ANA CAROLINA RIBEIRO DE OLIVEIRA

ANÁLISE BIOMÉTRICA DE ACESSOS DE *Capsicum chinense* Jacq. COM ÊNFASE NA DIVERSIDADE GENÉTICA

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2016

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T

O48a
2016

Oliveira, Ana Carolina Ribeiro de, 1991-

Análise biométrica de acessos de *Capsicum chinense* Jacq. com ênfase na diversidade genética / Ana Carolina Ribeiro de Oliveira. – Viçosa, MG, 2016.

xi, 52f. : il. ; 29 cm.

Inclui apêndice.

Orientador: Paulo Roberto Cecon.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.44-49.

1. *Capsicum chinense. 2.* Pimenta - Melhoramento genético - Métodos estatísticos. 3. Pimenta - Diversidade genética. 4. Análise multivariada. 5. Métodos de otimização. I. Universidade Federal de Viçosa. Departamento de Estatística. Programa de Pós-graduação em Estatística Aplicada e Biometria. II.Título.

CDD 22. ed. 635.643

ANA CAROLINA RIBEIRO DE OLIVEIRA

ANÁLISE BIOMÉTRICA DE ACESSOS DE *Capsicum chinense* Jacq. COM ÊNFASE NA DIVERSIDADE GENÉTICA

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 23 de fevereiro de 2016.

Antônio Policarpo Souza Carneiro

Mário Puiatti

Paulo Roberto Cecon
(Orientador)

Dedico este trabalho aos meus queridos pais, Antenor e Lucinea, exemplos de vida, amor, honestidade e luta; por todo incentivo e apoio.

"Jesus não olha tanto para a grandeza das obras, nem sequer para a sua dificuldade, mas para o amor com que se fazem."

Santa Terezinha do Menino Jesus

"Sucesso não tem a ver com o dinheiro que você ganha. Tem a ver com a diferença que você faz na vida das pessoas."

Michelle Obama

AGRADECIMENTOS

Agradeço a Deus, por me guiar ao longo desta jornada, fortalecendo-me frente aos obstáculos e pela companhia nos bons e maus momentos da vida.

Aos meus pais, Antenor e Lucinea, fontes inquebrantáveis de amor, confiança, proteção e persistência.

À minha irmã, Luana, pela amizade, amor, companheirismo e incentivo.

Aos meus familiares, em especial, minhas avós, Creusa e Carolina, por adoçarem minha vida.

Ao Vinícius, pelo carinho, confiança e incentivo em todos os momentos.

Aos amigos e colegas, pela amizade, incentivo e contribuições na realização deste trabalho, cada um em sua particularidade. Em especial, ao Rodrigo, Samuel, Tafarel, Leonardo e Jaqueline.

Ao professor e orientador, Paulo Roberto Cecon, pela paciência e ensinamentos transferidos, sobretudo, pela força demonstrada ao longo das dificuldades.

Ao professor e coorientador, Moysés Nascimento, pelas sugestões e contribuições neste trabalho.

Ao coorientador, Fernando Luiz Finger, pela contribuição à pesquisa.

Aos participantes da banca examinadora pela disponibilidade e sugestões enriquecedoras.

Aos professores do Programa de Pós Graduação em Estatística Aplicada e Biometria, pela minha formação acadêmica; e aos funcionários pelo auxílio, ao longo do curso.

À Universidade Federal de Viçosa e ao Programa de Pós Graduação em Estatística Aplicada e Biometria, pela oportunidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

A todos que de alguma forma tornaram este trabalho e consequente titulação uma conquista palpável.

Muito obrigada!

BIOGRAFIA

ANA CAROLINA RIBEIRO DE OLIVEIRA, filha de Antenor Henriques de Oliveira e Lucinea Rodrigues Ribeiro de Oliveira, nasceu em Cataguases, Minas Gerais, em 08 de maio de 1991.

Em março de 2009, ingressou no curso de Bacharelado em Agronomia na Universidade Federal do Espírito Santo, Alegre - ES, graduando-se em fevereiro de 2014.

Em março do mesmo ano, iniciou o curso de Mestrado no Programa de Pós-Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em 23 de fevereiro de 2016.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

OLIVEIRA, Ana Carolina Ribeiro de, M.Sc., Universidade Federal de Viçosa, fevereiro de 2016. **Análise biométrica de acessos de *Capsicum chinense* Jacq. com ênfase na diversidade genética**. Orientador: Paulo Roberto Cecon. Coorientadores: Fernando Luiz Finger e Moysés Nascimento.

Os estudos de divergência genética são fundamentais para subsidiar ações de conservação, de utilização dos recursos genéticos e posterior aplicação em programas de melhoramento, visando à obtenção de genótipos superiores. Sendo assim, este trabalho teve por objetivo avaliar a diversidade genética de acessos de pimenta, *Capsicum chinense* Jacq., por meio de técnicas multivariadas de agrupamentos utilizando os métodos hierárquicos (UPGMA e Ward) e de otimização (Tocher e Tocher modificado). O experimento foi conduzido na área experimental do setor de olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), sob delineamento inteiramente casualizado, com quatro repetições e uma planta por parcela. Foram avaliados 11 acessos de *C. chinense* registrados no Banco de Germoplasma de Hortaliças (BGH/UFV), com base em 11 caracteres. Os resultados indicaram pelos métodos hierárquicos a formação de dois grupos, sendo que 72,73% dos acessos pertenciam ao grupo I e 27,27% ao grupo II; e ambas as estruturas de agrupamento foram validadas pelo coeficiente de correlação cofenética (r). Os métodos de otimização, Tocher e Tocher modificado, reuniram os acessos em seis e quatro grupos, respectivamente, revelando maior diversidade dos acessos em relação aos métodos anteriores. Assim, independente do método utilizado foi possível identificar os acessos mais divergentes e, consequentemente, contribuir para futuras pesquisas de cruzamento buscando híbridos com maior efeito heterótico.

ABSTRACT

OLIVEIRA, Ana Carolina Ribeiro de, M.Sc., Universidade Federal de Viçosa, february, 2016. **Biometric analysis of *Capsicum chinense* Jacq access with emphasis on genetic diversity**. Adviser: Paulo Roberto Cecon. Co-advisers: Fernando Luiz Finger and Moysés Nascimento.

Studies of genetic diversity are essential to subsidize conservation actions, utilization of genetic resources and subsequent application in breeding programs aimed at obtaining superior genotypes. Thus, this study aimed to evaluate the genetic diversity of pepper accessions, *Capsicum chinense* Jacq., through multivariate clustering techniques using hierarchical methods (UPGMA and Ward) and optimization (Tocher and modified Tocher). The experiment was conducted in the experimental area of the horticulture sector of the Department of Plant Science at the Universidade Federal de Viçosa (UFV), under completely randomized design with four replications and one plant per plot. They evaluated 11 accessions of *C. chinense* recorded at the Germplasm Bank (BGH/UFV), based on 11 characters. The results indicated by hierarchical methods the formation of two groups, with 72.73% of the accessions belonging to the group I and 27.27% in group II; both grouping structures were validated by cophenetic correlation coefficient (r). The optimization methods, Tocher and modified Tocher, gathered the accessions in six and four groups, respectively, revealing greater diversity of accessions compared to the previous methods. Thus, regardless of the method used it was possible to identify the most divergent access and, consequently, contribute to future researches seeking cross hybrids with greater heterotic effect.

1 INTRODUÇÃO

As pimenteiras do gênero *Capsicum* spp. apresentam grande potencial para o melhoramento genético, em razão da variabilidade genética do fruto em termos de cor, tamanho, forma, composição química e grau de pungência, principalmente.

Entretanto, o sucesso de programas de melhoramento dependerá, sobretudo, da variabilidade existente na população em estudo e posterior recomendação de intercruzamentos entre indivíduos superiores e divergentes (CRUZ; CARNEIRO; REGAZZI, 2014).

A quantificação da divergência genotípica e fenotípica em várias espécies de hortaliças é realizada por técnicas biométricas multivariadas, sendo algumas dependentes de medidas de dissimilaridade (COSTA et al., 2006; SUDRÉ et al., 2005).

Dentre as técnicas multivariadas, destaca-se a análise de agrupamento, que corresponde ao processo de agregar objetos com base em suas características de modo que, os agrupamentos exprimam elevada homogeneidade interna e elevada heterogeneidade externa. Porém, a aplicabilidade da análise de agrupamentos em quase todas as áreas de interesse remonta a necessidade de uma melhor compreensão da técnica visando minimizar seu emprego irregular.

Os métodos de agrupamentos mais utilizados na determinação da divergência genética quantitativa são os hierárquicos e os de otimização (CRUZ; CARNEIRO; REGAZZI, 2014).

Os procedimentos hierárquicos envolvem uma série de n–1 decisões de agrupamentos que se combinam ou dividem e persistem na análise, conduzindo a uma estrutura denominada dendrograma, cujo funcionamento engloba processos simples e repetitivos. Em contrapartida, os procedimentos de otimização requerem a definição à priori do número de grupos, permitem a redesignação de indivíduos e não envolvem o processo de construção em árvore (HAIR JÚNIOR et al., 2009).

Diversos trabalhos para estimar a divergência genética em pimenteiras foram desenvolvidos, entre eles, o estudo da variabilidade genética quanto à atividade antioxidante e concentração de compostos antioxidantes em variedades crioulas de *Capsicum baccatum* (NEITZKE et al., 2015); a avaliação da divergência genética baseada em características relacionadas à qualidade fisiológica de sementes de *Capsicum annuum* L., em uma população F2 e seus pais (PESSOA et al., 2015); a

análise de divergência genética entre variedades crioulas de *Capsicum chinense* Jacq., a partir de descritores quantitativos e qualitativos multicategóricos (VASCONCELOS et al., 2014); o estudo da divergência genética entre os acessos de *Capsicum* spp., com base nos caracteres morfopolínicos (MARTINS et al., 2013); a análise de divergência genética de acessos de *C. chinense* Jacq., por meio de caracteres morfoagronômicos (FARIA et al., 2013); o estudo da divergência genética em germoplasmas de *Capsicum*, baseado em microssatélites (RAI et al., 2013); a análise da variabilidade e estrutura genética de populações selvagens, crioulas e híbridos, para determinar a consequência da domesticação em *Capsicum annuum* L., por meio de microssatélites (OLVERA et al., 2012); a avaliação do impacto humano na variação genética de pimentas selvagens, *Capsicum annuum* var. *glabriusculum* (JARA et al., 2011); entre outros.

Este trabalho teve por objetivo geral, aplicar os procedimentos estatísticos multivariados no estudo da diversidade genética em pimenteiras. Os objetivos específicos compreenderam: (i) avaliar a diversidade genética de 11 acessos de *C. chinense* Jacq., a fim de detectar os mais similares, (ii) avaliar a qualidade e obter o número final de grupos para os métodos hierárquicos e (iii) comparar os agrupamentos obtidos com os métodos hierárquicos (UPGMA e Ward) e os de otimização (Tocher e Tocher modificado).

2 REVISÃO BIBLIOGRÁFICA

2.1 Aspectos gerais da cultura da pimenteira

O gênero *Capsicum* pertencente à família Solanaceae, tribo Solaneae e subtribo Capsicinae, compreende uma grande diversidade de pimentas e pimentões, originários do continente americano. Das 35 espécies descritas, apenas cinco são domesticadas: *C. annuum* L., *C. chinense* Jacq., *C. frutescens* L., *C. baccatum* L. e *C. pubescens* Ruiz &Pav (PORTO; SILVA, 2013).

Dentre as espécies mencionadas, denota-se como mais brasileira a espécie *C. chinense*, cujo centro secundário de origem e diversidade é a bacia amazônica. Essa espécie apresenta potencial adaptativo às condições de clima equatorial e tropical, e possui grande variabilidade genética, sendo os representantes mais conhecidos as pimentas 'de cheiro', 'bode', 'cumari do Pará', 'murici', 'murupi', entre outras (REIFSCHNEIDER, 2000; EMBRAPA, 2007).

As pimentas, do latim *pigmentum*, possuem uma característica importante dos frutos, a pungência, responsável pela ardência decorrente da presença de substâncias denominadas capsaicinoides (REIFSCHNEIDER, 2000), com exceções as pimentas sem ardume. Sendo também fontes de antioxidantes como vitamina C (ácido ascórbico), carotenoides, antocianinas e vitamina E (tocoferol) (NEITZKE, 2012).

Estas espécies apresentam forte expressão na indústria alimentícia, farmacêutica, cosmética e paisagística. A área plantada com pimentas no mundo é estimada em 1.897.946 hectares com uma produção anual de 29.939.029 toneladas (t), com produtividade média de 15.77 toneladas por hectare ($t.ha^{-1}$) (FOOD AND AGRICULTURE ORGANIZATION, 2013).

A fisiologia para desenvolvimento das pimenteiras compreende temperaturas médias mensais entre 21 a 30 ºC, sendo sensível a baixas temperaturas e intolerantes à geadas. A época de semeadura varia de acordo com as condições climáticas de cada região. Em termos botânicos as plantas são autógamas, podendo também ocorrer fecundação cruzada. A altura e forma de crescimento variam de acordo com a espécie e as condições de cultivo. O sistema radicular é pivotante, podendo alcançar

profundidades entre 70 a 120 cm, com um número elevado de ramificações laterais (EMBRAPA, 2007).

2.2 Análises biométricas

2.2.1 Divergência genética

A divergência genética está relacionada ao grau de distanciamento entre genótipos ou populações quanto ao conjunto de caracteres de interesse, objetivando identificar as combinações híbridas com maior efeito heterótico e maior heterozigose, de modo que haja maior chance de obtenção de genótipos superiores nas gerações segregantes (CRUZ; REGAZZI; CARNEIRO, 2012).

