponível entre os genótipos avaliados. O interesse nesta avaliação reside na possibilidade de se descartar caracteres que contribuem pouco para a discriminação do material avaliado, reduzindo, dessa forma, mão-de-obra, tempo e custo despendidos na experimentação agrícola (Cruz e Carneiro, 2003).

Segundo Jolliffe (1972, 1973), quando se dispõe de grande número de descritores, é possível que muitos deles sejam redundantes, tornando-se útil a eliminação, porque, além de pouco informativos, ocorre acréscimo no trabalho de avaliação e não apresentam informação adicional.

A indicação de genótipos para combinações híbridas deve ser feita com base no potencial dos genótipos e na magnitude de suas dissimilaridades. Dessa forma, se há interesse em um determinado genótipo, em virtude de seu bom desempenho em relação a uma determinada característica, recomenda-se seu cruzamento com outros genótipos em função da maior distância gráfica apresentada em relação a esses. Assim, garante-se a maior variabilidade disponível para as gerações segregantes.

O emprego da técnica dos componentes principais para estudos sobre divergência genética em plantas, tem sido realizado com o propósito de manipular acessos em bancos de germoplasma, identificar parentais divergentes para hibridação, estimar a variabilidade total disponível em grupos geneticamente relacionados, avaliar a importância de cada caráter e promover a eliminação daqueles que contribuem pouco, em termos de variação, no grupo de indivíduos avaliados (Cruz, 2006).

A análise da divergência genética através da utilização de componentes principais vem sendo realizada em diversos trabalhos (Gouvêa, 2009; Gomes Filho, 2009; Oliveira *et al.*, 2009; Moreira *et al.*, 2009; Pezzopane *et al.*, 2009; Cardoso *et al.*, 2009b).

## 3.2. Análises de agrupamento

As técnicas de análise de agrupamento têm por objetivo dividir um grupo original de observações em vários grupos, segundo algum critério de similaridade ou dissimilaridade, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (Cruz *et al.*, 2004).

Na análise de agrupamento, várias questões surgem, como por exemplo, o número final de grupos desejados, a adequação da partição obtida e o tipo de medida de similaridade a ser utilizada. Com relação ao número de grupos desejados, o que se faz mais comumente é utilizar vários números de grupos e, por algum critério de otimização, selecionar o mais conveniente (Cruz *et al.*, 2004).

Para avaliação da adequação da partição, é comum a utilização da análise discriminante e, com relação às medidas de similaridade, várias são citadas, entretanto, as mais empregadas no melhoramento são as distâncias Euclidianas e a de Mahalanobis (Cruz e Carneiro, 2003).

Existe grande número de métodos de agrupamento disponíveis, dos quais o pesquisador tem que decidir qual o mais adequado ao seu trabalho, uma vez que as variadas técnicas podem levar a diferentes padrões de agrupamentos (Cruz *et al.*, 2004). Dentre os métodos de agrupamento mais comumente utilizados no melhoramento de variáveis quantitativas em plantas, citam-se os métodos baseados em distâncias, como os hierárquicos, e os de otimização.

Como métodos hierárquicos, destacam-se os métodos do vizinho mais próximo, vizinho mais distante, Ward e UPGMA (*Unweighted Pair Group Method With Arithmetic*). Um dos métodos de otimização mais comumente empregados no melhoramento genético é o proposto por Tocher (Rao, 1952).

A análise da divergência genética através da utilização de técnicas de análise de agrupamento, mais especificamente o método de otimização Tocher, vem sendo realizada em diversos trabalhos (Arnhold *et al.*, 2010; Coimbra *et al.*, 2010; Cardoso *et al.*, 2009b; Zuin *et al.*, 2009; Marim *et al.*, 2009; Barelli *et al.*, 2009; Peluzio *et al.*, 2009; Bertini *et al.*, 2009).

### 3.3. Análise discriminante de Anderson

Os estudos de diversidade por meio das técnicas de agrupamento são complementados com a utilização de análise discriminante, visando avaliar a adequação da partição dos elementos agrupados.

Segundo Cruz (2006), a análise discriminante de Anderson (1958) tem por finalidade conseguir a melhor discriminação entre os indivíduos, alocando-os em suas devidas populações, e permitir a classificação de novos materiais genéticos, de comportamento desconhecido, em populações já conhecidas.

De acordo com Anderson (1958), se medidas são realizadas em amostras aleatórias pertencentes a diversas categorias ou grupos, então cada grupo será visualizado como um conjunto de pontos em um espaço multidimensional. A análise discriminante reduz o número de medidas realizadas para um número menor de parâmetros, que são funções discriminantes linearmente dependentes das medidas originais. Desta forma, os grupos (conjuntos de pontos) poderão ser visualizados em um espaço multidimensional menor que o anterior e, os coeficientes das funções discriminantes indicarão a contribuição relativa das medidas originais, para cada função discriminante.

Para a análise discriminante de Anderson (1958) é necessário que, previamente, os acessos estejam divididos em populações ou grupos distintos. Com estas informações, geram-se funções, que são combinações lineares das características avaliadas, tendo por finalidade obter a melhor discriminação entre os indivíduos, alocando-os em seus devidos grupos ou populações. Estas funções passam a ser de grande valia por permitir classificar novos genótipos, de comportamento desconhecido, nas populações já conhecidas. A eficácia das variáveis utilizadas em promover a discriminação também é avaliada, permitindo conhecer a adequação da função estimada.

A análise discriminante de Anderson tem sido empregada em algumas culturas com objetivo de gerar funções discriminantes para classificar genótipos e discriminar espécies (Monteiro, 2008; Zanine *et al.*, 2007; Sudre *et al.*, 2006; Assis *et al.*, 2003; Abreu *et al.*, 2004), verificar a consistência dos agrupamentos propostos por outras técnicas (Mariot *et al.*, 2008), identificar e avaliar características para fins de diferenciação de cultivares (Lima Júnior *et al.*, 2009; Nogueira *et al.*, 2008).

### 3.4. Bioinformática

Pode-se definir a bioinformática como a ciência da computação aplicada à pesquisa biológica, ou ainda como a análise matemática do conteúdo informacional dos genomas. A bioinformática é especialmente importante em pesquisas genômicas, por causa da complexidade dos dados gerados (Borém e Miranda, 2005).

A Bioinformática combina conhecimentos de química, física, biologia, ciência da computação, informática, matemática e estatística para processar dados biológicos.

A interação entre a Biologia e a Informática não é mais uma novidade, e atualmente vem passando por um impulso renovado. A Biologia inspirou vários modelos computacionais, como, por exemplo, as redes neurais artificiais e os algoritmos genéticos (Martins, 2009).

No Brasil, a bioinformática foi utilizada de maneira sistemática e em larga escala pela primeira vez em 1997, no projeto de seqüenciamento do genoma da bactéria *Xylella fastidiosa*, causadora da doença do amarelinho nos laranjais (CIB, 2004).

Dada a riqueza das informações disponíveis na biologia molecular moderna, esta passa a ser vista em parte como uma ciência informacional onde a bioinformática tem papel central, e integrador, permitindo que experimentos sejam feitos *in silico* (no computador) e que o uso da bancada, para experimentos *in vitro* ou *in vivo*, seja adiado até que um foco maior seja dado à pesquisa sendo considerada. Esta tem sido a metodologia utilizada em importantes projetos de pesquisa na área de genética envolvendo o melhoramento de plantas (Martins, 2009).

As ferramentas de bioinformática e biotecnologia conduzem ao aceleração dos programas de melhoramento genético, à redução do número de ciclos de seleção, ao direcionamento para a melhoria de caracteres qualitativos e quantitativos de interesse e ao aumento do ganho genético e redução dos custos, com a operacionalização dos esquemas de seleção de novos genótipos (Valois, 1998).

É crescente a necessidade de viabilizar soluções para o desenvolvimento sustentável do agronegócio brasileiro, por meio da geração, adaptação e

transferência de conhecimentos e tecnologias de informação e de bioinformática, em benefício da sociedade.

Pesquisas propõem o desenvolvimento de métodos e técnicas nas áreas de modelagem matemática e simulação, para sua incorporação em processos com alto conteúdo tecnológico como previsão de safras, agroclimatologia, agricultura de precisão, sistemas especialistas para diagnósticos de pragas e doenças, entre outros (Ruggiero *et al.*, 2003).

Muitas são as técnicas de bioinformática que vêm sendo utilizadas no melhoramento de plantas (redes neurais artificiais, algoritmos genéticos, entre outras) e que podem auxiliar no desenvolvimento de novas variedades. Cabe ao melhorista avaliar o potencial dessas novas técnicas para o seu programa e adaptar aquelas, dentro de suas limitações de recursos, que contribuem para o aumento da produtividade e da eficiência.

Uma das técnicas mais utilizadas em inteligência artificial são as redes neurais artificiais. Essa tecnologia se enquadra no contexto da agricultura em qualquer situação na qual se deseja prever algo ou reconhecer algum padrão. As redes neurais possibilitam prever, por exemplo, de forma bastante precisa o comportamento de doenças e pragas, mediante análises de interação de dados climáticos (temperatura, umidade relativa do ar, precipitação, entre outros) e fitotécnicos (variedade, tipo de cultivo, região geográfica, entre outros) (Ruggiero *et al.*, 2003).

### **3.4.1. Redes neurais artificiais**

Redes neurais artificiais são técnicas computacionais que apresentam um modelo inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência (Braga *et al.*, 2000). São compostas por unidades de processamento simples, que calculam determinadas funções matemáticas, denominadas neurônios ou nós. O neurônio é uma unidade de processamento de informação fundamental para operação de uma rede neural artificial.

As redes neurais artificiais possuem como características a tolerância a falhas, adaptabilidade a novas condições, resolução de problemas com base no conhecimento passado e reconhecimento de padrões (Haykin, 2001).

A solução de problemas através de redes neurais artificiais é bastante atrativa, já que a forma como estes são apresentados internamente pela rede cria a possibilidade de um desempenho superior ao dos modelos convencionais (Braga *et al.*, 2000).

Segundo Galvão *et al.* (1999), em função de sua estrutura não-linear, as redes neurais artificiais conseguem captar características mais complexas dos dados, o que nem sempre é possível com a utilização das técnicas estatísticas tradicionais. Para Sudheer *et al.* (2003), a maior vantagem das redes neurais artificiais sobre os métodos convencionais é que elas não requerem informação detalhada sobre os processos físicos do sistema a ser modelado.

Na programação matemática ou otimização, a modelagem é feita pelo uso de relações lineares e/ou não-lineares. A modelagem em uma rede neural artificial é baseada em dados. Não existe a criação de restrições (Benite, 2003).

Não se pode negar a utilidade dos modelos tradicionais, porém os mesmos apresentam características que geram dificuldades operacionais para sua efetiva aplicação, como o problema inerente em se trabalhar com modelos estatísticos complexos, suas hipóteses restritivas a ser satisfeitas e a dificuldade em se alcançar conclusões e resultados por não especialistas (Benite, 2003).

De acordo com Mohammadi e Prasanna (2003), existem dois tipos básicos de métodos de agrupamento: os métodos baseados em distâncias e os métodos baseados em modelos.

Nos métodos baseados em distâncias, uma matriz de distâncias entre pares de indivíduos é utilizada como entrada de dados, na análise realizada por meio de um algoritmo de agrupamento específico. Nos métodos baseados em modelos, se assume que as observações de cada grupo constituem amostras aleatórias de algum modelo paramétrico particular, aspecto característico das redes neurais artificiais.

Existem diversos tipos de redes neurais e diferentes maneiras de classificá-las. Talvez a mais importante seja quanto à forma de aprendizado, que pode ser: supervisionado e não supervisionado.

No aprendizado supervisionado, são sucessivamente apresentados à rede conjuntos de padrões de entrada e seus correspondentes padrões de saída. Durante este processo, a rede realiza um ajustamento dos pesos das conexões entre os elementos de processamento, segundo uma determinada lei de

aprendizagem, até que o erro entre os padrões de saída gerados pela rede alcance um valor mínimo desejado (Braga *et al.*, 2000). Por exemplo, Perceptrons, Adaline e Madaline, Redes de Função de Base Radial, são algumas dentre as dezenas de redes com métodos de aprendizado supervisionado.

No aprendizado não supervisionado a rede “analisa” os conjuntos de dados apresentados, determina algumas propriedades dos conjuntos de dados e “aprende” a refletir estas propriedades na sua saída. A rede utiliza padrões, regularidades e correlações para agrupar os conjuntos de dados em classes. As propriedades que a rede vai “aprender” sobre os dados podem variar em função do tipo de arquitetura utilizada e da lei de aprendizagem (Braga *et al.*, 2000). Por exemplo, Mapa Auto-Organizável de Kohonen, Redes de Hopfield e Memória Associativa Bidirecional, são redes com métodos de aprendizado não supervisionado.

Muitos são os modelos de redes neurais artificiais, porém o modelo introduzido por Kohonen (1982), conhecido como Mapa Auto-Organizável de Kohonen, possui a vantagem de não exigir do pesquisador a aplicação de nenhuma teoria acerca da organização destes dados, e pode fornecer subsídios para a atualização ou mesmo formulação de novas teorias acerca do problema, o que possibilita seu emprego no trabalho em questão.

O modelo de Kohonen é uma rede de duas camadas que pode organizar os dados de entrada a partir de um início aleatório. O cenário resultante mostra os relacionamentos naturais entre os padrões que são fornecidos à rede. A rede combina uma camada de entrada com uma camada competitiva de unidades processadoras e é treinada pelo algoritmo de aprendizagem não supervisionada (Haykin, 2001).

Os padrões que chegam são classificados pelas unidades que eles ativam na camada competitiva. Similaridades entre os padrões são mapeadas em relacionamentos de proximidades sobre a grade da camada competitiva. Depois que o treino está completo, os relacionamentos e agrupamentos entre os padrões são observados na camada competitiva. O modelo de Kohonen fornece vantagens sobre as técnicas clássicas de reconhecimento de padrões porque utiliza a arquitetura paralela de uma rede neural artificial (Kovacs, 1997).

O uso da tecnologia de redes neurais artificiais vem se enquadrando no contexto da agricultura de diversas formas, como por exemplo, na identificação

dos estágios iniciais do desenvolvimento de pragas ou doenças (Silva *et al.*, 2005), em agrupamento de dados (Dutra *et al.*, 2009), na estimativa da produtividade de diferentes culturas (Picoli, 2007), na classificação das imagens de satélite para diversos fins (França, 2010; Schimith *et al.*, 2009; Vieira *et al.*, 2009; Watanabe *et al.*, 2009; Chagas *et al.*, 2009, 2007; Costa e Souza Filho, 2008), no controle de robôs (Pessin *et al.*, 2007), entre outras.

## MATERIAL E MÉTODOS

### 1. Área Experimental

O trabalho foi conduzido utilizando-se acessos do banco de germoplasma localizado na Fazenda Caliman Agrícola S/A, no município de Linhares, no Estado do Espírito Santo. A área comercial da empresa está localizada geograficamente entre os paralelos 19° 06'-19° 18' de latitude sul e entre os meridianos 39° 45'- 40° 19' oeste.

O solo é, em sua maioria, classificado como Podzólicos Vermelho Amarelo, textura argilarenosa, fase floresta sub-perenifólia, relevo plano a suavemente ondulado (platôs litorâneos). O clima da região é do tipo Awi de Koppen (tropical úmido), com chuvas no verão e inverno seco (Rolim *et al.*, 1999).

O plantio foi feito utilizando-se o delineamento em blocos casualizados, sendo 2 repetições com 20 plantas por parcela em fileira dupla, com espaçamento de 3,6 x 2 x 1,5m. Foram estudados 50 dos acessos que formam o banco, em duas épocas diferentes.

## 2. Material Genético

As informações sobre os acessos foram coletadas por Quintal (2009), em duas épocas distintas. Foram avaliados 50 acessos na primeira época (maio/2007) e 48 acessos na segunda época (agosto/2007).

No Quadro 1 é apresentada a relação e a descrição dos acessos avaliados. Na primeira época, foram avaliados todos os acessos listados. Porém, na segunda época, não foram avaliados os acessos Maradol (grande limão) e Americano, pois os mesmos foram perdidos. Nas análises envolvendo a média das duas épocas, foram avaliados os acesso em comum, logo, todos os acessos da segunda época.

No Apêndice 1 encontram-se listadas as médias dos acessos nas duas épocas de análise.

Quadro 1. Cinquenta acessos avaliados e mantidos na coleção de germoplasma Caliman/UENF da espécie *Carica papaya* L.

<b>Acesso</b>	<b>Denominação do Acesso</b>	<b>Grupo</b>	<b>Procedência</b>	<b>Categoria do material</b>	<b>Origem</b>
1	Caliman M5	Formosa	CALIMAN	Linhagem	Brasil
2	Sunrise Solo 783	Solo	UENF	Variedade	-
3	Costa Rica	Formosa	UENF	Linhagem	Costa Rica
4	Triwan Et	Formosa	CALIMAN	Linhagem	-
5	Diva	Solo	CALIMAN	Linhagem	Brasil
6	Grampola	Solo	CALIMAN	Linhagem	-
7	Sunrise Solo	Solo	CALIMAN	Variedade	Brasil
8	Caliman AM	Solo	Caliman Agrícola	Linhagem	Brasil
9	Caliman GB	Solo	Caliman Agrícola	Linhagem	Brasil
10	Caliman SG	Solo	Caliman Agrícola	Linhagem	Brasil

Quadro 1, cont.

<b>Acesso</b>	<b>Denominação do Acesso</b>	<b>Grupo</b>	<b>Procedência</b>	<b>Categoria do material</b>	<b>Origem</b>
11	Caliman G	Solo	Caliman Agrícola	Variedade	Brasil
12	Sunrise Solo 72/12 (202)	Solo	Papaya Ceres	Variedade	Havaí – EUA
13	Kapoho Solo (polpa amarela)	Solo	CNPMFT Cruz das Almas – BA	Linhagem	Havaí – EUA
14	Baixinho Santa Amália	Solo	Papaya Ceres	Variedade	Brasil
15	Sunrise Solo TJ	Solo	Papaya Ceres	Variedade	Brasil
16	Tailândia	Formosa	EBDA – BA	Linhagem	-
17	São Mateus	Solo	Papaya Ceres	Linhagem	Brasil
18	Kapoho Solo (polpa vermelha)	Solo	-	Linhagem	Havaí – EUA
19	Sunrise Solo (sel. Caliman)	Solo	Caliman Agrícola	Linhagem	Brasil
20	Waimanalo	Formosa	CNPMFT Cruz das Almas – BA	Linhagem	Havaí – EUA
21	Mamão Bené	Formosa	CALIMAN	Linhagem	Brasil
22	Mamão roxo	Solo	-	Linhagem	-
23	Maradol (orig. México)	Formosa	Cuba	Variedade	Cuba
24	Maradol (grande limão)	Formosa	Cuba	Variedade	Cuba
25	Sekati	Formosa	CALIMAN	Linhagem	-
26	Baixinho Super	Solo	UENF	Linhagem	Brasil
27	Americano	Formosa	CALIMAN	Linhagem	-
28	STZ-51	Formosa	CALIMAN	Linhagem	Brasil
29	STZ-52	Formosa	CALIMAN	Linhagem	Brasil
30	Calimosa	Formosa	CALIMAN	Híbrido	Brasil

Quadro 1, cont.

<b>Acesso</b>	<b>Denominação do Acesso</b>	<b>Grupo</b>	<b>Procedência</b>	<b>Categoria do material</b>	<b>Origem</b>
31	JS 12 (206)	Formosa	CNPMFT Cruz das Almas – BA	Linhagem	C. do Almeida
32	Cariflora 209	Formosa	UNESP Jaboticabal – SP	Dióico	EUA
33	Golden tipo Formosa	Formosa	CALIMAN	Linhagem	Brasil
34	STZ-03 - pecíolo curto	Formosa	CALIMAN	Linhagem	Brasil
35	STA Helena III Trat 12a plt 07X	Formosa	CALIMAN	Linhagem	Brasil
36	STA Helena III Trat 11a plt 08X	Formosa	CALIMAN	Linhagem	Brasil
37	STA Helena III Trat 50 plt 09X	Formosa	CALIMAN	Linhagem	Brasil
38	STA Helena III Trat 02 plt 01X	Formosa	CALIMAN	Linhagem	Brasil
39	STA Helena III Trat 14 plt 05X	Formosa	CALIMAN	Linhagem	Brasil
40	STA Helena III Trat 15 plt 04X	Formosa	CALIMAN	Linhagem	Brasil
41	STA Helena III Trat 04 plt 02X	Formosa	CALIMAN	Linhagem	Brasil
42	STA Helena III Trat 12b plt 06X	Formosa	CALIMAN	Linhagem	Brasil
43	Papaya 42 Formosa	Formosa	CALIMAN	Linhagem	Brasil
44	Papaya 45 Formosa roxo	Formosa	CALIMAN	Linhagem	Brasil
45	Papaya 46 claro	Formosa	CALIMAN	Linhagem	Brasil
46	Fruto médio verde	Formosa	CALIMAN	-	-
47	Golden robusto	Solo	CALIMAN	Linhagem	Brasil
48	Tainung	Formosa	CALIMAN	Híbrido	Taiwan
49	STZ 63	Formosa	CALIMAN	Linhagem	Brasil
50	Sekati (fruto longo macuco)	Formosa	CALIMAN	Linhagem	-

Fonte: Quintal (2009).

### 3. Caracteres analisados

No processo de caracterização foram analisadas 9 características quantitativas. Todas as medidas foram feitas a partir de três frutos no estágio I de maturação (uma pinta amarela) de três plantas por parcela. Sendo:

a) Massa de fruto (MF): foi obtida por meio de balança digital e expressa em gramas;

b) Comprimento do fruto (CF): foi medido a partir do comprimento longitudinal dos frutos, com utilização de paquímetro digital, e foi expresso em centímetros;

c) Diâmetro do fruto (DF): foi medido a partir do comprimento transversal dos frutos, com utilização de paquímetro digital, e foi expresso em centímetros;

d) Espessura da polpa (EP): foi determinada pelo comprimento lateral da polpa dos frutos, na região mediana dos mesmos, com a utilização de paquímetro digital, e foi expressa em centímetros;

e) Sólidos solúveis (SS): foi obtido por meio do suco extraído de uma amostra da polpa da região mediana do fruto, a partir da extração por prensa de mão. As leituras foram efetuadas em refratômetro portátil ATAGO N1, com leitura na faixa de 0º a 32º Brix;

f) Firmeza externa do fruto (FEF): foi determinada em três pontos equidistantes de cada face dos frutos, sendo que esses pontos estavam distanciados da casca de 0,5cm, por meio da resistência à penetração na polpa. Para tal foi utilizado um penetrômetro de bancada (*Fruit Pressure Tester, Italy*; modelo 53205) com adaptador de 3,0 x 3,0cm (altura x diâmetro), expressa em Newton (N);

g) Firmeza interna da polpa (FIF): foi feita dividindo-se os frutos ao meio, no sentido transversal, resultando em duas faces. A firmeza foi determinada em três pontos equidistantes de cada face dos frutos, por meio da resistência à penetração na polpa. Para tal foi utilizado um penetrômetro de bancada (*Fruit*

*Pressure Tester, Italy*; modelo 53205) com adaptador de 3,0 x 3,0cm (altura x diâmetro), expressa em N;

h) Incidência de pinta preta (*Asperisporium caricae*) e mancha fisiológica do mamoeiro (PP e MM): foram determinadas através de um sistema de notas variando de 0 a 5, de acordo com o grau de incidência. As médias foram obtidas através dos valores de três notas por cada amostra. As médias foram padronizadas, para evitar o efeito de escala, por meio da raiz quadrada.

#### **4. Análise dos dados**

No presente trabalho, para o estudo da divergência genética entre os acessos, foram empregadas as análises multivariadas por meio da técnica de componentes principais e por meio de métodos de agrupamento (Tocher e redes neurais artificiais). A consistência do agrupamento foi verificada através da análise discriminante de Anderson, conforme descrito por Cruz e Carneiro (2003).

Com base nas médias das características obtidas a partir dos acessos, foram obtidas as matrizes de distâncias Euclidianas para cada uma das duas épocas e para a média das épocas, que foram analisadas pelo método de componentes principais e pelo método de otimização Tocher.

Para a obtenção dos grupos, utilizando-se a rede neural artificial, foi utilizada a distância Euclidiana média como medida de dissimilaridade, para cada uma das duas épocas e para a média das épocas. As médias originais foram utilizadas como entradas da rede.

As análises foram conduzidas utilizando-se os recursos computacionais do programa Genes (Cruz, 2006) e da rede neural artificial implementada no presente trabalho.

##### **4.1. Distância Euclidiana**

Seja  $X_{ij}$  a observação referente a  $j$ -ésima característica ( $j = 1, 2, \dots, n$ ) no  $i$ -ésimo genótipo ( $i = 1, 2, \dots, p$ ), define-se, segundo Cruz e Carneiro (2003), a distância Euclidiana entre dois genótipos  $i$  e  $i'$  por meio da expressão:

$$d_{ii'} = \sqrt{\sum_j (x_{ij} - x_{i'j})^2}$$

Inconvenientes associados à distância Euclidiana dizem respeito ao fato de que a mesma é alterada com a mudança da escala das medições, com o número de caracteres estudados e não leva em conta o grau de correlação entre eles. Para contornar o problema de escala, tem sido recomendada a padronização dos dados, e, para contornar a influência do número de caracteres, utiliza-se a distância Euclidiana média, descrita em Cruz e Carneiro (2003) como segue:

$$\text{seja: } x_{ij} = \frac{X_{ij}}{S(X_j)}$$

em que  $S(X_j)$  é o desvio padrão dos dados do  $j$ -ésimo caracter, então

$$d_{ii'} = \sqrt{\frac{1}{n} \sum_j (x_{ij} - x_{i'j})^2}$$

é a distância Euclidiana média baseada em dados padronizados e  $n$  é o número de caracteres analisados.

As estimativas de  $d_{ii'}$  evidenciam o grau de dissimilaridade entre os genótipos analisados. Por meio do desempenho médio e dos valores de  $d_{ii'}$ , é possível identificar, dentre os selecionados, aqueles mais divergentes. Esses genótipos podem ser recomendados como progenitores destinados a hibridações nas etapas iniciais de um programa de melhoramento.

## 4.2. Componentes principais

A análise por componentes principais foi realizada de acordo com Cruz e Carneiro (2003).

Seja  $x_{ij}$  a média padronizada do  $j$ -ésimo caráter ( $j = 1, 2, \dots, n$ ) avaliada no  $i$ -ésimo progenitor ( $i = 1, 2, \dots, p$ ). A técnica de componentes principais consiste em

transformar o conjunto de  $n$  variáveis  $x_{i1}, x_{i2}, \dots, x_{in}$  em um novo conjunto  $Y_{i1}, Y_{i2}, \dots, Y_{in}$ , em que os  $Y_i$ 's são funções lineares dos  $x_i$ 's e independentes entre si.

Na obtenção dos componentes principais, as seguintes propriedades são consideradas:

a) Se  $Y_{ij}$  é um componente principal, então:

$$Y_{ij} = a_1x_{i1} + a_2x_{i2} + \dots + a_nx_{in}$$

b) Se  $Y_{ij'}$  é outro componente principal, então:

$$Y_{ij'} = b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in}$$

$$\sum_j a_j^2 = \sum_j b_j^2 = 1$$

$$\sum_j a_j b_j = 0, \text{ ou seja, os componentes são não-correlacionados.}$$

c) Entre os componentes principais,  $Y_{i1}$  apresenta a maior variância,  $Y_{i2}$  a segunda maior e assim sucessivamente.

Com base na propriedade c, objetiva-se, em estudos sobre divergência genética por meio dos componentes principais, avaliar a possibilidade de estudar a dispersão dos genótipos em sistemas de eixos cartesianos nos quais o aproveitamento da variabilidade disponível seja maximizada.

O primeiro componente é definido como o de maior importância, uma vez que retém a maior parte da variação total encontrada nos dados originais.

Em geral, os primeiros componentes principais em estudos da divergência genética têm sido utilizados quando esses envolvem pelo menos 80% da variação total. Nos casos em que este limite não é atingido nos dois primeiros componentes, a análise é complementada com a dispersão gráfica em relação ao terceiro e quarto componentes.

Com relação ao descarte de caracteres, tem sido comum descartar o caráter de maior coeficiente (em valor absoluto) a partir do último componente até aquele cujo autovalor não exceda 0,7.

### 4.3. Método de otimização Tocher

Esse método requer a obtenção da matriz de dissimilaridade, sobre a qual é identificado o par de indivíduos mais similares. Esses indivíduos formarão o grupo inicial. A partir daí é avaliada a possibilidade de inclusão de novos indivíduos, adotando-se o critério de que a distância média intragrupo deve ser menor que a distância média intergrupo.

A entrada de um indivíduo em um grupo sempre aumenta o valor médio da distância dentro do grupo. Assim, pode-se tomar a decisão de incluir o indivíduo em um grupo, por meio da comparação entre o acréscimo no valor médio da distância dentro do grupo e um nível máximo permitido, que pode ser estabelecido arbitrariamente, ou adotado, como tem sido geralmente feito, o valor máximo ( $\theta$ ) da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo (Cruz e Carneiro, 2003).

De acordo com Cruz e Carneiro (2003), a inclusão, ou não, do indivíduo  $k$  no grupo é, então, feita considerando:

$$\text{Se } \frac{d_{(\text{grupo})k}}{n} \leq \theta, \text{ inclui-se o indivíduo } k \text{ no grupo;}$$

$$\text{Se } \frac{d_{(\text{grupo})k}}{n} > \theta, \text{ o indivíduo } k \text{ não é incluído no grupo.}$$

Sendo  $n$  o número de indivíduos que constitui o grupo original.

Neste caso, a distância entre o indivíduo  $k$  e o grupo formado pelos indivíduos  $i$  e  $j$  é dada por:

$$d_{(ij)k} = d_{ik} + d_{jk}$$

A distância média dentro do grupo é a média das distâncias entre cada par de progenitores que o constitui. Pelo critério adotado, esta distância é sempre menor que as distâncias médias intergrupos, que são obtidas pela média das distâncias entre pares de genótipos pertencentes aos diferentes grupos (Cruz e Carneiro, 2003).

#### 4.4. Análise discriminante de Anderson

A técnica multivariada denominada função discriminante, proposta por Anderson (1958), tem por finalidade classificar um ou mais indivíduos em populações já conhecidas. No presente estudo, essa técnica foi utilizada com o propósito de verificar a consistência dos agrupamentos propostos pela técnica de componentes principais, pelo método de otimização de Tocher e pela rede neural artificial.

Supondo igualdade das matrizes de co-variâncias, então os componentes constantes para todo  $j$  ( $j = 1, 2, \dots, g$ ) podem ser retirados, e a função discriminante é:

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

Em que:

$p_j$  = probabilidade, a *priori*, de os indivíduos pertencerem a população  $\pi_j$ ;

$\tilde{x}$  = vetor de observações, ou vetor de variáveis representativas dos caracteres envolvidos na análise;

$\mu_j$  = vetor de médias dos descritores avaliados em  $\pi_j$ ;

$\Sigma_j$  = matriz de co-variâncias entre os descritores avaliados em  $\pi_j$ ;

Sendo  $\pi_j$  a população, classifica-se o  $i$ -ésimo material genético, com vetor de médias  $\tilde{x}_i$  na população  $\pi_j$  se, e somente se,  $D_j(\tilde{x}_i)$  for o menor entre os elementos do conjunto  $\{D_1(\tilde{x}_i), D_2(\tilde{x}_i), \dots, D_j(\tilde{x}_i)\}$ . Sendo  $D_j(\tilde{x}_i)$  o valor da função discriminante da população  $j$  para o vetor de características  $\tilde{x}_i$ .

Utilizando-se as funções discriminantes e os dados das próprias populações  $\pi_j$ , estimou-se a taxa de erro aparente (TEA) que mede a eficiência dessas funções em classificar os acessos, corretamente, nas populações previamente estabelecidas.

A taxa de erro aparente foi dada pela relação entre o número de classificações erradas e o número total de classificações (Cruz e Carneiro, 2003), de acordo com:

$$TEA(\%) = \frac{1}{N} \sum_{j=1}^g m_j$$

Sendo  $m_j$  o número de observações retiradas da população  $\pi_j$ , que foram, por meio das funções discriminantes obtidas, classificadas em outra população  $\pi_{j'}$ , sendo  $j' \neq j$  e  $j = 1, 2, \dots, g$  populações.

Considerando:

$$N = \sum_{j=1}^g n_j$$

onde  $n_j$  é o número de observações retiradas da população  $\pi_j$ .

#### 4.5. Rede neural artificial

Para o estudo da diversidade genética nos acessos estudados, foi implementado um programa computacional para aplicação da técnica de redes neurais artificiais. Sendo a metodologia aplicada baseada no modelo de Kohonen.

A abordagem tradicional do modelo de Kohonen foi derivada, de acordo com as características e necessidades do presente estudo.

A arquitetura da rede neural baseada no modelo de Kohonen, de acordo com o contexto deste estudo, pode ser visualizada na Figura 1.

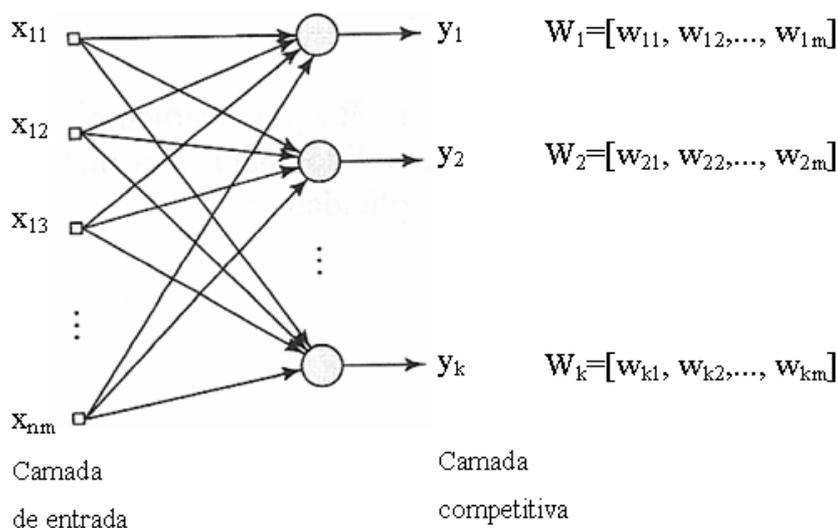


Figura 1. Arquitetura da rede neural artificial Kohonen composta por  $n$  acessos e  $m$  caracteres na camada de entrada, e por  $k$  neurônios na camada competitiva.

A mesma é composta por uma matriz de entrada  $n \times m$ , sendo  $n$  acessos e  $m$  elementos de entrada ou caracteres, que juntos representam o vetor de entrada  $X$ , e por  $k$  neurônios de saída, representando as classes às quais os acessos podem ser agrupados.

A metodologia aqui apresentada é baseada nos conceitos propostos por Haykin (2001).

Dada uma entrada, somente um neurônio de saída deve ser ativado, registrando assim, a classe à qual o acesso pertence. As classes devem agrupar acessos com características semelhantes. É feita, então, uma classificação por similaridade de valores.

#### 4.5.1. Processo Competitivo

Consiste em encontrar o neurônio com o melhor casamento por similaridade (vencedor)  $i(X)$ , no passo de tempo  $t$ , usando o critério da mínima distância entre os acessos.

Considerando-se que  $m = 9$  represente uma dimensão do espaço de entrada de dados, cada elemento do vetor de entrada corresponde às características referentes a um acesso.

Os dados utilizados foram: MF = massa de fruto; CF = comprimento do fruto; DF = diâmetro do fruto; EP = espessura da polpa; SS = sólidos solúveis;

FEF = firmeza externa do fruto; FIF = firmeza interna da polpa; PP = incidência de pinta preta; e MM = incidência de mancha fisiológica.

Como exemplo, na primeira época, sendo Caliman M5 representando o primeiro acesso:

$$X_1 = [433,33;13,42;8,49;2,62;11,87;80,67;59,32;1,41;1,34]$$

No processo de implementação da ferramenta, serão utilizados como vetores de entrada todos os acessos. Cada vetor de entrada representa os dados de um acesso. Logo, um padrão de entrada para a rede neural, na primeira época, por exemplo, é denotado aqui como:

$$X_n = [x_{n1}, x_{n2}, \dots, x_{n9}], n = 1, 2, \dots, 50$$

O conjunto de dados de entrada é alimentado por completo e repetidas vezes, no procedimento de treino da rede.

Considera-se que o vetor de peso sináptico inicial seja representado por:

$$W_k = [w_{k1}, w_{k2}, \dots, w_{km}]$$

Sendo,  $m = 9$  o número total de elementos da grade ou camada de entrada, e  $k = c$  o número total de elementos da camada de saída (classes) a ser definido posteriormente. Logo:

$$W_k = [w_{k1}, w_{k2}, \dots, w_{k9}], k = 1, 2, \dots, c$$

O vetor de peso sináptico é composto por dados pertencentes ao conjunto de dados de entrada. O vetor de peso sináptico de cada neurônio  $k$  da camada de saída tem a mesma dimensão que o espaço de entrada.

A localização espacial de uma vizinhança topológica de neurônios é realizada através da determinação dos acessos denominados centróides. Existe um centróide para cada grupo e o centróide é o acesso com as características mais representativas do grupo. Os centróides indicam o local mais típico de

qualquer indivíduo de um grupo particular. Na rede neural, o centróide é representado pelo vetor de peso sináptico.

O primeiro passo na operação de uma rede baseada no modelo de Kohonen é calcular o grau de similaridade para cada unidade na camada competitiva. Este valor quantifica o grau de igualdade entre os pesos de cada unidade e os valores correspondentes do padrão de entrada. Uma forma de se medir a similaridade entre a entrada e o neurônio é através de:

$$\|X_n - W_k\|$$

Que é a distância Euclidiana entre os vetores  $X_n$  e  $W_k$  e é calculada por:

$$\sqrt{\sum_{j=1}^m (x_{nj} - w_{kj})^2}$$

Para contornar a influência do número de caracteres, utiliza-se a distância Euclidiana média, de acordo com:

$$\sqrt{\frac{1}{m} \sum_{j=1}^m (x_{nj} - w_{kj})^2}$$

A unidade da camada de saída com a menor distância Euclidiana média vence a competição. Ao se usar o índice  $i(X)$  para identificar o neurônio com maior similaridade com o vetor  $X_n$  que está entrando na rede no momento,  $i(X)$  é tal que:

$$i(X) = \arg \min_k \|X_n - W_k\|$$

Onde o mínimo é tomado sobre todas as  $k$ -ésimas unidades na camada competitiva. Se duas unidades têm o mesmo valor da expressão  $\|X_n - W_k\|$ , então, por convenção, a unidade com o menor índice  $k$  é escolhida.

### 4.5.2. Processo Adaptativo

Consiste no ajuste dos vetores de pesos sinápticos dos neurônios, de acordo com a fórmula de atualização.

Dado o vetor de peso sináptico  $W_k(t)$  do neurônio  $k$  no tempo  $t$ , o vetor de peso atualizado  $W_k(t+1)$  no tempo  $t+1$  é definido por (Kohonen, 1982):

$$W_k(t+1) = W_k(t) + \eta(t)(X_n - W_k(t))$$

que é aplicado ao neurônio vencedor  $i$ , onde  $\eta(t)$  é o parâmetro da taxa de aprendizado e deve ser variável no tempo.

Essa equação tem o efeito de mover o vetor de peso sináptico  $W_i$  do neurônio vencedor  $i$  em direção ao vetor de entrada  $X_n$ . Através da apresentação repetida dos dados de treinamento, os vetores de peso sináptico tendem a seguir a distribuição dos vetores de entrada.

Deste ajustamento resulta a mudança dos pesos da unidade vencedora, tornando-a mais parecida com o padrão de entrada. A vencedora então se torna mais provável de (ou mais apta a) vencer a competição, seja ao se apresentar o mesmo ou um padrão semelhante de entrada.

O parâmetro da taxa de aprendizado  $\eta(t)$  é iniciada com um valor próximo a 0,1, devendo decrescer gradualmente, mas permanecendo acima de 0,01.

Resumindo, as regras básicas para implementação da rede neural podem ser descritas como segue:

1. Definir a arquitetura da rede: os vetores  $X_n$ , de acordo com a expressão  $X_n = [x_{n1}, x_{n2}, \dots, x_{nm}]$ , e  $W_k$ , de acordo com a expressão  $W_k = [w_{k1}, w_{k2}, \dots, w_{km}]$ ;
2. Inicializar os pesos  $W_k$ , do neurônio  $k$ , com valores aleatórios pertencentes ao conjunto de valores de entrada  $X$ ;
3. Apresentar uma entrada à rede;
4. Calcular a distância entre a entrada e os pesos para cada neurônio de saída,

de acordo com a expressão 
$$\sqrt{\frac{1}{m} \sum_{j=1}^m (x_{nj} - w_{kj})^2}$$
;

5. Selecionar o neurônio vencedor  $i(X)$ , ou seja, o neurônio com a menor distância do padrão de entrada apresentado, de acordo com a expressão
 
$$i(X) = \arg \min_k \|X_n - W_k\|;$$
6. Definir o parâmetro da taxa de aprendizagem  $\eta(t)$  e atualizar o vetor de peso sináptico do neurônio vencedor  $i(X)$ , de acordo com a expressão
 
$$W_k(t+1) = W_k(t) + \eta(t)(X_n - W_k(t));$$
7. Se existir, ainda, algum dado que faça parte do conjunto de treinamento e que não foi apresentado à rede, voltar ao passo 3;
8. Se o critério de parada foi satisfeito, encerrar essa fase do treinamento; caso contrário, realizar nova iteração do conjunto de treinamento, retornar ao passo 3.

O critério de parada da rede neural, de acordo com o paradigma utilizado (modelo de Kohonen), é implementado de tal forma que o algoritmo só deve parar caso não sejam observadas modificações significativas no mapa de características, por um número determinado de iterações.

#### 4.6. Gráfico Boxplot

O Boxplot é um gráfico que possibilita representar a distribuição de um conjunto de dados com base em alguns de seus parâmetros descritivos, quais sejam: a mediana (Q2), o quartil inferior (Q1), o quartil superior (Q3) e do intervalo interquartil (IQR = Q3 - Q1).

Os quartis são estatísticas que dividem os dados ordenados em quatro conjuntos com a mesma quantidade de dados. A mediana é o valor que divide os dados no centro. O primeiro quartil (Q1) divide no centro os valores anteriores à mediana e o terceiro quartil (Q3) divide no centro os valores posteriores à mediana. O intervalo interquartil (IQR) é a diferença entre Q3 e Q1.

O gráfico consiste em uma caixa cujos limites correspondem a Q1 e Q3, uma linha representando a mediana e duas linhas, também conhecidas como hastes, que demarcam duas possibilidades, descritas a seguir.

a) Caso não existam *outliers*, as hastes demarcam os valores máximo e mínimo do conjunto de dados;

b) Caso existam *outliers*, as hastes demarcam o maior valor não *outlier* e o menor valor não *outlier*.

Os *outliers* são valores extremamente altos ou baixos. A existência de *outliers* pode tanto indicar dados incorretos, como dados válidos que carecem de uma atenção especial. Dependendo do objetivo da análise é possível que justamente os *outliers* sejam o ponto de interesse.

A linha central da caixa marca a mediana do conjunto de dados. A parte inferior da caixa é delimitada pelo quartil inferior (Q1) e a parte superior pelo quartil superior (Q3). As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior a  $Q1 - 1,5IQR$  e do quartil superior até o maior valor não superior a  $Q3 + 1,5IQR$ . Os valores inferiores a  $Q1 - 1,5IQR$  e superiores a  $Q3 + 1,5IQR$  são representados individualmente no gráfico sendo estes valores caracterizados como *outliers*.

O Boxplot permite avaliar a simetria dos dados, sua dispersão e a existência ou não de *outliers* nos mesmos, sendo especialmente adequado para a comparação de dois ou mais conjuntos de dados correspondentes às categorias de uma característica.

## RESULTADOS

### 1. Componentes principais

#### 1.1. Primeira época

O resultado da análise de componentes principais, obtidos a partir de 9 caracteres quantitativos avaliados em 50 acessos (Quadro 2), mostra que foram necessários três componentes, para que a variância por eles explicada atingisse um mínimo de 80%.

Pode-se afirmar que a análise de componentes principais reduziu os nove descritores em três componentes principais, que explicaram 85,35% da variância total. O componente 1 explicou 62,08% da variância total, o componente 2 explicou 15,54% da variância total e o componente 3 explicou 7,72% da variância total.

Observou-se que massa de fruto foi o caráter de maior peso em CP 01 (0,3941), sendo identificado como o de maior contribuição para a variação acumulada, e o caráter de menor contribuição para a variação acumulada foi firmeza externa do fruto, com maior peso em CP 09 (-0,5818). Identificando-se assim, os caracteres de maior e menor importância no estudo da diversidade genética.

Quadro 2. Estimativa das variâncias, autovalores (AV), associados aos componentes principais (CP), e respectivos coeficientes de ponderação (autovetores) de 9 caracteres quantitativos avaliados em 50 acessos de mamoeiro (*Carica papaya* L.), na primeira época.

CP	AV	Variância	Caracteres *								
		Acumulada (%)	MF	CF	DF	EP	SS	FEF	FIF	PP	MM
CP 01	5,59	62,09	0,39	0,38	0,37	0,37	-0,30	0,39	0,38	0,02	0,20
CP 02	1,40	77,63	-0,17	-0,20	-0,01	0,24	0,22	0,00	-0,02	0,71	0,57
CP 03	0,69	85,36	0,04	-0,06	0,13	-0,04	0,61	0,08	0,08	-0,58	0,51
CP 04	0,50	90,96	0,24	-0,09	0,37	0,38	-0,18	-0,53	-0,55	-0,15	0,13
CP 05	0,42	95,64	0,16	-0,09	0,50	0,07	0,57	0,07	0,02	0,26	-0,56
CP 06	0,25	98,46	0,42	0,68	-0,26	-0,28	0,27	-0,24	-0,19	0,22	0,09
CP 07	0,11	99,73	-0,06	0,14	-0,53	0,76	0,24	0,05	-0,03	-0,13	-0,22
CP 08	0,02	99,90	0,55	-0,44	-0,25	0,00	0,00	-0,40	0,52	0,03	-0,01
CP 09	0,01	100,00	-0,49	0,35	0,22	0,09	0,03	-0,58	0,49	-0,02	-0,02

\*MF = massa de fruto; CF = comprimento do fruto; DF = diâmetro do fruto; EP = espessura da polpa; SS = sólidos solúveis; FEF = firmeza externa do fruto; FIF = firmeza interna da polpa; PP = incidência de pinta preta; e MM = incidência de mancha fisiológica.