A relevância acerca da conservação de recursos genéticos vegetais e animais em bancos de germoplasma estimulam a quantificação da divergência genética, importante parâmentro estimado pelo melhorista, para a obtenção de segregrantes transgressivos[1] e populações de ampla variabilidade genética, podendo ser avaliada por meio de características agronômicas, morfológicas, moleculares, entre outras.

Assim, os estudos de divergência genética têm permitido uma orientação inicial, corroborando na tomada de decisão sobre a escolha de parentais divergentes em programas de melhoramento, além de gerar informações úteis para preservação, uso dos acessos e consequente monitoramento de bancos de germosplamas (CRUZ; FERREIRA; PESSONI 2011).

Existem duas maneiras básicas de se inferir a diversidade genética, sendo uma de natureza quantitativa e a outra de natureza preditiva. Na quantitativa, através de análises dialélicas, determina-se a capacidade geral (CGC), e específica (CEC) de combinação e a heterose[2] manifestada nos híbridos. As avaliações de *p* genitores e de todas (ou amostras de) as suas combinações híbridas $p(p-1)/2$ é necessária, porém caracteriza-se, de acordo com a cultura, um processo difícil de execução, com baixa probabilidade de êxito na obtenção de híbridos, inviavél e onerosa quando *p* é elevado.

[1] Segregação transgressiva: obtenção de descendentes que se encontram fora dos limites dos progenitores, com relação a um ou mais caracteres.

[2] Heterose: vigor híbrido, ou seja, superioridade genética da média da geração F_1 em relação à média dos progenitores, com respeito a um ou vários caracteres.

Os métodos preditivos baseiam-se em diferenças morfológicas, fisiológicas e moleculares dos genótipos, quantificadas por medidas de dissimilaridade, dispensando a

obtenção das combinações híbridas entre eles, sendo os mais utilizados (CRUZ; REGAZZI; CARNEIRO, 2012).

Para o estudo da divergência genética quantitativa, os procedimentos mais empregados são aqueles que utilizam técnicas de análises multivariadas permitindo a avaliação de informações múltiplas (ou a análise de várias variáveis conjuntamente); citam-se, métodos agrupamento, sendo alguns dependentes da adoção de medidas de proximidade; análise por componentes principais e por variáveis canônicas.

Estudos de divergência genética têm sido utilizados em diversas culturas, tais como, pimenta (PESSOA et al., 2015; FARIA et al., 2013;), soja (LOPES et al., 2014; HAMAWAKI et al., 2012; PELUZIO et al., 2012), feijão (SOFI et al., 2014; COSTA et al., 2013), milho (RIGON; CAPUANI; RIGON, 2015; OLIBONI et al., 2012), pinus (SILVA et al., 2012), cana-de-açúcar (SOUZA, J. et al., 2013; DUTRA FILHO et al., 2011), entre outras.

2.2.3 Medidas de dissimilaridade

Alguns métodos de agrupamentos requerem uma medida de proximidade, similaridade ou dissimilaridade, que quantifique e informe o grau de semelhança ou de diferença entre os indivíduos a serem agrupados, respectivamente. As medidas são interrelacionadas, ou seja, obtêm-se uma medida de dissimilaridade a partir da similaridade e vice-versa (BUSSAB; MIAZAKI; ANDRADE, 1990; FERREIRA, 2011).

Um grande número de distâncias presentes na literatura têm sido propostas, e cada uma delas produzirá um determinado tipo de agrupamento. Logo, a escolha de qual medida adotar abrangerá vários fatores como: a natureza das variáveis, a escalas de medidas e o conhecimento da pesquisa (MINGOTI, 2005; POHLMANN, 2014).

As medidas de dissimilaridade são importantes nas análises de divergência genética, por expressarem as informações contidas no conjunto de dados, ou seja, representar a diversidade existente no conjunto de acessos analisados.

Estas medidas auxiliam na identificação de genitores uteis para programas de hibridização, de modo que os genitores possam apresentar características genéticas que proporcionam, na F1, maior heterose e, em posteriores gerações, indivíduos transgressivos (CRUZ; CARNEIRO; REGAZZI, 2014).

Das distâncias ou medidas, as mais utilizadas são: a distância euclidiana, distância euclidiana quadrática, a distância euclidiana média, a distância ponderada e a distância generalizada de Mahalanobis (1936) todas obtidas de variáveis quantitativas (CRUZ; FERREIRA; PESSONI, 2011; CRUZ; CARNEIRO; REGAZZI, 2014).

Cabe salientar que, a distância euclidiana quadrática padronizada por incorporar o procedimento de padronização sobre os dados utilizando a escala em termos de desvio padrão (X/S_x), é indicada para variáveis que possuem escalas diferentes (FERREIRA, 2011).

As razões para padronização deligenciam-se a evitar que as escalas escolhidas para medir as variáveis afetem arbritariamente a similaridade entre os indivíduos, além defazer com que as variáveis contribuam igualmente na avaliação da similaridade entre acessos, ou seja, impedindo que as variáveis com grande dispersão dominem a classificação das distâncias (CRUZ; FERREIRA; PESSONI, 2011).

2.2.4 Análise de Agrupamentos

A análise de agrupamentos, também denominada análise de conglomerados ou *clusters analysis* compreende um grupo de técnicas multivariadas, cujo objetivo é agrupar os elementos, por meio de algum critério de classificação, de modo que estabeleça homogeneidade dentro do grupo e heterogeneidade entre grupos (MINGOTI, 2005; CRUZ; REGAZZI; CARNEIRO, 2012).

Esta análise possui propriedades matemáticas, é descritiva, não teórica, não inferencial, e sua aplicação exige suporte conceitual, dado seu carácter exploratório, além de requerer cautela nas questões de representatividade da amostra e multicolinearidade entre variáveis (HAIR JÚNIOR et al., 2009).

Conforme BUSSAB, MIAZAKI e ANDRADE (1990), o processo de agrupamento envolve as seguintes etapas e estas não são independentes:

1. Planejamento do experimento;

2. Obtenção dos dados;
3. Tratamento dos dados;
4. Definição da medida de similaridade ou dissimilaridade;
5. Escolha e execução do método de agrupamento;
6. Apresentação, avaliação e validação dos resultados.

Dentre os métodos de agrupamento encontrados na literatura, os mais utilizados em divergência genética são os de otimização e os hierárquicos, que apresentam distintas particularidades.

Ressaltamos que, o processo de otimização requer uma pré-especificação, ou seja, um conhecimento a priori do número de grupos desejados, ao contrário das técnicas hierárquicas, que por sua vez permitem a criação de dendrogramas para representar e ilustrar os procedimentos de agrupamentos (MINGOTI, 2005).

Outra diferença relevante refere-se à capacidade de formar uma solução totalmente independente de qualquer outra nos métodos de otimização, em contrapartida, os métodos hierárquicos são baseados nas combinações dos agrupamentos anteriores (HAIR JÚNIOR et al., 2009).

2.2.4.1 Métodos Hierárquicos

Nos métodos hierárquicos, os indivíduos são agrupados por um processo iterativo em vários níveis, de modo hierárquico, até o estabelecimento do dendrograma ou diagrama de árvore (BUSSAB; MIAZAKI; ANDRADE, 1990; CRUZ, 2006).

Estas técnicas apresentam alguns pontos fortes, tais como, alta difusão do método, simplicidade do processo, rapidez, habilidade para examinar uma gama de soluções e facilidade de comparação entre elas, com variação de medidas de dissimilaridade (HAIR JÚNIOR et al., 2009).

Apresentam basicamente, duas etapas: a primeira refere-se à estimação de uma medida de similaridade ou dissimilaridade e, a segunda, a adoção de um algoritmo de agrupamento. A escolha do algoritmo exigirá o conhecimento de suas propriedades, aliado aos objetivos da pesquisa. Além disso, o resultado do agrupamento e a definição do número de grupos poderão ser influenciados por ambas as etapas (MINGOTI, 2005; FERREIRA, 2011).

De maneira geral, os métodos hierárquicos são divididos em: aglomerativos (*bottom-up*) e divisivos (*top-down*). Na Figura 1, os métodos aglomerativos se movem de baixo para cima e os divisivos em sentido contrário. No primeiro, o processo de agrupamento começa com cada indivíduo como seu próprio grupo e são sucessivos agrupamentos, até a obtenção de um grupo contendo todos os indivíduos; citam-se, o do vizinho mais próximo (*Single Linkage Method*); o do vizinho mais distante (*Complete Linkage Method*); o da ligação média (*Average Linkage*), o do centróide e o proposto por Ward (1963). No divisivo o processo é inverso, até a formação de grupos unitários, sendo o método mais conhecido o de Edwards e Cavalli-Sforza (1965) (HAIR JÚNIOR et al., 2009).

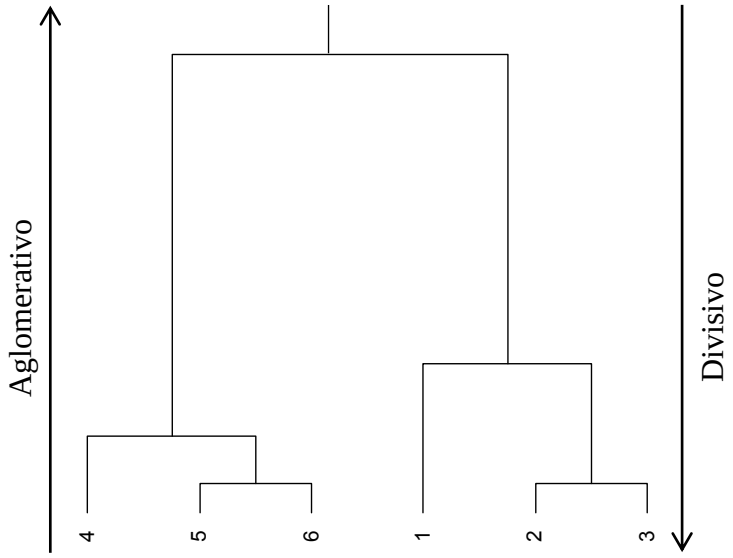

Figura 1. Dendrograma com abordagem hierárquica: aglomerativa e divisiva

Os métodos aglomerativos são mais utilizados face à eficiência, à disponibilidade de pacotes computacionais e menor aporte computacional exigido quando comparado ao divisivo; dentre eles, o ligação média e o proposto por Ward (1963) são, provavelmente, os mais preferidos (FERREIRA, 2011).

O método de ligação média (UPGMA) difere dos algoritmos de ligação simples e completa, por considerar as médias aritméticas das medidas de dissimilaridade e não depender dos valores extremos; logo, é menos afetado por variações atípicas. Além disso, tendem a produzir grupos com aproximadamente a mesma variância interna

(HAIR JÚNIOR et al., 2009) e a gerar valores maiores do coeficiente de correlação cofenética, caracterizando menor distorção na representação do agrupamento no dendrograma.

Conforme Ward (1963) e Hair Júnior et al. (2009), o método Ward procura minimizar a variação interna e permitir a visualização de grupos bem definidos, face à tendência de resultar em agrupamentos com um pequeno número de observações, pois a soma de quadrados está diretamente relacionada com o número de observações envolvidas. É uma técnica adequada quando se deseja formar grupos com aproximadamente o mesmo tamanho.

2.2.4.1.1 Determinação do número de grupos

A determinação do número de grupos é um dos principais entraves da aplicação dos métodos de agrupamentos, sobretudo nos algoritmos hierárquicos, pois ainda que o procedimento gere um conjunto de soluções de agrupamentos, o pesquisador deve escolher as soluções que melhor representam a estrutura de dados (regra de parada). Os nãos hierárquicos, por sua vez, requerem a escolha da melhor solução entre duas ou mais soluções (HAIR JÚNIOR et al., 2009).

Assim, a escolha de determinada regra de parada deve ser condizente com alguns critérios, são eles: razões práticas do pesquisador; análise visual das ramificações do dendrograma e critérios estatísticos, tais como: os índices RMSSTD (*Root-mean-square Standard Deviation*), BSS (*Between-group Sum of Squares*), SPRSQ (*Semipartial R-Square*) e RS (*R-Square*) (MINGOTI, 2005), e o método de Mojena (MOJENA, 1977). Este último foi proposto por Mojena (1977) e baseia-se no tamanho relativo dos níveis de fusão (distâncias) nos dendrogramas, e será explanado no decorrer do trabalho.

O embasamento em critérios empíricos de regras de parada deve ser complementado com conceituação de relações teóricas que podem sugerir um número natural de agrupamento (HAIR JÚNIOR et al., 2009).

2.2.4.1.2 Validação do agrupamento

A validação busca garantir que a solução de agrupamentos seja representativa do conjunto de dados, além de generalizável e estável ao longo do tempo, determinando qual o agrupamento proporcionou a melhor partição.

Diversas ferramentas são propostas para a validação do processo de agrupamento, são elas: validações externas - avaliam os agrupamentos com base em agrupamentos de referência (estrutura pré-definida, que reflete a intuição sobre a estrutura de dados), validações internas - analisam as informações contidas nos grupos obtidos e validações relativas - compara o agrupamento com outros agrupamentos, gerados pelo mesmo algoritmo, porém com diferentes parâmetros de entrada.

Dentre as técnicas de validação interna citam-se: coesão, acoplamento, coeficiente de correlação cofenética e coeficiente de silhueta. O coeficiente de correlação cofenética é fundamentado na semelhança entre a matriz de distâncias obtidas pelas medidas de parecença e a matriz de distâncias originais.

2.2.4.2 Métodos de Otimização

Os métodos não hierárquicos ou de partição produzem apenas uma solução de agrupamentos para um conjunto de *sementes* ou ponto de partida, que são empregadas para reunir indivíduos dentro de uma distância pré-especificada das *sementes*, ou seja, eles designam objetos a um grupo uma vez que o número de grupos é pré-especificado ou *a priori* (HAIR JÚNIOR et al., 2009; POHLMANN, 2014).