Pelo método de Singh (1981), com base na distância Euclidiana média, utilizado para avaliar a importância relativa dos nove caracteres quantitativos, demonstrou-se que somente a massa de fruto contribuiu com 98,82% para a divergência genética, enquanto que as demais características contribuíram com apenas 1,18%.

É justificável a característica massa de fruto proporcionar maior contribuição relativa, visto que a análise reuniu acessos de grupos heteróticos distintos, onde a variação de peso era extremamente perceptível.

A massa de fruto é uma característica importante para o mamoeiro, tendo em vista que é uma das principais características para identificação do grupo ao qual o fruto pertence (Solo ou Formosa), porém quando se pretende identificar frutos com características superiores e de melhor aceitação no mercado consumidor, o melhorista não pode considerar somente uma característica.

O fruto precisa ser doce (SS) e ter boa firmeza e espessura de polpa (FIF e EP), fatores que agradam ao paladar, e também, por exemplo, serem resistentes ao transporte (FEF). Logo, a análise tem que ser bem mais ampla do

que se considerar apenas uma característica como a mais importante. E ainda com valores tão altos de contribuição, minimizando a importância da contribuição das outras características.

Dessa forma, o método de Singh leva a concluir que os demais caracteres seriam pouco informativos para a avaliação da dissimilaridade genética entre os acessos, devido às estimativas de contribuição relativa de pequenas magnitudes. Afirmativa que deve ser considerada com ressalvas.

Como os dois primeiros componentes possuem uma variação acumulada de 77,63% nesse estudo, a dispersão gráfica foi realizada em duas dimensões. Segundo Amaral Júnior e Thiébaud (1999), a variação acumulada acima de 70% seria suficiente para a dispersão gráfica bidimensional.

Os componentes principais foram estimados de modo a dispor as cultivares em um gráfico utilizando a dispersão dos valores dos referidos acessos, em relação aos eixos representativos dos dois primeiros componentes principais. A dispersão dos escores referentes à posição de cada acesso, em eixos cartesianos, é apresentada na Figura 2.

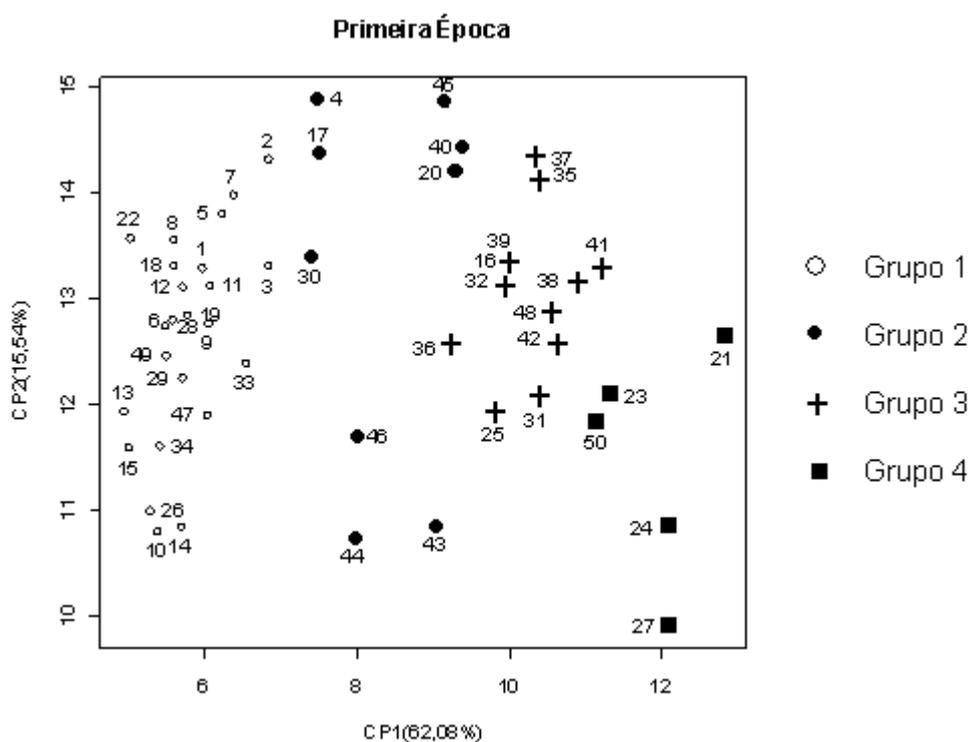


Figura 2. Dispersão dos 50 acessos de mamoeiro (*Carica papaya* L.) em dois componentes principais, com base nas distâncias Euclidianas médias, estimadas a partir de 9 caracteres quantitativos, na primeira época. A descrição dos acessos encontra-se na seção 2 do capítulo Material e Métodos.

A representação entre os dois primeiros componentes principais (CP 01 x CP 02), pode sugerir a formação de quatro grupos (Figura 2), e proporciona a distinção das principais características dos acessos que formam os grupos.

O grupo 1, formado pela maioria dos acessos, é composto por 17 acessos do grupo Solo e 7 acessos do grupo Formosa. É característico dos acessos que produzem frutos menores, com massa de fruto média de 395,90g, comprimento do fruto médio de 12,56cm, diâmetro do fruto médio de 8,39cm, espessura da polpa média de 2,40cm, conteúdo de sólidos solúveis médio de 11,18<sup>º</sup> Brix, firmeza externa do fruto média de 80,81N e firmeza interna da polpa média de 63,21N.

O grupo 2 é composto por 9 acessos do grupo Formosa. É característico dos acessos que produzem frutos com massa de fruto média de 826,48g, comprimento do fruto médio de 18,51cm, diâmetro do fruto médio de 9,82cm, espessura da polpa média de 2,83cm, conteúdo de sólidos solúveis médio de 10,38<sup>º</sup> Brix, firmeza externa do fruto média de 169,45N e firmeza interna da polpa média de 128,41N.

O grupo 3 é composto por 12 acessos do grupo Formosa. É característico dos acessos que produzem frutos com massa de fruto média de 1285,84g, comprimento do fruto médio de 22,73cm, diâmetro do fruto médio de 11,00cm, espessura da polpa média de 3,03cm, conteúdo de sólidos solúveis médio de 9,39<sup>º</sup> Brix, firmeza externa do fruto média de 184,34N e firmeza interna da polpa média de 140,30N.

O grupo 4 é composto por 5 acessos do grupo Formosa. É característico dos acessos que produzem frutos maiores, com massa de fruto média de 2329,34g, comprimento do fruto médio de 32,39cm, diâmetro do fruto médio de 12,39cm, espessura da polpa média de 3,18cm, conteúdo de sólidos solúveis médio de 8,11<sup>º</sup> Brix, firmeza externa do fruto média de 185,26N e firmeza interna da polpa média de 146,34N.

A identificação de pares de genótipos mais divergentes foi realizada com base na dispersão gráfica e confirmada nas medidas de dissimilaridade com base na distância Euclidiana média, garantindo maior confiança nos resultados.

De acordo com a visualização gráfica, confirmando nas medidas de dissimilaridade com base na distância Euclidiana média, foram identificados os acessos mais similares e os mais dissimilares. É esperado que entre os acessos

de grupos diferentes (Solo ou Formosa) a divergência genética seja maior que entre os acessos de um mesmo grupo.

Os acessos mais distantes geneticamente, com uma distância Euclidiana média de aproximadamente 2,73, são os acessos 10 - Caliman SG, grupo Solo, com massa de fruto de 288,33g e comprimento do fruto de 11,20cm, e 21 - Mamão Bené, grupo Formosa, com massa de fruto de 2441,67g e comprimento do fruto de 29,14cm.

Os acessos mais próximos geneticamente, com uma distância Euclidiana média de aproximadamente 0,17, são os acessos 12 - Sunrise Solo 72/12 (202), grupo Solo, com massa de fruto de 378,34g e comprimento do fruto de 11,55cm, e 19 - Sunrise Solo (sel. Caliman), grupo Solo, com massa de fruto de 425,00g e comprimento do fruto de 12,67cm.

## **1.2. Segunda época**

O resultado da análise de componentes principais, obtidos a partir de 9 caracteres quantitativos avaliados em 48 acessos (Quadro 3), mostra que foram necessários dois componentes, para que a variância por eles explicada atingisse um mínimo de 80%.

Pode-se afirmar que a análise de componentes principais reduziu os nove descritores em dois componentes principais, que explicaram 82,22% da variância total. O componente 1 explicou 69,23% da variância total e o componente 2 explicou 12,98% da variância total.

Observou-se que massa de fruto foi o caráter de maior peso em CP 01 (0,3841), sendo identificado como o de maior contribuição para a variação acumulada, e o caráter de menor contribuição para a variação acumulada foi firmeza interna do fruto, com maior peso em CP 09 (0,7081). Identificando-se assim, os caracteres de maior e menor importância no estudo da diversidade genética.

Pelo método de Singh (1981), com base na distância Euclidiana média, utilizado para avaliar a importância relativa dos nove caracteres quantitativos, demonstrou-se que somente a massa de fruto contribuiu com 97,97% para a divergência genética, enquanto que as demais características contribuíram com apenas 2,03%.

Quadro 3. Estimativa das variâncias, autovalores (AV), associados aos componentes principais (CP), e respectivos coeficientes de ponderação (autovetores) de 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro (*Carica papaya* L.), na segunda época.

CP	AV	Variância	Caracteres *								
		Acumulada (%)	MF	CF	DF	EP	SS	FEF	FIF	PP	MM
CP 01	6,23	69,23	0,38	0,38	0,36	0,38	-0,22	0,37	0,38	0,19	0,27
CP 02	1,17	82,22	-0,03	0,02	-0,18	-0,09	-0,63	-0,05	-0,01	0,60	-0,44
CP 03	0,66	89,56	-0,04	-0,09	0,02	0,08	0,31	-0,27	-0,23	0,70	0,51
CP 04	0,41	94,09	0,05	-0,21	0,28	0,09	-0,62	-0,36	-0,34	-0,30	0,39
CP 05	0,28	97,22	0,32	-0,11	0,54	0,28	0,26	-0,23	-0,28	0,07	-0,56
CP 06	0,14	98,81	0,49	0,66	-0,28	-0,26	0,05	-0,33	-0,25	-0,09	0,02
CP 07	0,07	99,56	-0,09	0,08	-0,51	0,81	-0,01	-0,02	-0,21	-0,10	-0,06
CP 08	0,03	99,87	0,71	-0,59	-0,34	-0,05	0,01	0,15	0,07	0,02	0,06
CP 09	0,01	100,00	0,01	-0,06	-0,05	0,13	0,02	-0,69	0,71	-0,04	-0,04

\*MF = massa de fruto; CF = comprimento do fruto; DF = diâmetro do fruto; EP = espessura da polpa; SS = sólidos solúveis; FEF = firmeza externa do fruto; FIF = firmeza interna da polpa; PP = incidência de pinta preta; e MM = incidência de mancha fisiológica.

Como os dois primeiros componentes possuem uma variação acumulada de 82,22% nesse estudo, a dispersão gráfica foi realizada em duas dimensões.

Os componentes principais foram estimados de modo a dispor as cultivares em um gráfico utilizando a dispersão dos valores dos referidos acessos, em relação aos eixos representativos dos dois primeiros componentes principais. A dispersão dos escores referentes à posição de cada acesso, em eixos cartesianos, é apresentada na Figura 3.

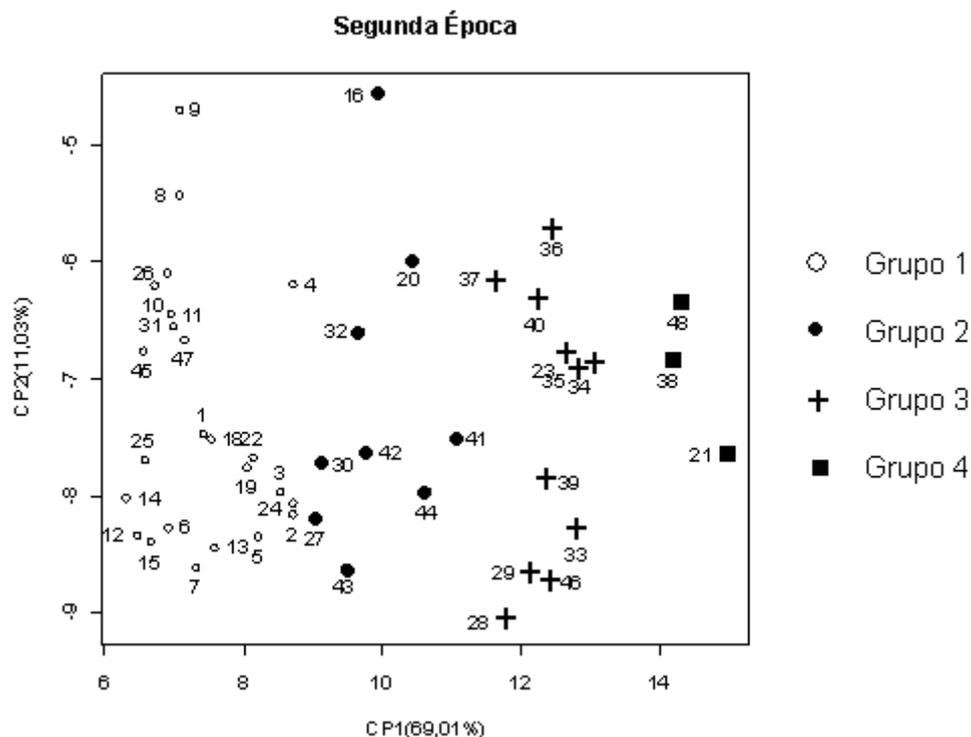


Figura 3. Dispersão dos 48 acessos de mamoeiro (*Carica papaya* L.) em dois componentes principais, com base nas distâncias Euclidianas médias, estimadas a partir de 9 caracteres quantitativos, na segunda época. A descrição dos acessos encontra-se na seção 2 do capítulo Material e Métodos.

A representação entre os dois primeiros componentes principais (CP 01 x CP 02), pode sugerir a formação de quatro grupos (Figura 3), e proporciona a distinção das principais características dos acessos que formam os grupos.

O grupo 1, formado pela maioria dos acessos, é composto por 17 acessos do grupo Solo e 8 acessos do grupo Formosa. É característico dos acessos que produzem frutos menores, com massa de fruto média de 444,27g, comprimento do fruto médio de 13,81cm, diâmetro do fruto médio de 8,16cm, espessura da polpa média de 1,99cm, conteúdo de sólidos solúveis médio de 10,99<sup>º</sup> Brix, firmeza externa do fruto média de 78,15N e firmeza interna da polpa média de 57,43N.

O grupo 2 é composto por 9 acessos do grupo Formosa. É característico dos acessos que produzem frutos com massa de fruto média de 767,59g, comprimento do fruto médio de 18,12cm, diâmetro do fruto médio de 9,47cm, espessura da polpa média de 2,36cm, conteúdo de sólidos solúveis médio de

10,40<sup>o</sup> Brix, firmeza externa do fruto média de 139,21N e firmeza interna da polpa média de 102,34N.

O grupo 3 é composto por 11 acessos do grupo Formosa. É característico dos acessos que produzem frutos com massa de fruto média de 1254,40g, comprimento do fruto médio de 23,17cm, diâmetro do fruto médio de 10,63cm, espessura da polpa média de 2,78cm, conteúdo de sólidos solúveis médio de 9,90<sup>o</sup> Brix, firmeza externa do fruto média de 179,32N e firmeza interna da polpa média de 141,15N.

O grupo 4 é composto por 3 acessos do grupo Formosa. É característico dos acessos que produzem frutos maiores, com massa de fruto média de 2069,44g, comprimento do fruto médio de 28,76cm, diâmetro do fruto médio de 12,01cm, espessura da polpa média de 3,05cm, conteúdo de sólidos solúveis médio de 9,46<sup>o</sup> Brix, firmeza externa do fruto média de 198,78N e firmeza interna da polpa média de 151,20N.

A identificação de pares de genótipos mais divergentes foi realizada com base na dispersão gráfica e confirmada nas medidas de dissimilaridade com base na distância Euclidiana média, garantindo maior confiança nos resultados.

De acordo com a visualização gráfica, confirmando nas medidas de dissimilaridade com base na distância Euclidiana média, foram identificados os acessos mais similares e os mais dissimilares. É esperado que entre os acessos de grupos diferentes (Solo ou Formosa) a divergência genética seja maior que entre os acessos de um mesmo grupo.

Os acessos mais distantes geneticamente, com uma distância Euclidiana média de aproximadamente 2,92, são os acessos 14 - Baixinho Santa Amália, grupo Solo, com massa de fruto de 311,67g e comprimento do fruto de 11,50cm, e 21 - Mamão Bené, grupo Formosa, com massa de fruto de 2213,33g e comprimento do fruto de 29,53cm.

Os acessos mais próximos geneticamente, com uma distância Euclidiana média de aproximadamente 0,21, são os acessos 10 - Caliman SG, grupo Solo, com massa de fruto de 300,00g e comprimento do fruto de 12,07cm, e 26 - STZ-51, grupo Solo, com massa de fruto de 385,00g e comprimento do fruto de 12,40cm.

### 1.3. Média das épocas

O estudo da média das duas épocas foi feito com o objetivo de verificar o impacto na redistribuição dos acessos e na reavaliação das características mais relevantes na predição da importância relativa das mesmas para a diversidade genética.

O resultado da análise de componentes principais, obtidos a partir de 9 caracteres quantitativos avaliados em 48 acessos (Quadro 4), mostra que foram necessários dois componentes, para que a variância por eles explicada atingisse um mínimo de 80%.

Quadro 4. Estimativa das variâncias, autovalores (AV), associados aos componentes principais (CP), e respectivos coeficientes de ponderação (autovetores) de 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro (*Carica papaya* L.), na média das épocas.

CP	AV	Variância		Caracteres *							
		Acumulada (%)	MF	CF	DF	EP	SS	FEF	FIF	PP	MM
CP 01	6,21	69,01	0,38	0,38	0,36	0,38	-0,25	0,37	0,38	0,16	0,25
CP 02	0,99	80,05	-0,12	-0,10	-0,14	0,07	-0,23	-0,18	-0,16	0,90	0,18
CP 03	0,92	90,31	-0,03	-0,11	0,13	0,10	0,68	-0,04	-0,06	0,01	0,70
CP 04	0,32	93,91	0,28	-0,14	0,51	0,30	-0,20	-0,49	-0,50	-0,16	-0,01
CP 05	0,30	97,29	-0,17	-0,02	-0,25	-0,14	-0,59	-0,07	-0,02	-0,35	0,64
CP 06	0,16	99,04	0,45	0,68	-0,36	-0,23	0,14	-0,28	-0,23	0,03	0,06
CP 07	0,07	99,78	-0,12	0,05	-0,52	0,81	0,05	0,03	-0,13	-0,14	-0,08
CP 08	0,02	99,95	0,71	-0,59	-0,33	-0,03	-0,01	0,00	0,20	0,01	0,03
CP 09	0,00	100,00	-0,12	0,06	0,02	0,12	0,04	-0,70	0,68	-0,02	-0,03

\*MF = massa de fruto; CF = comprimento do fruto; DF = diâmetro do fruto; EP = espessura da polpa; SS = sólidos solúveis; FEF = firmeza externa do fruto; FIF = firmeza interna da polpa; PP = incidência de pinta preta; e MM = incidência de mancha fisiológica.

Pode-se afirmar que a análise de componentes principais reduziu os nove descritores em dois componentes principais, que explicaram 80,04% da variância total. O componente 1 explicou 69,01% da variância total e o componente 2 explicou 11,03% da variância total.

Observou-se que massa de fruto foi o caráter de maior peso em CP 01 (0,3827), sendo identificado como o de maior contribuição para a variação acumulada, e o caráter de menor contribuição para a variação acumulada foi firmeza externa do fruto, com maior peso em CP 09 (-0,7049). Identificando-se assim, os caracteres de maior e menor importância no estudo da diversidade genética.

O método de Singh (1981), com base na distância Euclidiana média, utilizado para avaliar a importância relativa dos nove caracteres quantitativos, demonstrou que somente o caráter massa de fruto contribuiu com 98,25% para a divergência genética, enquanto que as demais características contribuíram com apenas 1,75%.

Fato observado em todas as épocas, e na média das épocas, é que as características massa de fruto, comprimento do fruto, diâmetro do fruto e espessura da polpa são altamente correlacionadas. O componente principal 01 é relacionado às características quantitativas de formato dos frutos (MF, CF, DF e EP) e de consistência dos frutos (FEF e FIF).

Como os dois primeiros componentes possuem uma variação acumulada de 80,04% nesse estudo, a dispersão gráfica foi realizada em duas dimensões. Segundo Amaral Júnior e Thiébaud (1999), a variação acumulada acima de 70% seria suficiente para a dispersão gráfica bidimensional.

Os componentes principais foram estimados de modo a dispor as cultivares em um gráfico utilizando a dispersão dos valores dos referidos acessos, em relação aos eixos representativos dos dois primeiros componentes principais. A dispersão dos escores referentes à posição de cada acesso, em eixos cartesianos, é apresentada na Figura 4.

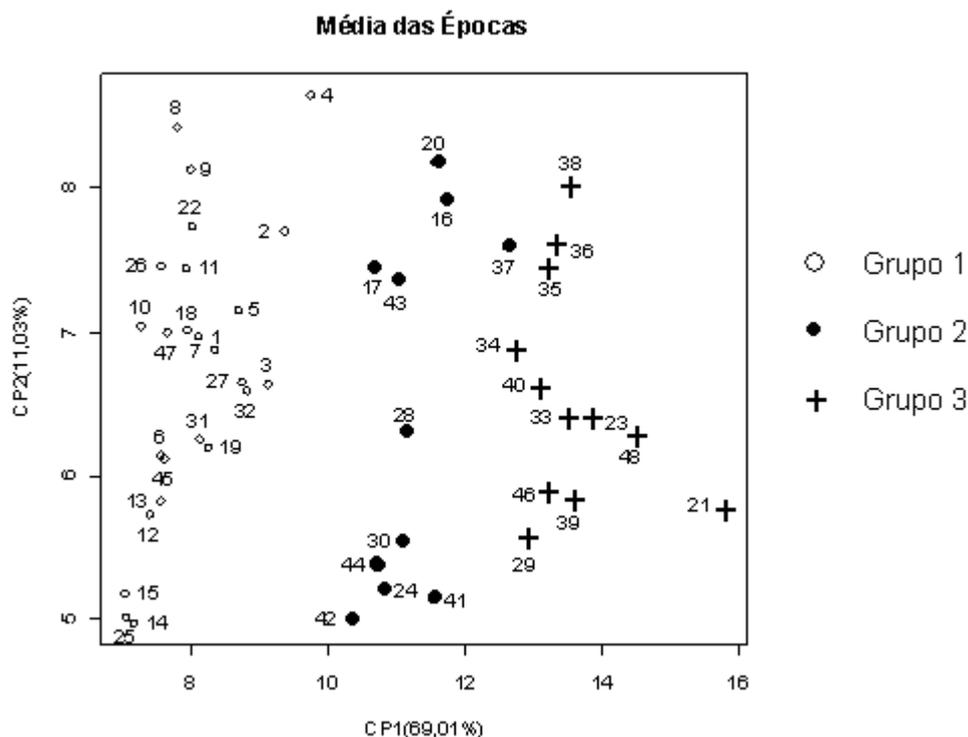


Figura 4. Dispersão dos 48 acessos de mamoeiro (*Carica papaya* L.) em dois componentes principais, com base nas distâncias Euclidianas médias, estimadas a partir de 9 caracteres quantitativos, na média das épocas. A descrição dos acessos encontra-se na seção 2 do capítulo Material e Métodos.