Estes métodos apresentam como vantagens, a possibilidade de se operar com uma grande quantidade de dados e, a capacidade de otimizar as soluções por meio da realocação de objetos até a obtenção de uma heterogeneidade mínima dentro dos grupos (HAIR JÚNIOR et al., 2009).

Os algoritmos de agrupamentos não hierárquicos são também denominados de k-médias ou *k-means clustering*, e eles são similares no método básico para designar indivíduos, mas diferenciam-se no grau em que cada indivíduo pode ser novamente designado entre clusters após a designação inicial, ou seja, na maneira que constituem a melhor partição (HAIR JÚNIOR et al., 2009).

Existem basicamente três tipos de procedimentos não hierárquicos: Referência Sequencial (*Sequential threshold*), Referência Paralela (*Paralled threshol*) e Otimização (*Optimization*) (POHLMANN, 2014).

No procedimento de otimização, ocorre à partição do conjunto de indivíduos em subgrupos não vazios e mutuamente exclusivos, permitindo a realocação dos objetos, por meio da maximização ou minimização de alguma medida preestabelecida, ou seja, objetiva-se alcançar uma partição que otimize alguma medida predefinida. Citam-se, os métodos de Tocher e Tocher modificado que correspondem aos mais utilizados na análise de divergência genética entre acessos (CRUZ, 2006).

O método de Tocher adota um critério de agrupamento que apresenta a distância média intragrupo menor que a distancia média intergrupo. No entanto, apresenta uma ineficiência, em casos de indivíduos com grande dissimilaridade, formando grupos com apenas um indivíduo, dada a influencia das distâncias dos indivíduos agrupados anteriormente. Em contrapartida, no Tocher modificado o processo de agrupamento é sequencial e não simultâneo, não existindo influencia dos indivíduos já agrupados, logo, mostra-se mais eficaz (VASCONCELOS et al., 2007).

Nos estudos de divergência genética de várias culturas têm-se utilizado métodos de otimização, dentre elas, pimenta (VASCONCELOS et al., 2012), soja (LOPES et al., 2014), maracujá (SOUSA et al., 2012), feijão (PEREIRA et al., 2011), açaizeiro (GALATE et al., 2012), entre outras.

3 MATERIAL E MÉTODOS

3.1 Descrição do experimento

O experimento foi conduzido na área experimental do setor de olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), município de Viçosa, localizado na Zona da Mata de Minas Gerais, cujas coordenadas são: 20º 45’ S e 42º 51’ W, com altitude média de 650 m.

Utilizou-se o delineamento experimental inteiramente casualizado, com quatro repetições e uma planta por parcela. Foram avaliados 11 acessos de *Capsicum chinense* registrados no Banco de Germoplasma de Hortaliças (BGH/UFV), conforme Tabela 1.

Tabela 1. Identificação dos 11 acessos de *Capsicum chinense* quanto aos genitores, a pungência e origem. UFV, Viçosa - Minas Gerais, 2007

Código	Acesso	Pungência	Local de coleta
1	BGH 1716	Sim	Pindaré - Mirim – MA
2	BGH 4289	Sim	Rondonópolis – MT
3	BGH 4733	Sim	Manaus – AM
4	BGH 5012	Sim	Água Branca – AL
5	BGH 6228	Sim	Brasília – DF
6	BGH 7295	Sim	Viçosa – MG
7	BGH 1716	Não	Pindaré - Mirim – MA
8	BGH 4201	Não	Belém – PA
9	BGH 4223	Não	IAC – SP
10	BGH 6233	Não	Brasília – DF
11	BGH 6378	Não	Boca do Janacanam – AM

3.2 Caracteres avaliados

Os caracteres de interesse que foram avaliados estão descritos abaixo:

- Peso total de frutos por planta (PTF)

Foi obtida mediante três colheitas mensais, no período de maior produtividade das plantas e os resultados foram expressos em gramas.

- Massa da matéria total do fruto maduro fresco (MTF)

Os resultados foram obtidos pela média de cinco frutos maduros frescos por planta, com a utilização de balança analítica. Os resultados foram expressos em gramas.

- Massa da matéria total do fruto maduro seco (MTS,)

Após a tomada dos dados de matéria fresca, descrito anteriormente, os frutos foram secos em estufa por 72 h à temperatura de 60 °C sob ventilação.

- Porcentagem de matéria seca (% MS)

A porcentagem de matéria seca foi determinada pela expressão:

$$\%\text{MS} = \frac{\text{MTS}}{\text{MTF}} \times 100$$

em que: MTS = massa da matéria total do fruto maduro seco e MTF = massa da matéria total do fruto maduro fresco.

- Teor de sólidos solúveis (SST)

Foi determinado em refratômetro Abbé com compensação de temperatura, conforme metodologia proposta pelo Instituto Adolfo Lutz (1985), obtido a partir da média de cinco frutos. Os resultados foram expressos em %.

- Teor de Vitamina C (VIT C)

Foi determinado conforme metodologia do Instituto Adolfo Lutz (1985) com modificações. Os resultados foram expressos em %.

- Comprimento do fruto maduro (CFM), largura do fruto maduro (LFM) e espessura da polpa (ESP)

Foram obtidos com a utilização de paquímetro na porção mediana de cinco frutos por planta com resultados expressos em milímetros.

- Número de sementes por fruto (NSF)

Foi obtido pela contagem das sementes de cinco frutos por planta.

- Pungência dos frutos (PUN)

A capsaicina total dos frutos foi avaliada por HPLC, de acordo com a metodologia proposta por Maillard et al. (1997), com modificações. Os resultados foram expressos em miligramas de capsaicina e dihidrocapsaicina por grama de matéria seca dos frutos ($mg.gMS^{-1}$).

3.3 Análises estatísticas

Os dados obtidos foram submetidos às análises de dissimilaridade (distância euclidiana quadrática padronizada), de agrupamento (UPGMA, Ward, Tocher e Tocher modificado) e, determinação do número de grupos e validação do agrupamento (Mojena e coeficiente de correlação cofenética), no caso hierárquico, utilizando os *softwares* R (R DEVELOPMENT CORE TEAM, 2015) e GENES (CRUZ, 2013).

Foram realizados os seguintes procedimentos:

3.3.1 Dissimilaridade entre acessos

A medida de dissimilaridade adotada foi a distância euclidiana quadrática padronizada, indicada para variáveis que possuem escalas diferentes, porém, não relacionadas, logo, considerou-se a inexistência de correlação entre as variáveis em estudo.

3.3.1.1 Distância euclidiana quadrática padronizada

As estimativas das distâncias quadráticas foram obtidas conforme a expressão abaixo:

$$d^2_{ii'} = \|y_i - y_{i'}\|^2_{\psi} = (y_i - y_{i'})^t \psi (y_i - y_{i'})$$

em que,

$d^2_{ii'}$: corresponde à distância quadrática entre os acessos i e i', onde i, i' = 1, 2,..., n;

y_i e $y_{i'}$: correspondem aos vetores p-dimensionais das médias das observações dos acessos i e i' (em casos de repetição);

ψ : métrica de interesse.

Sendo ψ definida por $\psi = d^{-1} = \text{diag}\left(1/s^2_k\right)$, onde $k = 1,2,...,p$, então a distância quadrática caracteriza-se como distância euclidiana quadrática padronizada. Sendo s^2_k a estimativa da variância da *k*-ésima variável da amostra de *n* objetos.

Exemplo hipotético: Avaliação de cinco acessos em delineamento inteiramente casualizado (DIC), com quatro repetições, por meio de três caracteres (Y_1, Y_2 e Y_3).

Tabela 2. Valores médios de cinco acessos, relativos aos caracteres Y_1, Y_2 e Y_3, conforme exemplo hipotético para obtenção da matriz de dissimilaridade

Acesso	Y_1	Y_2	Y_3
1	20,0	1,0	105
2	25,0	3,0	120
3	25,0	3,0	125
4	50,0	7,0	185
5	50,5	7,0	190

Com base nas informações da Tabela 2 é estimada a matriz de dissimilaridade. Inicialmente procede-se o cálculo da matriz de dispersão, cujos elementos da diagonal referem-se às variâncias (s^2_k) e, fora da diagonal, covariâncias.

A métrica de interesse ($\boldsymbol{\psi}$) corresponderá a uma matriz diagonal, ou seja, $\boldsymbol{\psi} = \text{diag}(1/s_k^2)$, logo,

$$\boldsymbol{\psi} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \end{array} \begin{array}{c} \begin{array}{ccc} \quad & \quad & \quad \end{array} \\ \begin{bmatrix} 1/s_1^2 & 0 & 0 \\ 0 & 1/s_2^2 & 0 \\ 0 & 0 & 1/s_3^2 \end{bmatrix}_{3x3} \end{array}$$

onde,

$$s_1^2 = \frac{\sum_{i=1}^{5}(y_i - \bar{y})^2}{n-1} = \frac{\left[(20-34{,}1)^2 + (25-34{,}1)^2 + \ldots + (50{,}5-34{,}1)^2\right]}{4} = 221{,}55$$

As demais variâncias são obtidas do mesmo modo e são dadas por: $s_2^2 = 7{,}20$ e $s_3^2 = 1562{,}50$. Assim, a matriz $\boldsymbol{\psi}$ é dada por:

$$\boldsymbol{\psi} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \end{array} \begin{array}{c} \begin{array}{ccc} \quad & \quad & \quad \end{array} \\ \begin{bmatrix} 1/221{,}55 & 0 & 0 \\ 0 & 1/7{,}20 & 0 \\ 0 & 0 & 1/1562{,}50 \end{bmatrix}_{3x3} \end{array}$$

As estimativas das distâncias euclidianas quadráticas padronizadas podem ser obtidas a partir de valores médios e da matriz $\boldsymbol{\psi}$. Considerando os acessos 1 e 2, a estimativa será:

$$d_{12}^2 = (\mathbf{y_1} - \mathbf{y_2})^t \boldsymbol{\psi} (\mathbf{y_1} - \mathbf{y_2})$$

$$d_{12}^2 = \left(\begin{bmatrix} 20 \\ 1 \\ 105 \end{bmatrix} - \begin{bmatrix} 25 \\ 3 \\ 120 \end{bmatrix} \right)^t_{1x3} \begin{bmatrix} 1/221{,}550 & & \\ & 1/7{,}200 & \\ & & 1/1562{,}500 \end{bmatrix}_{3x3} \left(\begin{bmatrix} 20 \\ 1 \\ 105 \end{bmatrix} - \begin{bmatrix} 25 \\ 3 \\ 120 \end{bmatrix} \right)_{1x3}$$

$$d_{12}^2 = \begin{bmatrix} -5 & -2 & -15 \end{bmatrix}_{1x3} \begin{bmatrix} 1/221{,}550 & & \\ & 1/7{,}200 & \\ & & 1/1562{,}500 \end{bmatrix}_{3x3} \begin{bmatrix} -5 \\ -2 \\ -15 \end{bmatrix}_{1x3}$$

$$d_{12}^2 = 0{,}8124$$

As demais estimativas das distâncias entre os acessos são obtidos conforme exposto acima. De tal forma que, a matriz $\mathbf{d^2}$ corresponderá a:

$$\mathbf{d^2} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 0 & 0{,}8124 & 0{,}9244 & 13{,}1583 & 13{,}8228 \\ & 0 & 0{,}0160 & 7.7472 & 8{,}2932 \\ & & 0 & 7{,}3272 & 7{,}8612 \\ & & & 0 & 0{,}0171 \\ & & & & 0 \end{bmatrix}_{5x5}$$

3.3.2 Métodos de Agrupamentos

3.3.2.1 Método UPGMA ou ligação média

Nesse algoritmo, a distância entre grupos é calculada pela distância média entre todos os pares de indivíduos dos dois diferentes grupos, ou seja, corresponde à dissimilaridade média de todos os indivíduos em um agrupamento com todos os indivíduos em outro. Este método evita caracterizar a dissimilaridade por valores extremos (máximo ou mínimo) e tende a combinar grupos com pequenas variações internas (FERREIRA, 2011).

A distância entre um indivíduo k e um grupo formado pelos indivíduos i e j é dada por:

$$d_{(ij)k} = \text{média}\left\{d_{ik}; d_{jk}\right\} = \frac{d_{ik} + d_{jk}}{2}$$

ou seja, $d_{(ij)k}$ é dada pela média do conjunto das distâncias dos pares de indivíduos (i, k) e (j, k).

A distância entre dois grupos é estabelecida por:

$$d_{(ij)(kl)} = \text{média}\left\{d_{ik}; d_{il}; d_{jk}; d_{jl}\right\} = \frac{d_{ik} + d_{il} + d_{jk} + d_{jl}}{4}$$

ou seja, $d_{(ij)(kl)}$ é dada pela média do conjunto, correspondente as distâncias entre os pares de indivíduos (i, k), (i, l), (j, k) e (j, l).