A representação entre os dois primeiros componentes principais (CP 01 x CP 02), pode sugerir a formação de três grupos (Figura 4), e proporciona a distinção das principais características dos acessos que formam os grupos.

O grupo 1, formado pela maioria dos acessos, é composto por 17 acessos do grupo Solo e 8 acessos do grupo Formosa. É característico dos acessos que produzem frutos menores, com massa de fruto média de 424,30g, comprimento do fruto médio de 13,18cm, diâmetro do fruto médio de 8,32cm, espessura da polpa média de 2,20cm, conteúdo de sólidos solúveis médio de 11,07<sup>º</sup> Brix, firmeza externa do fruto média de 78,20N e firmeza interna da polpa média de 59,42N.

O grupo 2 é composto por 10 acessos do grupo Formosa e 1 acesso do grupo Solo, mas com características de acesso do grupo Formosa. É característico dos acessos que produzem frutos com massa de fruto média de 890,00g, comprimento do fruto médio de 19,38cm, diâmetro do fruto médio de 9,87cm, espessura da polpa média de 2,62cm, conteúdo de sólidos solúveis

médio de 10,32<sup>o</sup> Brix, firmeza externa do fruto média de 162,60N e firmeza interna da polpa média de 122,18N.

O grupo 3 é composto por 12 acessos do grupo Formosa. É característico dos acessos que produzem frutos maiores, com massa de fruto média de 1516,25g, comprimento do fruto médio de 24,80cm, diâmetro do fruto médio de 11,19cm, espessura da polpa média de 3,01cm, conteúdo de sólidos solúveis médio de 9,40<sup>o</sup> Brix, firmeza externa do fruto média de 188,30N e firmeza interna da polpa média de 145, 83N.

A identificação de pares de genótipos mais divergentes foi realizada com base na dispersão gráfica e confirmada nas medidas de dissimilaridade com base na distância Euclidiana média, garantindo maior confiança nos resultados.

De acordo com a visualização gráfica, confirmando nas medidas de dissimilaridade com base na distância Euclidiana média, foram identificados os acessos mais similares e os mais dissimilares. É esperado que entre os acessos de grupos diferentes (Solo ou Formosa) a divergência genética seja maior que entre os acessos de um mesmo grupo.

Os acessos mais distantes geneticamente, com uma distância Euclidiana média de aproximadamente 2,99, são os acessos 10 - Caliman SG, grupo Solo, com massa de fruto média de 294,17g e comprimento do fruto médio de 11,63cm, e 21 - Mamão Bené, grupo Formosa, com massa de fruto de 2327,50g e comprimento do fruto médio de 29,34cm.

Os acessos mais próximos geneticamente, com uma distância Euclidiana média de aproximadamente 0,15, são os acessos 6 - Grampola, grupo Solo, com massa de fruto média de 348,33g e comprimento do fruto médio de 12,46cm, e 13 - Kapoho Solo (polpa amarela), grupo Solo, com massa de fruto de 386,67g e comprimento do fruto de 12,94cm.

## **2. Método de otimização Tocher**

### **2.1. Primeira época**

O agrupamento pelo método de Tocher, aplicado sobre a matriz de dissimilaridade, obtida a partir da distância Euclidiana, caracterizou a formação de sete grupos (Quadro 5).

Quadro 5. Grupos dos acessos de mamoeiro (*Carica papaya* L.), na primeira época, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos avaliados em 50 acessos.

Grupos	Acessos
1	1 - Caliman M5, 2 - Sunrise Solo 783, 3 - Costa Rica, 5 - Diva, 6 - Grampola, 7 - Sunrise Solo, 8 - Caliman AM, 9 - Caliman GB, 10 - Caliman SG, 11 - Caliman G, 12 - Sunrise Solo 72/12 (202), 13 - Kapoho Solo (polpa amarela), 14 - Baixinho Santa Amália, 15 - Sunrise Solo TJ, 17 - São Mateus, 18 - Kapoho Solo (polpa vermelha), 19 - Sunrise Solo (sel. Caliman), 22 - Mamão roxo, 26 - Baixinho Super, 28 - STZ-51, 29 - STZ-52, 33 - Golden tipo Formosa, 34 - STZ-03 - pecíolo curto, 47 - Golden robusto, 49 - STZ 63
2	16 - Tailândia, 20 - Waimanalo, 25 - Sekati, 31 - JS 12 (206), 32 - Cariflora 209, 35 - STA Helena III Trat 12a plt 07X, 36 - STA Helena III Trat 11a plt 08X, 37 - STA Helena III Trat 50 plt 09X, 38 - STA Helena III Trat 02 plt 01X, 39 - STA Helena III Trat 14 plt 05X, 40 - STA Helena III Trat 15 plt 04X, 41 - STA Helena III Trat 04 plt 02X, 42 - STA Helena III Trat 12b plt 06X, 45 - Papaya 46 claro, 48 - Tainung, 50 - Sekati (fruto longo macuco)
3	43 - Papaya 42 Formosa, 44 - Papaya 45 Formosa roxo, 46 - Fruto médio verde
4	21 - Mamão Bené, 23 - Maradol (orig. México), 24 - Maradol (grande limão)
5	27 - Americano
6	4 - Triwan Et
7	30 - Calimosa

Acessos com características semelhantes integram os mesmos grupos. O primeiro grupo reuniu todos os acessos do grupo Solo, porém apresentou alguns acessos do grupo Formosa. Esses acessos do grupo Formosa possuem características semelhantes às características dos acessos do grupo Solo. Comportamento também observado na análise por componentes principais.

O segundo grupo é formado por 16 acessos do grupo Formosa, indicando a similaridade genética existente entre esses acessos. Porém, os 9 acessos restantes foram distribuídos em vários grupos (3, 4, 5, 6, 7) e fora do grupo 2, representativo dos acessos do grupo Formosa.

O fato de esses poucos acessos estarem distribuídos em vários grupos, indica a existência de diversidade genética entre esses acessos e no grupo Formosa. Principalmente pelo fato de os acessos do grupo Solo estarem reunidos em um único grupo.

Cabe ressaltar a existência dos grupos unitários 5, 6 e 7, formados pelos acessos 27 - Americano, 4 - Triwan Et e 30 - Calimosa, respectivamente.

O grupo 1, formado por 50% dos acessos, é composto por 17 acessos do grupo Solo e 8 acessos do grupo Formosa. É característico dos acessos que produzem frutos menores, com massa de fruto média de 407,67g, comprimento do fruto médio de 12,60cm, diâmetro do fruto médio de 8,44cm, espessura da polpa média de 2,42cm, conteúdo de sólidos solúveis médio de 11,22° Brix, firmeza externa do fruto média de 82,87N e firmeza interna da polpa média de 64,42N.

O grupo 2 é composto por 16 acessos do grupo Formosa, com massa de fruto média de 1272,50g, comprimento do fruto médio de 22,38cm, diâmetro do fruto médio de 10,91cm, espessura da polpa média de 3,09cm, conteúdo de sólidos solúveis médio de 9,41° Brix, firmeza externa do fruto média de 187,89N e firmeza interna da polpa média de 142,53N.

O grupo 3 é composto por 3 acessos do grupo Formosa, com massa de fruto média de 815,00g, comprimento do fruto médio de 20,56cm, diâmetro do fruto médio de 9,62cm, espessura da polpa média de 2,52cm, conteúdo de sólidos solúveis médio de 9,49° Brix, firmeza externa do fruto média de 173,65N e firmeza interna da polpa média de 137,19N.

O grupo 4 é composto por 3 acessos do grupo Formosa, com massa de fruto média de 2398,34g, comprimento do fruto médio de 30,76cm, diâmetro do fruto médio de 13,19cm, espessura da polpa média de 3,24cm, conteúdo de sólidos solúveis médio de 7,79° Brix, firmeza externa do fruto média de 176,78N e firmeza interna da polpa média de 140,43N.

O grupo 5 é unitário e composto pelo acesso 27 - Americano do grupo Formosa, com massa de fruto de 2386,67g, comprimento do fruto de 41,77cm, diâmetro do fruto de 10,62cm, espessura da polpa de 2,81cm, conteúdo de sólidos solúveis médio de 7,54° Brix, firmeza externa do fruto de 203,96N e firmeza interna da polpa de 168,76N.

O grupo 6 é unitário e composto pelo acesso 4 - Triwan Et do grupo Formosa, com massa de fruto de 605,00g, comprimento do fruto de 13,95cm, diâmetro do fruto de 9,77cm, espessura da polpa de 3,11cm, conteúdo de sólidos solúveis médio de 10,70º Brix, firmeza externa do fruto de 95,54N e firmeza interna da polpa de 71,80N.

O grupo 7 é unitário e composto pelo acesso 30 - Calimosa do grupo Formosa, com massa de fruto de 833,34g, comprimento do fruto de 19,95cm, diâmetro do fruto de 9,39cm, espessura da polpa de 2,32cm, conteúdo de sólidos solúveis médio de 13,70º Brix, firmeza externa do fruto de 174,22N e firmeza interna da polpa de 123,46N.

As estimativas das distâncias intragrupos e intergrupos, obtidas a partir do método de agrupamento de Tocher, encontram-se no Quadro 6. A maior distância intergrupos foi encontrada entre os grupos 1 e 5, com distância estimada de 2,4183. Outras distâncias intergrupos estimadas, que podem ser consideradas, foram entre os grupos 5 e 6, 1 e 4, 5 e 7 e 4 e 7, com valores de dissimilaridade 2,3965, 2,2498, 2,1553 e 2,0305, respectivamente. Indicando provavelmente as melhores combinações intergrupos para cruzamentos.

O valor mínimo da distância foi obtido entre os grupos 1 e 6 (1,0800), evidenciando que o intercruzamento dos acessos componentes destes grupos pode não ser muito indicado para a obtenção de genótipos superiores nas gerações segregantes.

Quadro 6. Estimativa das distâncias genéticas intragrupos e intergrupos dos grupos dos acessos de mamoeiro (*Carica papaya* L.), na primeira época, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos e avaliados em 50 acessos.

Grupos	1	2	3	4	5	6	7
1	0,7683	1,6433	1,2796	2,2498	2,4183	1,0800	1,1964
2		0,8295	1,2056	1,2092	1,5386	1,2480	1,3706
3			0,6416	1,6229	1,5789	1,5093	1,1687
4				0,8654	1,1300	1,9185	2,0305
5					0	2,3965	2,1553
6						0	1,2456
7							0

## 2.2. Segunda época

O agrupamento pelo método de Tocher, aplicado sobre a matriz de dissimilaridade, obtida a partir da distância Euclidiana, caracterizou a formação de dez grupos (Quadro 7).

Quadro 7. Grupos dos acessos de mamoeiro (*Carica papaya* L.), na segunda época, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.

Grupos	Acessos
1	1 - Caliman M5, 2 - Sunrise Solo 783, 3 - Costa Rica, 4 - Triwan Et, 5 - Diva, 6 - Grampola, 7 - Sunrise Solo, 8 - Caliman AM, 10 - Caliman SG, 11 - Caliman G, 12 - Sunrise Solo 72/12 (202), 13 - Kapoho Solo (polpa amarela), 14 - Baixinho Santa Amália, 15 - Sunrise Solo TJ, 18 - Kapoho Solo (polpa vermelha), 19 - Sunrise Solo (sel. Caliman), 22 - Mamão roxo, 24 - Sekati, 25 - Baixinho Super, 26 - STZ-51, 31 - Golden tipo Formosa, 45 - Golden robusto, 47 - STZ 63, 27 - STZ-52
2	44 - Fruto médio verde, 41 - Papaya 42 Formosa, 46 - Tainung, 29 - JS 12 (206), 37 - STA Helena III Trat 14 plt 05X, 36 - STA Helena III Trat 02 plt 01X, 39 - STA Helena III Trat 04 plt 02X, 23 - Maradol (orig. México), 40 - STA Helena III Trat 12b plt 06X, 35 - STA Helena III Trat 50 plt 09X, 34 - STA Helena III Trat 11a plt 08X
3	21 - Mamão Bené, 38 - STA Helena III Trat 15 plt 04X
4	30 - Cariflora 209, 32 - STZ-03 - pecíolo curto, 42 - Papaya 45 Formosa roxo
5	16 - Tailândia, 20 - Waimanalo
6	17 - São Mateus, 28 - Calimosa
7	43 - Papaya 46 claro
8	33 - STA Helena III Trat 12a plt 07X
9	48 - Sekati (fruto longo macuco)
10	9 - Caliman GB

Acessos com características semelhantes integram os mesmos grupos. O primeiro grupo reuniu a maioria dos acessos do grupo Solo (14 acessos dos 16 acessos), porém apresentou alguns acessos do grupo Formosa (10 acessos). Esses acessos do grupo Formosa possuem características semelhantes às características dos acessos do grupo Solo. Comportamento também observado na análise por componentes principais.

O segundo grupo é formado por 11 acessos do grupo Formosa, indicando a similaridade genética existente entre esses acessos. Porém, os 13 acessos restantes foram distribuídos em vários grupos (3, 4, 5, 6, 7, 8 e 9) e fora do grupo 2, representativo dos acessos do grupo Formosa.

O fato de esses poucos acessos estarem distribuídos em vários grupos, indica a existência de diversidade genética entre esses acessos e no grupo Formosa. Principalmente pelo fato de os acessos do grupo Solo estarem reunidos praticamente em um único grupo.

Cabe ressaltar a existência dos grupos unitários 7, 8, 9 e 10, formados pelos acessos 43 - Papaya 46 claro, 33 - STA Helena III Trat12a plt 07X, 48 - Sekati (fruto longo macuco), 9 - Caliman GB, respectivamente.

O grupo 1, formado por 50% dos acessos, é composto por 16 acessos do grupo Solo e 8 acessos do grupo Formosa. É característico dos acessos que produzem frutos menores, com massa de fruto média de 444,38g, comprimento do fruto médio de 13,80cm, diâmetro do fruto médio de 8,20cm, espessura da polpa média de 1,99cm, conteúdo de sólidos solúveis médio de 11,16° Brix, firmeza externa do fruto média de 75,93N e firmeza interna da polpa média de 55,48N.

O grupo 2 é composto por 11 acessos do grupo Formosa, com massa de fruto média de 1190,76g, comprimento do fruto médio de 23,17cm, diâmetro do fruto médio de 10,23cm, espessura da polpa média de 2,71cm, conteúdo de sólidos solúveis médio de 9,75° Brix, firmeza externa do fruto média de 180,42N e firmeza interna da polpa média de 140,83N.

O grupo 3 é composto por 2 acessos do grupo Formosa, com massa de fruto média de 2121,67g, comprimento do fruto médio de 26,94cm, diâmetro do fruto médio de 12,50cm, espessura da polpa média de 3,06cm, conteúdo de sólidos solúveis médio de 9,12° Brix, firmeza externa do fruto média de 193,70N e firmeza interna da polpa média de 144,45N.

O grupo 4 é composto por 3 acessos do grupo Formosa, com massa de fruto média de 749,45g, comprimento do fruto médio de 17,62cm, diâmetro do fruto médio de 9,31cm, espessura da polpa média de 2,23cm, conteúdo de sólidos solúveis médio de 10,27º Brix, firmeza externa do fruto média de 147,41N e firmeza interna da polpa média de 106,51N.

O grupo 5 é composto por 2 acessos do grupo Formosa, com massa de fruto média de 727,51g, comprimento do fruto médio de 18,77cm, diâmetro do fruto médio de 8,99cm, espessura da polpa média de 2,30cm, conteúdo de sólidos solúveis médio de 9,29º Brix, firmeza externa do fruto média de 135,21N e firmeza interna da polpa média de 110,04N.

O grupo 6 é composto por 1 acesso do grupo Formosa e 1 acesso do grupo Solo, com massa de fruto média de 985,84g, comprimento do fruto médio de 19,67cm, diâmetro do fruto médio de 10,14cm, espessura da polpa média de 2,73cm, conteúdo de sólidos solúveis médio de 11,87º Brix, firmeza externa do fruto média de 171,31N e firmeza interna da polpa média de 122,62N.

O grupo 7 é unitário e composto pelo acesso 43 - Papaya 46 claro do grupo Formosa, com massa de fruto de 700,00g, comprimento do fruto de 15,87cm, diâmetro do fruto de 9,50cm, espessura da polpa de 2,50cm, conteúdo de sólidos solúveis médio de 10,73º Brix, firmeza externa do fruto de 86,29N e firmeza interna da polpa de 63,21N.

O grupo 8 é unitário e composto pelo acesso 33 - STA Helena III Trat 12a plt 07X do grupo Formosa, com massa de fruto de 1396,67g, comprimento do fruto de 20,34cm, diâmetro do fruto de 13,80cm, espessura da polpa de 2,96cm, conteúdo de sólidos solúveis médio de 9,90º Brix, firmeza externa do fruto de 169,96N e firmeza interna da polpa de 130,53N.

O grupo 9 é unitário e composto pelo acesso 48 - Sekati (fruto longo macuco) do grupo Formosa, com massa de fruto de 1965,00g, comprimento do fruto de 32,40cm, diâmetro do fruto de 11,03cm, espessura da polpa de 3,04cm, conteúdo de sólidos solúveis médio de 10,14º Brix, firmeza externa do fruto de 208,94N e firmeza interna da polpa de 164,69N.

O grupo 10 é unitário e composto pelo acesso 9 - Caliman GB do grupo Solo, com massa de fruto de 278,34g, comprimento do fruto de 11,45cm, diâmetro do fruto de 7,35cm, espessura da polpa de 1,79cm, conteúdo de sólidos solúveis

médio de 8,53<sup>o</sup> Brix, firmeza externa do fruto de 60,50N e firmeza interna da polpa de 50,27N.

As estimativas das distâncias intragrupos e intergrupos, obtidas a partir do método de agrupamento de Tocher, encontram-se no Quadro 8. A maior distância intergrupos foi encontrada entre os grupos 3 e 10, com distância estimada de 2,6859. Outras distâncias intergrupos estimadas, que podem ser consideradas, foram entre os grupos 9 e 10, 1 e 3, 1 e 9 e 8 e 10, com valores de dissimilaridade 2,6755, 2,4501, 2,4417 e 2,3628, respectivamente. Indicando provavelmente as melhores combinações intergrupos para cruzamentos.

O valor mínimo da distância foi obtido entre os grupos 3 e 9 (0,9292), evidenciando que o intercruzamento dos acessos componentes destes grupos pode não ser muito indicado para a obtenção de genótipos superiores nas gerações segregantes.

Quadro 8. Estimativa das distâncias genéticas intragrupos e intergrupos dos grupos dos acessos de mamoeiro (*Carica papaya* L.), na segunda época, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos e avaliados em 48 acessos.

Grupos	1	2	3	4	5	6	7	8	9	10
1	0,7293	1,7028	2,4501	1,0240	1,3476	1,5161	1,0817	2,0033	2,4417	1,1201
2		0,7421	1,1119	1,0840	1,1404	1,0259	1,3104	1,0189	1,1911	1,9699
3			0,5238	1,8461	1,7695	1,5125	1,8748	0,9723	0,9292	2,6859
4				0,6586	1,0042	1,1830	1,0989	1,4080	1,8440	1,3274
5					0,8704	1,3227	1,3904	1,6327	1,6428	1,2252
6						0,9035	1,1005	1,2514	1,4999	2,0992
7							0	1,4717	2,1349	1,6442
8								0	1,4196	2,3628
9									0	2,6755
10										0

### 2.3. Média das épocas

O estudo da média das duas épocas foi feito com o objetivo de verificar o impacto na redistribuição dos acessos.

O agrupamento pelo método de Tocher, aplicado sobre a matriz de dissimilaridade, obtida a partir da distância Euclidiana, caracterizou a formação de nove grupos (Quadro 9).

Quadro 9. Grupos dos acessos de mamoeiro (*Carica papaya* L.), na média das épocas, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.

<b>Grupos</b>	<b>Acessos</b>
1	1 - Caliman M5, 3 - Costa Rica, 5 - Diva, 6 - Grampola, 7 - Sunrise Solo, 10 - Caliman SG, 11 - Caliman G, 12 - Sunrise Solo 72/12 (202), 13 - Kapoho Solo (polpa amarela), 14 - Baixinho Santa Amália, 15 - Sunrise Solo TJ, 18 - Kapoho Solo (polpa vermelha), 19 - Sunrise Solo (sel. Caliman), 22 - Mamão roxo, 25 - Baixinho Super, 26 - STZ-51, 27 - STZ-52, 31 - Golden tipo Formosa, 32 - STZ-03 - pecíolo curto, 45 - Golden robusto, 47 - STZ 63
2	23 - Maradol (orig - México), 29 - JS 12 (206), 33 - STA Helena III Trat 12a plt 07X, 34 - STA Helena III Trat 11a plt 08X, 35 - STA Helena III Trat 50 plt 09X, 36 - STA Helena III Trat 02 plt 01X, 37 - STA Helena III Trat 14 plt 05X, 38 - STA Helena III Trat 15 plt 04X, 39 - STA Helena III Trat 04 plt 02X, 40 - STA Helena III Trat 12b plt 06X, 46 - Tainung
3	24 - Sekati, 30 - Cariflora 209, 41 - Papaya 42 Formosa, 42 - Papaya 45 Formosa roxo, 44 - Fruto médio verde
4	8 - Caliman AM, 9 - Caliman GB
5	2 - Sunrise Solo 783, 4 - Triwan Et, 17 - São Mateus, 43 - Papaya 46 claro
6	16 - Tailândia, 20 - Waimanalo
7	28 - Calimosa
8	48 - Sekati (fruto longo macuco)
9	21 - Mamão Bené

Acessos com características semelhantes integram os mesmos grupos. O primeiro grupo reuniu todos os acessos do grupo Solo, porém apresentou alguns acessos do grupo Formosa. Esses acessos do grupo Formosa possuem características semelhantes às características dos acessos do grupo Solo. Comportamento também observado na análise por componentes principais.

O grupo 1, formado pela maioria dos acessos, é composto por 18 acessos do grupo Solo e 3 acessos do grupo Formosa. É característico dos frutos com massa de fruto média de 418,93g, comprimento do fruto médio de 13,15cm, diâmetro do fruto médio de 8,27cm, espessura da polpa média de 2,18cm, conteúdo de sólidos solúveis médio de 11,26<sup>o</sup> Brix, firmeza externa do fruto média de 78,40N e firmeza interna da polpa média de 59,29N.

O grupo 2 é composto por 11 acessos do grupo Formosa. É característico dos frutos com massa de fruto média de 1353,64g, comprimento do fruto médio de 23,86cm, diâmetro do fruto médio de 10,74cm, espessura da polpa média de 2,96cm, conteúdo de sólidos solúveis médio de 9,38<sup>o</sup> Brix, firmeza externa do fruto média de 185,11N e firmeza interna da polpa média de 143,63N.

O grupo 3 é composto por 5 acessos do grupo Formosa. É característico dos frutos com massa de fruto média de 882,50g, comprimento do fruto médio de 19,40cm, diâmetro do fruto médio de 9,99cm, espessura da polpa média de 2,52cm, conteúdo de sólidos solúveis médio de 10,10<sup>o</sup> Brix, firmeza externa do fruto média de 170,02N e firmeza interna da polpa média de 127,68N.

O grupo 4 é composto por 2 acessos do grupo Solo. É característico dos frutos com massa de fruto média de 335,42g, comprimento do fruto médio de 12,32cm, diâmetro do fruto médio de 7,72cm, espessura da polpa média de 2,11cm, conteúdo de sólidos solúveis médio de 9,52<sup>o</sup> Brix, firmeza externa do fruto média de 72,27N e firmeza interna da polpa média de 56,75N.

O grupo 5 é composto por 2 acessos do grupo Solo e 2 acessos do grupo Formosa. É característico dos frutos com massa de fruto média de 653,13g, comprimento do fruto médio de 14,66cm, diâmetro do fruto médio de 9,70cm, espessura da polpa média de 2,69cm, conteúdo de sólidos solúveis médio de 10,95<sup>o</sup> Brix, firmeza externa do fruto média de 111,08N e firmeza interna da polpa média de 80,72N.

O grupo 6 é composto por 2 acessos do grupo Formosa. É característico dos frutos com massa de fruto média de 937,09g, comprimento do fruto médio de

20,02cm, diâmetro do fruto médio de 10,08cm, espessura da polpa média de 2,69cm, conteúdo de sólidos solúveis médio de 9,24<sup>o</sup> Brix, firmeza externa do fruto média de 147,06N e firmeza interna da polpa média de 114,72N.

O grupo 7 é unitário e composto pelo acesso 28 - Calimosa do grupo Formosa, com massa de fruto de 992,50g, comprimento do fruto de 21,79cm, diâmetro do fruto de 9,67cm, espessura da polpa de 2,45cm, conteúdo de sólidos solúveis médio de 12,75<sup>o</sup> Brix, firmeza externa do fruto de 176,87N e firmeza interna da polpa de 133,13N.

O grupo 8 é unitário e composto pelo acesso 48 - Sekati (fruto longo macuco) do grupo Formosa, com massa de fruto de 2015,00g, comprimento do fruto de 30,14cm, diâmetro do fruto de 11,40cm, espessura da polpa de 3,22cm, conteúdo de sólidos solúveis médio de 9,89<sup>o</sup> Brix, firmeza externa do fruto de 200,48N e firmeza interna da polpa de 153,16N.

O grupo 9 é unitário e composto pelo acesso 21 - Mamão Bené do grupo Formosa, com massa de fruto de 2327,50g, comprimento do fruto de 29,34cm, diâmetro do fruto de 13,44cm, espessura da polpa de 3,11cm, conteúdo de sólidos solúveis médio de 8,95<sup>o</sup> Brix, firmeza externa do fruto de 210,05N e firmeza interna da polpa de 163,90N.

As estimativas das distâncias intragrupos e intergrupos, obtidas a partir do método de agrupamento de Tocher, encontram-se no Quadro 10. A maior distância intergrupos foi encontrada entre os grupos 4 e 9, com distância estimada de 2.8848. Outra distância intergrupos estimada, que pode ser considerada, foi entre os grupos 1 e 9, com valor de dissimilaridade 2.7186. Indicando provavelmente as melhores combinações intergrupos para cruzamentos.

O valor mínimo da distância foi obtido entre os acessos dos grupos 2 e 8 (0,9155), evidenciando que o intercruzamento dos acessos componentes destes grupos pode não ser muito indicado para a obtenção de genótipos superiores nas gerações segregantes.

Os grupos gerados pelo método de otimização de Tocher facilitam a seleção de material divergente para compor o melhoramento por meio da geração de híbridos, uma vez que permitem a seleção de genótipos pertencentes a grupos heteróticos distintos, a serem indicados para cruzamento.

Quadro 10. Estimativa das distâncias genéticas intragrupos e intergrupos dos grupos dos acessos de mamoeiro (*Carica papaya* L.), na média das épocas, formados através do método de agrupamento de Tocher, a partir da distância Euclidiana, baseando-se em 9 caracteres quantitativos e avaliados em 48 acessos.

Grupos	1	2	3	4	5	6	7	8	9
1	0.6597	1.8970	1.2902	0.9502	1.1346	1.5084	1.4226	2.3139	2.7186
2		0.6859	1.1440	1.9994	1.32861	0.9659	1.2052	0.9155	1.1460
3			0.6450	1.5325	1.2678	1.1279	1.1123	1.4552	1.8337
4				0.4650	1.3192	1.3665	1.7807	2.3878	2.8848
5					0.7240	1.0400	1.0521	1.8200	2.1105
6						0.8053	1.1988	1.3135	1.7485
7							0	1.4821	1.8683
8								0	0.9275
9									0

### 3. Rede neural artificial

#### 3.1. Determinação do número ótimo de grupos

Um aspecto importante na interpretação da análise de agrupamento é a determinação do número de grupos que melhor descreva a real estrutura dos dados analisados (Amaral Júnior *et al.*, 2010).

A determinação do número de grupos é um parâmetro ajustável no programa desenvolvido. Porém, para Souza *et al.* (1997), há métodos que podem ser adotados para se determinar o número de grupos. Todavia, o que se faz comumente é utilizar vários números de grupos e, por algum critério de otimização, selecionar o mais conveniente, como, por exemplo, a análise discriminante, usada neste estudo para verificação da adequação dos agrupamentos obtidos.

Segundo a análise discriminante de Anderson, o melhor resultado, com relação aos agrupamentos, é apresentado quando são utilizados três grupos para a primeira época, segunda época e média das épocas, resultando em 96,00%, 93,75% e 93,75% de percentual de classificação correta em cada agrupamento, respectivamente (Quadro 11).

Observa-se que, quanto menor o número de grupos, maior é a confiabilidade do agrupamento. Essa afirmativa é justificada pelo fato de que a população é formada por indivíduos de apenas dois grupos de mamoeiros hermafroditas, o grupo Solo e o grupo Formosa. Ou seja, a classificação é feita entre indivíduos de uma mesma espécie.

Quadro 11. Resultado da análise discriminante de Anderson, segundo percentual de classificação correta dos grupos propostos pela rede neural artificial, para *Carica papaya* L.

Número de grupos da rede neural artificial	Percentual de classificação correta		
	Primeira época	Segunda época	Média das épocas
3	96,00	93,75	93,75
4	92,00	91,67	89,59
5	92,00	87,50	79,17
6	80,00	79,17	81,25
7	70,00	72,92	79,17
8	72,00	77,09	79,17
9	72,00	77,09	68,75
10	70,00	75,00	77,09

### 3.2. Primeira época

O resultado da classificação da rede neural artificial, obtido a partir de 9 caracteres quantitativos avaliados em 50 acessos, se encontra no Quadro 12. As entradas da rede neural foram compostas pelas médias originais. Foram estipulados quatro grupos para a classificação, que proporcionam a distinção das principais características dos acessos que formam os grupos.

A definição do número de grupos foi dada com base nos grupos propostos pela técnica de componentes principais. Dessa forma, foi estipulado um cenário mais apropriado para comparação dos métodos.

Quadro 12. Grupos dos acessos de mamoeiro (*Carica papaya* L.), na primeira época, formados através da rede neural artificial, a partir das médias originais, baseando-se em 9 caracteres quantitativos avaliados em 50 acessos.

Grupos	Acessos
1	1 - Caliman M5, 2 - Sunrise Solo 783, 3 - Costa Rica, 5 - Diva, 6 - Grampola, 7 - Sunrise Solo, 8 - Caliman AM, 9 - Caliman GB, 10 - Caliman SG, 11 - Caliman G, 12 - Sunrise Solo 72/12 (202), 13 - Kapoho Solo (polpa amarela), 14 - Baixinho Santa Amália, 15 - Sunrise Solo TJ, 18 - Kapoho Solo (polpa vermelha), 19 - Sunrise Solo (sel. Caliman), 22 - Mamão roxo, 26 - Baixinho Super, 28 - STZ-51, 29 - STZ-52, 33 - Golden tipo Formosa, 34 - STZ-03 - pecíolo curto, 47 - Golden robusto, 49 - STZ 63
2	16 - Tailândia, 25 - Sekati, 31 - JS 12 (206), 32 - Cariflora 209, 35 - STA Helena III Trat 12a plt 07X, 36 - STA Helena III Trat 11a plt 08X, 37 - STA Helena III Trat 50 plt 09X, 38 - STA Helena III Trat 02 plt 01X, 39 - STA Helena III Trat 14 plt 05X, 41 - STA Helena III Trat 04 plt 02X, 42 - STA Helena III Trat 12b plt 06X, 48 - Tainung
3	4 - Triwan Et, 17 - São Mateus, 20 - Waimanalo, 30 - Calimosa, 40 - STA Helena III Trat 15 plt 04X, 43 - Papaya 42 Formosa, 44 - Papaya 45 Formosa roxo, 45 - Papaya 46 claro, 46 - Fruto médio verde
4	21 - Mamão Bené, 23 - Maradol (orig. México), 24 - Maradol (grande limão), 27 - Americano, 50 - Sekati (fruto longo macuco)

O grupo 1 é composto por 17 acessos do grupo Solo e 7 acessos do grupo Formosa. É característico dos acessos que produzem frutos menores, com massa de fruto média de 395,90g, comprimento do fruto médio de 12,56cm, diâmetro do fruto médio de 8,39cm, espessura da polpa média de 2,40cm, conteúdo de sólidos solúveis médio de 11,18<sup>o</sup> Brix, firmeza externa do fruto média de 80,81N e firmeza interna da polpa média de 63,21N.

O grupo 2 é composto por 12 acessos do grupo Formosa. É característico dos frutos com massa de fruto média de 777,78g, comprimento do fruto médio de 18,51cm, diâmetro do fruto médio de 9,71cm, espessura da polpa média de 2,75cm, conteúdo de sólidos solúveis médio de 10,46<sup>o</sup> Brix, firmeza externa do fruto média de 163,11N e firmeza interna da polpa média de 123,06N.

O grupo 3 é composto por 8 acessos do grupo Formosa e 1 acesso do grupo Solo. É característico dos frutos com massa de fruto média de 1322,36g,

comprimento do fruto médio de 22,73cm, diâmetro do fruto médio de 11,08cm, espessura da polpa média de 3,09cm, conteúdo de sólidos solúveis médio de 9,32<sup>º</sup> Brix, firmeza externa do fruto média de 189,10N e firmeza interna da polpa média de 144,31N.

O grupo 4 é composto por 5 acessos do grupo Formosa. É característico dos acessos que produzem frutos maiores, com massa de fruto média de 2329,34g, comprimento do fruto médio de 32,39cm, diâmetro do fruto médio de 12,39cm, espessura da polpa média de 3,18cm, conteúdo de sólidos solúveis médio de 8,11<sup>º</sup> Brix, firmeza externa do fruto média de 185,26N e firmeza interna da polpa média de 146,34N.

Quanto à indicação de cruzamentos, a análise multivariada possibilita a predição da heterose, servindo como base para a recomendação de cruzamentos. Porém, as possibilidades são muitas, desde que se obedeça aos princípios de se cruzar os acessos mais distantes geneticamente e com as melhores características agronômicas desejadas.

### **3.3. Segunda época**

O resultado da classificação da rede neural artificial, obtido a partir de 9 caracteres quantitativos avaliados em 48 acessos, se encontra no Quadro 13. As entradas da rede neural foram compostas pelas médias originais. Foram estipulados quatro grupos para a classificação, que proporcionam a distinção das principais características dos acessos que formam os grupos.

A definição do número de grupos foi dada com base nos grupos propostos pela técnica de componentes principais. Dessa forma, foi estipulado um cenário mais apropriado para comparação dos métodos.

O grupo 1 é composto por 16 acessos do grupo Solo e 5 acessos do grupo Formosa. É característico dos acessos que produzem frutos menores, com massa de fruto média de 398,18g, comprimento do fruto médio de 13,26cm, diâmetro do fruto médio de 7,98cm, espessura da polpa média de 1,94cm, conteúdo de sólidos solúveis médio de 11,01<sup>º</sup> Brix, firmeza externa do fruto média de 70,56N e firmeza interna da polpa média de 51,90N.

Quadro 13. Grupos dos acessos de mamoeiro (*Carica papaya* L.), na segunda época, formados através da rede neural artificial, a partir das médias originais, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.

Grupos	Acessos
1	1 - Caliman M5, 4 - Triwan Et, 5 - Diva, 6 - Grampola, 7 - Sunrise Solo, 8 - Caliman AM, 9 - Caliman GB, 10 - Caliman SG, 11 - Caliman G, 12 - Sunrise Solo 72/12 (202), 13 - Kapoho Solo (polpa amarela), 14 - Baixinho Santa Amália, 15 - Sunrise Solo TJ, 18 - Kapoho Solo (polpa vermelha), 19 - Sunrise Solo (sel. Caliman), 22 - Mamão roxo, 25 - Baixinho Super, 26 - STZ-51, 31 - Golden tipo Formosa, 45 - Golden robusto, 47 - STZ 63
2	21 - Mamão Bené, 38 - STA Helena III Trat 15 plt 04X, 48 - Sekati (fruto longo macuco)
3	23 - Maradol (orig. México), 28 - Calimosa, 29 - JS 12 (206), 33 - STA Helena III Trat 12a plt 07X, 34 - STA Helena III Trat 11a plt 08X, 35 - STA Helena III Trat 50 plt 09X, 36 - STA Helena III Trat 02 plt 01X, 39 - STA Helena III Trat 04 plt 02X, 40 - STA Helena III Trat 12b plt 06X, 41 - Papaya 42 Formosa, 46 - Tainung
4	2 - Sunrise Solo 783, 3 - Costa Rica, 16 - Tailândia, 17 - São Mateus, 20 - Waimanalo, 24 - Sekati, 27 - STZ-52, 30 - Cariflora 209, 32 - STZ-03 - pecíolo curto, 37 - STA Helena III Trat 14 plt 05X, 42 - Papaya 45 Formosa roxo, 43 - Papaya 46 claro, 44 - Fruto médio verde

O grupo 2 é composto por 3 acessos do grupo Formosa. É característico dos acessos que produzem frutos maiores, com massa de fruto média de 2069,44g, comprimento do fruto médio de 28,76cm, diâmetro do fruto médio de 12,01cm, espessura da polpa média de 3,05cm, conteúdo de sólidos solúveis médio de 9,46<sup>º</sup> Brix, firmeza externa do fruto média de 198,78N e firmeza interna da polpa média de 151,20N.

O grupo 3 é composto por 11 acessos do grupo Formosa. É característico dos frutos com massa de fruto média de 1264,40g, comprimento do fruto médio de 23,23cm, diâmetro do fruto médio de 10,74cm, espessura da polpa média de 2,77cm, conteúdo de sólidos solúveis médio de 9,93<sup>º</sup> Brix, firmeza externa do fruto média de 179,11N e firmeza interna da polpa média de 141,01N.

O grupo 4 é composto por 11 acessos do grupo Formosa e 2 acessos do grupo Solo. É característico dos frutos com massa de fruto média de 734,10g,

comprimento do fruto médio de 17,64cm, diâmetro do fruto médio de 9,26cm, espessura da polpa média de 2,35cm, conteúdo de sólidos solúveis médio de 10,52º Brix, firmeza externa do fruto média de 132,86N e firmeza interna da polpa média de 97,58N.

Quanto à indicação de cruzamentos, a análise multivariada possibilita a predição da heterose, servindo como base para a recomendação de cruzamentos. Porém, as possibilidades são muitas, desde que se obedeça aos princípios de se cruzar os acessos mais distantes geneticamente e com as melhores características agronômicas desejadas.

### **3.4. Média das épocas**

O estudo da média das duas épocas foi feito com o objetivo de verificar o impacto na redistribuição dos acessos.

O resultado da classificação se encontra no Quadro 14. Foram avaliados 9 caracteres quantitativos em 48 acessos. As entradas da rede neural foram compostas pelas médias das duas épocas. Foram estipulados três grupos para a classificação, que proporcionam a distinção das principais características dos acessos que formam os grupos.

O grupo 1, formado pela maioria dos acessos, é composto por 21 acessos do grupo Solo, 6 acessos do grupo Formosa. É característico dos acessos desse grupo uma massa de fruto média de 445,22g, comprimento do fruto médio de 13,48cm, diâmetro do fruto médio de 8,40cm, espessura da polpa média de 2,23cm, conteúdo de sólidos solúveis médio de 11,02º Brix, firmeza externa do fruto média de 84,45N e firmeza interna da polpa média de 64,16N.

O grupo 2, foi composto por 7 acessos do grupo Formosa, os acessos que produzem os maiores frutos. Os acessos deste grupo apresentaram uma massa de fruto média de 1710,71g, comprimento do fruto médio de 25,68cm, diâmetro do fruto médio de 11,78cm, espessura da polpa média de 3,11cm, conteúdo de sólidos solúveis médio de 9,55º Brix, firmeza externa do fruto média de 186,99N e firmeza interna da polpa média de 143,23N.

Quadro 14. Grupos dos acessos de mamoeiro (*Carica papaya* L.) formados através da rede neural artificial, a partir das médias das épocas, baseando-se em 9 caracteres quantitativos avaliados em 48 acessos.

Grupos	Acessos
1	1 - Caliman M5, 2 - Sunrise Solo 783, 3 - Costa Rica, 4 - Triwan Et, 5 - Diva, 6 - Grampola, 7 - Sunrise Solo, 8 - Caliman AM, 9 - Caliman GB, 10 - Caliman SG, 11 - Caliman G, 12 - Sunrise Solo 72/12 (202), 13 - Kapoho Solo (polpa amarela), 14 - Baixinho Santa Amália, 15 - Sunrise Solo TJ, 18 - Kapoho Solo (polpa vermelha), 19 - Sunrise Solo (sel. Caliman), 22 - Mamão roxo, 25 - Baixinho Super, 26 - STZ-51, 27 - STZ-52, 31 - Golden tipo Formosa, 32 - STZ-03 - pecíolo curto, 43 - Papaya 46 claro, 44 - Fruto médio verde, 45 - Golden robusto, 47 - STZ 63
2	21 - Mamão Bené, 23 - Maradol (orig. México), 33 - STA Helena III Trat 12a plt 07X, 38 - STA Helena III Trat 15 plt 04X, 39 - STA Helena III Trat 04 plt 02X, 46 - Tainung, 48 - Sekati (fruto longo macuco)
3	16 - Tailândia, 17 - São Mateus, 20 - Waimanalo, 24 - Sekati, 28 - Calimosa, 29 - JS 12 (206), 30 - Cariflora 209, 34 - STA Helena III Trat 11a plt 08X, 35 - STA Helena III Trat 50 plt 09X, 36 - STA Helena III Trat 02 plt 01X, 37 - STA Helena III Trat 14 plt 05X, 40 - STA Helena III Trat 12b plt 06X, 41 - Papaya 42 Formosa, 42 - Papaya 45 Formosa roxo

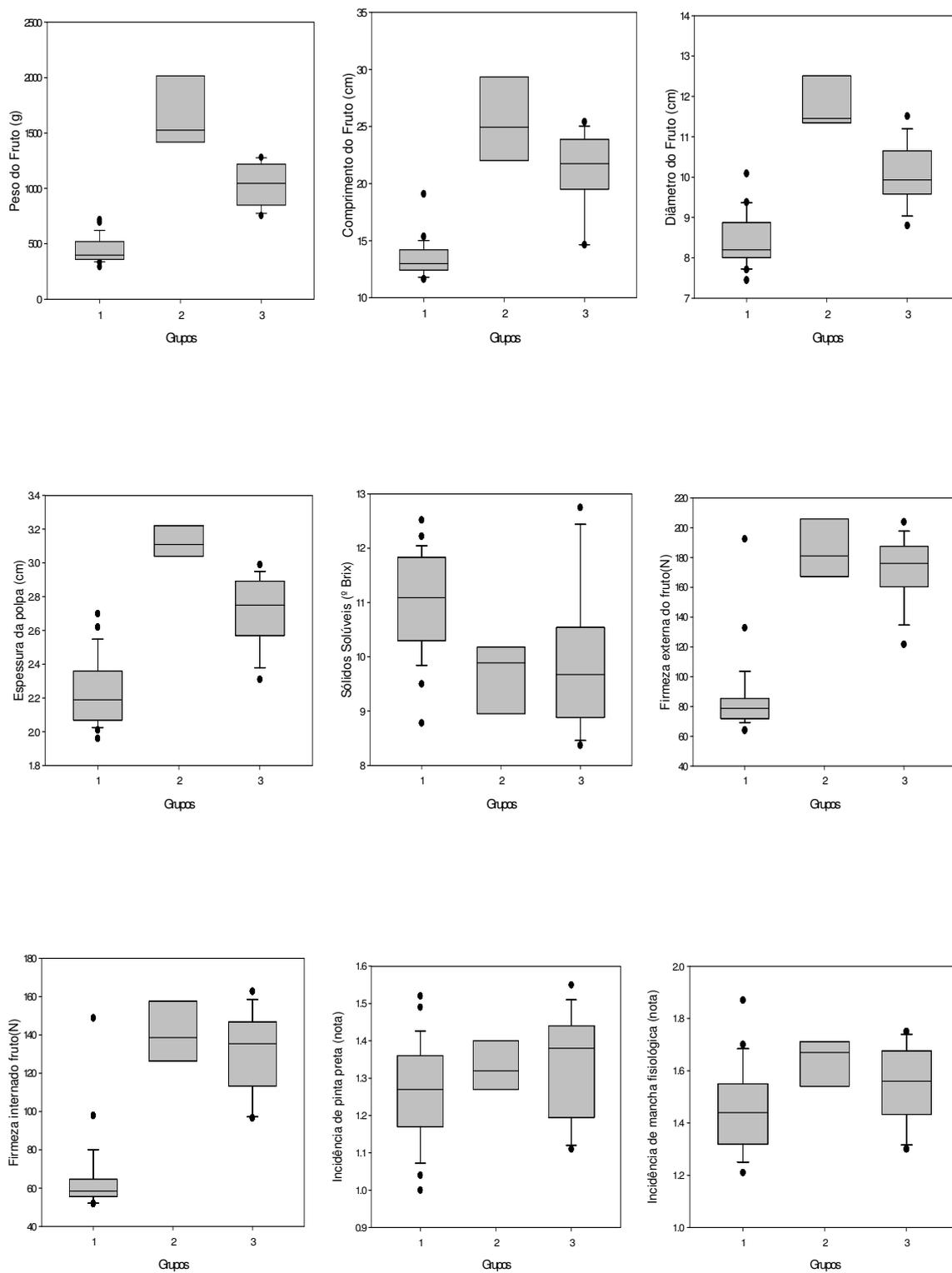
O grupo 3, foi composto por 1 acesso do grupo Solo e 13 acessos do grupo Formosa. É característico dos acessos com massa de fruto média de 1042,62g, comprimento do fruto médio de 21,18cm, diâmetro do fruto médio de 10,10cm, espessura da polpa média de 2,72cm, conteúdo de sólidos solúveis médio de 9,91<sup>º</sup> Brix, firmeza externa do fruto média de 172,44N e firmeza interna da polpa média de 131,74N.

Os grupos gerados pela rede neural artificial facilitam a seleção de material divergente para compor o melhoramento por meio da geração de híbridos, uma vez que permitem a seleção de genótipos pertencentes a grupos heteróticos distintos, a serem indicados para cruzamento.

### 3.5. Gráficos Boxplot

O gráfico Boxplot foi utilizado para uma melhor análise do agrupamento proposto pela rede neural artificial na média das épocas, conforme Figura 5.