Exemplo hipotético: Considerando as dissimilaridades expressas pela matriz de distâncias euclidianas padronizadas quadráticas (item 3.3.1.1), aplica-se o método UPGMA, conforme as etapas abaixo:

Etapa 1: Identificar na matriz de dissimilaridade original ($\mathbf{d^2}$), dada a seguir, os acessos mais similares que formarão o grupo inicial e, posteriormente, calcular a distância em relação aos demais acessos.

$$\mathbf{d^2} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 0 & 0{,}8124 & 0{,}9244 & 13{,}1583 & 13{,}8228 \\ & 0 & \mathbf{0{,}0160} & 7.7472 & 8{,}2932 \\ & & 0 & 7{,}3472 & 7{,}8612 \\ & & & 0 & 0{,}0171 \\ & & & & 0 \end{bmatrix}_{5x5}$$

- Grupo I: acessos 2 e 3
- Nível de fusão: 0,0160
- Distância entre os demais acessos:

$$(d_{23})_1 = \frac{d_{21} + d_{31}}{2} = \frac{0{,}8124 + 0{,}9244}{2} = 0{,}8684$$

$$(d_{23})_4 = \frac{d_{24} + d_{34}}{2} = \frac{7{,}7472 + 7{,}3472}{2} = 7{,}5472$$

$$(d_{23})_5 = \frac{d_{25} + d_{35}}{2} = \frac{8{,}2932 + 7{,}8612}{2} = 8{,}0772$$

- Nova matriz de dissimilaridade ($\mathbf{d_1^2}$)

$$\mathbf{d_1^2} = \begin{matrix} 1 \\ (2,3) \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 0 & 0{,}8684 & 13{,}1583 & 13{,}8228 \\ & 0 & 7{,}5472 & 8{,}0772 \\ & & 0 & \mathbf{0{,}0171} \\ & & & 0 \end{bmatrix}_{4x4}$$

Etapa 2: Identificar na matriz de dissimilaridade ($\mathbf{d_1^2}$), os acessos mais similares e proceder conforme Etapa 1.

- Grupo II: acessos 4 e 5.
- Nível de fusão: 0,0171
- Distâncias entre os acessos restantes:

$$d_{(45)1} = \frac{d_{41} + d_{51}}{2} = \frac{13{,}1583 + 13{,}8228}{2} = 13{,}4905$$

$$d_{(45)23} = \frac{d_{42} + d_{43} + d_{52} + d_{53}}{4} = \frac{7{,}7472 + 7{,}3472 + 8{,}2932 + 7{,}8612}{4} = 7{,}8122$$

- Nova matriz de dissimilaridade ($\mathbf{d_2^2}$)

$$\mathbf{d_2^2} = \begin{matrix} \\ 1 \\ (2,3) \\ (4,5) \end{matrix} \begin{bmatrix} 0 & \mathbf{0,8684} & 13,4905 \\ & 0 & 7,8122 \\ & & 0 \end{bmatrix}_{3x3}$$

Etapa 3: Idem Etapa 2, identificar na matriz ($\mathbf{d_2^2}$) os acessos mais similares.

- Grupo III: acessos (2,3) e 1
- Nível de fusão: 0,8684
- Distância em relação ao acesso (4,5)

$$d_{(231)45} = \frac{d_{24} + d_{34} + d_{14} + d_{25} + d_{35} + d_{15}}{6} =$$

$$d_{(231)45} = \frac{7,7472 + 7,3472 + 13,1583 + 8,2932 + 7,8612 + 13,8228}{6} = 9,7050$$

- Nova matriz de dissimilaridade ($\mathbf{d_3^2}$)

$$\mathbf{d_3^2} = \begin{matrix} \\ (2,3,1) \\ (4,5) \end{matrix} \begin{bmatrix} 0 & 9,7050 \\ 9,7050 & 0 \end{bmatrix}$$

Etapa 4: Reproduzir a etapa anterior até que o útimo indivíduo seja incluído no grupo final (2,3,1,4,5).

Este método, quando aplicado à matriz de dissimilaridade expressa pela distância euclidiana quadrática padronizada proporciona o resultado apresentado na Tabela 3 e, no dendrograma (Figura 2).

Tabela 3. Resultado do agrupamento pelo método UPGMA, para o exemplo em consideração

Etapa	Fusão		Nível de Fusão	Distância (%)	Números de acessos
	Acesso	Acesso			
1	2	3	0,0160	0,1649	2
2	4	5	0,0171	0,1763	2
3	(2,3)	1	0,8684	8,9480	3
4	(2,3,1)	(4,5)	9,7050	100,00	5

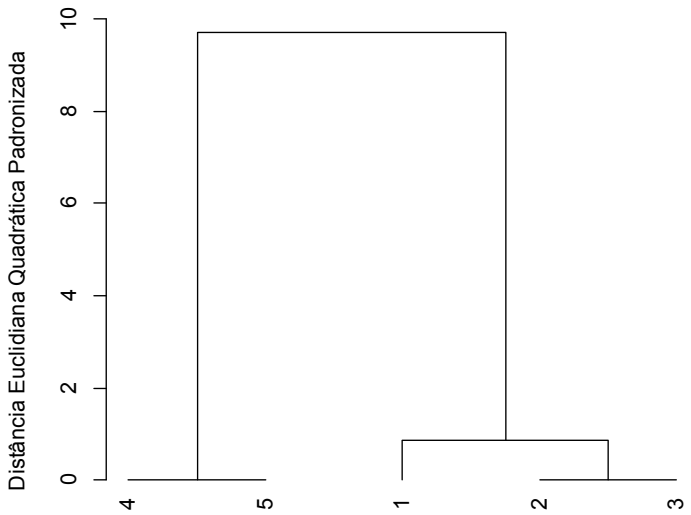

Figura 2. Dendrograma obtido pelo método UPGMA, a partir das medidas de dissimilaridade entre cinco acessos hipotéticos

3.3.2.2 Métodos da Variância Mínima de Ward

O método de Ward também conhecido como "Mínima Variância" ou método do incremento das somas de quadrados, baseia-se na análise de variância. As somas de quadrados entre e dentro dos grupos são critérios de agrupamento. Esse método é indicado para variáveis quantitativas, pois têm como base vetores de médias e também leva em consideração a diferença dos tamanhos dos *clusters* comparados. Além disso, tende à produção de grupos com, aproximadamente, o mesmo número de elementos (MINGOTI, 2005; FERREIRA, 2011).

Conforme Mingoti (2005) o procedimento fundamenta-se nas seguintes etapas: (1) cada elemento é considerado como um único *cluster;* (2) em cada passo do agrupamento calcula-se a soma de quadrados dentro de cada *cluster*. A ideia é agrupar os elementos que minimizam o incremento dessa soma de quadrados. A soma de quadrados é dada por:

$$SS_i = \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^t (Y_{ij} - \overline{Y}_{i.})$$

em que,

SS_i : corresponde a soma de quadrados do *cluster* C_i ;

n_i : é o número de elementos do *cluster* C_i ;

Y_{ij} : é o vetor de observações do j-ésimo elemento amostral pertencente ao i-ésimo *cluster*;

$\overline{Y}_{i.}$: é o vetor de médias do *cluster* C_i .

A soma de quadrados total (SSR) dentro dos *clusters* no passo *k* é definida como:

$$SSR = \sum_{i=1}^{g_k} SS_i$$

onde,

g_k : é o número de *clusters* no passo k.

A distância entre os *clusters* C_h e C_i refere-se à soma de quadrados entre os *clusters* e é dada por:

$$d(C_h, C_i) = \left[\frac{n_h n_i}{n_h + n_i} \right] (Y_{h.} - \overline{Y}_{i.})^t (Y_{h.} - \overline{Y}_{i.})$$

em que,

n_h : é o número de elementos do *cluster* C_h ;

n_i : é o número de elementos do *cluster* C_i ;

$\overline{Y}_h$: é o vetor de médias do *cluster* C_h ;

$\overline{Y}_i$: é o vetor de médias do *cluster* C_i .

Assim, a distância entre os *clusters*, dada por $d(C_h, C_i)$, corresponde à diferença entre o valor da SSR depois e antes de se combinar os *clusters* C_h e C_i, num único *cluster*. Logo, em cada etapa, o método combina os dois *clusters* que fornecem menor valor de SSR (MINGOTI, 2005).

Lance e Williams (1967) propuseram uma parametrização da fórmula de distâncias adaptavél aos diversos métodos de agrupamentos, adotando os parâmetros (α, β, γ), no caso, a implementação do método Ward possibilitou uma atualização da distância dos *clusters* em cada etapa do agrupamento, otimizando a função.

O algoritmo é dado por:

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma \left| d_{ik} - d_{jk} \right|$$

$$\therefore d_{(ij)k} = \frac{n_i + n_k}{n_i + n_j + n_k} d_{ik} + \frac{n_j + n_k}{n_i + n_j + n_k} d_{jk} + \frac{-n_k}{n_i + n_j + n_k} d_{ij}$$

onde,

$d_{(ij)k}$: distância entre os acessos i, j, grupo recém-formado, e o acesso k.

n_i, n_j, n_k: representam o número de acessos envolvidos.

Assumindo: $\alpha_i + \alpha_j + \beta = 1$, $\alpha_i = \alpha_j$, $\beta < 0$ e $\gamma = 0$.

Exemplo hipotético: Considerando a matriz de dissimilaridade original (d^2) obtida anteriormente (item 3.3.1.1), aplica-se o método Ward, conforme as etapas abaixo:

Etapa 1: Identificar na matriz de disimiliaridade (**d^2**) os acessos mais similares, que irão compor o grupo inicial e, posteriormente calcular a distância em relação aos demais acessos.

$$\mathbf{d^2} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & 0{,}8124 & 0{,}9244 & 13{,}1583 & 13{,}8228 \\ & 0 & \mathbf{0{,}0160} & 7.7472 & 8{,}2932 \\ & & 0 & 7{,}3472 & 7{,}8612 \\ & & & 0 & 0{,}0171 \\ & & & & 0 \end{bmatrix}_{5x5}$$

- Grupo I: acessos 2 e 3.
- Nível de fusão: 0,0160
- Distância entre os demais acessos:

$$d_{(23)1} = \frac{n_2 + n_1}{n_2 + n_3 + n_1} d_{21} + \frac{n_3 + n_1}{n_2 + n_3 + n_1} d_{31} + \frac{-n_1}{n_2 + n_3 + n_1} d_{23}$$

$$= \frac{1+1}{1+1+1} 0{,}8124 + \frac{1+1}{1+1+1} 0{,}9244 + \frac{-1}{1+1+1} 0{,}0160 = 1{,}1525$$

$$d_{(23)4} = \frac{n_2 + n_4}{n_2 + n_3 + n_4} d_{24} + \frac{n_3 + n_4}{n_2 + n_3 + n_4} d_{34} + \frac{-n_4}{n_2 + n_3 + n_4} d_{23}$$

$$= \frac{1+1}{1+1+1} 7{,}7472 + \frac{1+1}{1+1+1} 7{,}3472 + \frac{-1}{1+1+1} 0{,}0160 = 10{,}0576$$

$$d_{(23)5} = \frac{n_2 + n_5}{n_2 + n_3 + n_5} d_{25} + \frac{n_3 + n_5}{n_2 + n_3 + n_5} d_{35} + \frac{-n_5}{n_2 + n_3 + n_5} d_{23}$$

$$= \frac{1+1}{1+1+1} 8{,}2932 + \frac{1+1}{1+1+1} 7{,}8612 + \frac{-1}{1+1+1} 0{,}0160 = 10{,}7643$$

- Nova matriz de dissimilaridade ($\mathbf{d_1^2}$)

$$\mathbf{d_1^2} = \begin{matrix} 1 \\ (2,3) \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 0 & 1{,}1525 & 13{,}1583 & 13{,}8228 \\ & 0 & 10{,}0576 & 10{,}7643 \\ & & 0 & \mathbf{0{,}0171} \\ & & & 0 \end{bmatrix}_{4x4}$$

Etapa 2: Identificar na matriz de dissimilaridade ($\mathbf{d_1^2}$) obtida na etapa anterior os acessos mais similares.

- Grupo II: acessos 4 e 5.
- Nível de fusão: 0,0171
- Distância entre os demais acessos

$$d_{(45)1} = \frac{n_4 + n_1}{n_4 + n_5 + n_1} d_{41} + \frac{n_5 + n_1}{n_4 + n_5 + n_1} d_{51} + \frac{-n_1}{n_4 + n_5 + n_1} d_{45}$$

$$= \frac{1+1}{1+1+1} 13{,}1583 + \frac{1+1}{1+1+1} 13{,}8228 + \frac{-1}{1+1+1} 0{,}0171 = 17{,}9817$$

$$d_{(45)(23)} = \frac{n_4 + n_{23}}{n_4 + n_5 + n_{23}} d_{4(23)} + \frac{n_5 + n_{23}}{n_4 + n_5 + n_{23}} d_{5(23)} + \frac{-n_{23}}{n_4 + n_5 + n_{23}} d_{45}$$

$$= \frac{1+2}{1+1+2} 10{,}0576 + \frac{1+2}{1+1+2} 10{,}7643 + \frac{-2}{1+1+2} 0{,}0171 = 15{,}6079$$

- Nova matriz de dissimilaridade ($\mathbf{d_2^2}$)

$$\mathbf{d_2^2} = \begin{matrix} 1 \\ (2,3) \\ (4,5) \end{matrix} \begin{bmatrix} 0 & \mathbf{1{,}1525} & 17{,}9817 \\ & 0 & 15{,}6079 \\ & & 0 \end{bmatrix}_{3x3}$$

Etapa 3: Idem Etapa 2.

- Grupo III: acessos (2,3) e 1.
- Nível de fusão: 1,1525

- Distância entre o acesso restante (4,5)

$$d_{(2,3,1)(4,5)} = \frac{n_{45} + n_{23}}{n_{23} + n_1 + n_{45}} d_{(2,3)(4,5)} \frac{n_1 + n_{45}}{n_{23} + n_1 + n_{45}} d_{1(4,5)} + \frac{-n_{45}}{n_{23} + n_1 + n_{45}} d_{(2,3,1)}$$

$$= \frac{2+2}{2+1+2} 15{,}6079 + \frac{1+2}{2+1+2} 17{,}9817 + \frac{-2}{2+1+2} 1{,}1525 = 22{,}8143$$

Etapa 4: Reproduzir a etapa anterior até que o útimo indivíduo seja incluído no grupo final (2,3,1,4,5).

Na Tabela 4 são apresentados os resultados do método Ward, quando aplicado à matriz de dissimilaridade expressa pela distância euclidiana quadrática padronizada. Ainda é possível ilustrar as etapas, conforme Figura 3.

Tabela 4. Resultado do agrupamento pelo método Ward, para o exemplo em consideração

Etapa	Fusão		Nível de Fusão	Distância (%)	Números de acessos
	Acesso	Acesso			
1	2	3	0,0160	0,070	2
2	4	5	0,0171	0,075	2
3	(2,3)	1	1,1524	5,050	3
4	(2,3,1)	(4,5)	22,8143	100,00	5

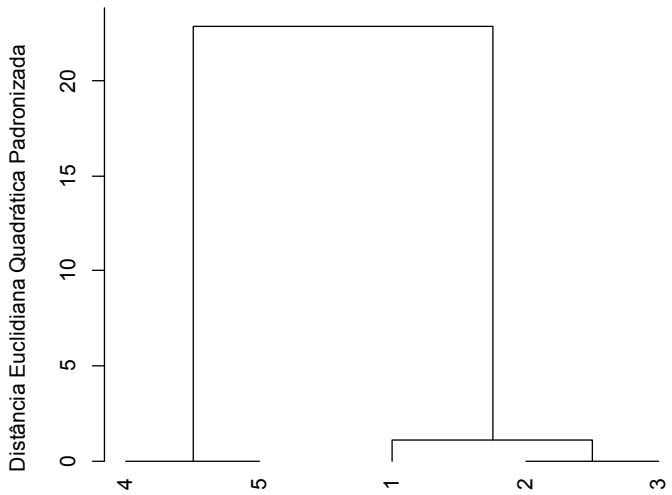

Figura 3. Dendrograma obtido pelo método Ward, a partir das medidas de dissimilaridades entre cinco acessos hipotéticos

3.3.2.3 Método de Tocher

O método requer a obtenção da matriz de dissimilaridade, na 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. Sendo que, a entrada de um indivíduo em um grupo sempre aumenta o valor médio da distância dentro do grupo (CRUZ et al., 2011, CRUZ; CARNEIRO; REGAZZI, 2014).

Assim, a inclusão ou não, do indivíduo k no grupo é validada pela seguinte comparação:

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

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

em que,

$d_{(grupo)k}$: é a distância média entre o indivíduo k e um determinado grupo;

n: é o número de indivíduos que constitui o grupo original;

θ: é o critério de agrupamento ou nível máximo permitido, que corresponde a maior dentre as menores distâncias envolvendo cada acesso.

Onde a distância entre o indivíduo k e o grupo formado pelos indivíduos ij é dada por: $d_{(ij)k} = d_{ik} + d_{jk}$.

Exemplo hipotético: Considerando a matriz de dissimilaridade obtida anteriormente (item 3.3.1.1), pode-se aplicar o método de Tocher, conforme as etapas abaixo:

Etapa 1: Determinar o critério de agrupamento (θ)

Será necessário identificar na matriz de dissimilaridade ($\mathbf{d^2}$) as menores distâncias envolvendo os acessos.

$$\mathbf{d^2} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 0 & 0{,}8124 & 0{,}9244 & 13{,}1583 & 13{,}8228 \\ & 0 & 0{,}0160 & 7.7472 & 8{,}2932 \\ & & 0 & 7{,}3472 & 7{,}8612 \\ & & & 0 & 0{,}0171 \\ & & & & 0 \end{bmatrix}_{5x5}$$

As menores distâncias correspondem a: (1) 0,8124, (2) 0,0160, (3) 0,0160, (4) 0,0171 e (5) 0,0171.

Assim, é estabelecido $\theta = 0{,}8124$ (maior entre as menores distâncias) como o nível máximo para formação ou inclusão de um novo acesso no grupo.

Etapa 2: Identificar na matriz de dissimilaridade ($\mathbf{d^2}$) os acessos mais similares, que irão formar o grupo I e, posteriormente calcular a distância entre este grupo e os demais acessos:

- Acessos: 2 e 3
- Nível de fusão: 0,0160
- Distância entre os demais acessos:

$(d_{23})_1 = d_{21} + d_{31} = 0{,}8124 + 0{,}9244 = 1{,}7368$

$$(d_{23})_4 = d_{24} + d_{34} = 7{,}7472 + 7{,}3472 = 15{,}0944$$

$$(d_{23})_5 = d_{25} + d_{35} = 8{,}2932 + 7{,}8612 = 16{,}1544$$

Com base nas distâncias entre os acessos e o grupo (2,3), verifica-se que o acesso 1 possui menor distância, ou seja, é o mais similar. Avalia-se então a possibilidade de inclusão do mesmo.

$$\frac{(d_{23})_1}{2} = \frac{1{,}7368}{2} = 0{,}8684$$

Logo, $\frac{(d_{23})_1}{2} > \theta$, conclui-se que o acesso 1 não poderá ser incluído no grupoI.

Etapa 3: Obter o grupo II, repetindo a etapa anterior com os acessos não agrupados (1,4 e 5). Considera-se agora a seguinte matriz de dissimilaridade:

$$\mathbf{d^2} = \begin{matrix} 1 \\ 4 \\ 5 \end{matrix}\begin{bmatrix} 0 & 13{,}1583 & 13{,}8228 \\ & 0 & \mathbf{0{,}0171} \\ & & 0 \end{bmatrix}_{3x3}$$

- Acessos: 4 e 5
- Nível de fusão: 0,0171
- Distância entre os demais acessos:

$$(d_{45})_1 = d_{41} + d_{51} = 13{,}1583 + 13{,}8228 = 26{,}9811$$

A inclusão do acesso neste grupo é avaliada da seguinte forma:

$$\frac{(d_{45})_1}{2} = \frac{26{,}9811}{2} = 13{,}4905$$

Como $13{,}4905 > \theta$, o acesso 1 não comporá este grupo. Dessa forma, haverá a formação de um terceiro grupo que incluirá somente o acesso 1.

O estabelecimento de grupos, pelo método Tocher segundo o critério de dissimilaridade expresso pela distância euclidiana quadrática padronizada é apresentado na Tabela 5.

Tabela 5. Grupos de acessos estabelecidos pelo método de Tocher, para o exemplo em consideração

Grupo	Acessos	Distância intragrupo	Distância intergrupo*	θ^{**}
I	2 e 3	0,0160	7,8122	0,8124
II	4 e 5	0,0171	13,4905	
III	1	0,0000		

*Em relação ao grupo posterior; ** Critério de agrupamento simultâneo

Sendo, as distâncias intragrupos (1) e as distâncias intergrupos (2), obtidas conforme as expressões abaixo (CRUZ, 2006):

$$\bar{d}_i = \frac{2\sum_{j>}^{n}\sum_{j'}^{n} d_{jj'}}{n(n-1)} \qquad (1)$$

em que,

j e j': correspondem ao acesso do grupo i e i', respectivamente;

n é o número de acessos dentro do grupo i.

$$\bar{d}_{ii'} = \frac{\sum_{j=1}^{n_1}\sum_{j'=1}^{n_2} d_{jj'}}{n_1 n_2} \qquad (2)$$

em que, n_1 e n_2 correspondem ao número de acessos dentro dos grupos i e i', respectivamente (CRUZ, 2006).

3.3.2.4 Método de Tocher Modificado

Este método, proposto por Vasconcelos et al. (2007), difere do original por adotar critério de aglomeração inverso, de modo que o processo de agrupamento deixa de ser simultâneo e passa a ser sequencial. O grupo inicial é formado da mesma forma que no método original. Já a decisão de incluir um indivíduo no grupo é tomada por meio da comparação entre o acréscimo no valor médio da distância dentro do grupo e um nível máximo (θ_1) da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo.

Na formação do próximo grupo, o procedimento é similar, diferindo apenas pelo fato de que é o valor máximo (θ_2) da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo, porém excluindo as informações daqueles anteriormente agrupados, e assim sucessivamente.

Exemplo hipotético: Considerando a matriz de dissimilaridade (**d**^{2}) obtida anteriormente (item 3.3.1.1), pode-se aplicar o método de Tocher modificado, conforme abaixo:

Etapa 1: Determinar o critério de agrupamento inicial (θ_1)

Identificar a maior entre as menores distâncias envolvendo os acessos, no caso, 0,8124 que corresponderá a θ_1, ou seja, o nível máximo para formação ou inclusão de um novo acesso no grupo.

Etapa 2: Formação do grupo I

Agrupam-se os acessos que apresentam a menor distância. Assim, o primeiro grupo também será constituído pelos acessos 2 e 3 ($d_{23} = 0{,}0160$).

Avalia-se a distância entre o grupo recém-formado e os demais acessos, sendo a menor distância estabelecida entre (2,3) e 1.

$$(d_{23})_1 = d_{21} + d_{31} = 0{,}8124 + 0{,}9244 = 1{,}7368$$

Verifica-se a inclusão ou não do acesso 1, de acordo com o critério estabelecido, no caso, $\frac{d_{(2,3)1}}{2} > (\theta_1)$, logo, o acesso 1 não será incluido ao grupo (2,3).

Etapa 3: Definição do novo critério de agrupamento (θ_2)

Corresponde a maior entre as menores distâncias envolvendo cada acesso, porém excluindo-se as informações do grupo anterior (2,3).

No caso, restando os acessos 1, 4 e 5, cujas menores distâncias são: 13,1583, 0,0171 e 0,0171, respectivamente, o novo critério de agrupamento (θ_2) será 13,1583.

Etapa 4: Formação do grupo II.

Entre os acessos remanescentes, os mais similares são o 4 e 5, cuja distância é 0,0171. Logo, o grupo II será composto por esses acessos.

Avalia-se a distância do acesso 1 em relação ao grupo recém formado.

$$(d_{45})_1 = d_{41} + d_{51} = 13{,}1583 + 13{,}8228 = 26{,}9811$$

A inclusão do acesso 1 ao grupo II não será efetivada, pois $\frac{(d_{45})_1}{2} = 13{,}4905 > \theta_2$.

O processo termina de forma idêntica ao método de Tocher original, com a formação de um terceiro grupo composto somente pelo acesso 1 (Tabela 6).

Tabela 6. Grupos de acessos estabelecidos pelo método de Tocher modificado, para o exemplo em consideração

Grupo	Acessos	Distância intragrupo	Distância intergrupo*	θ_i **
I	2 e 3	0,0160	7,8122	0,8124
II	4 e 5	0,0171	13,4905	13,1583
III	1	0,0000		

*Em relação ao grupo posterior;** Critério de agrupamento sequencial, i=1, 2

3.3.3 Determinação do número de grupos em algoritmos hierárquicos

Existem alguns critérios para a determinação do número final de grupos. São eles: razões práticas do pesquisador, análise visual das ramificações do dendrograma, critérios pré-determinados sob fundamentos teóricos, tais como, Mojena, R^2, Pseudo F, Pseudo T^2, correlação semiparcial e outros (POHLMANN, 2014; MINGOTI, 2005).

O critério de Mojena foi proposto por Mojena (1977) e baseia-se no tamanho relativo dos níveis de fusão (distâncias) no dendrograma, cuja finalidade é determinar um número de grupos que otimize a qualidade do ajuste do agrupamento aos dados (FERREIRA, 2011).

O número de grupos é determinado dado pelo primeiro estágio no dendrograma no qual:

$$\alpha_j > \theta_k$$

$$\theta_k = \bar{\alpha} + kS_\alpha$$

em que,

j: corresponde as etapas do processo de agrupamento (j=1, 2,..., n-1, sendo n referente ao número de acessos);

α_j: corresponde aos níveis de fusão;

$\bar{\alpha}$ e S_α: corresponde a média e o desvio padrão, respectivamente, dos α's. Obtidas conforme as expressões abaixo:

$$\bar{\alpha} = \frac{1}{n-1}\sum_{j=1}^{n}\alpha_j \qquad S_\alpha = \sqrt{\frac{\sum_{j=1}^{n}\alpha^2 \frac{1}{n-1}(\sum_{j=1}^{n}\alpha_j)^2}{n-2}}$$

onde,

k: é uma constante, conforme Milligan e Cooper (1985) deve assumir valor 1,25, baseado em simulação.

Exemplo hipotético: Considerando a matriz de distância euclidiana quadrática padronizada (item 3.3.1.1) e os métodos de agrupamento UPGMA e Ward, itens 3.3.2.1 e 3.3.2.2, respectivamente, determinaremos o número de grupos por meio do critéro de Mojena.

Caso 1: Método UPGMA, considerando a Etapa 4.

$$\theta_k = \bar{\alpha} + kS_\alpha$$

$$\bar{\alpha} = \frac{0{,}0160 + 0{,}0171 + 0{,}8683 + 9{,}7050}{4} = 2{,}6516$$

$$S_\alpha = \sqrt{\frac{0{,}0160^2 + \ldots + 9{,}7050^2 - \frac{(0{,}0160 + \ldots + 9{,}7050)^2}{4}}{3}} = \sqrt{22{,}2725} = 4{,}7194$$

$$\therefore \theta_k = 2{,}6516 + 1{,}25 \times 4{,}7194 = 8{,}5508$$

A determinação do número de grupos, confome Mojena aplicada ao método UPGMA é relatada na Tabela 7.

Tabela 7. Resultado da determinação do número de grupos pelo método UPGMA, para o exemplo em consideração

Etapa	α	$\bar{\alpha}$	S_α	$\theta_k\,(k=1{,}25)$	$\theta_k\,(k=1{,}00)$
1	0,0160	---	---	---	---
2	0,0171	---	---	---	---
3	0,8684	0,3005	0,4918	0,9152	0,7922
4	9,7050	2,6516	4,7194	8,5508*	7,3710

*ponto de corte no dendrograma

Caso 2: Método Ward, considerando a Etapa 4.

$$\theta_k = \bar{\alpha} + kS_\alpha$$

$$\bar{\alpha} = \frac{0{,}0160 + 0{,}0171 + 1{,}1524 + 22{,}8143}{4} = 5{,}9999$$

$$S_\alpha = \sqrt{\frac{0{,}0160^2 + \ldots + 22{,}8143^2 - \frac{(0{,}0160 + \ldots + 22{,}8143)^2}{4}}{3}} = \sqrt{125{,}9400} = 11{,}2223$$

$$\therefore \theta_k = 5{,}9999 + 1{,}25 \times 11{,}2223 = 20{,}0279$$

Em suma, os resultados da determinação pelo método Ward são apresentados na Tabela 8.