Figura 5. Boxplot dos valores mínimos, máximos e mediana para os três grupos propostos pela rede neural artificial, na média das épocas, a partir de 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro (*Carica papaya L.*).



O Boxplot permite avaliar a simetria dos dados, sua dispersão e a existência ou não de *outliers* nos mesmos, sendo especialmente adequado para a comparação de dois ou mais conjuntos de dados correspondentes às categorias de uma característica.

Através da análise dos gráficos, pode-se observar que o grupo 1 possui os menores valores de medianas e as menores freqüências para todas as características, exceto SS, característica onde apresentou, inclusive, a maior mediana e as maiores freqüências. Ou seja, o grupo 1 tende a agrupar os indivíduos com as menores médias para todas as características, exceto SS, característica onde tende a agrupar os indivíduos com as maiores médias.

O grupo 1 agrupa todos os acessos do grupo Solo, exceto o acesso 17 - São Mateus que está presente no terceiro grupo, e alguns acessos do grupo Formosa. Porém, o referido acesso apresenta médias superiores às dos indivíduos do grupo Solo e os acessos do grupo Formosa apresentam médias características dos acessos do grupo Solo, justificando a composição do grupo. O grupo se apresenta homogêneo, reunindo os acessos que produzem frutos menores. Dessa forma, observa-se que para a característica MF, no grupo 1, a menor freqüência é justificável.

Essas afirmativas são corroboradas pela análise apresentada em Quintal (2009), quando considerado o agrupamento proposto pelo método UPGMA.

Dessa forma, ao se considerar o mercado consumidor, principalmente o mercado externo, que valoriza os frutos menores, o grupo 1 apresenta indivíduos com características importantes e que devem ser consideradas.

O grupo 2 possuiu os maiores valores de mediana e as maiores freqüências para a maioria das características, sendo: MF, CF, DF, EP, FEF, FIF e MM. Ou seja, o grupo 2 tende a agrupar os indivíduos com as maiores médias nas características citadas. Fato justificado pelo grupo 2 reunir os acessos do grupo Formosa que produzem frutos maiores.

Com relação à característica PP, o grupo 3 apresentou a maior mediana e o grupo 1 apresentou a menor mediana, se destacando dos demais. Porém, o grupo 1 apresentou valores *outliers*, indicando a presença de indivíduos com valores acima do limite máximo para a característica em questão. Dessa forma, o grupo 1 pode ser considerado quando a característica de interesse é a maior resistência à pinta preta, porém deve-se ter o cuidado de não escolher os

indivíduos que apresentaram valores *outliers* altos (8 - Caliman AM), dando preferência aos indivíduos que apresentaram valores *outliers* baixos (14 - Baixinho Santa Amália).

Os *outliers* são valores extremamente altos ou baixos. A existência de *outliers* pode tanto indicar dados incorretos, como dados válidos que carecem de uma atenção especial.

Com relação à característica MM, o grupo 2 apresentou a maior mediana e o grupo 1 apresentou a menor mediana, se destacando dos demais. Porém, o grupo 1 apresentou valores *outliers*, indicando a presença de indivíduos com valores acima do limite máximo para a característica em questão. Dessa forma, o grupo 1 pode ser considerado quando a característica de interesse é a maior resistência à mancha fisiológica, porém deve-se ter o cuidado de não escolher os indivíduos que apresentaram valores *outliers* altos (2 - Sunrise Solo 783), dando preferência aos indivíduos que apresentaram valores *outliers* baixos (10 - Caliman SG).

O grupo 2 foi o único que não apresentou valores *outliers* em todas as características estudadas. A uniformidade apresentada se justifica pelo grupo ser formado por apenas 7 acessos do grupo Formosa.

Através do gráfico também é possível analisar a heterogeneidade e a homogeneidade intragrupos, através da dispersão apresentada. Quanto maior a dispersão observada no gráfico, maior é a heterogeneidade entre os indivíduos do grupo, para os valores da característica em questão.

Segundo Cruz et al. (2004), as técnicas de análise de agrupamento, como as redes neurais artificiais, por exemplo, têm por objetivo dividir um grupo original de observações em vários grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos.

O grupo 1 apresentou a maior homogeneidade intragrupos para as características MF, CF, DF, FEF e FIF. O grupo 2 apresentou a maior homogeneidade intragrupos para as características EP, SS, PP e MM.

As maiores dispersões de valores, distribuídas ao longo das médias, indicando maior heterogeneidade intragrupos, foram apresentadas no grupo 2, para as características MF, CF, DF e FEF, e no grupo 3, para as características EP, SS, FIF, PP e MM.

Essa análise auxilia o melhorista no processo de escolha de indivíduos para uso em programas de melhoramento, em futuras hibridações, na tentativa de obtenção de indivíduos com as características de interesse.

## **4. Análise discriminante de Anderson**

### **4.1. Primeira época**

Esse procedimento foi adotado partindo do princípio de que se conhecia a qual grupo pertenciam os acessos. Dessa forma, a consistência do agrupamento foi verificada através da análise discriminante de Anderson, conforme descrito por Cruz e Carneiro (2003).

As tabelas que contêm essas informações devem ser analisadas da seguinte maneira: na diagonal principal encontra-se a porcentagem de classificação correta para cada um dos grupos. Todos os demais campos correspondem à classificação incorreta. Para verificar a porcentagem de classificação incorreta para um determinado grupo, deve-se observar a linha da tabela.

Ao se considerar o agrupamento proposto pela análise por componentes principais, a taxa de erro aparente, de acordo com a análise discriminante de Anderson, foi de 8,00%. Ou seja, com base na análise discriminante de Anderson 92,00% dos acessos foram classificados corretamente nos grupos previamente definidos.

O Quadro 15 apresenta a porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson, com base na classificação proposta pela técnica de componentes principais.



Ao se considerar o agrupamento proposto pela rede neural artificial, a taxa de erro aparente, de acordo com a análise discriminante de Anderson, foi de 8,00%. Ou seja, com base na análise discriminante de Anderson 92,00% dos acessos foram classificados corretamente nos grupos previamente definidos.

O Quadro 17 apresenta a porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson, com base na classificação proposta pela rede neural artificial.

Quadro 17. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela rede neural artificial, considerando 9 caracteres quantitativos avaliados em 50 acessos de mamoeiro (*Carica papaya* L.), na primeira época.

Pop	1	2	3	4
1	100,00	0,00	0,00	0,00
2	0,00	91,67	8,33	0,00
3	11,11	22,22	66,67	0,00
4	0,00	0,00	0,00	100,00

## 4.2. Segunda época

Esse procedimento foi adotado partindo do princípio de que se conhecia a qual grupo pertenciam os acessos. Dessa forma, a consistência do agrupamento foi verificada através da análise discriminante de Anderson, conforme descrito por Cruz e Carneiro (2003).

As tabelas que contêm essas informações devem ser analisadas da seguinte maneira: na diagonal principal encontra-se a porcentagem de classificação correta para cada um dos grupos. Todos os demais campos correspondem à classificação incorreta. Para verificar a porcentagem de classificação incorreta para um determinado grupo, deve-se observar a linha da tabela.

Ao se considerar o agrupamento proposto pela análise por componentes principais, a taxa de erro aparente, de acordo com a análise discriminante de Anderson, foi de 12,50%. Ou seja, com base na análise discriminante de

Anderson 87,50% dos acessos foram classificados corretamente nos grupos previamente definidos.

O Quadro 18 apresenta a porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson, com base na classificação proposta pela técnica de componentes principais.

Quadro 18. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela técnica de componentes principais, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro (*Carica papaya* L.), na segunda época.

<b>Pop</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
1	92,00	8,00	0,00	0,00
2	22,22	77,78	0,00	0,00
3	0,00	18,18	81,82	0,00
4	0,00	0,00	0,00	100,00

Ao se considerar o agrupamento proposto pelo método de otimização Tocher, a taxa de erro aparente, de acordo com a análise discriminante de Anderson, foi de 18,75%. Ou seja, com base na análise discriminante de Anderson 81,25% dos acessos foram classificados corretamente nos grupos previamente definidos.

O Quadro 19 apresenta a porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson, com base na classificação proposta pelo método de otimização Tocher.

Quadro 19. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pelo método de otimização Tocher, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro (*Carica papaya* L.), na segunda época.

Pop	1	2	3	4	5	6	7	8	9	10
1	83,33	0,00	0,00	0,00	4,17	0,00	12,50	0,00	0,00	0,00
2	0,00	72,73	0,00	18,18	9,09	0,00	0,00	0,00	0,00	0,00
3	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
4	33,33	0,00	0,00	66,67	0,00	0,00	0,00	0,00	0,00	0,00
5	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
6	0,00	50,00	0,00	0,00	0,00	50,00	0,00	0,00	0,00	0,00
7	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00
9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00

O Quadro 20 apresenta a porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson, com base na classificação proposta pela rede neural artificial.

Quadro 20. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela rede neural artificial, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro (*Carica papaya* L.), na segunda época.

Pop	1	2	3	4
1	90,48	0,00	0,00	9,52
2	0,00	100,00	0,00	0,00
3	0,00	0,00	100,00	0,00
4	15,38	0,00	0,00	84,62

Ao se considerar o agrupamento proposto pela rede neural artificial, a taxa de erro aparente, de acordo com a análise discriminante de Anderson, foi de

8,33%. Ou seja, com base na análise discriminante de Anderson 91,67% dos acessos foram classificados corretamente nos grupos previamente definidos.

Resultado superior ao obtido pelo agrupamento proposto pela técnica de componentes principais e pelo método de otimização Tocher.

### **4.3. Média das épocas**

Esse procedimento foi adotado partindo do princípio de que se conhecia a qual grupo pertenciam os acessos. Dessa forma, a consistência do agrupamento foi verificada através da análise discriminante de Anderson, conforme descrito por Cruz e Carneiro (2003).

As tabelas que contêm essas informações devem ser analisadas da seguinte maneira: na diagonal principal encontra-se a porcentagem de classificação correta para cada um dos grupos. Todos os demais campos correspondem à classificação incorreta. Para verificar a porcentagem de classificação incorreta para um determinado grupo, deve-se observar a linha da tabela.

Ao se considerar o agrupamento proposto pela análise por componentes principais, a taxa de erro aparente, de acordo com a análise discriminante de Anderson, foi de 12,50%. Ou seja, com base na análise discriminante de Anderson 87,50% dos acessos foram classificados corretamente nos grupos previamente definidos.

O Quadro 21 apresenta a porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson, com base na classificação proposta pela técnica de componentes principais.



Ao se considerar o agrupamento proposto pela rede neural artificial, a taxa de erro aparente, de acordo com a análise discriminante de Anderson, foi de 6,25%. Ou seja, com base na análise discriminante de Anderson, 93,75% dos acessos foram classificados corretamente nos grupos previamente definidos.

O Quadro 23 apresenta a porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson, com base na classificação proposta pela rede neural artificial.

Quadro 23. Porcentagem de classificação correta e incorreta de cada grupo, obtida mediante a análise discriminante de Anderson (1958), com base na classificação proposta pela rede neural artificial, considerando 9 caracteres quantitativos avaliados em 48 acessos de mamoeiro (*Carica papaya* L.), na média das épocas.

<b>Pop</b>	<b>1</b>	<b>2</b>	<b>3</b>
1	96,30	0,00	3,70
2	0,00	85,71	14,29
3	0,00	7,14	92,86

## DISCUSSÃO

Os recursos genéticos de fruteiras tropicais estão sendo utilizados em programas de melhoramento genético e pesquisas correlatas, aproveitando-se o fato de o Brasil ser um dos mais importantes centros de diversidade genética de várias espécies frutíferas tropicais (Ferreira, 1997). Os trabalhos com mamoeiro nesta área têm sido relativamente recentes, verificando-se que, pela estreita base genética, com poucos genótipos disponíveis, as possibilidades de melhoramento genético da cultura tornam-se restritas (Pereira *et al.*, 2010; Quintal *et al.*, 2007a, 2007b).

Apesar da facilidade de cultivo e do grande consumo de mamão, a cultura tem problemas relacionados com pragas, doenças, características agrônômicas e reduzida disponibilidade de genótipos melhorados para exploração comercial (Pereira *et al.*, 2010; Foltran *et al.*, 1993), com produtividades limitadas. Com a perspectiva de aumento de produtividade, o melhoramento genético do mamoeiro requer melhor aproveitamento dos métodos de melhoramento, por meio de seleção de variedades agronomicamente superiores e síntese de linhagens e, ou, híbridos resistentes às doenças e pragas, com características comerciais favoráveis (Dantas *et al.*, 2002).

Em estudos considerando a mesma população avaliada no presente trabalho (Quintal *et al.*, 2007a), foram estimados os parâmetros genéticos (variâncias), observando-se variabilidade fenotípica, para todos os caracteres em

questão, confirmando, dessa forma, a possibilidade de identificação de genótipos superiores.

Em Quintal (2009) realizou-se a análise de variância univariada. Os resultados indicam, para todas as características, a presença de variabilidade entre os acessos.

Quintal *et al.* (2007b) também evidenciaram a presença de variabilidade genética entre os mesmos acessos, observando diferença significativa para o efeito de genótipos, nos mesmos caracteres estudados no presente trabalho.

Com relação à técnica de componentes principais, o gráfico permite a identificação dos genótipos com as características mais divergentes, proporcionando a noção da variabilidade genética presente na população estudada. Os genótipos mais discrepantes para as características estudadas são os de maior interesse em programas de inter cruzamentos (Amaral Júnior e Thiébaud, 1999). Porém, não é suficiente que os genótipos de interesse sejam distantes geneticamente, eles precisam apresentar boas médias para as características de interesse.

Os acessos que apresentaram maior dispersão dos escores podem ser utilizados em programas de melhoramento, em futuras hibridações, na tentativa de obtenção de materiais superiores.

Através da análise de componentes principais, verificou-se que a variação acumulada distribuiu-se entre os dois ou três primeiros componentes, ou seja, retendo o máximo de variação nos componentes iniciais, semelhante ao que vem sendo observado em diversos trabalhos.

Gomes Filho (2009) estudando a diversidade genética existente entre acessos de goiabeira, encontrou 85,00% da variação total nos três primeiros componentes principais. Em pequi, Oliveira *et al.* (2009) encontram 90,70% da variação total nos dois primeiros componentes. Moreira *et al.* (2009) encontram 90,20% de variação total nos três primeiros componentes avaliando a divergência de cultivares de feijão-vagem. Dias *et al.* (1994), observando alguns trabalhos dessa natureza, atribuíram a dispersão da variação, em vários ou poucos componentes, ao tipo de descritor utilizado.

O método de otimização de Tocher apresenta uma inconveniência no agrupamento dos genótipos com maior dissimilaridade. Na maioria dos casos, cada genótipo forma um grupo específico, de apenas um genótipo, em virtude do

agrupamento ser influenciado pela distância dos genótipos já agrupados. Outro problema é a formação do primeiro grupo com um número elevado de genótipos.

O método de Tocher modificado (seqüencial) proposto por Vasconcelos *et al.* (2007), difere do original pelo fato de ser adotado um critério diferenciado para a inclusão de indivíduos em cada grupo que é formado, podendo ser considerado uma alternativa para contornar o problema citado.

Segundo Pereira *et al.* (2009), quanto ao número de grupos, o método de Tocher seqüencial apresenta, de forma geral, um número menor, já que este método não recebe influência de indivíduos já agrupados, no agrupamento de indivíduos mais distantes.

Porém, ao se analisar os dados com base no método de Tocher modificado, a formação dos grupos se manteve praticamente a mesma, apresentando a redução de um ou dois grupos nas épocas estudadas.

A formação destes grupos é de fundamental importância na escolha dos progenitores, pois as novas combinações híbridas a serem estabelecidas devem ser baseadas na magnitude de suas dissimilaridades e no potencial *per se* dos genitores (Peluzio *et al.*, 2009).

As cultivares reunidas em grupos mais distantes dá um indicativo de serem dissimilares, podendo ser consideradas como promissoras em cruzamentos artificiais. Entretanto, além de dissimilares, é necessário que os genitores associem média elevada e variabilidade nas características que estejam sendo melhoradas.

Cruz *et al.* (2004) sugerem o não envolvimento de indivíduos de mesmo padrão de dissimilaridade nos cruzamentos, de modo a não restringir a variabilidade genética e, assim, evitar reflexos negativos nos ganhos a serem obtidos pela seleção.

Diversos trabalhos com aplicação do método de otimização de Tocher, também afirmam que o mesmo auxilia o melhorista na seleção de material divergente para compor o melhoramento por meio da geração de híbridos (Arnhold *et al.*, 2010; Coimbra *et al.*, 2010; Cardoso *et al.*, 2009b; Zuin *et al.*, 2009; Marim *et al.*, 2009; Barelli *et al.*, 2009; Peluzio *et al.*, 2009; Bertini *et al.*, 2009).

Corroborando a possibilidade de escolha de progenitores, em programas de melhoramento genético, visando à obtenção de genótipos superiores em

populações de *Carica papaya* L., Cardoso *et al.* (2009a) observaram elevada divergência genética para atributos relacionados à qualidade de sementes de mamoeiro, e estudos publicados por Silva *et al.* (2008, 2007b, 2007d) também indicam a disponibilidade de variabilidade genética entre os acessos de *Carica papaya* L.

Em relação aos grupos formados pela estimativa da divergência genética entre genótipos na população de estudo, deve-se considerar as possibilidades existentes para o caso de plantas perenes com propagação sexuada, com a possibilidade de obtenção de linhagens via autofecundação, e ou, obtenção de novos híbridos.

Em relação à indicação de cruzamentos, Dias *et al.* (1994) explicam que em estudos envolvendo elevado número de genótipos, surge a possibilidade de considerar-se não o genótipo individualmente, mas sim o grupo de genótipos. Nesse caso, os cruzamentos indicados são preferencialmente entre genótipos pertencentes aos grupos que apresentam as maiores distâncias intergrupos, e com características agrônômicas de interesse.

Para a indicação de novos cruzamentos, a fim de se explorar a heterose e a diversidade de alelos presentes, deve-se levar em consideração também o desempenho agrônômico dos genótipos, devendo-se indicar cruzamentos entre os genótipos mais divergentes e com melhor desempenho agrônômico para as características estudadas, tendo em vista a possibilidade de melhor complementação alélica e conseqüente melhora no desempenho dos genótipos, nas próximas gerações.

Outra possibilidade é recomendar diretamente genótipos promissores para novas avaliações em conjunto com outros acessos, para se aferir as possibilidades de uso dos mesmos como novas variedades a serem recomendadas.

A utilização das redes neurais como métodos de classificação é um caminho promissor. Esses classificadores têm a vantagem de ser não-paramétricos, necessitar de pequenas amostras para treinamento (Kavzoglu e Mather, 2003) e ser tolerantes a dados falhos (Bishop, 1995).

No desenvolvimento da rede neural, foram testadas diferentes bases de dados para alimentação da camada de entrada da rede. Médias originais e médias padronizadas, através de diversas fórmulas propostas pela literatura.

Tendo em vista que, no início do treinamento, os pesos sinápticos são gerados aleatoriamente e que esses valores iniciais poderiam influenciar o resultado final do treinamento, foram utilizados dois tipos de pesos sinápticos. Foram utilizados pesos aleatórios e pesos representando as médias dos grupos propostos pela técnica de componentes principais, não registrando diferença na classificação.

Observou-se que a rede neural não é influenciada pela escala dos dados de entrada ou dos vetores de peso. A utilização dos dados originais resultou na mesma classificação proposta quando dados padronizados, através das diversas fórmulas propostas pela literatura, foram utilizados, contrariando Braga (2000) que, segundo trabalho publicado, afirma que a utilização de vetores de pesos com magnitudes diferentes prejudica a competição entre os nodos, por desviar a atenção da rede para a magnitude dos mesmos.

Fator importante, e que deve ser ressaltado, é a necessidade dos dados serem bem característicos dos grupos. Quando os dados se apresentam mais heterogêneos, caracterizando os indivíduos com relação aos seus grupos, a rede neural tende ao seu melhor desempenho. Estes argumentos são relevantes e demonstram a efetividade da utilização da técnica das redes neurais na avaliação da diversidade genética em recursos genéticos, para fins de utilização em programas de melhoramento de diversas culturas.

Quanto à classificação dos acessos, as redes neurais artificiais e a técnica de componentes principais se mostraram métodos similares.

Estudos publicados de redes neurais artificiais com algoritmo de aprendizado supervisionado, como exemplo, Hirayama (2004), demonstram resultados satisfatórios com acertos próximos de 100%. Logo, presume-se que em redes neurais artificiais com algoritmo de aprendizado não supervisionado, a garantia de sucesso é assegurada, obedecendo-se rigorosamente as características que regem a metodologia.

Com relação à análise discriminante de Anderson, sua utilização foi considerada viável, como um método para verificar a consistência do agrupamento proposto pelas técnicas aqui avaliadas. Sudre *et al.* (2006) também afirmaram que a função discriminante de Anderson demonstrou ter um grande potencial de uso, na caracterização e manejo de bancos de germoplasma, visto

ser uma ferramenta adicional para se aferir a correta classificação fornecida pelos vários métodos de análise multivariada utilizados.

Com relação à margem dos valores encontrados para a taxa de erro aparente, Mariot *et al.* (2008), trabalhando com acessos de *M. ilicifolia* e *M. aquifolium*, encontraram taxa de erro aparente de 10,48%, quando os acessos foram agrupados pelo método de otimização de Tocher. Valor considerado adequado pelos autores e similar aos encontrados no presente trabalho.

A consistência de agrupamentos de espécies do gênero *Brachiaria*, foi verificada por Assis *et al.* (2003), através da análise discriminante de Anderson, onde taxas de erro aparente foram encontradas com valores elevados, como por exemplo, 69,33% para *B. brizantha*.

## RESUMO E CONCLUSÕES

A espécie *Carica papaya* L. é amplamente cultivada. Porém, a cultura tem problemas relacionados com a reduzida disponibilidade de genótipos melhorados para exploração comercial (Pereira *et al.*, 2010; Foltran *et al.*, 1993). O estudo da diversidade genética é de fundamental importância na seleção preliminar de acessos com características superiores e a utilização desses materiais com sucesso em programas de melhoramento genético (Dantas *et al.*, 2002).

Dessa forma, o estudo foi realizado com o objetivo geral de avaliar estratégias de análise da diversidade genética em *Carica papaya* L., a partir de acessos do gênero mantidos no banco de germoplasma localizado na Fazenda Caliman Agrícola S/A, no município de Linhares, no Estado do Espírito Santo, em duas diferentes épocas do ano. Nove caracteres quantitativos foram considerados em cinquenta acessos, utilizando-se o delineamento em blocos casualizados com 2 repetições. Foi empregada a análise multivariada por meio da técnica de componentes principais e por meio de métodos de agrupamento (Tocher e redes neurais artificiais). A consistência do agrupamento foi verificada através da análise discriminante de Anderson.

Observou-se na presente população de estudo, a presença de diversidade genética entre os acessos avaliados, possibilitando a seleção de novos indivíduos com características superiores e favoráveis para a obtenção de novas cultivares de mamão, prevendo progresso genético.

Verificou-se que a técnica de componentes principais foi capaz de estabelecer classificações adequadas da diversidade genética, identificando claramente os grupos. As informações resultantes são úteis no entendimento da diversidade dentro da coleção e como a mesma está estruturada.