Tabela 8. Resultado da determinação do número de grupos pelo método Ward, para o exemplo em consideração

Etapa	α_j	$\bar{\alpha}$	S_α	$\theta_k\,(k=1{,}25)$	$\theta_k\,(k=1{,}00)$
1	0,0160	---	---	---	---
2	0,0171	---	---	---	---
3	1,1524	0,3952	0,4300	0,9327	0,8252
4	22,8143	5,9999	11,2223	20,0279*	17,2223

*ponto de corte no dendrograma

Na Figura 4 são apresentados os dendrogramas produzidos com a utilização dos métodos UPGMA (1) e Ward (2), e o ponto de corte através do método proposto acima.

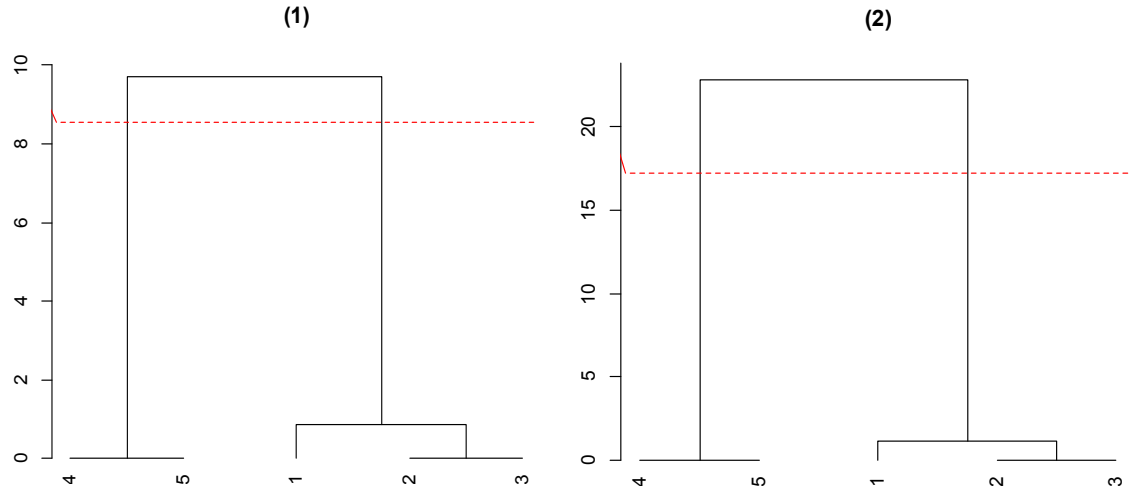

Figura 4. Dendrogramas referentes aos métodos UPGMA (1) e Ward (2), com ênfase no ponte de corte determinado pelo critério de Mojena, conforme exemplo hipotético

3.3.4 Validação do agrupamento em algoritmos hierárquicos

Existem diversos métodos de validação e, dentre eles, destaca-se o critério coeficiente de correlação cofenética (r) que mede o grau de ajuste entre a matriz de dissimilaridade (D) e a matriz cofenética (C), ou seja, a preservação das distâncias resultante do agrupamento em relação às distâncias originais. A matriz cofenética ou de distâncias recuperadas, corresponde a matriz resultante da simplificação proporcionada pelo respectivo método de agrupamento (SNEATH; SOKAL, 1973).

A ideia é obter a correlação entre os elementos acima da diagonal das matrizes de dissimilaridade (D) e da matriz cofenética (C), sendo calculado a partir da seguinte expressão:

$$r = \frac{\hat{Cov}(D,C)}{\sqrt{\hat{V}(D)\hat{V}(C)}} = \frac{\sum_{i=1}^{n} d_i c_i - \frac{\sum_{i=1}^{n} d_i \sum_{i=1}^{n} c_i}{n}}{\sqrt{\sum_{i=1}^{n} d_i^2 - \frac{(\sum_{i=1}^{n} d_i)^2}{n}} \sqrt{\sum_{i=1}^{n} c_i^2 - \frac{(\sum_{i=1}^{n} c_i)^2}{n}}}$$

em que,

r: coeficiente de correlação;

c_i : matriz cofenética;

d_i : matriz de dissimilaridade.

Quanto mais próximo de 1 for o coeficiente de correlação, menor será a distorção ocasionada pelo agrupamento dos acessos face à utilização de determinado método.

Exemplo hipotético: Considerando a matriz de distâncias euclidinas padronizadas quadráticas obtidas anteriormente (item 3.3.1.1), e os métodos de agrupamento UPGMA (item 3.3.2.1) e Ward (item 3.3.2.2) obteremos os respectivos coeficientes de correlação cofenética.

Caso 1: Método UPGMA

- Matriz cofenética (com base nos níveis de fusão)

$$\mathbf{C} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 0 & 0{,}8684 & 0{,}8684 & 9{,}7050 & 9{,}7050 \\ & 0 & 0{,}0160 & 9{,}7050 & 9{,}7050 \\ & & 0 & 9{,}7050 & 9{,}7050 \\ & & & 0 & 0{,}0171 \\ & & & & 0 \end{bmatrix}_{5x5}$$

- Matriz de dissimilariade ($\mathbf{d^2}$)

$$\mathbf{d^2} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 0 & 0{,}8124 & 0{,}9244 & 13{,}1583 & 13{,}8228 \\ & 0 & 0{,}0160 & 7{,}7472 & 8{,}2932 \\ & & 0 & 7{,}3472 & 7{,}8612 \\ & & & 0 & 0{,}0171 \\ & & & & 0 \end{bmatrix}_{5x5}$$

- Coeficiente de correlação cofenética (r)

$$r = \frac{\sum_{i=1}^{n} d_i c_i - \frac{\sum_{i=1}^{n} d_i \sum_{i=1}^{n} c_i}{n}}{\sqrt{\sum_{i=1}^{n} d_i^2 - \frac{(\sum_{i=1}^{n} d_i)^2}{n}} \sqrt{\sum_{i=1}^{n} c_i^2 - \frac{(\sum_{i=1}^{n} c_i)^2}{n}}}$$

onde,

$$\sum_{i=1}^{n} c_i d_i = \left[(0{,}8684 \times 0{,}8124) + (0{,}8684 \times 0{,}9244) + \ldots + (0{,}0171 \times 0{,}0171)\right] = 566{,}6299$$

$$\sum_{i=1}^{n} c_i = (0{,}8684 + 0{,}8684 + \ldots + 0{,}0171) = 59{,}9999 \; ; \; \sum_{i=1}^{n} c_i^2 = 566{,}6309$$

$$\sum_{i=1}^{n} d_i = (0{,}8124 + 0{,}9244 + \ldots + 0{,}0171) = 59{,}9998 \; ; \; \sum_{i=1}^{n} d_i^2 = 610{,}3018$$

$$\therefore r = \frac{566{,}6299 - 359{,}7558}{\sqrt{206{,}6321}\sqrt{250{,}3042}} = 0{,}9086$$

Caso 2: Método Ward, $r = 0{,}9083$.

Obs.: Os scripts das análises descritas anteriormente são apresentados no Apêndice A.

4 RESULTADOS E DISCUSSÃO

Inicialmente, as estimativas das distâncias mostram que o menor grau de dissimilaridade baseado na distância euclidiana quadrática padronizada, foi obtido entre os acessos 1 (BGH 1716) e 5 (BGH 6228), com $d^2 = 1{,}78$. Em contrapartida, os acessos 2 (BGH 4289) e 10 (BGH 6233) apresentaram o maior grau de dissimilaridade, com distância de 72,16 (Tabela 9).

Tabela 9. Medidas de dissimilaridade baseada na distância euclidiana quadrática padronizada (d^2) dos 11 acessos de *Capsicum chinense* baseada em 11 caracteres quantitativos

	2	3	4	5	6	7	8	9	10	11
1	8,79	10,58	2,24	1,78	8,46	8,45	7,36	24,58	41,89	27,82
2		34,11	6,41	15,14	13,13	21,05	23,35	48,09	72,16	47,13
3			16,85	6,68	24,11	12,89	9,56	13,43	23,90	24,72
4				4,17	6,92	11,84	9,53	34,65	52,59	35,65
5					11,43	6,09	4,22	19,63	36,46	26,31
6						25,07	22,02	50,08	68,54	52,12
7							10,92	11,07	26,70	20,67
8								21,79	33,93	20,22
9									10,40	11,05
10										11,19

A dissimilaridade encontrada é condizente com as diferenças médias dos caracteres entre os acessos em estudo, com ênfase nos seguintes caracteres: peso total do fruto (PTF), massa total do fruto fresco (MTF), percentagem de matéria seca (MTS), sólidos solúveis totais (SST), comprimento do fruto (CMP), largura do fruto (LAR) e pungência dos frutos (PUN).

Observa-se que os acessos mais similares são provenientes de regiões geográficas distantes, (1) Pindaré-Mirim/MA e (5) Brasília/DF; e a menor similaridade foi obtida entre acessos pertencentes à mesma região geográfica, (2) Rondonópolis/MT e (10) Brasília/DF; ou seja, não se verifica uma relação entre distância geográfica e distância genética, conforme observado por Moura et al. (2010), decorrentes dos

processos de dispersão desta cultura pelos animais (pássaros, morcegos, entre outros) e seres humanos.

A distância média das medidas de dissimilaridade observada entre os acessos foi $22{,}00 \pm (16{,}50)$. Faria et al. (2012) analisando 49 acessos de pimenta, com base em 10 descritores, obtiveram distâncias que variam entre 9,47 a 463, 59, com média de 236,53. Entretanto, Moura et al. (2010) observaram que entre os 56 acessos de *C. chinense*, com base em 43 caracteres, a distância média obtida foi de $0{,}40 \pm (0{,}08)$. Essa variação de distâncias é influenciada pela origem dos acessos; pela medida de dissimilaridade adotada, aliada aos caracteres envolvidos, dentre outros fatores.

Adotando o método de agrupamento de ligação média entre grupos (UPGMA) baseado na distância euclidiana quadrática padronizada (d^2), a ordem de agrupamento observada é apresentada na Tabela 10.

Tabela 10. Agrupamento dos 11 acessos de *C. chinense* do BGH/UFV segundo o método UPGMA expressa pela distância euclidiana quadrática padronizada

Etapas	Fusão		Nível de Fusão	Distância (%)	Número de acessos
	Acesso	Acesso			
1	1	5	1,78	5,12	2
2	(1,5)	4	3,20	9,20	3
3	(1,5,4)	8	7,04	20,25	4
4	(1,5,4,8)	7	9,33	26,84	5
5	9	10	10,40	29,92	2
6	(9, 10)	11	11,12	31,99	3
7	(1,5,4,8,7)	3	11,31	32,54	6
8	2	6	13,13	37,77	2
9	(1,5,8,7,3)	(2,6)	17,24	49,60	7
10	(1,5,4,8,7,3,2,6)	(9,10,11)	34,76	100,00	11

Posteriormente, a definição do número de grupos, com base no critério de Mojena (1977) estabeleceu o ponto de corte (θ) em 23,45. Assim, os acessos foram divididos em dois grupos, como pode ser visto na Figura 5.

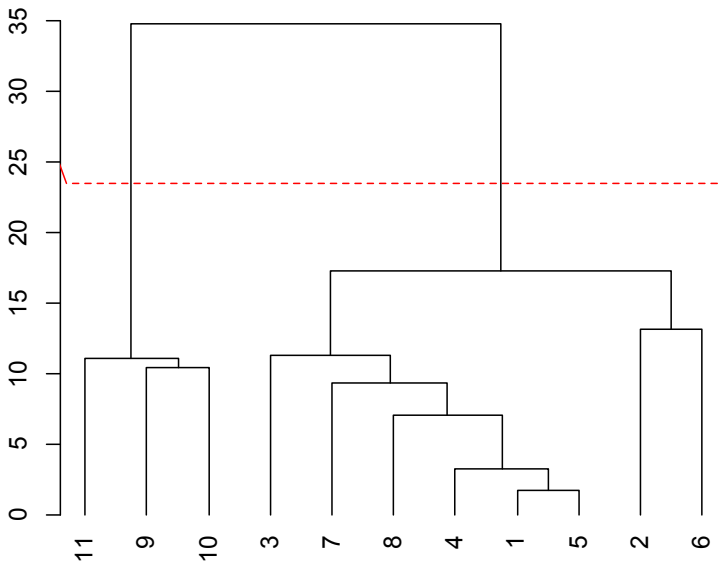

Figura 5. Dendrograma estabelecido pelo método UPGMA, baseado na distância euclidiana quadrática padronizada, delimitado pelo ponte de corte (θ)

O grupo I reuniu oito acessos (1, 2, 3, 4, 5, 6, 7, 8), ou seja, 72,73% dos indivíduos e o grupo II englobou três acessos (9 10 e 11), com 27,27% dos indivíduos (Figura 5).

Em estudos sobre dissimilaridade genética em pimenteiras foram encontrados resultados próximos para o método UPGMA, conforme Souza et al. (2015) utilizando 49 acessos de *C. chinense* constituíram três grupos pelo método UPGMA, adotando a distância de Mahalanobis, baseado em 23 parâmetros. Vasconcelos et al. (2014) analisando 22 acessos de *C. chinense*, também formaram três grupos pelo método UPGMA, utilizando a distância de euclidiana, baseado em 6 caracteres quantitativos. O primeiro grupo composto por 19 acessos, o segundo por dois acessos e o terceiro exclusivamente um acesso.

No presente trabalho, no grupo I observam-se menores valores médios para os caracteres: peso total do fruto (874,75 g), massa total do fruto fresco (3,21 g), massa total do fruto seco (0,49 g), largura do fruto (14,52 mm), comprimento do fruto (33,89 mm) e espessura do fruto (1,86 mm); e maiores médias nos quesitos pungência dos frutos (4,04 mg.gMS^{-1}), percentagem de matéria seca (16,44%) e sólidos solúveis totais (9,81 ºBrix).

Os acessos pertencentes ao grupo II apresentaram pequenas variações internas, sendo semelhante para os seguintes caracteres, com maiores valores médios, peso total

do fruto (1475,83 g), massa total do fruto fresco (11,23 g), massa total do fruto seco (1,26 g), largura do fruto (28,81 mm) e espessura do fruto (2,85 mm).

O método Ward apresentou as seguintes etapas de agrupamento (Tabela 11).

Tabela 11. Agrupamento dos 11 acessos de *C. chinense* do BGH/UFV segundo o método Ward expressa pela distância euclidiana quadrática padronizada

Etapas	Fusão		Nível de Fusão	Distância (%)	Número de acessos
	Acesso	Acesso			
1	1	5	1,78	1,58	2
2	(1,5)	4	3,68	3,27	3
3	(1,5,4)	8	9,19	8,17	4
4	9	10	10,40	9,25	2
5	(9,10)	11	11,35	10,09	3
6	(1,5,4,8)	7	11,99	10,66	5
7	2	6	13,13	11,68	2
8	(1,5,4,8,7)	3	14,41	12,81	6
9	(1,5,8,7,3)	(2,6)	31,61	28,11	8
10	(1,5,4,8,7,3,2,6)	(9,10,11)	112,44	100,00	11

Adotando o critério Mojena (1977) houve a delimitação em $\theta = 62{,}96$. Dessa forma, o método reuniu os acessos em dois grupos; o grupo I composto por oito acessos (1, 2, 3, 4, 5, 6, 7, 8) e o grupo II por três acessos (9 10 e 11), de modo semelhante ao método UPGMA (Figura 6).

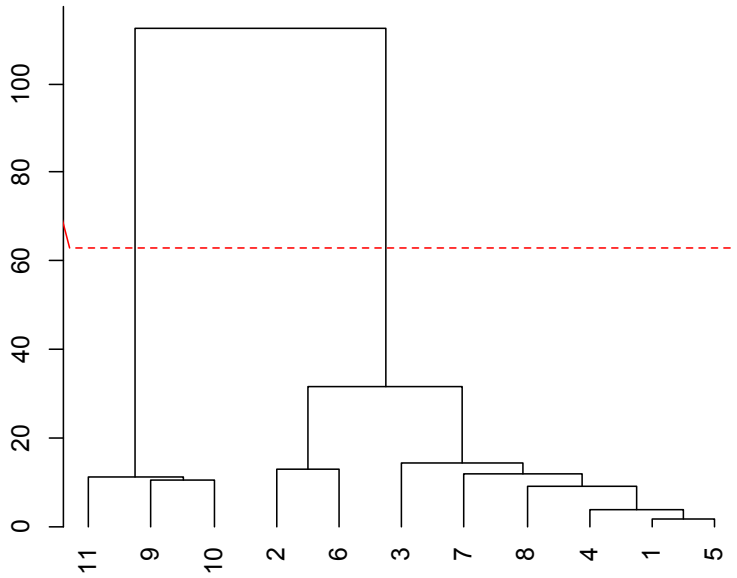

Figura 6. Dendrograma referente ao método Ward, baseado na distância euclidiana quadrática padronizada, delimitado pelo ponte de corte (θ)

Embora as estruturas dos agrupamentos sejam similares entre os dois métodos, pode-se observar que há alterações nos níveis de fusão (distâncias) em que os acessos são agrupados, face às especificidades inerentes a cada método.

Avaliando a solução de agrupamentos, por meio dos coeficientes de correlação cofenética para os métodos UPGMA e Ward obteve-se os valores de 0,72 e 0,71, respectivamente, isso mostra que o ajuste entre a matriz de dissimilaridade (D) e a matriz cofenética (C) pode ser considerado adequado, conforme Rohlf (1970) e Vaz Patto et al. (2004).

Segundo Rohlf (1970), na prática, coeficientes menores que 0,7 refletem a inadequação do método de agrupamento para resumir a informação do conjunto de dados. Vaz Patto et al. (2004) determina que $r \geq 0{,}56$ é considerado adequado, por retratar menor distorção provocada pelo agrupamento. Logo, independente do método utilizado, a representação é válida, contribuindo para o aumento da confiabilidade das conclusões frente à interpretação dos dendrogramas, ou seja, os grupos realmente diferem uns dos outros.

O método de agrupamento Tocher (RAO, 1952) permitiu a formação de seis grupos distintos para os caracteres quantitativos, sendo quatro deles compostos por um único acesso (Tabela 12).

Tabela 12. Agrupamento dos 11 acessos de *C. chinense* do BGH/UFV segundo o método Tocher expressa pela distância euclidiana quadrática padronizada

Grupos	Acessos	Distância intragrupo	Distância intergrupo*	θ^{**}
I	1,5,4,8,7	6,6614	30,3307	11,05
II	9,10	10,4057	11,1180	-
III	11	0,0000	24,7238	-
IV	3	0,0000	24,1134	-
V	6	0,0000	13,1301	-
VI	2	0,0000	-	-

*Em relação ao grupo posterior; ** Critério de agrupamento simultâneo

No grupo I estão incluídos os acessos 1 (BGH 1716), 5 (BGH 6228), 4 (BGH 5012), 8 (BGH 4201) e 7 (BGH 1716); no grupo II, os acessos 9 (BGH 4223) e 10 (BGH 6233) e nos demais grupos: III, IV, V e VI, apenas um acesso, conforme demonstrado na Tabela 12.

Pesquisas anteriores apresentam resultados semelhantes, a saber: Vasconcelos et al. (2014), em 22 acessos utilizando o método Tocher obteveram quatro grupos, sendo o grupo I constituído por 15 acessos, o grupo II por cinco acessos, o grupo III e IV por apenas um acesso cada. Faria et al. (2012), analisando 49 acessos, baseado em 10 caracteres, pelo método Tocher distribuiu os acessos em quatro grupos.

A determinação do número final de grupos, pelo método Tocher, por adotar um critério de agrupamento fixo, em certos casos, culminará com um maior número de grupos, conforme observado. Resultados semelhantes foram encontrados por Vasconcelos et al. (2014) e Faria et al. (2012). Portanto, esse método possui maior poder de discriminação, permitindo a identificação de mais grupos contendo acessos similares.

Na aplicação da técnica de Tocher modificado, os acessos foram separados em quatro grupos, o grupo I reuniu cinco acessos (45,45%), de forma idêntica ao método Tocher original; o grupo II reuniu três acessos (27,27%) sendo constituído pelos mesmos acessos do grupo II para os métodos UPGMA e Ward, o grupo III reuniu dois acessos (18,18%) e o grupo IV um único acesso (Tabela 13).

Tabela 13. Agrupamento dos 11 acessos de *C. chinense* do BGH/UFV segundo o método Tocher modificado expressa pela distância euclidiana quadrática padronizada

Grupos	Acessos	Distância intragrupo	Distância intergrupo*	θ_i^{**}
I	1,5,4,8,7	6,6614	28,9327	11,05
II	9,10,11	10,8805	56,3538	13,43
III	2,6	13,1301	29,1129	24,11
IV	3	0,0000	-	-

*Em relação ao grupo posterior; ** Critério de agrupamento sequencial, i=1, 2, 3

Como exposto, o método Tocher modificado proporcionou a redução do número de grupos, cerca de 33% em relação ao método Tocher, pois utiliza um critério de agrupamento que possui proporcionalidade quanto à dissimilaridade existente entre os acessos do grupo, ou seja, que varia de acordo com os acessos remanescentes decorrente do aumento do valor criterial ao longo do processo de agrupamento.

Esse método, além de manter a característica de menor distância intragrupo em relação a intergrupo proporcionou um agrupamento dos acessos mais próximos em contrapartida ao Tocher, por exemplo, a união do acesso 11 ao grupo II (9, 10), além da formação de novo grupo com os acessos 2 e 6. Esses resultados eram esperados pela proximidade das medidas de distâncias entre os acessos envolvidos.

Analisando os métodos de otimização, observa-se que o acesso 3 (BGH 4733) mostrou-se divergente dos demais visto que formou um grupo exclusivo.

De modo geral, todos os acessos que constituíram os pares mais dissimilares com base na distância euclidiana quadrática padronizada, foram designados a grupos distintos para os métodos adotados (UPGMA, Ward, Tocher e Tocher modificado).

Por fim, detectou-se divergência genética entre os acessos de pimenta, conforme exposto em outros trabalhos com *C. chinense*, tais como, Souza et al. (2015), Vasconcelos et al. (2014), Souza, C. et al. (2013), Faria et al. (2012), Finger et al. (2010), Moura et al. (2010). Segundo Souza et al. (2015), a análise da divergência entre acessos da mesma espécie constituem uma opção para identificar características efetivas em diversos processos, tais como, regulação do crescimento, biomassa e produção de frutos.

5 CONCLUSÕES

Os métodos hierárquicos, UPGMA e Ward, apresentaram resultados concordantes formando dois grupos distintos; o grupo I, composto por 72,73% dos acessos (1, 2, 3, 4, 5, 6, 7 e 8) e o grupo II por 27,27% dos acessos (9, 10 e 11); também obtiveram coeficientes de correlação cofenética (r) similares e suficientes para validação das estruturas de agrupamento.

Os métodos não hierárquicos, Tocher e Tocher modificado, formaram maior número de grupos, seis e quatro, respectivamente. Adotando o método de Tocher o grupo I reuniu os acessos 1, 5, 4, 8 e 7, o grupo II os acessos 9 e 10, e nos demais grupos (III, IV, V e VI) um único acesso cada. O método Tocher modificado elencou os acessos nos seguintes grupos: I (1, 5, 4, 8, 7), II (9, 10, 11), III(2, 6) e IV (3).

As diferenças de tamanhos de grupos obtidos pelos processos hierárquicos e não hierárquicos, pode ser associado à restrição dos resultados imposta pelos hierárquicos ao impedir a redesignação de acessos gerando agrupamentos que explicam menos variações. Dentre os métodos empregados, o de Tocher mostra-se mais discriminatório.

Por fim, independente do método adotado, foi possível identificar os acessos mais divergentes e contribuir para futuras pesquisas de cruzamento buscando híbridos com maior efeito heterótico.

6 REFERÊNCIAS BIBLIOGRÁFICAS

BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 105p., 1990.

COSTA, E. N.; SOUZA, B. H. S. de; BOTTEGA, D. B.; OLIVEIRA, F. Q. de; RIBEIRO, Z. A.; BOIÇA JÚNIOR, A. L. Genetic divergence of bean genotypes to infestation of *Zabrotes subfasciatus* (Bohemann) (Coleoptera: Bruchidae). **Semina**: Ciências Agrárias, Londrina, v.34, n.6, p.2737-2752, 2013.

COSTA, F. R.; PEREIRA, T. N. S.; VITÓRIA, A. P.; CAMPOS, K. P. de; RODRIGUES, R.; SILVA, D. H. da; PEREIRA, M. G. Genetic diversity among *Capsicum* accessions using RAPD markers. **Crop Breeding and Applied Biotechnology**, Viçosa, v.6, n.1, p.18-23, 2006.

CRUZ, C.D. GENES: a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**, v.35, n.3, p.271-276, 2013.

CRUZ, C. D. **Programa genes: análise multivariada e simulação**. Viçosa: Ed. UFV, 175p., 2006.

CRUZ, C.D.; CARNEIRO, P.C.S.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: Ed UFV, v.2, 668p., 2014.

CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Ed. Suprema, 620p., 2011.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 4. ed. Viçosa: Ed. UFV, v. 1, 480p., 2012.

DUTRA FILHO, J. de A.; OLIVEIRA, L. J.; MELO, T. de.; RESENDE, L. V.; ANUNCIAÇÃO FILHO, J. da; BASTOS, G. Q. Aplicação de técnicas multivariadas no estudo da divergência genética em cana-de-açúcar. **Revista Ciência Agronômica**, Fortaleza, v.42, n.1, p.185-192, 2011.

EDWARDS, A. W. F; CAVALLI-SFORZA, L. L. A method for cluster analysis. **Biometrics**, v.21, n.2, p.362–375, 1965.

EMBRAPA. **Pimenta** (*Capsicum* spp.). Disponível em: < http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Pimenta/Pimenta_capsicum_spp/botanica.html > Acesso em: 10/05/2015.

FARIA, P. N.; CECON, P. R.; FINGER, F. L.; SILVA, A. R.; SILVA, F. F.; CRUZ, C. D.; SAVIO, F. L. Métodos de agrupamento em estudo de divergência genética de pimentas. **Horticultura Brasileira**, Brasília, v. 30, n.3, p. 428-432, 2012.

FARIA, P. N.; LAIA, G. A.; CARDOSO, K. A.; FINGER, F. L.; CECON, P. C. Estudo da variabilidade genética de amostras de pimenta (*Capsicum chinense* Jacq.) existentes num banco de germoplasma: um caso de estudo*. **Revista de Ciências Agrárias**, Recife, v.36, n.1, p.17- 22, 2013.

FERREIRA, D. F. **Estatística multivariada**. 2. ed. rev. ampl. Lavras: Ed. UFLA, 676p., 2011.