Os grupos gerados pelo método de otimização de Tocher facilitam a seleção de material divergente para compor o melhoramento por meio da geração de híbridos, uma vez que permitem a seleção de genótipos pertencentes a grupos heteróticos distintos, a serem indicados para cruzamento.

A técnica de rede neural artificial se demonstrou viável na classificação dos acessos e na avaliação da diversidade genética em recursos genéticos, para fins de utilização em programas de melhoramento da cultura.

A utilização da análise discriminante de Anderson foi considerada viável, como um método para verificar a consistência do agrupamento proposto pelas técnicas utilizadas nesse trabalho.

Dessa forma, pode-se concluir, de acordo com os valores das taxas de erro aparente, que as técnicas de agrupamento foram adequadas na identificação da variabilidade genética entre os acessos estudados.

Segundo a análise discriminante de Anderson, a rede neural artificial demonstrou maior consistência nos agrupamentos na segunda época e na média das épocas, superando os resultados apresentados pela técnica de componentes principais e pelo método de otimização de Tocher.

Considerando as sugestões mencionadas na análise dos resultados, vale ressaltar que a utilização de determinado acesso vai depender do objetivo de cada programa de melhoramento. Contudo, como afirmado por Valois *et al.* (1998), o mais importante é que uma vez avaliado o germoplasma e os dados disponibilizados, sua utilização fica mais acessível e com maior possibilidade de sucesso.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Abreu, F. B., Leal, N. R., Rodrigues, R., Amaral Júnior, A. T., Silva, D. J. H. (2004) Divergência genética entre acessos de feijão-de-vagem de hábito de crescimento indeterminado. *Horticultura Brasileira*, Brasília, 22 (3): 547-552.
- Amaral Júnior, A. T. do, Thiébaud, J. T. de L. (1999) *Análise multivariada na avaliação da diversidade em recursos genéticos vegetais*. Apostila. Campos dos Goytacazes: UENF, 55p.
- Amaral Júnior, A. T. do, Viana, A. P., Gonçalves, L. S. A., Barbosa, C. D. (2010) Recursos genéticos vegetais e o melhoramento de plantas. *In: Pereira, T. N. S. (ed.) Germoplasma: conservação, manejo e uso no melhoramento de plantas*. Viçosa: Arca, 250p.
- Anderson, T. W. (1958) *An introduction to multivariate statistical analysis*. New York: John Wiley, 221p.
- Andreotti, C. M. (1998) *A cultura do mamão*. Brasília: EMBRAPA-SPI, 80p.
- Arnhold, E., Silva, R., Viana, J. (2010) Seleção de linhagens S5 de milho-pipoca com base em desempenho e divergência genética. *Acta Scientiarum. Agronomy*, Maringá, 32 (2): 279-283.

- Arriel, E. F., Paula, R. C., Rodrigues, T. J. D., Bakke, A. O., Arriel, N. H. C. (2006) Divergência genética entre progênies de *Cnidoscolus phyllacanthus* submetidas a três regimes hídricos. *Científica*, 34 (2): 229-237.
- Assis, G. M. L., Euclides, R. F., Cruz, C. D., Valle, C. B. do (2003) Discriminação de espécies de *Brachiaria* baseada em diferentes grupos de caracteres morfológicos. *Revista Brasileira de Zootecnia*, Viçosa, 32 (3): 576-584.
- Awada, M. (1953) Effects of moisture on yield and sex expression of the papaya plants (*Carica papaya* L.). *Agr. Expt. Sta. Prog.* Hawaii: 97.
- Badillo, V. M. (1971) *Monografía de la familia Caricaceae*. Venezuela: Editorial Nuestra America C. A., 221p.
- Badillo, V. M. (1993) Caricaceae: segundo esquema. *Revista de la Facultad de Agronomía*, Venezuela, 43: 1-111.
- Badillo, V. M. (2000) *Carica* L. vs *Vasconcella* St. Hil. (Caricaceae) con la rehabilitación de este ultimo. *Ernstia*, 10: 74-79.
- Barelli, M. A. A., Gonçalves-Vidigal, M. C., Vidigal Filho, P. S., Neves, L. G., Silva, H. T. da (2009) Divergência genética em cultivares tradicionais de feijão do Estado de Mato Grosso do Sul. *Semina: Ciências Agrárias*, Londrina, 30 (1): 1061-1072.
- Barros, A. M., Faleiro, F. G., Karia, C. T., Shiratsuchi, L. S., Andrade, R. P. de, Lopes, G. K. B. (2005) Variabilidade genética e ecológica de *Stylosanthes macrocephala* determinadas por RAPD e SIG. *Pesquisa Agropecuária Brasileira*, Brasília, 40 (9): 899-909.
- Benite, M. (2003) *Aplicação de modelos de redes neurais na elaboração e análise de cenários macroeconômicos*. Dissertação (Mestrado em Engenharia), São Paulo, Escola de Engenharia de São Carlos – Universidade de São Paulo, 123p.
- Bertini, C. H. C. de M., Teófilo, E. M., Dias, F. T. C. (2009) Divergência genética entre acessos de feijão-caupi do banco de germoplasma da UFC. *Revista Ciência Agronômica*, 40 (1): 99-105.

- Bishop, C. M. (1995) *Neural networks for pattern recognition*. Oxford, New York: Oxford University Press, 482p.
- Borém, A., Miranda, G. V. (2005) *Melhoramento de Plantas*. Viçosa: UFV, 525p.
- Braga, A. P., Carvalho, A. P. L. F., Ludemir, T. B. (2000) *Redes neurais artificiais: teoria e aplicações*. Rio de Janeiro: LTC, vol. 1, 250p.
- Bruckner, C. H. (ed.) (2002) *Melhoramento de fruteiras tropicais*. Viçosa: UFV, 422p.
- Cardoso, D. L., Silva, R. F. da, Pereira, M. G., Viana, A. P., Araújo, E. F. (2009a) Diversidade genética e parâmetros genéticos relacionados à qualidade fisiológica de sementes em germoplasma de mamoeiro. *Rev. Ceres*, Viçosa, 56 (5): 572-579.
- Cardoso, W. S., Paes, M. C. D., Galvão, J. C. C., Rios, S. de A., Guimarães, P. E. de O., Schaffert, R. E., Borém, A. (2009b) Variabilidade de genótipos de milho quanto à composição de carotenoides nos grãos. *Pesquisa Agropecuária Brasileira*, Brasília, 44 (2): 164-173.
- Carvalho, L. P. de, Lanza, M. A., Fallieri, J., Santos, J. W. dos (2003) Análise da diversidade genética entre acessos de banco ativo de germoplasma de algodão. *Pesquisa Agropecuária Brasileira*, Brasília, 38 (10): 1149-1155.
- Cattaneo, L. F. (2001) *Avaliação da divergência genética e análise de gerações de mamoeiro (Carica papaya L.)*. Tese (Doutorado em Produção Vegetal), Campos dos Goytacazes – RJ, Universidade Estadual do Norte Fluminense – UENF, 95p.
- Cattaneo, L. F. (2003) Melhoramento genético do mamoeiro assistido por marcadores moleculares. In: Martins, D. dos S. (ed.) *Papaya Brasil: Qualidade do Mamão para o Mercado Interno*. Vitória: Incaper, 191-208.
- Cattaneo, L. F., Costa, A. F. S., Pereira, M. G., Santana, E. N. (2007) Produtividade de cultivares de mamoeiro do grupo Formosa no Estado do Espírito Santo. In: Martins, D. dos S., Costa, A. N. da, Costa, A. de F. S. da

- (eds). *Papaya Brasil: manejo, qualidade e mercado do mamão*, Vitória – ES, DCM/INCAPER, 1: 366-368.
- Chagas, C. da S., Vieira, C. A. O., Fernandes Filho, E. I., Carvalho Junior, W. (2009) Utilização de redes neurais artificiais na classificação de níveis de degradação em pastagens. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 13 (3): 319-327.
- Chagas, C. S., Fernandes Filho, E. I., Vieira, C. A. O., Carvalho Júnior, W. (2007) Utilização de redes neurais artificiais para predição de classes de solo em uma bacia hidrográfica no Domínio de Mar de Morros. *XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, INPE: 2421-2428.
- CIB (2004) O biólogo das telas de computador. *BioTech: Informação Científica sobre Biotecnologia*, São Paulo, 7: 4. Página mantida pelo Conselho de Informações sobre Biotecnologia - CIB. Disponível em <http://www.cib.org.br>, acesso em 05/2009.
- Coimbra, R. R., Miranda, G. V., Cruz, C. D., Melo, A. V. de, Eckert, F. R. (2010) Caracterização e divergência genética de populações de milho resgatadas do Sudeste de Minas Gerais. *Rev. Ciênc. Agron.*, 41 (1): 159-166.
- Costa, F. H. dos S., Souza Filho, C. R. de (2008) Aplicação de redes neurais artificiais para reconhecimento de padrões em solos. *XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, INPE: 5177-5183.
- Cruz, C. D. (1990) *Aplicação de algumas técnicas multivariadas no melhoramento de plantas*. Tese (Doutorado em Agronomia), Piracicaba, Universidade de São Paulo, 188p.
- Cruz, C. D. (2006) *Programa Genes: estatística experimental e matrizes*. Viçosa: UFV, 285p.
- Cruz, C. D., Carneiro, P. C. S. (2003) *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: UFV, vol. 2, 585p.
- Cruz, C. D., Regazzi, A. J., Carneiro, P. C. S. (2004) *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: UFV, vol. 1, 480p.

- Dantas, J. L. L. (2000a) Introdução. *In: Trindade, V. T. (org) Mamão. Produção: aspectos técnicos*. Brasília: Embrapa, 9.
- Dantas, J. L. L. (2000b) Cultivares. *In: Trindade, V. T. (org) Mamão. Produção: aspectos técnicos*. Brasília: Embrapa, 15-17.
- Dantas, J. L. L., Dantas, A. C. V. L., Lima, J. F. de (2002) Mamoeiro. *In: Bruckner, C. H. (ed.) Melhoramento de fruteiras tropicais*. Viçosa: UFV, 309-349.
- Dantas, J. L. L., Junghans, D. T., Lima, J. F. de (eds) (2003) *Mamão: o produtor pergunta, a Embrapa responde*. Brasília: Embrapa, 151p.
- Dantas, J. L. L., Morales, C. F. (1996) Caracterização de germoplasma de mamão (*Carica spp.*). *Revista Brasileira de Fruticultura*, Cruz das Almas, 18 (2): 295-309.
- Dantas, J. L. L., Neto, M. T. de C. (2000) Aspectos botânicos e fisiológicos. *In: Trindade, V. T. (org) Mamão. Produção: aspectos técnicos*. Brasília: Embrapa, 11-14.
- Dias, L. A. dos S., Kageyama, P. Y., Castro, G. C. T. (1994) Divergência genética multivariada na preservação de germoplasma de cacau (*Theobroma cacao L.*). *Agrotrópica*, 9: 29-40.
- Dutra, R. M. O., Sperandio, M., Coelho, J. (2009) O método Ward de agrupamento de dados e sua aplicação em associação com os Mapas Auto-Organizáveis de Kohonen. Disponível em [www.labplan.ufsc.br/sperandio/ward22\\_final.pdf](http://www.labplan.ufsc.br/sperandio/ward22_final.pdf), acesso em 06/2009.
- Fehr, W. R. (1987) *Principles of cultivar development*. New York: Macmillan, 536p.
- Ferreira, F. R. (1997) Recursos genéticos de fruteiras tropicais e subtropicais no Brasil. *Simpósio Brasileiro de Melhoramento de Frutíferas*, Jaboticabal: FCAVJ/UNESP, 1: 5.
- Foltran, D. E., Gonçalves, P. D. S., Sabino, J. C., Vilela, R. C. F. (1993) Estimates of genetic and phenotypic parameters in papaya. *Bragantia*, Campinas, 52 (1): 7-15.

- França, M. M. (2010) análise do uso da terra no município de Viçosa-MG mediado por classificações supervisionadas com redes neurais artificiais e Maxver. *Revista Brasileira de Geografia Física, América do Norte*, 2 (3): 92-101.
- Galvão, C. O., Valença, M. J. S., Vieira, V. P. P. B., Diniz, L. S., Lacerda, E. G. M., Carvalho, A. C. P. L. F, Ludermir, T. B. (1999) *Sistemas inteligentes: Aplicações a recursos hídricos e ciências ambientais*. Porto Alegre: UFRGS/ABRH, 246p.
- Gomes Filho, A. (2009) *Diversidade genética em acessos de goiabeiras (Psidium guajava L.) provenientes de Bom Jesus do Itabapoana - RJ*. Tese (Doutorado em Genética e Melhoramento de Plantas), Campos dos Goytacazes – RJ, Universidade Estadual do Norte Fluminense – UENF, 121p.
- Gouvêa, L. R. L. (2009) *Divergência genética em seringueira estimada através de técnicas multivariadas e marcadores moleculares microssatélites*. Dissertação (Mestrado em Melhoramento Genético Vegetal), Campinas, IAC, 89p.
- Haykin, S. (2001) *Redes neurais: princípios e prática*. Porto Alegre: Bookman, 900p.
- Hirayama, V. (2004) *Classificador de Qualidade de Álcool Combustível e Poder Calorífico de Gás GLP*. Dissertação (Mestrado em Engenharia Microeletrônica), São Paulo, Universidade de São Paulo, Escola Politécnica, 82p.
- Hofmeyer, I. D. J. (1938) Genetical studies of *Carica papaya* L. Dept. Agr. For. *Sci. Bul.*, Africa, 187: 1-64.
- Horovitz, S., Zerpa, D. M., Arnal, H. (1953) Frecuencias de equilibrio de las formas sexuales em poblaciones de *Carica papaya* L. *Agricultura Tropical, Venezuela*, 3 (3): 149-174.
- IBGE (2008) Produção Agrícola Municipal, 2008. Rio de Janeiro. Disponível em <http://www.ibge.gov.br>, acesso em 01/2010.
- Jolliffe, I. T. (1972) Discarding variables in a principal component analysis. I. Artificial data. *Appl. Stat.*, 21 (2): 160-173.

- Jolliffe, I. T. (1973) Discarding variables in a principal component analysis. II. Real data. *App. Stat.*, 21 (1): 21-31.
- Joly, A. B. (1993) *Botânica: Introdução à taxonomia vegetal*. São Paulo: Nacional, 777p.
- Kavzoglu, T, Mather, P. M. (2003) The use of backpropagation artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24 (13): 4907-4938.
- Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, Germany, 43: 59-69.
- Kovacs, Z. L. (1997) *Redes neurais artificiais: fundamentos e aplicações*. São Paulo: Collegium Cognitio, 174p.
- Lima Júnior, L., Ferreira, R., Silva, J., Santos, E., Lira Júnior, M., Ferraz, I. (2009) Utilização de técnicas multivariadas na classificação de fases de crescimento de *Leucaena leucocephala* (LAM.) de Wit. *Floresta*, Curitiba, 39 (4): 921-935.
- Lorenzi, H., Bacher, L., Lacerda, M., Sartori, S. (2006) *Frutas brasileiras e exóticas cultivadas*. São Paulo: Instituto Plantarum, 640p.
- Luna, J. V. V. (1980) *Instruções práticas para a cultura do mamoeiro*. Salvador: EPABA, 14p.
- Marim, B. G., Silva, D. J. H. da, Carneiro, P. C. S, Miranda, G. V., Mattedi, A. P., Caliman, F. R. B. (2009) Variabilidade genética e importância relativa de caracteres em acessos de germoplasma de tomateiro. *Pesquisa Agropecuária Brasileira*, 44 (10): 1283-1290.
- Marin, S. L. D, Gomes, J. A., Silva, J. G. F., Salgado, J. S. (1994) Comportamento de preços de mamão do grupo Solo na região Norte do Espírito Santo destinado aos mercados nacional e internacional. *Congresso Brasileiro de Fruticultura*. Salvador: Sociedade Brasileira de Fruticultura, 665.
- Marin, S. L. D, Ruggiero, C. (1988) Toxicidade de inseticidas, acaricidas e fungicidas ao mamoeiro cv. Solo. *Simpósio Brasileiro sobre a cultura do mamoeiro*. São Paulo: FCAV/UNESP, 219-228.

- Marin, S. L. D., Pires, A. V., Torezani, E. G., Frigini, J. M. N., Modenesi, G. A. (2005) melhoramento genético do mamoeiro (*Carica papaya* L.): desenvolvimento e recomendação de híbridos para a região norte do Estado do Espírito Santo. *In: Martins, D. dos S. (ed.) Papaya Brasil: Mercado e inovações tecnológicas para o mamão*. Vitória: Incaper, 313-316.
- Marin, S. L. D., Silva, J. G. F. (1996) Aspectos econômicos e mercados para a cultura do mamoeiro do grupo solo na região norte do Espírito Santo. *In: Mendes, L. G., Dantas, J. L. L., Morales, C. F. G. (eds.) Mamão no Brasil*. Cruz das Almas: EUFBA/EMBRAPA, 3-20.
- Mariot, M. P., Barbieri, R. L., Sinigaglia, C., Ribeiro, M. V. (2008) Variabilidade em matrizes de acessos de espinheira-santa. *Ciência Rural*, 38 (2): 351-357.
- Martins, D. dos S., Costa, A. de F. S. da (2003) *A cultura do mamoeiro: tecnologias de produção*. Vitória: Incaper, 497p.
- Martins, W. S. (2009) Bioinformática. Brasília: Associação Brasileira das Universidades Comunitárias. Página mantida pela ABRUC. Disponível em <http://www.abruc.org.br>, acesso em 05/2009.
- Melo Júnior, A. F., Carvalho, D., Pova, J. S. R., Bearzoti, E. (2004) Estrutura genética de populações naturais de pequi ( *Caryocar brasiliense* Camb.). *Scientia Forestalis*, 66: 56-65.
- Mohammadi, S. A., Prasanna, B. M. (2003) Analysis of genetic diversity in crop plants: salient statistical tools and considerations. *Crop Science*, 432: 1235-1248.
- Monteiro, E. R. (2008) *Identificação botânica e divergência genética em pimentas do gênero Capsicum spp.* Dissertação (Mestrado em Agronomia), Teresina, Piauí, Universidade Federal do Piauí, 66p.
- Moreira, R. M. P., Ferreira, J. M., Takahashi, L. S. A., Vanconcelos, M. E. C., Geus, L. C., Botti, L. (2009) Potencial agrônomico e divergência genética entre genótipos de feijão-vagem de crescimento determinado. *Semina: Ciências Agrárias*, Londrina, 30 (1): 1051-1060.

- Nakasone, H. Y. (1980) Melhoramento de mamão no Havaí. *Simpósio Brasileiro sobre a cultura do mamoeiro*, Jaboticabal, FCAV/UNESP: 275-287.
- Nakasone, H. Y., Paull, R. E. (1998) Tropical fruits. *Crop Production Science in Horticulture*. New York: Cab International, 445p.
- Nogueira, A. P. O., Sedyama, T., Cruz, C. D., Reis, M. S., Pereira, D. G., Jangarelli, M. (2008) Novas características para diferenciação de cultivares de soja pela análise discriminante. *Cienc. Rural*, Santa Maria, 38 (9): 2427-2433.
- Oliveira, A. M. G., Farias, A. R. N., Santos Filho, H. P., Oliveira, J. R. P., Dantas, J. L. L., Santos, L. B. dos, Oliveira, M. de A., Souza Júnior, M. T., Silva, M. J., Almeida, O. A. de, Nickell, O., Medina, V. M., Cordeiro, Z. J. M. (1994) *Mamão para exportação: aspectos técnicos da produção*. Brasília: EMBRAPA-SPI, 52p.
- Oliveira, C. A. M., Silva, E. F., Molina, S. G., Ferreira, R., Lira, D. A. S., Barros Junior, J. A. B. (2006) Diversidade e estrutura genética em populações de *Caesalpinia achinata* (Lam.) na Estação Ecológica do Tapacurá-PE. *Scientia Forestalis*, 70: 77-83.
- Oliveira, M. A. B. (1999) *Variações de algumas características fisiológicas dos frutos de mamoeiro (Carica papaya L.) em função de diferentes épocas de colheita*. Tese (Doutorado em Produção Vegetal), Campos dos Goytacazes – RJ, Universidade Estadual do Norte Fluminense – UENF, 72p.
- Oliveira, M. E. B. de, Guerra, N. B., Maia, A. de H. N., Alves, R. E., Xavier, D. da S., Matos, N. M. dos S. (2009) Caracterização física de frutos do pequizeiro nativo da Chapada do Araripe – CE. *Revista Brasileira de Fruticultura*, Jaboticabal - SP, 31 (4): 1196-1201.
- Oliveira, M. S. P., Ferreira, D. F., Santos, J. B. (2007) Divergência genética entre acessos de açazeiro fundamentada em descritores morfoagronômicos. *Pesquisa Agropecuária Brasileira*, 42 (4): 501-506.
- Paiva, J. R., Cavalcanti, J. J. V., Sabry Neto, H., Freitas, A. S. M., Sousa, F. H. L. (2001) Variabilidade genética em caracteres morfológicos de populações de

- plantas jovens de acerola (*Malpighia emarginata*). *Revista Brasileira de Fruticultura*, 23 (2): 350-352.
- Peluzio, J., Vaz-De-Melo, A., Afférri, F., Silva, R., Barros, H., Nascimento, I., Fidelis, R. (2009) Variabilidade genética entre cultivares de soja, sob diferentes variações edafoclimáticas, na região centro-sul do estado do Tocantins. *Revista Brasileira de Tecnologia Aplicada nas Ciências Agrárias*, 2: 55-64.
- Pereira, M. G., Silva, F. F. da, Pereira, T. N. S. (2010) Recursos genéticos vegetais e o melhoramento de plantas. In: Pereira, T. N. S. (ed.) *Germoplasma: conservação, manejo e uso no melhoramento de plantas*. Viçosa: Arca, 250p.
- Pereira, T. M., Espósito, D. P., Souza, A. O., Couto, M. F., Cruz, C. D. (2009) Estudo comparativo dos métodos de agrupamento otimizados de Tocher e Tocher modificado. *XIII Encontro Latino Americano de Iniciação Científica e IX Encontro Latino Americano de Pós-Graduação*: 1-3.
- Pessin, G., Osório, F., Musse, S., Nonnemacher, V. (2007) Utilizando redes neurais artificiais no controle de robôs móveis aplicados no combate de incêndios. *XVI Seminário de Computação*, Blumenau: 19-30.
- Pezzopane, C. de G., Favarin, J. L., Maluf, M. P., Pezzopane, J. R. M., Guerreiro Filho, O. (2009) Atributos fenológicos e agronômicos em cultivares de cafeeiro arábica. *Ciência Rural*, 39 (3): 711-717.
- Picoli, M. C. A. (2007) *Estimativa da produtividade agrícola da cana-de-açúcar utilizando agregados de redes neurais artificiais: estudo de caso usina Catanduva*. Dissertação (Mestrado em Sensoriamento Remoto), São José dos Campos – SP, INPE, 90p.
- Quintal, S. S. R. (2009) *Caracterização e avaliação de um banco de germoplasma de mamoeiro para estudos dos parâmetros genéticos e diversidade genética*. Dissertação (Mestrado em Genética e Melhoramento de Plantas), Campos dos Goytacazes – R.J., Universidade Estadual do Norte Fluminense – UENF, 154p.