FINGER, F. L.; LANNES, S. D.; SCHUELTER, A. R.; DOEGE, J.; COMERLATO, A. P.; GONÇALVES, L. S. A.; FERREIRA, F. R. A.; CLOVIS, L. R.; SCAPIM, C. A. Genetic diversity of *Capsicum chinense* (Solanaceae) accessions based on molecular markers and morphological and agronomic traits. **Genetics and Molecular Research**, v.9, n.3, p.1852-1864, 2010.

FOOD AND AGRICULTURE ORGANIZATION (FAO). **Faostat**: premilinary 2011. Disponível em: <http://faostat.fao.org/site/567/DesktopDefault.aspx?Page ID=567#ancor>.Acesso em 03/09/2015.

GALATE, R. dos S.; MOTA, M. G. da C.; GAIA, J. M. D.; COSTA, M. do S. S. Morphoagronomic characterization of assai palm's germplasm from Eastern Amazon. **Revista Brasileira de Fruticultura**, Jaboticabal, v.34, n.2, p.540-551, 2012.

HAIR JÚNIOR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 688p., 2009.

HAMAWAKI, O. T.; SOUSA, L. B. de; ROMANATO, F. N.; NOGUEIRA, A. P. O.; SANTO JÚNIOR, C. D.; POLIZEL, A. C. Genetic parameters and variability in soybean genotypes. **Comunicata Scientiae**, Teresina, v.3, n.2, p.76- 83, 2012.

LANCE, G. N.; WILLIAMS, W. T. A general theory of classificatory sorting strategies: 1. hierarchical systems. **The Computer Journal**, v.9, n.4, p. 373-380, 1967.

JARA, P. G.; LETELIER, A. M.; FRAILE, A.; PIÑERO, D.; ARENAL, F. G. Impact of human management on the genetic variation of wild pepper, *Capsicum annuum* var. *glabriusculum*. **PLoS ONE**, v. 6, n.12: e28715, 2011.

LOPES, L. A.; PELUZIO, J. M.; AFFÉRRI, F. S.; CARVALHO, E. V. de. Variabilidade genética entre cultivares de soja, quanto ao rendimento de óleo, no estado do Tocantins. **Comunicata Scientiae**, Teresina, v.5, n.3, p.279-285, 2014.

MARTINS, K. C.; SOUZA, S. A. M.; PEREIRA, T. N. S.; RODRIGUES, R.; PEREIRA, M. G.; CUNHA, M. da. Palynological characterization and genetic divergence between accessions of chilli and sweet peppers. **Horticultura Brasileira**, Brasília, v.31, n.4, p.568-573, 2013.

MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of cluster in a data set. **Psychometrika**, New York, v.50, p.159-179, 1985.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: Editora UFMG, 297p., 2005.

MOJENA, R. Hierárquical grouping method and stopping rules: an evaluation. **Computer Journal**, v.20, p.359-363, 1977.

MOURA, M. C. C. L.; GONÇALVES, L. S. A.; SUDRÉ, C. P.; RODRIGUES, R.; AMARAL JÚNIOR, A. T.; PEREIRA, T. N. S. Algoritmo de Gower na estimativa da divergência genética em germoplasma de pimenta. **Horticultura Brasileira**, Brasília, v.28, n.2, p.155-161, 2010.

NEITZKE R. S. **Recursos genéticos de pimentas do gênero Capsicum: explorando a multiplicidade de usos**. Tese (Doutorado em Fitomelhoramento) - Universidade Federal de Pelotas. Orientadora: Rosa Lía Barbieri. 115f., 2012.

NEITZKE, R. S.; VASCONCELOS, C. S.; BARBIERI, R. L.; VIZZOTTO, M.; FETTER, M. R.; CORBELINI, D. D. Variabilidade genética entre compostos antioxidantes em variedades crioulas de pimentas (*Capsicum baccatum*). **Horticultura Brasileira**, Brasília, v.33, n.4, p.415-421, 2015.

OLIBONI, R.; FARIA, M. V.; NEUMANN, M.; BATTISTELLI, G. M.; TEGONI, R. G.; RESENDE, J. T. V. de. Genetic divergence among maize hybrids and correlations with heterosis and combining ability. **Acta Scientiarum. Agronomy**, Maringá, v.34, n.1, p.37-44, 2012.

OLVERA, A. P.; VERDUGO, S. H.; RAMÍREZ, V. R.; RODRÍGUEZ, A. G.; OYAMA, K. Genetic diversity and structure of pepper (*Capsicum annuum* L.) from northwestern Mexico analyzed by microsatellite markers. **Crop Science Society of America**, Madison, v.52, n.1, p.231-241, 2012.

PELUZIO, J. M.; PIRES, L. P. M.; CANCELLIER, L. L.; AFFÉRRI, F. S.; COLOMBO, G. A.; TEXEIRA JÚNIOR, T.; RIBEIRO, G. R. dos S. Genetic divergence among soybean cultivars in irrigated lowland in the State of Tocantins. **Ciência Rural**, Santa Maria, v.42, n.3, p.395-400, 2012.

PEREIRA, T.; COELHO, C. M. M.; SANTOS, J. C. P. dos; BOGO, A.; MIQUELLUTI, D. J. Diversidade no teor de nutrientes em grãos de feijão crioulo no Estado de Santa Catarina. **Acta Scientiarum. Agronomy**, Maringá, v.33, n.3, p.477-485, 2011.

PESSOA, A. M. S.; BARROSO, P.A., RÊGO, E. R. do; MEDEIROS, G. A.; BRUNO, R. L.A; RÊGO, M.M. Genetic divergence of physiological-quality traits of seeds in a population of peppers. **Genetics and Molecular Research**, v.14 n.4, p.12479-12488, 2015.

POHLMANN, M. C. Análise de Conglomerados. In: CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. (Coord.). **Análise multivariada**: para os cursos de administração, ciências contábeis e economia. 1. ed. São Paulo: Atlas, 568p., 2014.

PORTO, F. R. C.; SILVA, J. C. Etnobotânica e uso medicinal da pimenta malagueta (Capsicum frutescens) pelos horticultores e consumidores da horta comunitária da Vila Poty, Teresina, Piauí, Brasil. **Revista FSA**: Periódico Científico da Faculdade Santo Agostinho, v. 9, n. 1, p. 139-152, 2013.

R DEVELOPMENT CORE TEAM. **R**: A language and environment for statistical computing. Vienna, Austria: R Fundation for Statistical Computing, 2015. Disponível em: <http://r(project .org>.

RAI, V. P; KUMAR, R.; KUMAR, S.; RAI, A.; KUMAR, S.; SINGH, M.; SINGH, S. P.; RAI, A. B.; PALIWAL, R. Genetic diversity in *Capsicum* germplasm based on microsatellite and random amplified microsatellite polymorphism markers. **Physiology and Molecular Biology of Plants**, New Delhi, v.19, n.4, p.575-586, 2013.

RAO C.R. **Advanced statistical methods in biometric research**. New York: John Wiley & Sons, 390p., 1952.

REIFSCHNEIDER, F. J. B. (Org.). **Capsicum**: pimentas e pimentões no Brasil. Brasília: Embrapa Comunicação para Transferência de Tecnologia: Embrapa Hortaliças, 2000.

RIGON, J. P. G.; CAPUANI, S.; RIGON, C. A. G. Genetic divergence among maize hybrids by morphological descriptors. **Bragantia**, Campinas, v.74, n.2, p.156-160, 2015.

ROHLF, F. J. Adaptive Hierarchical clustering schemes. **Systematic Zoology**, v.18, p.58-82, 1970.

SILVA, J. M. da; AGUIAR, A. V. de ; MORI, E. S.; MORAES, M. L. T. de. Divergência genética entre progênies de *Pinus caribaea* var. *caribaea* com base em caracteres quantitativos. **Pesquisa Florestal Brasileira**, Colombo, v.32, n.69, p.69-77, 2012.

SNEATH, P. H. A.; SOKAL, R.R. **Numerical taxonomy**. New York: Hafner, 1975.

SOFI, P. A.; ZARGAR, M. Y.; SHEIKH, F.A..; IRAM, S., SHAFI, T. Genetic variability and factor analysis in common bean (*Phaseolus vulgaris* L.) germplasm collection for yield related traits. **Electronic Journal of Plant Breeding,** Coimbatore, v.5, n.2, p.254-259, 2014.

SOUSA, L. B. de; SILVA, E. M.; GOMES, R. L. F.; LOPES, A. C. de A.; SILVA, I. C. V. Characterization and genetic divergence of access of *Passiflora edulis* and *P. cincinnata* based on physical and chemical characteristics of fruits. **Revista Brasileira de Fruticultura,** Jaboticabal, v.34, n.3, p.832-839, 2012.

SOUZA, C. S. de; SCHUELTER, A. R.; FINGER, F. L.; CASALI, V. W. D. Variabilidade genética em acessos de *Capsicum chinense* por meio de marcadores isoenzimáticos e RAPD. **Scientia Agraria**, Curitiba, v.14, n.1, p.9-23, 2013.

SOUZA, J. R. de; BOIÇA JÚNIOR, A. L.; PERECIN, D.; CARGNELUTTI FILHO, A.; COSTA, J. T. da. Divergência genética de cultivares de cana-de-açúcar quanto à resistência a *Diatraea saccharalis*.Semina: **Ciências Agrárias**, Londrina, v.34, n.6, p.3367-3376, 2013.

SOUZA, L. R. de; SCOSSA, F.; CHAVES, I. S.; KLEESSEN, S.; SALVADOR, L. F. D.; MILAGRE, J. C.; FINGER, F.L.; BHERING, L. L.; SULPICE, R.; ARAÚJO, W. L.; NIKOLOSKI, Z.; FERNIE, A. R.; NESI, A. N. Exploring natural variation of

photosynthetic, primary metabolism and growth parameters in a large panel of *Capsicum chinense* accessions. **Planta**, Heidelberg, v. 242, n.3, p.677-691, 2015.

SUDRÉ, C. P. RODRIGUES, R.; RIVA, E. M.; KARASAWA, M.; AMARAL JÚNIOR, A. T. Divergência genética entre acessos de pimenta e pimentão utilizando técnicas multivariadas. **Horticultura Brasileira**, Brasília, v.23, n.1, p.22-27, 2005.

VASCONCELOS, C. S.; BARBIERI, R. L.; NEITZKE, R. S.; PRIORI, D.; FISCHER, S. Z.; MISTURA, C. C. Determinação da dissimilaridade genética entre acessos de *Capsicum chinense* com base em características de flores. **Revista Ceres**, Viçosa, v.59, n.4, p.493-498, 2012.

VASCONCELOS, C. S.; BARBIERI, R. L.; NEITZKE, R. S.; PRIORI, D.; FISCHER, S. Z.; MISTURA, C. C. Distância genética entre variedades crioulas de *Capsicum chinense*. **Magistra**, Cruz das Almas, v.26, n.2, p.178-185, 2014.

VASCONCELOS, E. S.; CRUZ, C. D.; BHERING, L. L.; RESENDE JÚNIOR, M. F. R. Método alternativo para análise de agrupamento. **Pesquisa Agropecuária Brasileira**, Brasília, v.42, n.10, p.1421-1428, 2007.

VAZ PATTO, M. C.; SATOVIC, Z.; PÊGO, S.; FEVEREIRO, P. Assessing the genetic diversity of Portuguese maize germoplasm using microsatellite markers. **Euphytica**, Wageningen, v.137, n.1, p.63-72, 2004.

WARD, J. H. Hierarquical grouping to optimize an objective function. **Journal of the American Statistical Association**, v.58, p.236-244, 1963.

APÊNDICE

APÊNDICE A

Script das análises - Software R

```
*****************************************************************
1. Distância Euclidiana Quadrática Padronizada (X/Sx)
*****************************************************************

dados<-read.table("C:\\Users\\Ana Carolina\\Desktop\\
dist.eq.pdesvio.txt", header=F)#distância obtida no software
Genes
D1<-as.dist(dados)

*****************************************************************
2. Método UPGMA
*****************************************************************

UPGMA<-hclust(D1, method="average") #método UPGMA
UPGMA
plot(UPGMA, hang=-1, cex=0.8, ylab="Distâncias Euclidianas
Padronizadas Quadráticas", main="UPGMA", xlab="Acessos")
UPGMA$height #pontos de fusão


*****************************************************************
3. Método Ward
*****************************************************************

WARD<-hclust(D1, method="ward") #método Ward
WARD
plot(WARD, hang=-1, cex=0.8, ylab="Distâncias Euclidianas
Padronizadas Quadráticas", main="Ward", xlab="Acessos")
WARD$height #pontos de fusão


*****************************************************************
4. Método Tocher
*****************************************************************

library(biotools) #pacote no R
T1<-tocher(D1, algorithm = "original")
T1$distClust #distância entre os clusters

*****************************************************************
5. Método Tocher modificado
*****************************************************************

T2<-tocher(D1, algorithm = "sequential")
T2$distClust #distância entre os clusters


*****************************************************************
6. Critério de Mojena
*****************************************************************
```

```
De acordo com Milligan e Cooper (1985), k=1,25.


## 6.1 UPGMA ##
mojena<-mean(UPGMA$height)+1.25*sd(UPGMA$height) #valor Mojena
abline(h=mojena, v=NULL, col=2, lty=2) #linha de corte
(grupos<-cutree(UPGMA,k=k)) #número de grupos


## 6.2 WARD ##
mojena<-mean(WARD$height)+1.25*sd(WARD$height) #valor Mojena
abline(h=mojena, v=NULL, col=2, lty=2) #linha de corte
(grupos<-cutree(WARD,k=k)) #número de grupos

**********************************************************************
7. Coeficiente de Correlação Cofenética
**********************************************************************

## 7.1 UPGMA ##
D3<-cophenetic(UPGMA)
D3
cor(D1,D3)


## 7.2 WARD ##
D4<-cophenetic(WARD)
D4
cor(D1,D4)

**********************************************************************
```