- Quintal, S. S. R., Viana, A. P., Pereira, M. G., Ferreguetti, G. A. (2007a) Parâmetros genéticos em genótipos de mamoeiros cultivados no norte do Espírito Santo. *Boletim Técnico da III Reunião de Pesquisa de Frutimamão*, 1: 180-182.
- Quintal, S. S. R., Viana, A. P., Pereira, M. G., Ferreguetti, G. A. (2007b) Estimativa da diversidade genética em genótipos de mamoeiro com base em estudos morfoagronômicos. *Boletim Técnico da III Reunião de Pesquisa de Frutimamão*, Campos dos Goytacazes, UENF, 1: 183-184.
- Ramos, H. C., Pinto, F. O., Silva, F. F. da, Pereira, M. G. (2007a) Monitoramento do nível de homozigose e análise genômica na geração de S2 de mamoeiro (*Carica papaya* L.) via Marcador Microsatélite. *Boletim Técnico da III Reunião de Pesquisa do Frutimamão*, Campos dos Goytacazes, UENF, 1: 214-217.
- Ramos, H. C., Pinto, F. O., Silva, F. F. da, Pereira, M. G. (2007b) Desenvolvimento de marcadores microsatélites para mamão (*Carica papaya* L.). *Boletim Técnico da III Reunião de Pesquisa do Frutimamão*, Campos dos Goytacazes, UENF, 1: 218-222.
- Ramos, H. C., Pinto, F. O., Silva, F. F. da, Pereira, M. G. (2007c) Genotipagem de população segregante de retrocruzamento em mamoeiro (*Carica papaya* L.) via marcadores microsatélites. *III Simpósio do Papaya Brasileiro. Papaya Brasil: Manejo, Qualidade e Mercado do Mamão*, Vitória, Incaper, 1: 360-362.
- Rao, R. C. (1952) *Advanced statistical methods in biometrical research*. New York: John Willey and Sons, 390p.
- Rech Filho, E. L. (2009) Entrevista cedida por Elíbio Leopoldo Rech Filho, pesquisador da Embrapa Recursos Genéticos e Biotecnologia, ao site Criar e Plantar. Disponível em <http://www.criareplantar.com.br>, acesso em 06/2009.
- Rolim, S. G., Couto, H. T. Z., Jesus, R. M. (1999) Mortalidade e retrocruzamento de árvores na floresta atlântica de Linhares (ES). *Scientia Forestalis*, 55: 49-69.

- Ruggiero, C., Durigan, J. F., Goes, A. de, Natale, W., Benassi, A. C. (2003) Panorama da cultura do mamão no Brasil e no mundo: situação atual e tendências. In: Martins, D. dos S. (ed.) *Papaya Brasil: qualidade do mamão para o mercado interno*. Vitória: Incaper, 13-34.
- Sampaio, H. S. de V., Luna, J. V. U., Sampaio, L. S. de V. (1983) Comportamento de linhas endógamas de mamão (*Carica papaya* L.) e seus híbridos, em solo infestado com *Phytophthora* sp. *Magistra*, Cruz das Almas, 1: 36-45.
- Schimith, R., Gleriani, J. M., Soares, V. P., Silva, E., Martins, S. V., Pinto, F. de A. de C. (2009) Generalização temporal de redes neurais artificiais na classificação espectro-temporal de culturas agrícolas. *XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, INPE: 459-466.
- Silva, E. F., Martins, L. S. S., Oliveira, V. R. (2009) Diversity and genetic structure in cajá tree (*Spondias mombin* L.) populations in northern Brazil. *Revista Brasileira de Fruticultura*, 31 (1): 171-181.
- Silva, F. F. da, Pereira, M. G., Campos, W. F., Damasceno Junior, P. C., Pereira, T. N. S., Souza Filho, G. A. de, Ramos, H. C., Viana, A. P., Ferregueti, G. A. (2007a) DNA marker-assisted sex conversion in elite papaya genotype (*Carica papaya* L.). *Crop Breeding and Applied Biotechnology*, 7: 52-58.
- Silva, F. F. da, Pereira, M. G., Damasceno Junior, P. C., Daher, R. F., Pereira, T. N. S., Souza Filho, G. A. de, Viana, A. P., Ferregueti, G. A. (2007b) Monitoring of the genetic variability in papaya parent 'Formosa' of 'UENF/CALIMAN01' hybrid via RAPD. *Crop Breeding and Applied Biotechnology*, 7: 36-42.
- Silva, F. F. da, Pereira, M. G., Damasceno Junior, P. C., Pereira, T. N. S., Viana, A. P., Daher, R. F., Ramos, H. C., Ferregueti, G. A. (2007c) Evaluation of the sexual expression in a segregating BC1 papaya population. *Crop Breeding and Applied Biotechnology*, 7: 16-23.
- Silva, F. F. da, Pereira, M. G., Ramos, H. C., Damasceno Junior, P. C., Pereira, T. N. S., Ide, C. D. (2007d) Genotypic correlations of morpho-agronomic traits in

- papaya and implications for genetic breeding. *Crop Breeding and Applied Biotechnology*, 7: 345-352.
- Silva, F. F. da, Pereira, M. G., Ramos, H. C., Damasceno Junior, P. C., Pereira, T. N. S., Viana, A. P., Daher, R. F., Ferregueti, G. A. (2008) Estimation of genetic parameters related to morpho-agronomic and fruit quality traits of papaya. *Crop Breeding and Applied Biotechnology*, 8: 65-73.
- Silva, S. H. M. G. da, Moraes, W., Marque, E. S. H., Jorge, L. A. de C. (2005) Processamento digital de imagens para identificação da Sigatoka Negra. V *Congresso Brasileiro de Agroinformática*, SBI-AGRO, Londrina, 1:-.
- Silva, S., Tassara, H. (2005) *Frutas Brasil Frutas*. São Paulo: Empresa das Artes, 323p.
- Singh, D. (1981) The relative importance of characters affecting genetic divergence. *The Indian Journal of Genetic and Plant Breeding*, New Delhi, 41: 237-245.
- Souza, A. L., Ferreira, R. L. C., Xavier, A. (1997) Análise de agrupamento aplicada à área florestal. *Boletim Técnico SIF*, 16: 109p.
- Souza, J. da S. (2000) Comercialização. In: Trindade, V. T. (org) *Mamão. Produção: aspectos técnicos*. Brasília: Embrapa, 69-70.
- Storey, W. B. (1938) The primary flowers types of papaya and the primary fruit types that develop from them. *Proc. Amer. Soc. Hort. Sci.*, 83-85.
- Storey, W. B. (1941) The botany and sex relations of the papaya. In: Storey, W. B., Jones, W. V. (eds.) *Papaya production in the Hawaiian Islands, part 1*. Hawaii: Hawaii Agricultural Experiment Station, 5-22.
- Sudheer, K. P., Gosain, A. K., Ramasastri, K. S. (2003) Estimating actual evapotranspiration from limited climatic data using neural computing technique. *Journal of Irrigation and Drainage Engineering*, 129 (3): 214-218.
- Sudre, C. P., Rodrigues, R., Riva, E. M., Karasawa, M., Amaral Júnior, A. T. (2005) Divergência genética entre acessos de pimenta e pimentão utilizando técnicas multivariadas. *Horticultura Brasileira*, Brasília, 23 (1): 22-27.

- Sudre, C. P., Cruz, C. D., Rodrigues, R., Riva, E. M., Amaral Junior, A. T., Silva, D. J. H., Pereira, T. N. S. (2006) Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. *Horticultura Brasileira*, Campinas, 24 (1): 88-93.
- Telles, M. P. C., Valva, F. D., Bandeira, L. F., Coelho, A. S. G. (2003) Caracterização genética de populações naturais de araticunzeiro (*Annona crassiflora* Mart.- Annonaceae) no Estado de Goiás. *Revista Brasileira de Botânica*, 25 (1):123-129.
- Valois, A. C. C. (1998) Biodiversidade, biotecnologia e propriedade intelectual. *Cadernos de Ciência & Tecnologia*, Brasília, 15: 21-31.
- Vasconcelos, E. S. de, Cruz, C. D., Bhering, L. L., Resende Júnior, M. F. R. (2007) Método Alternativo para Análise de Agrupamento. *Pesquisa Agropecuária Brasileira*, 42 (10): 1421-1428.
- Viana, A. P., Pereira, T. N. S., Pereira, M. G., Souza, M. M., Maldonado, J. F. M., Amaral Júnior, A. T. (2006) Genetic diversity in yellow passion fruit populations. *Crop Breeding and Applied Biotechnology*, 6: 87-94.
- Vieira, T. G. C, Lacerda, W. S., Botelho, T. G. (2009) Mapeamento de áreas cafeeiras utilizando redes neurais artificiais: Estudo de caso na região de Três Pontas, Minas Gerais. *XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, INPE: 7947-7954.
- Watanabe, F. S. Y., Imai, N. N., Samizava, T. M., Rocha, P. C. (2009) Classificação da vegetação de áreas úmidas baseada em redes neurais artificiais: estudo de caso da planície fluvial do alto rio Paraná. *XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, INPE: 5515-5522.
- Yamanishi, O. K., Mello, R. M., Martins, V. A., Lima, L. A., Fagundes, G. R. (2006) Comportamento do mamoeiro Sekati nas condições do oeste da Bahia. *Revista Brasileira de Fruticultura*, Cruz das Almas, 28 (1): 79-82.
- Zanine, A. de M., Dias, P. F., Souto, S. M., Ferreira, D. de J., Santos, E. M., Pinto, L. F. B. (2007) Uso de funções discriminantes para comparação de cultivares dos gêneros *Cynodon* e *Digitaria* quanto à produção de matéria seca e teores

de macronutrientes. *Asociación Latinoamericana de Producción Animal*, 15 (4): 152-156.

Zuin, G., Vidigal Filho, P., Kvitschal, M., Gonçalves-Vidigal, M., Coimbra, G. (2009) Divergência genética entre acessos de mandioca-de-mesa coletados no município de Cianorte, região Noroeste do Estado do Paraná. *Semina: Ciências Agrárias*, Londrina, 30 (1): 21-30.

## APÊNDICE 1

Quadro 24. Médias dos acessos avaliados e mantidos na coleção de germoplasma Caliman/UENF da espécie *Carica papaya* L., na primeira época.

ACESSO	MF	CF	DF	EP	SS	FEF	FIF	PP	MM
1 - Caliman M5	433,33	13,42	8,49	2,62	11,87	80,67	59,32	1,41	1,34
2 - Sunrise Solo 783	528,34	13,14	9,20	2,65	10,60	87,51	68,75	1,52	1,56
3 - Costa Rica	520,00	13,34	8,95	2,78	11,04	94,74	72,07	1,29	1,53
4 - Triwan Et	605,00	13,95	9,77	3,11	10,70	95,54	71,80	1,63	1,51
5 - Diva	478,33	13,47	8,88	2,59	11,84	73,08	61,89	1,41	1,51
6 - Grampola	328,34	11,85	8,17	2,39	12,07	87,12	64,61	1,30	1,35
7 - Sunrise Solo	465,00	13,42	8,88	2,78	10,97	71,95	50,69	1,49	1,45
8 - Caliman AM	330,00	11,85	7,95	2,47	10,47	75,95	57,37	1,63	1,17
9 - Caliman GB	388,33	13,53	8,07	2,36	9,04	83,42	67,74	1,53	1,17
10 - Caliman SG	288,33	11,20	7,57	2,18	8,30	74,52	62,43	1,24	1,00
11 - Caliman G	408,33	12,54	8,59	2,61	10,44	78,61	64,26	1,51	1,20
12 - Sunrise Solo 72/12 (202)	378,34	11,55	8,67	2,49	12,74	72,12	60,40	1,22	1,52
13 - Kapoho Solo (polpa amarela)	358,34	12,92	8,27	1,94	12,64	81,43	67,62	1,28	1,18
14 - Baixinho Santa Amália	383,34	11,80	8,48	2,23	10,64	85,84	68,37	1,00	1,27
15 - Sunrise Solo TJ	418,33	13,05	8,28	2,04	13,20	85,60	62,81	1,15	1,22
16 - Tailândia	1410,00	21,42	12,39	3,19	8,63	134,79	105,34	1,47	1,47
17 - São Mateus	690,00	13,60	9,58	2,93	12,32	132,18	93,51	1,47	1,54
18 - Kapoho Solo (polpa vermelha)	391,67	12,09	8,39	2,36	12,50	86,27	66,21	1,41	1,35
19 - Sunrise Solo (sel. Caliman)	425,00	12,67	8,54	2,41	12,30	78,08	61,11	1,23	1,47

Quadro 24, cont.

<b>ACESSO</b>	<b>MF</b>	<b>CF</b>	<b>DF</b>	<b>EP</b>	<b>SS</b>	<b>FEF</b>	<b>FIF</b>	<b>PP</b>	<b>MM</b>
20 - Waimanalo	883,34	21,13	9,93	2,96	9,74	183,02	133,45	1,63	1,45
21 - Mamão Bené	2441,67	29,14	14,08	3,17	8,64	212,13	174,88	1,31	1,65
22 - Mamão roxo	365,00	13,17	7,90	1,98	12,30	63,05	55,34	1,49	1,41
23 - Maradol (orig. México)	2286,67	33,05	11,85	3,24	8,00	150,08	123,14	1,41	1,35
24 - Maradol (grande limão)	2466,67	30,10	13,63	3,30	6,73	168,12	123,28	1,20	1,29
25 - Sekati	1351,67	21,92	11,02	3,08	9,24	170,06	126,48	1,19	1,41
26 - Baixinho Super	405,00	11,90	8,19	2,34	12,37	76,23	65,27	1,00	1,22
27 - Americano	2386,67	41,77	10,62	2,81	7,54	203,96	168,76	1,08	1,34
28 - STZ-51	383,34	12,30	8,32	2,30	11,30	84,67	62,85	1,47	1,15
29 - STZ-52	376,67	12,24	8,18	2,50	10,63	74,60	62,70	1,34	1,18
30 - Calimosa	833,34	19,95	9,39	2,32	13,70	174,22	123,46	1,50	1,32
31 - JS 12 (206)	1353,34	26,69	10,47	2,80	9,27	211,64	159,87	1,24	1,48
32 - Cariflora 209	1096,67	14,78	13,37	3,01	10,90	197,26	150,10	1,34	1,45
33 - Golden tipo Formosa	368,34	14,15	8,32	2,54	9,77	107,75	68,52	1,27	1,37
34 - STZ-03 - pecíolo curto	281,67	11,10	7,90	2,26	9,84	70,11	60,89	1,22	1,20
35 - STA Helena III Trat 12A plt 07X	1376,67	23,37	11,22	3,29	10,47	190,24	143,01	1,53	1,57
36 - STA Helena III Trat 11A plt 08X	1055,00	20,03	10,72	2,91	9,70	186,00	135,89	1,46	1,19
37 - STA Helena III Trat 50 plt 09X	1116,67	23,62	9,87	2,97	8,40	193,04	162,65	1,57	1,68
38 - STA Helena III Trat 02 plt 01X	1358,34	25,65	10,04	3,13	7,60	194,40	141,33	1,32	1,72
39 - STA Helena III Trat 14 plt 05X	1163,34	25,47	8,70	2,98	8,84	198,89	159,92	1,41	1,67
40 - STA Helena III Trat 15 plt 04X	806,67	21,49	9,22	3,01	9,17	182,80	141,15	1,64	1,53
41 - STA Helena III Trat 04 plt 02X	1646,67	22,44	12,29	3,46	10,37	204,99	165,03	1,39	1,49
42 - STA Helena III Trat 12B plt 06X	1360,00	22,48	10,82	3,15	8,67	204,18	155,90	1,32	1,47
43 - Papaya 42 Formosa	960,00	21,25	10,53	2,70	8,60	177,64	143,02	1,22	1,10
44 - Papaya 45 Formosa roxo	921,67	22,47	9,83	2,44	9,43	147,08	107,77	1,22	1,08
45 - Papaya 46 claro	736,67	14,83	10,67	2,90	10,07	179,25	132,62	1,53	1,73
46 - Fruto médio verde	563,34	17,95	8,49	2,41	10,44	196,24	160,78	1,27	1,20
47 - Golden robusto	413,33	12,27	8,95	2,39	9,94	93,85	62,34	1,32	1,15
48 - Tainung	1580,00	24,94	12,02	3,15	9,77	183,73	126,16	1,35	1,49
49 - STZ 63	385,00	12,42	8,30	2,34	11,40	72,36	63,47	1,35	1,22
50 - Sekati (fruto longo macuco)	2065,00	27,89	11,78	3,40	9,64	192,02	141,62	1,28	1,29

MF = massa de fruto; CF = comprimento do fruto; DF = diâmetro do fruto; EP = espessura da polpa; SS = sólidos solúveis; FEF = firmeza externa do fruto; FIF = firmeza interna da polpa; PP = incidência de pinta preta; e MM = incidência de mancha fisiológica.

Quadro 25. Médias dos acessos avaliados e mantidos na coleção de germoplasma Caliman/UENF da espécie *Carica papaya* L., na segunda época.

<b>ACESSO</b>	<b>MF</b>	<b>CF</b>	<b>DF</b>	<b>EP</b>	<b>SS</b>	<b>FEF</b>	<b>FIF</b>	<b>PP</b>	<b>MM</b>
1 - Caliman M5	446,67	14,57	8,07	2,09	11,80	78,12	55,50	1,22	1,47
2 - Sunrise Solo 783	586,67	14,30	9,52	2,40	11,07	69,79	52,16	1,21	1,83
3 - Costa Rica	571,67	15,22	8,83	2,03	10,60	85,27	66,60	1,15	1,83
4 - Triwan Et	558,34	15,92	9,00	2,12	10,17	74,99	61,22	1,41	1,63
5 - Diva	530,00	14,89	9,00	2,14	11,84	78,31	52,76	1,22	1,78
6 - Grampola	368,33	13,07	8,05	1,79	11,90	77,33	54,88	1,08	1,57
7 - Sunrise Solo	528,34	14,57	8,85	2,00	12,37	70,45	52,99	1,08	1,53
8 - Caliman AM	345,00	12,44	7,52	1,81	10,03	69,21	51,62	1,35	1,35
9 - Caliman GB	278,34	11,45	7,35	1,79	8,53	60,50	50,27	1,29	1,35
10 - Caliman SG	300,00	12,07	7,34	1,89	10,70	53,84	41,67	1,29	1,41
11 - Caliman G	388,34	12,87	7,85	1,78	10,47	65,58	47,50	1,21	1,41
12 - Sunrise Solo 72/12 (202)	320,00	12,09	7,35	1,68	11,70	78,07	51,63	1,00	1,57
13 - Kapoho Solo (polpa amarela)	415,00	12,97	8,08	2,09	10,97	74,63	59,70	1,00	1,73
14 - Baixinho Santa Amália	311,67	11,50	7,37	1,85	11,54	57,22	51,03	1,00	1,47
15 - Sunrise Solo TJ	376,67	12,93	7,79	1,88	11,84	73,04	45,95	1,00	1,52
16 - Tailândia	753,34	17,62	9,37	2,41	8,47	108,63	87,86	1,46	1,41
17 - São Mateus	820,00	15,70	10,33	2,87	11,93	163,10	102,44	1,41	1,73
18 - Kapoho Solo (polpa vermelha)	395,00	12,78	8,37	2,10	11,40	71,38	53,63	1,22	1,58
19 - Sunrise Solo (sel. Caliman)	525,00	15,20	8,83	2,19	11,20	85,09	53,14	1,15	1,58
20 - Waimanalo	701,67	19,92	8,60	2,19	10,10	161,78	132,21	1,47	1,63
21 - Mamão Bené	2213,33	29,53	12,80	3,05	9,27	207,98	152,91	1,29	1,91
22 - Mamão roxo	451,67	15,22	8,22	2,15	11,17	64,52	48,87	1,28	1,82
23 - Maradol (orig. México)	1481,67	24,03	10,85	2,89	8,90	178,00	129,21	1,22	1,73
24 - Sekati	760,00	17,00	8,79	2,40	11,70	128,21	90,02	1,08	1,41
25 - Baixinho Super	376,67	12,92	7,67	1,87	11,63	66,22	47,21	1,08	1,41
26 - STZ-51	385,00	12,40	8,02	1,81	10,63	56,18	47,31	1,29	1,35
27 - STZ-52	663,33	17,47	9,39	2,14	11,64	117,92	83,13	1,21	1,71
28 - Calimosa	1151,67	23,64	9,95	2,59	11,80	179,52	142,80	1,22	1,96
29 - JS 12 (206)	1185,00	20,74	10,79	2,64	11,17	196,14	165,79	1,21	1,85
30 - Cariflora 209	626,67	14,50	9,65	2,23	10,60	141,18	106,83	1,08	1,52
31 - Golden tipo Formosa	351,67	13,27	7,60	2,02	10,10	80,93	48,38	1,08	1,32
32 - STZ-03 - pecíolo curto	925,00	18,07	9,44	2,29	10,14	120,02	90,61	1,28	1,52
33 - STA Helena III Trat 12A plt 07X	1396,67	20,34	13,80	2,96	9,90	169,96	130,53	1,15	1,77
34 - STA Helena III Trat 11A plt 08X	1350,00	23,33	10,50	2,91	9,50	197,04	149,20	1,35	1,78
35 - STA Helena III Trat 50 plt 09X	1276,67	25,70	9,94	2,69	9,27	185,40	146,28	1,29	1,78
36 - STA Helena III Trat 02 plt 01X	1201,67	25,17	9,74	2,65	9,13	168,06	140,49	1,53	1,77

Quadro 25, cont.

<b>ACESSO</b>	<b>MF</b>	<b>CF</b>	<b>DF</b>	<b>EP</b>	<b>SS</b>	<b>FEF</b>	<b>FIF</b>	<b>PP</b>	<b>MM</b>
37 - STA Helena III Trat 14 plt 05X	911,67	23,08	8,90	2,54	9,60	175,29	134,21	1,47	1,78
38 - STA Helena III Trat 15 plt 04X	2030,00	24,35	12,20	3,07	8,97	179,41	135,99	1,41	1,90
39 - STA Helena III Trat 04 plt 02X	1188,34	21,59	10,62	2,99	9,90	206,67	149,90	1,15	1,73
40 - STA Helena III Trat 12B plt 06X	1183,34	22,35	10,59	2,83	9,13	165,53	137,23	1,35	1,68
41 - Papaya 42 Formosa	1021,67	23,65	10,10	2,44	9,90	172,89	132,65	1,08	1,56
42 - Papaya 45 Formosa roxo	696,67	20,30	8,83	2,18	10,07	181,04	122,10	1,00	1,53
43 - Papaya 46 claro	700,00	15,87	9,50	2,50	10,73	86,29	63,21	1,15	2,00
44 - Fruto médio verde	826,67	20,24	9,25	2,31	10,17	188,66	137,10	1,08	1,73
45 - Golden robusto	323,34	12,75	7,45	1,79	9,97	64,10	48,88	1,00	1,35
46 - Tainung	1471,67	24,95	11,24	2,88	10,60	150,95	127,03	1,15	1,91
47 - STZ 63	386,67	12,67	7,73	1,83	11,33	82,03	65,66	1,29	1,41
48 - Sekati (fruto longo macuco)	1965,00	32,40	11,03	3,04	10,14	208,94	164,69	1,51	1,61

MF = massa de fruto; CF = comprimento do fruto; DF = diâmetro do fruto; EP = espessura da polpa; SS = sólidos solúveis; FEF = firmeza externa do fruto; FIF = firmeza interna da polpa; PP = incidência de pinta preta; e MM = incidência de mancha fisiológica.

## **ANEXO 1**

Disco com os arquivos necessários para a utilização da rede neural artificial e as instruções para instalação e uso.